Vehicle Detection and Tracking in Car Video Based on Motion Model

Amirali Jazayeri, Hongyuan Cai, *Member, IEEE*, Jiang Yu Zheng, *Senior Member, IEEE*, and Mihran Tuceryan, *Senior Member, IEEE*

***Abstract*—This paper aims at real-time in-car video analysis to detect and track vehicles ahead for safety, autodriving, and target tracing. This paper describes a comprehensive approach to localizing target vehicles in video under various environmental conditions. The extracted geometry features from the video are continuously projected onto a 1-D profile and are constantly tracked. We rely on temporal information of features and their motion behaviors for vehicle identification, which compensates for the complexity in recognizing vehicle shapes, colors, and types. We probabilistically model the motion in the field of view according to the scene characteristic and the vehicle motion model. The hidden Markov model (HMM) is used to separate target vehicles from the background and track them probabilistically. We have investigated videos of day and night on different types of roads, showing that our approach is robust and effective in dealing with changes in environment and illumination and that real-time processing becomes possible for vehicle-borne cameras.**

***Index Terms*—Dynamic target identification, feature detection, hidden Markov model (HMM), in-car video, probability, tracking, vehicle motion, 1-D profiling.**

I. Introduction

SENSING vehicles ahead and traffic situations during driving are important aspects in safe driving, accident avoidance, and automatic driving and pursuit. We designed a system that is capable of identifying vehicles ahead, moving in the same direction as our car, by tracking them continuously with an in-car video camera. The fundamental problem here is to identify vehicles in changing environment and illumination. Although there have been numerous publications on general object recognition and tracking, or a combination of them, not many of these techniques could successfully be applied in real time for in-car video, which has to process the input on-the-fly during vehicle movement. This paper introduces an effort to design and implement such real-time oriented algorithms and systems that are highly adaptive to the road and traffic scenes based on domain-specific knowledge on road, vehicle, and control.

The in-car video is from a camera facing forward (see Fig. 1), which is the simplest and most widely deployed system on police cars. It records various traffic and road situations ahead and is particularly important to safe driving, traffic recording, and vehicle pursuit. Our objective is to detect vehicles ahead or those being pursued and continuously track them on video. It is not easy for a single moving camera to quickly yield the information from dynamic scenes without stereo or other sensors' assistance [1]. The main difficulties are, first, the numerous variations of vehicles in color, shape, and type. The vast amount of vehicle samples is difficult to model or learn [2]. Second, the vehicle detection must be done automatically along with the video tracking. Many tracking algorithms only assume easily detectable targets or known initial positions [3]. Third, the in-car video may confront drastic variation of environment and illumination [4]. Transition through shadowy and sunny locations in urban areas, dim lighting at night, loss of color on a cloudy day, highlights on vehicles, occlusions between vehicles, and so on make the feature extraction, recognition, and tracking unstable.

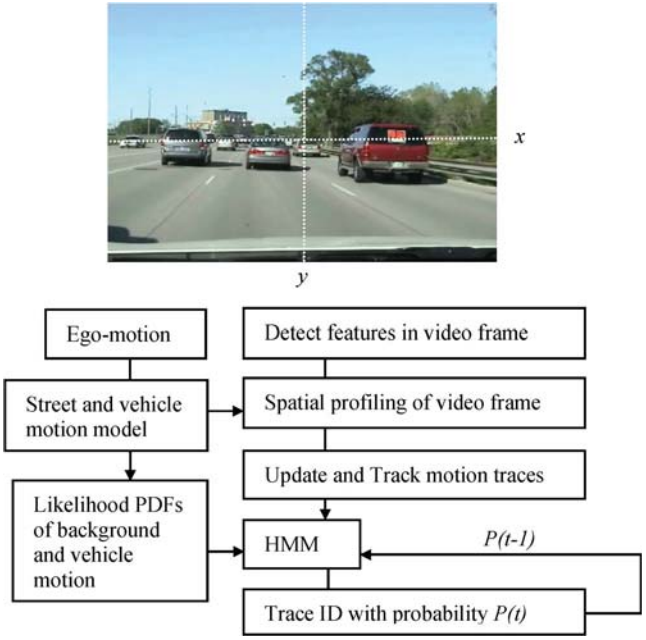
x
y
Ego-motion
Street and vehicle motion model
Likelihood PDFs of background and vehicle motion
Detect features in video frame
Spatial profiling of video frame
Update and Track motion traces
HMM
P(t-1)
Trace ID with probability P(t)

Fig. 1. Typical frame of in-car video and the processing diagram. The video resolution is 640 × 480 pixels.

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 vehicle and scene modeling and learning, which are mostly unstable, time consuming, and nongeneral. Second, we focus

Manuscript received December 1, 2009; revised March 14, 2010, May 18, 2010, and October 19, 2010; accepted January 30, 2011. Date of publication March 3, 2011; date of current version June 6, 2011. This work was supported by the National Institute of Justice, USA, under Grant 2007-DE-BX-K007. The Associate Editor for this paper was A. Hegyi.

The authors are with Indiana University–Purdue University Indianapolis, Indianapolis, IN 46202 USA (e-mail: amirali.jazayeri@gmail.com; hocai@cs.iupui.edu; jzheng@cs.iupui.edu; tuceryan@cs.iupui.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2011.2113340

on the horizontal scene movement for fast processing based on the configuration of camera and vehicle driving mechanism. One-dimensional profiles are accumulated from video frames to show the horizontal motion directly in traces. We track feature trajectories in such temporal profiles so that the real-time vehicle and background classification can be done in a reduced dimension. Third, we put more weight on the understanding of motion behavior of the vehicles than object shape analysis in identifying targets. We use the hidden Markov model (HMM), which has widely been used in describing sequential data [5], [6], to model the continuous movement of features so that the results will be less influenced by thresholding during low-level preprocessing. Identification of target vehicles will be given in terms of probabilities.

Many related works of in-car video or images to identify cars are shape-based methods [7], [8] that usually suffer from brittleness due to the variety of vehicles and backgrounds. For example, the symmetric property is employed in [7] and [9]–[11]. Deformable frame models are used in [12]–[14] and [27]. Because of the lack of robustness in feature extraction from complex scenes, subsequent decisions can rarely be applied to video in real time. Several works have employed probability in their formulation, but they are mainly applied to the occurrence and appearance (intensity distribution) of vehicles [2], [11], [15], [16], [26], or some of them are from a static camera [17]–[19].

The significance of this paper lies in its temporal processing in target identification. We focus on the time domain that displays the vehicle generated motion. The motion is general for all types of vehicles. The motion behavior of scenes in videos is modeled with the HMM, and this makes the vehicle identification with certainty measure. In addition, the profiled 1-D arrays reduce the dimensionality of data, thus achieving real-time processing to simultaneously identify and track vehicles.

In the following, we will give an overview and describe general assumptions in Section II. We introduce feature detection and tracking in Section III. We investigate the motion properties of scenes in terms of likelihoods in Section IV. We address the identification of feature trajectory as a vehicle or background in Section V. Experimental results and discussion are given in Section VI.

II. Motion Information In-Car Video

A. Overview of the Work

By showing the continuous motion of extracted points in car video without color and shape information to human subjects, we have confirmed that humans are capable of separating vehicles from background after knowing where the source is from. As the observer vehicle moves on the road, the relative background motion is determined. The motion projected to the camera frame is further determined from object distances. Such motion shows unique properties coherent to the vehicle ego-motion. On the other hand, the target vehicles moving in the same direction as the camera have different motion against the background and, thus, show different optical flow from that of the background in the video. We look into the motion characteristics of tracked objects and separate them as static background or moving cars in the process shown in Fig. 1.

Different from many other works that put more effort into vehicle shape analysis in individual video frames, this paper only extracts low-level features such as corners, intensity peaks, and horizontal line segments, as will be described in Section III. These features are profiled to the temporal domain to ensure the compactness of data and the robustness of processing against the camera/vehicle shaking, noises, and irrelevant features. The motion of scenes appears as traces in the temporal profiles. The horizontal image velocities of features are obtained through trace tracking.

To identify tracked traces as cars and background, we model the motion properties of scenes in the video in Section IV. Because the background flow is ego-motion orientated and position dependent in the image frame, we examine the image flow with respect to the position during ego-motion and describe them by their joint probability. We estimate likelihood probability distributions for cars and background, respectively, given the probability density functions (pdfs) of scenes, ego-motion, and pursuit vehicles. With such initially calculated likelihood pdfs, we can obtain high responses to the expected motion events.

In Section V, we estimate the posterior probabilities of motion traces as target vehicles or background based on observations, i.e., the detected image position and velocity at each moment. To achieve a stable classification of vehicles, we evaluate their motion process by means of the HMM when traces are tracked continuously. The precalculated likelihood pdfs and the transition probabilities are used in the HMM for the state estimation of traces. The continuous vehicle motion provides more robust identification than shape features that are affected by occlusion, illumination, and background changes.

B. Dynamic Environment of In-Car Video

The general assumptions that we made are the following: 1) The observer car should not turn away completely or stray too far from the target vehicles on the road. 2) A translation of the observer vehicle is required. For the vehicle-borne camera, a pure rotation without significant translation, e.g., turning at a street corner, does not provide sufficient information to separate the targets and background motion, because such a rotation generates almost the same optical flow in the entire frame. Another reasonable assumption is the continuity of the vehicle and camera motion, which is guaranteed in general according to the driving mechanism of four-wheeled vehicles.

Assume that the camera-centered coordinate system is O-XYZ (or O-xyz), with the X-axis toward right, the Y-axis facing down, and the Z-axis in the vehicle's direction of motion. Denote the 3-D position of a scene point by (X, Y, Z) in O-XYZ and its projected image coordinates by (x, y). A typical view of the in-car video is depicted in Fig. 2, where the focus of expansion (FOE) is located in the center part of the image $I(x, y)$. On a curved road, a steering adds a rotation to the ego-motion [see Fig. 2(b)].

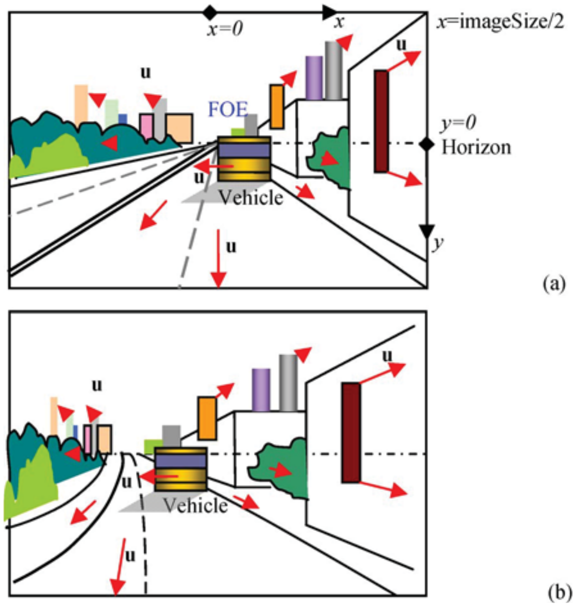

Fig. 2. Typical views of in-car video with red arrows indicating optical flow or image velocity. (a) Straight road. (b) Mildly curved road.

A target car that is moving on the road may change its horizontal position and scale when it changes lanes and speed but may still retain its position within the road even when its instantaneous image velocity $\boldsymbol{u}(x, y)$ dramatically changes. The background, however, has a sort of motion coherence, with the flow spreading out gradually from FOE toward the sides of the image frame. The image velocity increases as the scene moves closer. This cue is sufficient for humans to classify vehicles and background even without knowing the shapes of objects. We will model this motion coherence for automatic background and vehicle separation.

According to the perspective projection of the camera, the image position of a moving scene point $P(X, Y, Z)$ is

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

where f is the camera focal length. Denote the relative translation of scene point $P(X, Y, Z)$ to the camera by $(T_x(t), T_y(t), T_z(t))$ in the system O-XYZ and denote the rotation velocities of the observer vehicle/camera in pitch, yaw, and roll by $(R_x(t), R_y(t), R_z(t))$ in radians per second, where the pitch and roll of the observer vehicle have $R_x(t) \approx 0$ and $R_z(t) \approx 0$ on a flat road. The relative speed of the scene point P to the camera is

$$\begin{aligned}(V_x(t), V_y(t), V_z(t)) &= (T_x(t), T_y(t), T_z(t)) \\ &\quad + (X, Y, Z) \times (R_x(t), R_y(t), R_z(t))\end{aligned} \tag{2}$$

according to [20]. By differentiating (1) with respect to t and replacing related terms in the result using (1) again, the horizontal component of the image velocity of P becomes

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

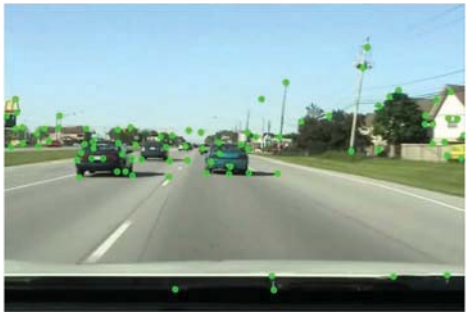

Fig. 3. Corner detection in video sequence. Corners are marked with green circles.

Replacing $V_x(t)$ and $V_z(t)$ with (2) and setting $R_x(t) = 0$ and $R_z(t) = 0$, we obtain

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

if we denote

$$v^{(t)}(t) = \frac{fT_x(t) - x(t)T_z(t)}{Z(t)} \quad v^{(r)}(t) = -\frac{x^2(t) + f^2}{f}R_y(t) \tag{5}$$

to be the components of horizontal image velocity from translation and rotation, respectively. If the observer vehicle moves on a straight road, i.e., $R_y(t) = 0$, then we simply have $V_x(t) = T_x(t)$ and $V_z(t) = T_z(t)$.

III. Feature Extraction and Temporal Tracking

A. *Real-Time Feature Extraction in Video Frames*

The segmentation of vehicles from background is difficult due to the complex nature of the scenes, including occlusion by other moving vehicles, complicated shapes and textures, coupled with the ever changing background. The presence of specular reflection on the metallic surfaces and back windows, and shadows on most cars (deformed and spread), make the color and shape information unreliable. To cope with these variations, we select three types of low-level features for reliable vehicle detection: horizontal line segments, corner points, and intensity peaks.

1) Corner Detection: The important features in the video are the corners appearing on high contrast and high curvature points (excluding edges). In daytime videos, corner points on a vehicle surface or background scenes keep their positions stable over time on surfaces so that they provide coherent motion information of regions. At occluding contours of vehicles (usually on two sides of the vehicles), however, the corner points detected are formed by vehicle and background, which do not physically exist and are extremely unstable during the relative motion of vehicles and background scenes. We use the Harris feature for detecting corners in the video frames [21] (see the results in Fig. 3).

2) Line Segment Detection: We have noticed that the backs of vehicles typically contain many horizontal edges formed by vehicle tops, windows, bumpers, and shadows. Most of

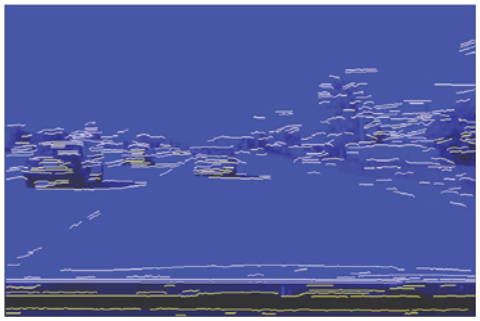

Fig. 4. Tracked edge points form horizontal line segments in video frames that can characterize vehicles.

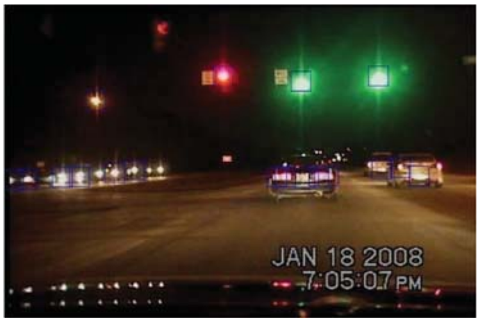

Fig. 5. Intensity peaks (marked in squares) extracted on traffic lights, tail lights, and front lights of vehicles.

them are visible during daylight, which indicates the existence of a vehicle. For vehicles with partially occluded backs, the detection of horizontal line segments is still stable. The vertical line segments, however, are not guaranteed to be visible due to a curved vehicle body, frequent occlusion by other cars, and occlusion over a changing background.

We convolve each video frame with a vertical differential operator $\partial I(x,y)/\partial y$, which results in the image $I'_y(x,y)$. Then, an edge-following algorithm searches peaks with contrasts higher than threshold δ_1 in $I'_y(x,y)$ to produce horizontal line segments. The horizontal search has a vertical tolerance of ± 2 pixels, and it selects point candidates with the same edge sign. It also uses another threshold $\delta_2 < \delta_1$ to bridge line segments at weak contrast points by allowing a 3-pixel trial after reaching the end by using δ_1. The tracked edge points form a line segment if it satisfies constraints on the minimum and maximum lengths and near horizontal orientation. These thresholds are loosely determined to yield more segments. We pass more information to the later probabilistic processing rather than abruptly cutting it off at this stage.

Line tracking may break down at highlights, insufficient resolution on distant cars, and scale changes when the depth changes. The results may also contain segments in a static background such as long wires, markings painted on the road, and other building structures. Fig. 4 shows a frame of line extraction overlapped with the intensity image.

3) Intensity Peak Detection for Night Scenes: The intensity peaks from the tail and head lights of moving vehicles and from the street lamps are used as features when the lighting conditions are poor. We detect intensity peaks from vehicle lights to obtain more evidence of the vehicle presence. We use a large Gaussian filter to smooth the image and find the local maxima beyond a threshold adaptively determined. These peaks will further be tracked across frames for the direction and speed of the moving targets. Fig. 5 shows examples of the intensity peak detection and tracking in video. After preprocessing and feature extraction, it is noted that a single data source alone cannot guarantee reliable results.

B. Vertical Profiling of Features to Realize Data Reduction

To speed up the processing for real-time target tracking and to obtain robust results against shaking of a vehicle on changeable road slope, we vertically project the intensity/color $I(x,y)$ in each video frame to form a 1-D profile $T(x)$. Consecutive profiles along the time axis generate a condensed spatiotemporal image $T(x,t)$ used for analyzing the scene movement [22], [23]. The vertical projection of intensities through a weight mask $w(x,y)$ is implemented as

$$T(x,t) = \sum_{y=-h/2}^{h/2} w(x,y) I(x,y,t) \tag{6}$$

where h is the image height. The weight distribution, which will be described in Section IV-C, is high at regions with high probability of containing vehicles in the image. The traces in the condensed spatiotemporal image show movements of long or strong vertical lines in the video. Slanted lines and short edges in the individual video frames will not be distinct in it. Fig. 6(a) shows such an image, which is a compact and robust representation for recording and estimating vehicle motion.

In addition to the color/intensity profiling, we also vertically profile features extracted in the video to generate feature traces in the spatiotemporal images. For example, we profile the number of horizontal line segments at each position of x by

$$T_l(x,t) = \sum_{y=-h/2}^{h/2} w(x,y) L(x,y,t) \tag{7}$$

where $L(x,y,t)$ takes value 1 on a horizontal line segment and 0 otherwise in frame t. Such a result can be found in Fig. 6(b), where the bright stripes accumulated from many line segments show the motion of vehicles. Long and horizontal lines, such as road paints and wires in the scene, equally add to all the positions in the profile without influencing the traces of vehicles. 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 many cases when entering and exiting a tunnel, penetrating shady and sunny locations, and being illuminated by other vehicles. Such changes alter the intensity of the entire frame. In the spatiotemporal profiles, such illumination changes noticeably appear as horizontal lines over the entire image width. Its influence on the vehicle position is filtered out by taking a horizontal derivative of the profiles.

Compared with the intensity profiles with smooth traces from vertical lines, the profiling of horizontal line segments yields

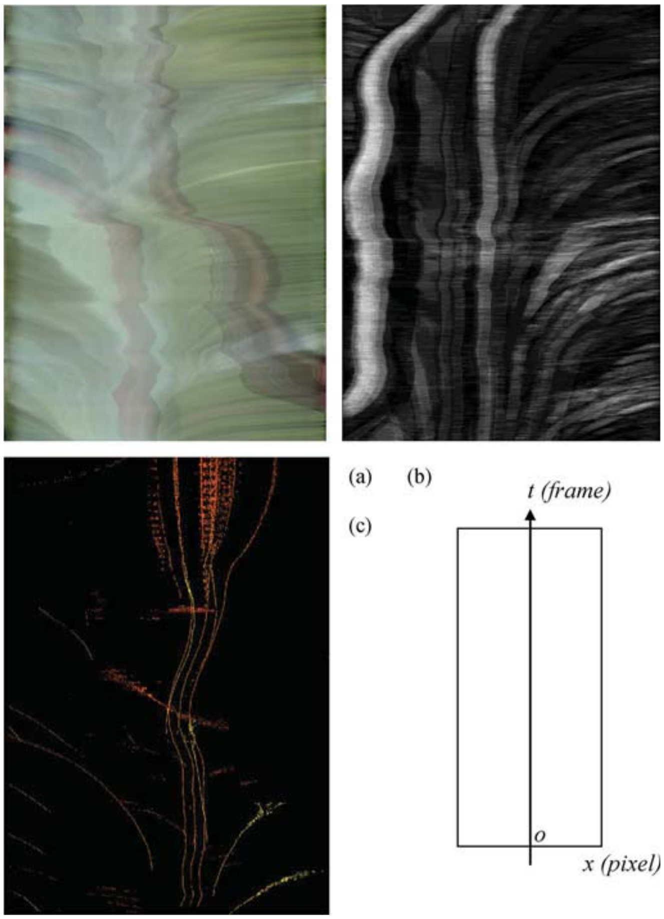

Fig. 6. Examples of different profiles. (a) An intensity profile from tree background and a red car. (b) A profile from horizontal line segments. (c) A profile of intensity peaks in a night scene before a car stops (tail break light in red).

noisy traces due to the instability of line detection. Because the lighting conditions, background, and camera noise vary frame by frame, the line segments found in the images do not always appear at consistent locations and in the same length, which has to be postprocessed according to their temporal coherence. In addition, as is evident in Fig. 6(b), many traces may not belong to a vehicle but to the background, which is left for the trace classification.

C. Tracking Features in Temporal Profiles for Motion

The intensity profile shows the horizontal positions of vertical features and ignores the horizontal features. This produces a compact image for understanding the horizontal motions in the video. The effects of the image changes caused by jitter in camera roll and tilt on uneven roads are significantly reduced in the profiles.

Tracking intensity profiles is done by horizontal differentiation of $T(x, t)$. The edges are marked in another image $E(x, t)$, and a maximum span is set in searching consecutive trace points. At the same time, $\partial T(x, t)/\partial t$ is computed to find horizontal edges, because a fast moving trace is very slanted in $T(x, t)$. We thus can link horizontal edges in $E(x, t)$ for traces with a high image velocity. The very long horizontal segments in $E(x, t)$, mainly from the instantaneous illumination changes, are ignored in the result. This processing results in the image position $x(t)$ and the image velocity $v(t) = x(t) - x(t - 1)$ for a trace.

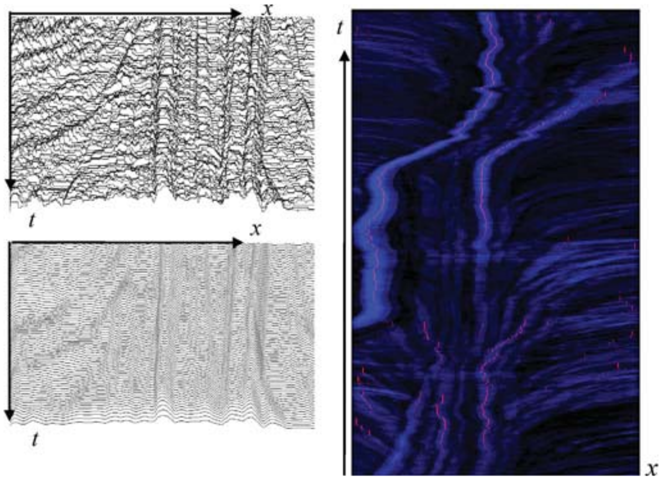

Fig. 7. Tracing centers of line segment clusters. (Left) Profiled distribution of line segments and its smoothed distribution. (Right) An example of trace center (in purple) overlapped with traces.

To preserve the continuity of motion, we compare the image velocities $v(t - 1)$ from tracked points and $v(t)$ from new candidates and select a successive point such that $|v(t) - v(t - 1)|$ is kept minimum. In addition to the requirement of high contrast, the sign of trace at $E(x, t)$ is used as reference in tracking. The approach to examine the motion continuity is also applied in tracking intensity peaks and corners in the profiles.

Among all types of traces, the horizontal line segment piles provide the most salient clue of vehicle presence. Tracking traces of line segments in $T_l(x, t)$ is done by locating the center of each trace, because the endpoints of the line segments are unreliable, particularly for distant vehicles. We first horizontally smooth $T_l(x, t)$ to obtain major traces, as shown in Fig. 7. These traces are marked at their peaks $x(t)$ above a threshold for the centers. The threshold is adaptively decided at each t according to the average and standard deviation of $T_l(x, t)$.

Random line segments on background and instantaneous light changes over the entire frame (tail braking lights) will cause a long horizontal line in $T_l(x, t)$. However, these will be ignored in the processes of finding peaks (they have no distinct peak) and tracking over long periods of time. In the daytime, we use line traces in the car identification and refine the car width by the intensity profile. At night and under low visibility conditions, such as snow and fog, the profiles from intensity peaks of tail lights can be used. The profile from corner points is used to enhance the identification in any case. In addition, the horizontal line may not detect small targets such as motorbikes. The intensity and corner profiles have to be used then.

IV. Modeling Motion of Scenes in Likelihoods

Let us model the motion with probability to determine the state of a tracked object with the HMM. We assign two hidden states to a trace at any time instance t: *car* and *background*, which are denoted by C_t and B_t. The observations are image

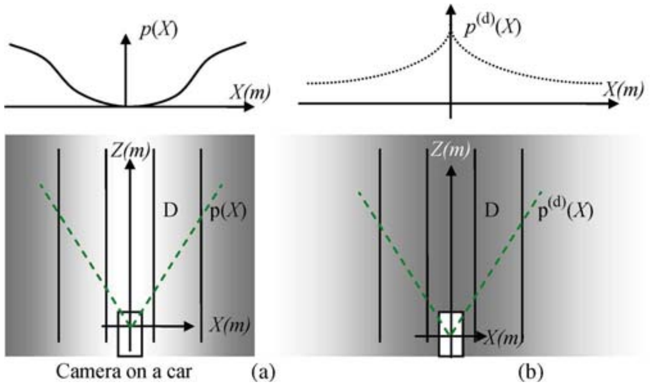

Fig. 8. Probability distribution of background beside the road and its detectability displayed in dark intensity. (a) Background feature distribution on road sides. (b) Detectability of features reduced because of occlusion in urban areas as the distance to the road increases.

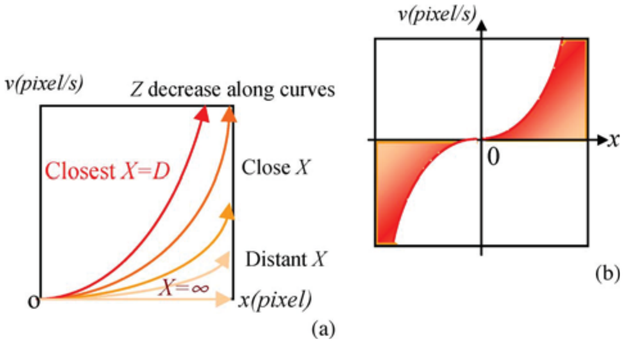

Fig. 9. Relation of the image position and the horizontal image velocity on background scenes. (a) Motion traces of background points on the right side of the road. As the depth Z is reduced, the background moves fast in the outward direction. (b) Traces for background on both sides of the road. The colors correspond to different X's from the camera axis (or simply road).

position $x(t)$ and the horizontal image velocity component $v(t)$ of $\boldsymbol{u}(x, y)$, both are continuous distributions rather than discrete events in the conventional HMM [5]. Array $(x(t), v(t))$ is obtained from each trajectory tracked in the condensed profiles. The two parameters are not independent; neither one will determine the identity of object feature alone. We first calculate their joint likelihood pdfs $\mathrm{P}(x, v|B)$ and $\mathrm{P}(x, v|C)$ for background and vehicles in this section to estimate the posterior probabilities $\mathrm{P}(B|(x(t), v(t)))$ and $\mathrm{P}(C|(x(t), v(t)))$ later in the next section based on observations.

A. Likelihood Function of Background in Video

1) Traveling Along Straight Roads: In the camera coordinate system O-XYZ, X is the horizontal distance of an object point from the camera axis (from road if the camera is on a straight lane). For background, we have $|X| > D$, where $2D$ is the average road width. We can assume its probability distribution in O-XYZ, as shown in Fig. 8(a). The heights of background features are homogeneously distributed to include high buildings and trees. Assuming a pure translation of the vehicle first (i.e., $R_y = 0$), the absolute speed V of the camera (in Z-direction) follows a normal distribution $p(V) \sim G(S, \sigma^2)$, where its average value S can be set properly such as 50 km/h for pursuing, and the standard deviation σ can be 10 km/h.

For a background point, its $V_x = 0$ and $V_z = -V$ approaching the camera, and its X is fixed. Then, the image velocity in (3) becomes

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

This draws dense function curves in $(x(t), v(t))$ space for various X, as shown in Fig. 9(a). If the vehicle roughly has a constant speed, then the trajectory is a hyperbola $x(t) = fX/(Z_0 - Vt)$ in profile $T(x, t)$ from an initial depth Z_0, according to (1). As to the distribution of background features along the X-direction, we can assume that it follows a flipped Gaussian distribution, i.e., $p(X) \sim 1 - \exp(-X^2/2D^2)$, as Fig. 8(a) depicts.

In general, if the pdf of a random variable χ is $p_\chi(\chi)$, and β is a monotonic function of χ, i.e., $\beta = f(\chi)$, then the pdf of β can be calculated as

$$p_\beta(\beta) = p_\chi\left(f^{-1}(\beta)\right)\left|\frac{\partial f^{-1}(\beta)}{\partial \beta}\right| \text{ or } p_\beta(\beta) = \frac{p_\chi\left(f^{-1}(\beta)\right)}{\left|\frac{\partial f(\chi)}{\partial \chi}\right|} \tag{9}$$

according to [24]. The meaning behind these formulas is as follows: First, for discrete random events, a value of β is mapped from χ. Therefore, the occurrence probability of β is exactly the same as that of χ, i.e., $p_\beta(\beta) = p_\chi(\chi)$. Inversely mapping β to χ in the domain can obtain $p_\beta(\beta)$ from the corresponding $p_\chi(f^{-1}(\beta))$. Second, for a continuous function f, the derivatives $|\cdot|$ in (9) adjust the local range of the pdf between domain χ and function β in representing the same random event. Further, for a multivariate function f, i.e., β and χ are vectors of the same length, the pdf of the function can also be derived similarly with $|\cdot|$ when it changed to a Jacobian format.

Now, we compute the likelihood $\mathrm{p}(x, v|\mathrm{B})$ of image motion behavior for background point $(X, Z) \in \mathrm{B}$. Here, the original variables are X, Z, and V in 3-D space, and their pdfs are preset. The image parameters x and v are their dependent functions, as described by (1) and (8). Because the inverse mapping from x, v to X, Z, V is not a unique transformation between the same number of variables, we have to use Bayes' theorem and conditional probability to include all the possible cases of X, Z, V in producing the joint probability of (x, v). We map pair (x, v) to variables Z and V in the 3-D space but loop X over its whole scope, which is detailed as

$$\begin{aligned} p(x, v|B) &= p\left(x, v|(X, Z) \in B\right) && (10)\\ &= \int_X p(X) p\left(x, v|X, (X, Z) \in B\right) \\ &\quad \times dX && \textit{Cond. Prob.} \\ &= \int_X p(X) p\left(Z(x, v), V(x, v)|X\right) \\ &\quad \times \left|\frac{fX}{x^2}\right|^2 dX && \text{[see (9), (1), and (8)]} \end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad Replace\ p(x,v)\ with\ p(Z,V)\ and\\
&\qquad\qquad compute\ Jacobian\\
&= \int_X p(X)p\left(Z(x,v)|X\right)p\left(V(x,v)|X\right)\\
&\quad\times\left|\frac{fX}{x^2}\right|^2 dX \qquad\qquad Z,V\ independent\\
&= \int_X p(X)p\left(Z=\frac{fX}{x}\ |X\right)\\
&\quad\times p\left(V=\frac{vfX}{x^2}\ |X\right)\left|\frac{fX}{x^2}\right|^2 dX \quad \text{[see (1) and (8)]}.
\end{aligned}
$$

Because $p(Z)$ is invariant in the Z-direction, as depicted in Fig. 8(a), $p(Z)$ can be moved out of the integral above as a constant. With input original probability distributions of X and V, the likelihood pdf for background becomes

$$
\begin{aligned}
p(x,v|B) &= C_b \int_X p(X)p\left(V=\frac{vfX}{x^2}\ |X\right)\\
&\quad\times\left|\frac{fX}{x^2}\right|^2 dX \qquad\qquad \text{[see (10)]}\\
&= C_b \int_X \left(1-e^{\frac{-X^2}{2D^2}}\right) e^{\frac{-\left(\frac{vfX}{x^2}-S\right)^2}{2\sigma^2}}\\
&\quad\times\left|\frac{fX}{x^2}\right|^2 dX
\end{aligned}
\tag{11}
$$

where C_b is a constant, and a Jacobian $|\cdot|$ is included.

In real situations, we further consider the detectability of background scenes as a function $p^{(\mathrm{d})}(X) \propto 1/(|X|+1)$. The objects at the roadside have the highest visibility. The further the object is away from the road ($|X|$ increases), the greater the chance of it to be occluded by front objects. $p(X)$ modified by $p^{(d)}(X)$ is

$$
p(X) = \frac{1-\exp(-X^2/2D^2)}{1+|X|}. \tag{12}
$$

The likelihood pdf of the background is then

$$
p(x,v|B) = C_1 \int_X \frac{1-e^{\frac{-X^2}{2D^2}}}{|X|+1} e^{\frac{-\left(\frac{vfX}{x^2}-S\right)^2}{2\sigma^2}} \left|\frac{fX}{x^2}\right|^2 dX \tag{13}
$$

where C_1 is a constant for normalization, i.e., $\int_{\mathrm{v}}\int_{\mathrm{x}} p(x,v|\mathrm{B})dxdv = 1$. After $p(x,v|\mathrm{B})$ is established, we normalize it in the entire scope to determine C_1 and create a pdf, as shown in Fig. 10. The values form a look-up table for use in real-time tracking and computation using the HMM.

2) Traveling on Curved Roads: During smooth driving of the observer vehicle along a straight and mildly curved road, its steering change should be small. We can describe the vehicle rotation $R_y(t)$ in a normal distribution with a small variance. According to (4) and (5), we estimate a general distribution $p(x,v)$, including rotation by adding component $v^{(r)}(t)$ to the image velocity that was from translation $v^{(t)}(t)$ so far. As illustrated in Fig. 11, this is a vertical shift of $p(x,v)$ from $v^{(t)}(t)$ by a term $v^{(r)}(t)$ in a quadratic form. The updated scope of feature trajectories is intuitively drawn in Fig. 11 for a certain steering angle R_y. If the rotation parameter $R_y(t)$ is not provided from the encoder of the observer vehicle, then we have to include all the possible values of $R_y(t)$ in normal distribution to compute $p(x,v|\mathrm{B})$ as

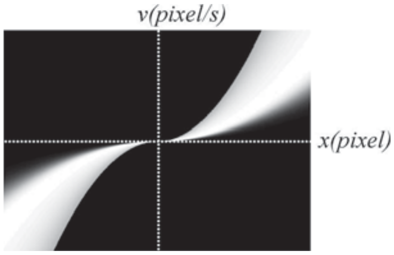

Fig. 10. Probability distribution $p(x,v)$ of background in motion.

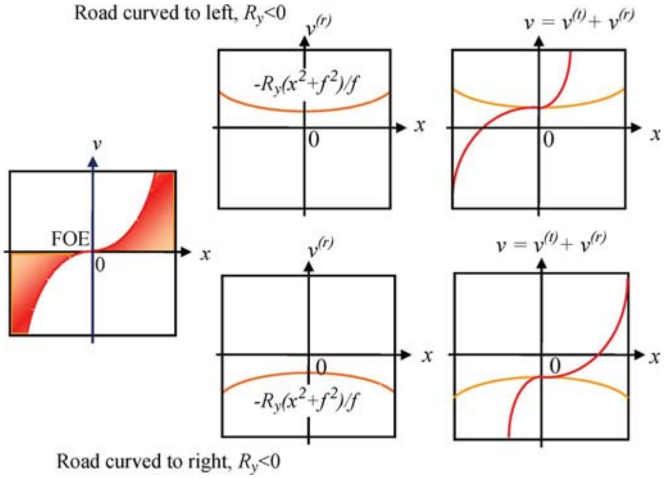

Fig. 11. Adding a rotation component to translation to cope with the curved road. (Left) $p(x,v)$ from translation of camera. (Middle) Motion component caused by rotation from vehicle steering. (Right) Shifting distribution $p(x,v)$ from translation yields a new distribution (between two curves) including rotation.

$$
\begin{aligned}
p(x,v|B) &= \int_{R_y} p(R_y)p(x,v|R_y)dR_y \qquad\qquad (14)\\
&= \int_{R_y} p(R_y)p\left(x,v\ |v=\frac{x^2V}{fX}-\frac{x^2+f^2}{f}R_y\right)dR_y\\
&\qquad\qquad\qquad\qquad\qquad\qquad Eq.\ 4\\
&= \int_{R_y}\int_X p(R_y)p^{(d)}(X)p(X)p\left(Z=\frac{fX}{x}\ |X\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad Eqs.\ 3, 13\\
&\quad\times p\left(V=\left(v+\frac{x^2+f^2}{f}R_y\right)\frac{fX}{x^2}\ |X\right)dXdR_y\\
&= C_{1r}\int_{R_y}\int_X e^{\frac{-R_y^2}{2\sigma_r^2}}\frac{1-e^{\frac{-X^2}{2D^2}}}{|X|+1}\\
&\quad\times e^{\frac{-\left(\left(v+\frac{x^2+f^2}{f}R_y\right)\frac{fX}{x^2}-S\right)^2}{2\sigma^2}}\left|\frac{fX}{x^2}\right|^2 dXdR_y
\end{aligned}
$$

where C_{1r} is a constant to be determined in the normalization of $\mathrm{p}(x,v|\mathrm{B})$. Fig. 12(a) shows such a result that is basically a

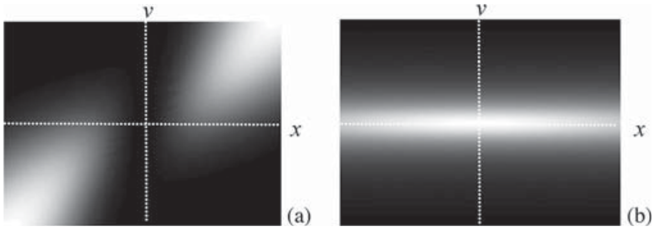

Fig. 12. Probability distributions of background and vehicle in color. (a) Background when the observer vehicle involves rotation. (b) Probability distribution $p(x, v)$ of vehicle features appearing in continuous space (x, v).

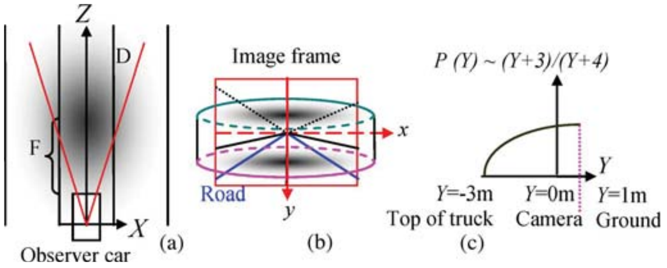

Fig. 13. Probability density of target vehicle position relative to the camera shown in the intensity. The darker the intensity, the higher the probability. (a) Target distribution in top view. (b) Projection of the distribution onto image frame. (c) Vehicle feature distribution in height, which is from the ground to the top of highest vehicle in about 4 m. The Y-axis is downward to fit with the Z-axis in the forward direction.

Gaussian blur of the pdf of vehicle translation (see Fig. 10) in the velocity direction.

B. Probability of Target Vehicle Motion in Video

Assuming a pure translation of the observer vehicle, as shown in Fig. 13(a), we can assume the 3-D position of a target vehicle on the road $(\mathrm{X}(t), \mathrm{Z}(t))$ following 2-D normal distribution $\mathrm{G}((0, F), (D^2, (2F)^2))$, where F is its average distance from the camera. Here, $X(t)$ and $Z(t)$ are independent, and D and $|2F|$ are used as the standard deviation. The projection of the normal distribution onto the image frame is depicted in Fig. 13(b), which is distributed vertically in the cylinder according to a height probability distribution $H(Y)$ of vehicle features, as shown in Fig. 13(c). In the Y-direction, the features are guaranteed to be detected near the ground due to the uniformity and high contrast of the vehicle shadow, tire, and bumper against the homogeneous ground. However, features may not be found reliably at higher Y positions due to a low height of target vehicle, highlights on its metal top, and changing reflections on the back window. A gray car may have its top color mixed with the sky. Here, we design a feature distribution function in Y position as $p(Y) \sim (Y + 3)/(Y + 4)$ in system O-XYZ, where $Y \in [-3\text{ m}, 1\text{ m}]$, assuming that the Y-axis is facing down and that the camera position $(Y = 0)$ on the vehicle is 1 m above the ground $(Y = 1\text{ m})$. Moreover, we assume that the relative speed of the target vehicle to the camera, which is denoted by (T_x, T_z), also follows a normal distribution $\mathrm{G}((0, 0), (\sigma_x^2, \sigma_z^2))$ in a stable pursuing period, where T_x and T_z are independent.

For a tracked car, we can thus compute its image behavior $\mathrm{p}(x, v|\mathrm{C})$ according to (1) and (3). For a point on the target vehicle

$$\begin{aligned}
p(x, v|C) &= p\left(x, v|(X, Z, T_x, T_z) \in C\right)\\
&= p\left(x = \frac{fX}{Z}, v = \frac{fT_x - xT_z}{Z}\right) \qquad (15)\\
&= \int_Z p(Z)p\left(x = \frac{fX}{Z}, v = \frac{fT_x - xT_z}{Z}\,|Z\right) dZ\\
&\hspace{20em} Bayesian\\
&= \int_Z p(Z)p\left(X = \frac{xZ}{f}, T_z, T_x = \frac{Zv + xT_z}{f}\,|Z\right)\\
&\quad \times \left|\frac{Z}{f}\right|^2 dZ \qquad \text{[see (9)]}\\
&= \int_Z p(Z)p\left(X = \frac{xZ}{f}\,|Z\right)\\
&\quad \times p\left(T_z, T_x = \frac{Zv + xT_z}{f}\,|Z\right)\left|\frac{Z}{f}\right|^2 dZ\\
&\hspace{16em} X, T_z, T_x\ independent\\
&= \int_Z p(Z)p\left(X = \frac{xZ}{f}\,|Z\right)\\
&\quad \times \left\{\int_{T_z} p(T_z)p\left(T_x = \frac{Zv + xT_z}{f}\,|T_z, Z\right) dT_z\right\}\\
&\quad \times \left|\frac{Z}{f}\right|^2 dZ. \qquad Bayesian
\end{aligned}$$

Filling in probability distributions of $p(T_x)$, $p(Z)$, and $p(T_z)$, we can obtain $p(x, v|C)$ as

$$\begin{aligned}
p(x, v|C) &= C_2 \int_Z \int_{T_z} e^{\frac{-(Z-F)^2}{2(2F)^2}} e^{\frac{-\left(\frac{xZ}{f}\right)^2}{2D^2}} e^{\frac{-T_z^2}{2\sigma_z^2}}\\
&\qquad \times e^{\frac{-\left(\frac{Zv + xT_z}{f}\right)^2}{2\sigma_x^2}} \left|\frac{Z}{f}\right|^2 dT_z dZ\\
&= C_2 \int_Z \int_{T_z} e^{-\frac{(Z-F)^2}{2(2F)^2} - \frac{\left(\frac{xZ}{f}\right)^2}{2D^2} - \frac{T_z^2}{2\sigma_z^2} - \frac{\left(\frac{Zv + xT_z}{f}\right)^2}{2\sigma_x^2}}\\
&\qquad \times \left|\frac{Z}{f}\right|^2 dT_z dZ \qquad (16)
\end{aligned}$$

where C_2 is a constant for normalization, i.e., $\int_{\mathrm{v}} \int_{\mathrm{x}} p(x, v|\mathrm{C}) dx dv = 1$. Fig. 12(b) shows this likelihood pdf for target vehicles. Moreover, the likelihood pdf containing the observer vehicle rotation can similarly be derived like (14) as

$$\begin{aligned} p(x,v|C) &= \int_{R_y} p(R_y) \int_{Z} p(Z) p\left(X = \frac{xZ}{f} \,|Z\right) \\ &\quad \times \left\{ \int_{T_Z} p(T_z) p\left(T_x = \frac{Zv + xT_z}{f} \right.\right. \\ &\qquad\qquad \left. + \frac{x^2+f^2}{f} Z R_y \,|T_z, Z, R_y\right) \\ &\qquad \left. \times \left|\frac{Z}{f}\right|^2 dT_z \right\} dZ dR_y. \end{aligned} \tag{17}$$

C. Occurrence of Target Vehicles in Camera Frame

The vertical profiling of intensity and features in video frames has facilitated target tracking. Here, we examine the profiling weight in the image frame such that the spatiotemporal representation obtained will best reflect the motion behaviors of target vehicles. Using the same vehicle pdf in the previous subsection, we will compute the probability of feature at each image position, i.e., $p(x,y|C)$ or simply $w(x,y)$.

Because the mapping from X, Y, Z to x, y is not a one-to-one relation, the probability $p(x,y|C)$ is computed by

$$\begin{aligned} p(x,y|C) &= \int_{Z} p(Z) p(x,y|Z) dZ \qquad\qquad Bayesian \\ &= \int_{Z} p(Z) p\left(X = \frac{xZ}{f}, Y = \frac{yZ}{f} \,|Z\right) \left|\frac{Z}{f}\right|^2 dZ \\ &\qquad\qquad\qquad\qquad\qquad [\text{see } (9)] \\ &= \int_{Z} p(Z) p\left(X = \frac{xZ}{f} \,|Z\right) \\ &\qquad \times p\left(Y = \frac{yZ}{f} \,|Z\right) \left|\frac{Z}{f}\right|^2 dZ \\ &\qquad\qquad\qquad\qquad X, Y \; independent \\ &= C_3 \int_{Z} e^{\frac{-(Z-F)^2}{2(2F)^2}} e^{\frac{-\left(\frac{xZ}{f}\right)^2}{2D^2}} \left(\frac{yZ}{f} + 3\right) \\ &\qquad \times \left(\frac{yZ}{f} + 4\right)^{-1} \left|\frac{Z}{f}\right|^2 dZ \\ &= C_3 \int_{Z} e^{\frac{-(Z-F)^2}{2(2F)^2}} e^{\frac{-\left(\frac{xZ}{f}\right)^2}{2D^2}} \frac{yZ+3f}{yZ+4f} \left|\frac{Z}{f}\right|^2 dZ \end{aligned} \tag{18}$$

where C_3 is a constant for normalization. One result is displayed in Fig. 14. We use it as $w(x,y)$ in (6) and (7) in profiling to enhance the extraction of vehicles and to ignore majority of irrelevant features from the background.

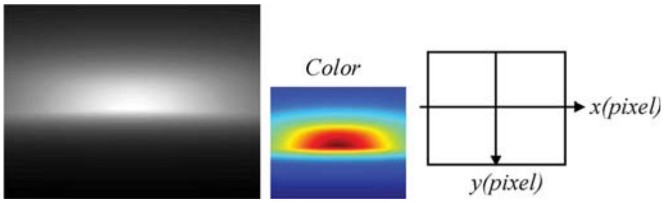

Fig. 14. Probability of vehicle features $w(x,y)$ in the image frame for the vertical profiling of image features.

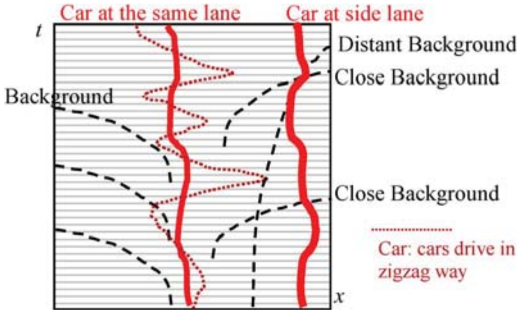

Fig. 15. Motion behaviors of vehicles and background in trajectories shown in stacked 1-D-profiles along the time axis. The position and tangent of curves show the motion information with respect to the camera.

V. Motion-Based Identification of Target

A. Motion Behaviors of Scenes in Different Traces

The motion properties in the spatiotemporal image are the key evidence in our vehicle identification. The background motion is determined by the camera ego-motion and its distance. As the forward-looking camera undergoes translation in the Z-direction, we can obtain the condensed temporal image composed of 1-D profiles. In such spatiotemporal representations illustrated in Fig. 15, 1) backgrounds pursue hyperbolic trajectories expanding from the projection of FOE (features move sideways in the view with increasing image velocities), 2) the curvature of a trajectory is high for objects close to the road and is low (flat trace) for objects far from the road, and 3) the image velocity is proportional to the image position $x(t)$. The image velocity is high at scenes passing by and is low at scenes ahead. On the other hand, the vehicles tracked within the road may stay in the image frame even if they drive irregularly in a zigzag way.

How to classify traces to vehicles and background? By tracking the trajectories for a short while, our method will be able to identify the traces with certain probabilities. This is done by computing the probability at each moment using the HMM based on (x,v) sequences. Continuous tracking of traces further enhances the confidence of the decision. By formulating the classification problem in a probabilistic framework, we can avoid the accumulation of nonlinear decisions from multiple thresholds in segmentation and tracking and, thus, improve the quality of vehicle identification.

B. Computing Posterior Probability Using HMM

After initially obtaining the likelihoods $p(x,v|B)$ and $p(x,v|C)$ in two tables, we can now estimate the identities of features based on their motion behaviors. The posterior probabilities of states under observation $(x(t), v(t))$ are denoted by $\mathrm{P}(C_t|x(t), v(t))$ and $\mathrm{P}(B_t|x(t), v(t))$, respectively, or $\mathrm{P}(C_t)$ and $\mathrm{P}(B_t)$ for short. At any time t, we should keep

$$\mathrm{P}(C_t) + \mathrm{P}(B_t) = 1. \tag{19}$$

The probability of state transition from frame $t-1$ to frame t is defined as

$$\begin{array}{ll} \mathrm{P}(C_t|B_{t-1}) = 0.5 & \mathrm{P}(B_t|B_{t-1}) = 0.5 \\ \mathrm{P}(C_t|C_{t-1}) = 0.8 & \mathrm{P}(B_t|C_{t-1}) = 0.2. \end{array} \tag{20}$$

The transition from car to car, i.e., $\mathrm{P}(C_t|C_{t-1}) = 0.8$, emphasizes the continuity of car motion. Once a car is detected, it will not be missed easily. A background may be identified as a car later, i.e., $\mathrm{P}(C_t|B_{t-1}) = 0.5$, because there may not be a strong clue to determine a car in the image region near the FOE, where both background and vehicle have a small image velocity. When a trajectory is initially detected $(t = 0)$, its probabilities as a car and background are set empirically as $\mathrm{P}(C_0) = 0.7$ and $\mathrm{P}(B_0) = 0.3$ according to our detectability of vehicle features in the weighted image area.

Using Viterbi algorithm [5] in the HMM, the probability of a trace to be assigned as car at time t is optimized as

$$\begin{aligned} \mathrm{P}(C_t) = \max \big[& \mathrm{P}(B_{t-1})\mathrm{P}(C_t|B_{t-1})\mathrm{p}\left(x(t), v(t)|C_t\right) \\ & \mathrm{P}(C_{t-1})\mathrm{P}(C_t|C_{t-1})\mathrm{p}\left(x(t), v(t)|C_t\right)\big]. \end{aligned} \tag{21}$$

In addition, the probability as background is

$$\begin{aligned} \mathrm{P}(B_t) = \max \big[& \mathrm{P}(B_{t-1})\mathrm{P}(B_t|B_{t-1})\mathrm{p}\left(x(t), v(t)|B_t\right) \\ & \mathrm{P}(C_{t-1})\mathrm{P}(B_t|C_{t-1})\mathrm{p}\left(x(t), v(t)|B_t\right)\big]. \end{aligned} \tag{22}$$

If $\mathrm{P}(C_t) > \mathrm{P}(B_t)$, then the trace is considered as a car at time t and as background otherwise. The identity of a trace is formally output or displayed 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. If a new trace is found, it assembles a new HMM. The calculated trace identity may be uncertain at the beginning due to lack of evidence. The probability will increase as the trace is constantly tracked and updated.

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

$$P(C_t) \leftarrow \frac{P(C_t)}{P(C_t) + P(B_t)}, \qquad P(B_t) \leftarrow \frac{P(B_t)}{P(C_t) + P(B_t)} \tag{23}$$

to avoid a quick decreasing of the $\mathrm{P}(C_t)$ and $\mathrm{P}(B_t)$ values to zero in (21) and (22), which is caused by multiplying a series of probability values less than 1.

VI. Experiments and Discussion

Because the introduced probability distributions of scenes 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 the forward direction, horizon, and the visible portion of self-car in video frame in advance. These parameters are invariant to the dynamic scenes during vehicle motion.

We examined our method on eight video shots lasting 1 h and 20 min collected from rural and urban roads under various illumination conditions. The videos consist of both night-time and day-time clips. We implemented our method in C++ using OpenCV on AVI video format. The likelihood maps were computed offline using Matlab.

The thresholds used in this paper are usually low to allow more candidates to be picked up in the processing; we avoid abrupt cutting off of features and traces in feature detection and tracking. Their classification and identification as real targets and noises are handled later probabilistically in the HMM. This solves the problem of threshold selection that drastically affects the results.

In feature detection, the length of a horizontal line segment is not very important, since we are counting numbers of line segments at each image location x and then finding peaks after smoothing; a long line may contribute everywhere horizontally and will not influence the peak locally. As one can observe, the traces in the profiles of line segments provide strong evidence of vehicle presence. On the other hand, the traces from intensity and corner profiles are much more precise and smooth. Hence, our strategy is to combine these features for a correct decision on vehicle width. By referencing to the left and right traces in $T(x, t)$ that bound a vehicle trace in $T_l(x, t)$, the width search of vehicles is narrowed in the video frame. Hence, a box can be located on a vehicle for display in every video frame.

In feature tracking using vertically condensed profiles, we need target vehicles to be present in a certain number of frames before we can conduct a successful identification. This number is calculated empirically depending on the targeted objects. In our tests, we defined the minimum tracking duration for a targeted vehicle as 50 frames (roughly 1 s of video). We remove the tracking history after a trace is lost for at least 20 frames. The minimum strength of the trace for thresholding purposes is set to the top 67 percentile; this will remove the noise and weak traces.

We do not use any real data to train parameters in the HMM; rather, we define the basic probability distribution according to the physical environment and derive the probabilities in images and video. Our assumptions of background and vehicle are reasonably general to cover a large scope of variations, which even includes mild turning on a curved path. We have used the parameters in Table I for likelihood computation; the resulting likelihood pdfs of the two events have major differences (see Fig. 12), although some parts of the pdf where $|x|$ is small are mutually overlapped. An observation that falls into this area may have a low certainty in the car and background classification, which has to be disambiguated after a period of time as their motions demonstrate sufficient differences.

TABLE I
PARAMETER SELECTION IN PROBABILITY COMPUTATION

	Parameters	Physical Environment	Values
D	Average road width	As wide as three lanes	6m
F	Distance to target	Minimum safe distance	10m
σ_F	Standard deviation of target distance		20m
σ_x	Standard deviation of relative horizontal speed T_x of target vehicle	Maximum cutting of three lanes, tolerant for moving on curved path	6m/s
σ_z	Standard deviation of relative translation speed T_z	T_z is zero if target is pursued	10m /s
S	Average pursuing speed of observer vehicle	50km/h	15m /s
σ	Standard deviation of the speed	10km/h	5m /s
f	Camera focal length	Through offline calibration	900 pixel
$\int_z$	Range for integration	From camera position to distance close to infinity	0~200m
$\int_x$	Range for integration	Wider than a road to include all backgrounds in video	-50~50m
$\int_{Tz}$	Range for relative speed		-40~40m/s
$\int_{Ry}$	Range of integration		-10~10 degree/s
H	The maximum height of vehicle	As high as a truck, but mostly for cars	4m
σ_r	Standard deviation of steering angle of R_y	From the maximum tuning radius of a vehicle and road curvature.	5 degree/s

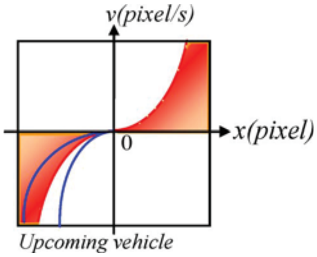

Fig. 16. Traces and motion of opposite vehicles bounded by blue curves.

The oncoming vehicles traveling on the opposite lanes are treated as background. In addition to their positions $|X|$ that are beyond the boundaries of the lanes of the observer vehicle, their relative velocities, which are denoted as V_{op}, are faster than that of the background $(-V)$. The traces of opposite vehicles will be even more slanted than the background traces in the condensed profile. As in Fig. 9, these traces will be located close to the v-axis in a strongly curved way, as illustrated in Fig. 16. Their likelihood of motion is not overlapped with that of the forward moving vehicles, except in the area where $|x|$ is small. This means that the oncoming vehicle is far away and, thus, small in the image; it has little influence on the safety tracking of vehicles. On the other hand, a stationary vehicle at roadside has the same behavior as the background and is treated as a static obstacle.

The pdf and parameters defined in this paper are wide to include all possible cases. If the ego-motion could be read from its mechanical/electric system, then the pdfs of many parameters would be narrowed down dynamically to proper ranges, and the likelihood maps will be precise for target tracking and identification. Fig. 17 shows examples of trace identification. Fig. 18 gives vehicle detection and tracking results. The program can successfully track the targets and is invariant to lighting condition. In Fig. 18(f), the tracking program locates two overlapping boxes on a single car as the scaling up of the car increases detailed line segments and then traces. Fig. 18(b) shows the presence of multiple distant cars moving together. The line segmenting algorithm recognizes them as a single group. As the vehicles separate their movements, the mixture of horizontal line segments has dispatched, and thus, the algorithm separates them. Moreover, the shadow with a vehicle body also creates edges in the video, and we just consider it to be a part of the vehicle.

We have obtained the confusion matrix shown in Table II. The classified result of each trace is compared with the truth in the video by humans and is verified according to the trace points

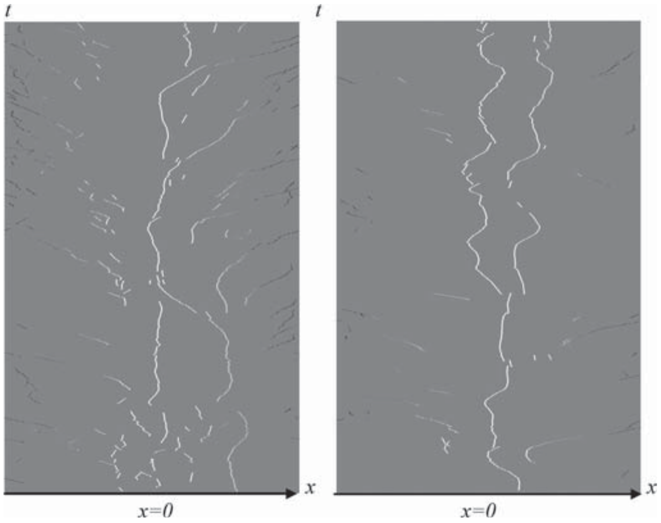

Fig. 17. Probabilities of traces displayed in gray level. Car: white. Background: black. No feature and uncertain traces: gray. A feature may be uncertain at the beginning of its trace and turn to be confident as it is tracked for a long time.

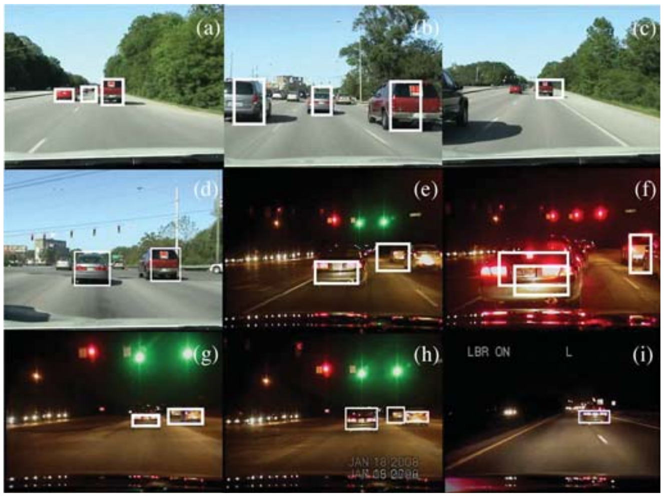

Fig. 18. Vehicle detection and tracking results.

TABLE II
CONFUSION MATRIX OF CAR AND BACKGROUND IDENTIFICATION

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

after the initial uncertain period determined by the minimum number of frames for identification. In most cases of correct detection, the duration of tracking lasts as long as the vehicle remains in the scene. If a detected vehicle moves too far from the observer car, the tracking is stopped.

The limitations of this paper are as follows: 1) Although the method functions well in mild driving conditions, more complex pdfs for unsteady motion should be investigated under drastic driving conditions. 2) Distant vehicles are difficult to detect because of their small sizes in the vertical profiling, although they have no influence on safety driving. The loss of target tracking due to a far distance can still be picked up if the target gets close again. 3) The necessity in collecting motion evidence for a period of time when the observer vehicle has a reasonably high speed. The longer we track an object, the higher the probability of a correct identification is by the HMM. Although HMM provides probability from the very beginning as a trace is detected, it will be uncertain if the time for displaying motion tendency is insufficient. This time depends on the vehicle speed and how wide the environment is. The distant background in the forward direction requires a long time to classify. Fortunately, they are far away and are mostly ignored due to their small sizes in images or are not urgent obstacles to avoid. An examination of their scale changes can help the decision making. Even if the observer vehicle moves fast, wide scenes (large $|X|$) may still be uncertain for a while due to their low image velocities, according to (8). Such a case will not affect safety either. In contrast, close-by objects and narrow environments usually show fast image velocities during the ego-motion, which can be identified quickly.

VII. CONCLUSION

This paper has focused on an important task to detect and track vehicles ahead with an in-car video camera. Our approach is mainly based on motion information. Several general features that characterize the vehicles ahead are robustly extracted in the video. Their positions are profiled vertically to yield condensed profiles to speed up vehicle tracking and identification. Feature motions are displayed as observable traces. We introduced the HMM to vehicle identification during tracking such that the decision follows a probability framework that is less influenced by ad hoc thresholds. The joint probability of image positions and the velocities of traces are calculated for the HMM to dynamically separate target vehicles from background. The use of temporal motion coherence of features enhanced the identification and tracking of vehicles. Experimental results show the effectiveness of the system design and implementation. The computation is implemented in real time and is easy to embed into a hardware for real vehicle-borne video.

ACKNOWLEDGMENT

The authors would like to thank H. Blitzer for his contribution in the project.

REFERENCES

[1] H. Takizawa, K. Yamada, and T. Ito, "Vehicles detection using sensor fusion," in *Proc. IEEE Intell. Vehicle*, 2004, pp. 238–243.

[2] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," in *Proc. IEEE CVPR*, 2000, pp. 746–751.

[3] R. Lakaemper, S. S. Li, and M. Sobel, "Correspondences of point sets using Particle Filters," in *Proc. ICPR*, Dec. 2008, pp. 1–5.

[4] M. Betke and H. Nguyen, "Highway scene analysis from a moving vehicle under deduced visibility conditions," in *Proc. IEEE Intell. Vehicle*, 1998, pp. 131–136.

[5] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[6] X. Huang, M. Jack, and Y. Ariki, *Hidden Markov Models for Speech Recognition*. Edinburgh, U.K.: Edinburgh Univ. Press, 1990.

[7] J. Chu, L. Ji, L. Guo, B. Li, and R. Wang, "Study on method of detecting preceding vehicle based on monocular camera," in *Proc. IEEE Intell. Vehicle*, 2004, pp. 750–755.

[8] D. Alonso, L. Salgado, and M. Nieto, "Robust vehicle detection through multidimensional classification for on board video based systems," in *Proc. IEEE ICIP*, Sep. 2007, vol. 4, pp. 321–324.

[9] P. Parodi and G. Piccioli, "A feature-based recognition scheme for traffic scenes," in *Proc. IEEE Intell. Vehicle*, 1995, pp. 229–234.

[10] C. Hoffman, T. Dang, and C. Stiller, "Vehicle detection fusing 2D visual features," in *Proc. IEEE Intell. Vehicle*, 2004, pp. 280–285.

[11] L. Gao, C. Li, T. Fang, and Z. Xiong, "Vehicle detection based on color and edge information," in *Proc. Image Anal. Recog.*, vol. 5112, *Lect. Notes Comput. Sci.*, 2008, pp. 142–150.

[12] J. Dubuisson, S. Lakshmanan, and A. K. Jain, "Vehicle segmentation and classification using deformable templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 293–308, Mar. 1996.

[13] T. N. Tan and K. D. Baker, "Efficient image gradient based vehicle localization," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1343–1356, Aug. 2000.

[14] Y. Guo, Y. Shan, H. Sawhney, and R. Kumar, "PEET: Prototype embedding and embedding transition for matching vehicles over disparate viewpoints," in *Proc. IEEE CVPR*, 2007, pp. 17–22.

[15] C. R. Wang and J.-J. Lien, "Automatic vehicle detection using local features—A statistical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 83–96, Mar. 2008.

[16] H. Cheng, N. Zheng, and C. Sun, "Boosted Gabor features applied to vehicle detection," in *Proc. 18th ICPR*, 2006, vol. 2, pp. 662–666.

[17] W. Zhang, X. Z. Fang, and X. K. Yang, "Moving vehicles segmentation based on Bayesian framework for Gaussian motion model," *Pattern Recognit. Lett.*, vol. 27, no. 9, pp. 956–967, Jul. 2006.

[18] Z. Zhu, H. Lu, J. Hu, and K. Uchimura, "Car detection based on multicues integration," in *Proc. 17th ICPR*, 2004, vol. 2, pp. 699–702.

[19] J. Kato, T. Watanabe, S. Joga, J. Rittscher, and A. Blake, "An HMM-based segmentation method for traffic monitoring movies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1291–1296, Sep. 2002.

[20] A. Bruss and B. K. P. Horn, "Passive navigation," *Comput. Vis. Graph. Image Process.*, vol. 21, no. 1, pp. 3–20, Jan. 1983.

[21] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vis. Conf.*, 1988, pp. 147–151.

[22] G. Flora and J. Y. Zheng, "Adjusting route panoramas with condensed image slices," in *Proc. ACM Conf. Multimedia*, Augsburg, Germany, 2007, pp. 815–818.

[23] J. Y. Zheng, Y. Bhupalam, and H. Tanaka, "Understanding vehicle motion via spatial integration of intensities," in *Proc. 19th ICPR*, Dec. 2008, pp. 1–5.

[24] D. Stirzaker, *Elementary Probability*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[25] C. Demonceaus, A. Potelle, and D. Kachi-Akkouche, "Obstacle detection in a road scene based on motion analysis," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1649–1656, Nov. 2004.

[26] S. Sivaraman and M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 267–276, Jun. 2010.

[27] N. Ghosh and B. Bhanu, "Incremental unsupervised three-dimensional vehicle model learning from video," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 423–440, Jun. 2010.

Amirali Jazayeri received the B.S. degree from the University of Indianapolis, Indianapolis, IN, and the M.S. degree from Indiana University–Purdue University Indianapolis in 2010.

He is currently with Precise Path Robotics, IN, as a software engineer and researcher. He has been working on video processing and human motion understanding in the project.

Hongyuan Cai (M'10) received the B.E. degree in computer science and technology from Beijing Forestry University, Beijing, China, in 2007 and the M.S. degree in computer science from Indiana University–Purdue University Indianapolis (IUPUI), Indianapolis, IN, in 2009. He is currently pursuing the Ph.D degree with the Department of Computer and Information Science, IUPUI.

His research interest includes image, video, multimedia, and Geo-spatial data analysis, computer vision, and spatial-temporal analysis of video and its visualization. His current research focus includes scene representation for indoor and urban environments and video analysis and visualization.

Jiang Yu Zheng (M'90–SM'05) received the B.S. degree from Fudan University, Shanghai, China, in 1983 and the Ph.D. degree from Osaka University, Osaka, Japan, in 1990.

He is currently a Professor with the Department of Computer Science, Indiana University–Purdue University Indianapolis, Indianapolis, IN. His research interests include image and video processing, computer vision, multimedia, virtual reality, robotics, and digital forensics. He has published over 100 research papers as a primary author in journals and refereed international conferences in these areas.

Dr. Zheng received best paper awards from the Information Processing Society of Japan in 1991 and the ACM Virtual Reality Software and Technology Award in 2004 for creating the first digital panorama and scene tunnel images, respectively.

Mihran Tuceryan (M'86–SM'05) received the B.S. degree from the Massachusetts Institute of Technology, Cambridge, in 1978 and the Ph.D. degree from the University of Illinois at Urbana-Champaign, in 1986.

He is currently an Associate Professor with the Department of Computer Science, Indiana University–Purdue University Indianapolis, Indianapolis, IN. His research interests include computer vision and visualization, augmented reality, and pattern recognition.