Overtaking Vehicle Detection Using Dynamic and Quasi-Static Background Modeling

Junxian Wang[1], George Bebis[1] and Ronald Miller[2]
[1] Computer Vision Laboratory, University of Nevada, Reno, NV
[2]Vehicle Design R&A Department, Ford Motor Company, Dearborn, MI
(junxian,bebis)@cse.unr.edu, rmille47@ford.com

***Abstract*— Robust and reliable detection of overtaking vehicles is an important component of any on-board driver assistance system. Optical flow, with the abundant motion information present in image sequences, has been studied extensively for vehicle detection. However, using dense optical flow for vehicle detection is sensitive to shocks and vibrations of the mobile camera; image outliers caused by illumination changes; and high computational complexity. To improve vehicle detection performance and reduce computational complexity, we propose an efficient and robust methodology for overtaking vehicle detection based on homogeneous sparse optical flow and eigenspace modeling. Specifically, our method models the background into dynamic and quasi-static regions. Instead of using dense optical flow to model the dynamic parts of the background, we employ homogeneous sparse optical flow, which makes detection more robust to camera shocks and vibrations. Moreover, to make detection robust to illumination changes, we employ a block-based eigenspace approach to represent quasi-static regions in the background. A region-based hysteresis-thresholding approach, augmented by a localized spatial segmentation procedure, attains a good tradeoff between true detections and false positives. The proposed methodology has been evaluated using challenging traffic scenes illustrating good performance.**

I. Introduction

Overtaking vehicle detection is an important component of any on-board driver assistance system. It can be used to alert the driver about driving conditions, possible collision with other vehicles, or trigger the automatic control of the vehicle for collision avoidance and mitigation. Overtaking vehicle detection based on active sensors such as laser, radar, and sonar has several drawbacks due to sensor interferences between different vehicles in a limited space. Passive sensor-based detection approaches, such as vision-based methods, are becoming widely used due to their low cost and less interferences between vehicles.

In vision-based overtaking vehicle detection systems, a single camera is usually mounted on the host vehicle to capture rear-view image sequences. Various approaches have been proposed to detect moving vehicles assuming dynamic background [1]. These methods can be classified into two main categories: appearance-based [2][3][4][5] and motion-based [6][7]. Since overtaking vehicles might be partially occluded and need to be detected as soon as possible (i.e., before entering the blind spot of the host vehicle) appearance-based methods are not very appropriate [8]. In this case, the relative motion information obtained via the calculation of optical flow becomes an important cue for detecting moving vehicles. A common approach for overtaking vehicle detection using optical flow is to compare a predicted flow field, calculated by projecting vehicle velocity in 2D, with the actual image flow, calculated from motion estimation [9][10][11]. From a practical point of view, overtaking vehicle detection using optical flow has the following three difficulties: (1) noise due to camera shocks and vibrations will cause errors in the computation of the temporal derivatives; (2) lack of texture in the road regions and small gray-level variations introduce significant instabilities in the computation of the spatial derivatives and (3) structured noise and strong illumination changes cause spurious image features and unreliable flow estimates. Given these inherent difficulties, getting reliable dense optical flow estimates is not an easy task for overtaking vehicle detection. Some researchers have tried to improve the performance of optical flow computation methods by introducing different techniques. Zhu [8] used variable bandwidth density fusion with dynamic scene modeling to achieve reliable flow estimation for passing vehicle detection and estimate the trajectory of the moving vehicle. However, their system uses a forward-facing CCD camera, which makes it difficult to observe overtaking vehicle in the blind spot. Moreover, it has high time requirements due to using the mean-shift algorithm to iteratively compute the dominant motion based on traditional dense optical flow.

To address the above difficulties, we propose an overtaking vehicle detection method based on homogeneous sparse optical flow. Our method models the background into dynamic regions (i.e., abundant texture such as passing trees) and a quasi-static regions (i.e., lack of texture such as road regions). Instead of using dense optical flow, we employ homogeneous sparse optical flow to model the dynamic background, which allows detection to be robust to camera shocks. Also, a block-based eigenspace approach is used to model the quasi-static background, providing good robustness to illumination changes. To determine candidate overtaking vehicles in an image, we apply background subtraction. Then, we classify each candidate (i.e., overtaking vs passing) by calculating its dominant motion. This approach reduces the influence of unreliable motion outliers. Our experimental results show that the proposed method yields reliable dynamic scene analysis and has good detection performance. In addition, it is robust to changes in lighting conditions and robust to partial occlusion. Compared to traditional optical flow methods, sparse optical

flow has lower time requirements.

The rest of the paper is organized as follows: Section II provides a brief overview of the proposed methodology. Section III presents our approach for dynamic and quasi-static background modeling using homogeneous sparse optical flow and block-based eigenspace analysis, respectively. In Section IV, we discuss the process for generating the target candidates, removing false positives due to noise, and classifying targets using dominant motion analysis. Finally, our experimental results are presented in Section V while our conclusions and directions for future research are given in Section VI.

II. Overtaking vehicle detection: Methodology

The proposed methodology for overtaking vehicle detection consists of four phases: (i) scene modeling, (ii) vehicle detection, (iii) vehicle tracking, and (iv) velocity verification. The block diagram of our algorithm is shown in Figure 1.

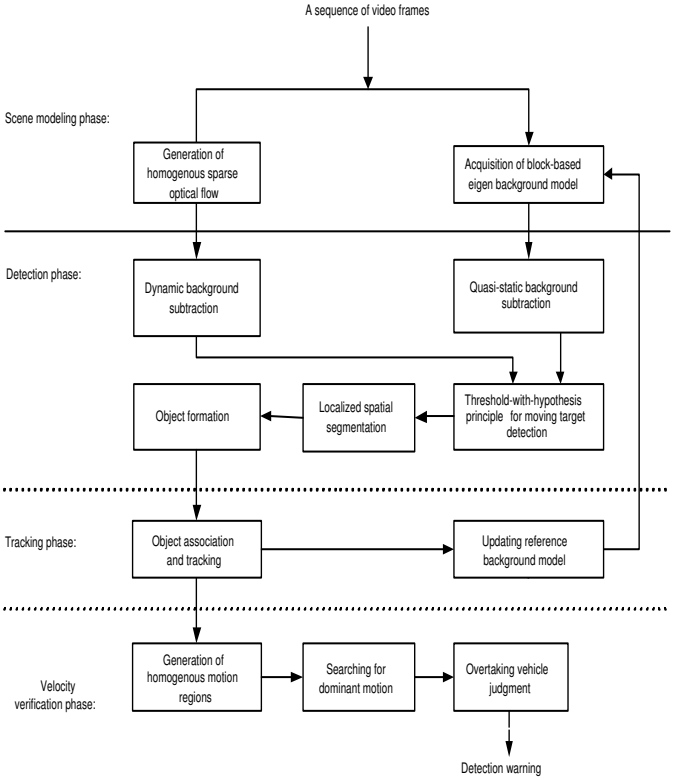


Fig. 1. The proposed overtaking vehicle detection algorithm.

In the scene modeling phase, the dynamic and quasi-static background are modeled using a homogeneous block-based (i.e., sparse) optical flow and block-based eigenspace analysis respectively. In the detection phase, the moving candidates are segmented out from the background. This phase consists of two steps: (i) generation of candidate solutions and (ii) removing false positives due to noise. To generate the overtaking vehicle candidates, we subtract the dynamic and quasi-static backgrounds from the current frame and threshold the results. To determine a suitable threshold, we have adopted a region-based hysteresis-thresholding approach, followed by a localized spatial segmentation process. This approach has shown to be very effective in remove false positives. Finally, we apply connected component analysis to form the candidate solutions. In the tracking phase, candidate solutions are associated between frames and tracked over time. Finally, in the velocity verification phase, the motion distribution of the moving targets is statistically analyzed using the x and y motion directions and the amplitude of the optical flow. Information about the dominant motion of a moving target is useful for issuing appropriate warnings depending on the status of the vehicle (e.g., when a vehicle is in the blind spot of the host vehicle).

III. Scene modeling

In this paper, rear-view traffic scene images are segmented into three main regions: (i) dynamic background, (ii) quasi-static background, and (iii) moving targets. Usually, texture abundant large scale background moves consistently in the field of view as the camera moves along with the ego-vehicle. We call this type of background as “dynamic background”. In contrast to dynamic background which exhibits consistent motion, special types of background (e.g., road, sky) show only small gray-level variations due to lack of texture and behave as “quasi-static” background relative to the host vehicle. To model the dynamic background of a scene, we propose using a homogeneous sparse optical flow approach which also reduces noise interferences due to camera shocks. Using a similar approach to model the quasi-static background would lead to significant instabilities due to small gray-level variations. Therefore, we use a block-based eigenspace approach which also provides a representation less sensitive to illumination changes.

A. Dynamic background modeling

As mentioned in the previous section, the dynamic background of a scene moves consistently in the field of view as the camera moves along with the ego-vehicle. Detecting moving targets in this case becomes more complex since the image motion caused by the moving targets can be confused with the dynamic background motion. Optical flow estimation has become an indispensable component for many computer vision applications where the objective is to extract the prime source of motion of moving targets. However, it is not always easy to extract reliable velocity differences between moving targets. On the other hand, it is usually more reliable to detect dominant velocity differences between the boundary of a moving target and the dynamic background. Therefore, we have decided to model the motion of the dynamic background instead of modeling the motion of the moving targets.

It is well known that the optical flow vectors of the dynamic background have opposite direction with respect to the motion of the ego-vehicle. However, in a real road scene, the direction of the estimated optical flow of the dynamic background is not only caused by the ego-motion of the vehicle, but also by shocks and vibrations experienced by the vehicle. In addition,

dense optical flow vector computations are time quite consuming when the resolution of the image is high. To deal with these issues, we model the dynamic background by considering homogeneous regions of "*sparse optical flow*". To improve robustness, we only consider the horizontal component of "sparse" optical flow field and discard the vertical component. This is because most vertical deviation errors of the optical flow vectors are caused by the vertical shocks of the camera which results in unreliable motion estimates [6].

It should be mentioned that, vehicle vibrations introduce high frequency components in the image motion. This represents a significant source of error in the computation of the temporal derivatives of optical flow [6]. Usually, spatial smoothing is be used to reduce the effect of such high frequency components. In our case, we compute the sparse optical field using block-based spatial smoothing on low resolution (i.e., sub-sampled) images. The proposed sparse optical flow scheme reduces computation time while suppressing spatial and temporal derivative computation errors due to camera shocks and vibrations.

Modeling the dynamic background of a scene involves the following three steps: (i) image sub-sampling using a block-based approach and sparse optical flow computation, (ii) ordering the optical flow vectors by thresholding the angle difference between vectors within neighboring blocks at every spatial position, and (iii) clustering the ordered sparse optical flow field vectors into homogeneous regions and excluding disordered optical flow vectors.

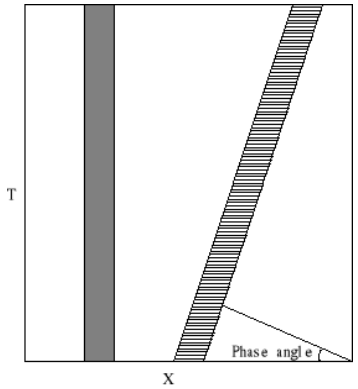


Fig. 2. Different phase angles of a gray-level distribution of a moving and a stationary target with time. The vertical bar is the gray-level distribution of the stationary target with time. The slope bar is the gray-level distribution of the moving target with time. The phase angle is the angle between the x-axis and the normal to the slope bar.

A.1 Estimation of sparse optical flow

A wide range of techniques have been developed for optical flow field estimation including differential, matching, energy-based, and phase-based approaches. Fleet and Jepson [12] have shown that the temporal evolution of contours of constant phase provides a better approximation to local velocity. Figure 2 shows different phase angles of the gray-level distribution of a moving and a stationary target with time [13]. Phase contours are more robust with respect to smooth shading and lighting variations, and more stable with respect to small deviations due to image translations. Using phase-based optical flow, we have developed a sparse optical flow technique to model the dynamic background for overtaking vehicle detection.

Let us consider a sequence of q frames (i.e., from $(t-q-1)$th to tth), each with size $H \times L$. Also, let us denote the intensity value of the tth frame at spatial position (n_1, n_2) as $S_p(n_1, n_2, t)$. Each frame is divided into $N_H \times N_L$ non-overlapping square blocks with size $s \times s$, where $N_H = H/s$ and $N_L = L/s$. Let $S_b(x_1, x_2, t)$ be a block-based image which is obtained by sub-sampling frame t as follows:

$$S_b(x_1, x_2, t) = \frac{1}{s^2} \sum_{k=-s/2}^{s/2} \sum_{l=-s/2}^{s/2} S_p(n_1 + k, n_2 + l, t). \quad (1)$$

where $1 < x_1 < N_H$, $1 < x_2 < N_L$ and $S_p(n_1 + k, n_2 + l, t)$ is among the eight neighboring pixels around $S_p(n_1, n_2, t)$, $x_1 \times s < n_1 + k < (x_1 + 1) \times s$ and $x_2 \times s < n_2 + l < (x_2 + 1) \times s$.

Let $\phi(x_1, x_2, t)$ denote the phase responses of the q block-based image frames $S_b(x_1, x_2, t-q-1), \cdots, S_b(x_1, x_2, t)$ which are obtained by spatially filtering each block-based image with a set of quadrature filter pairs. Assuming phase constancy [12], in a small region with the motion field satisfied, $\phi(x_1, x_2, t) = c$. Differentiating with respect to t, we have:

$$\frac{d\phi(x_1, x_2, t)}{dt} = 0. \quad (2)$$

The phase-based optical flow vector at a given image location is computed by solving the linear equation:

$$\nabla\phi(x, t) \cdot (v, 1) = (\phi_x, \phi_t) \cdot (v, 1) = 0. \quad (3)$$

where $\nabla\phi(x,t) \doteq [\frac{\partial\phi(x,t)}{\partial x_1} \ \frac{\partial\phi(x,t)}{\partial x_2} \ \frac{\partial\phi(x,t)}{\partial t}] = (\phi_x, \phi_t)$, $v = [\frac{dx_1}{dt}, \frac{dx_2}{dt}]$, and $\phi_x = [\phi_{x_1}, \phi_{x_2}]$ is the spatial phase gradient with respect to x_1 axis and x_2 axis. Also, ϕ_t is the temporal phase gradient and $< \cdot >$ denotes vector inner product.

From equation (3), due to the aperture problem, we can only estimate the component of the optical flow vector which is in the direction of the spatial phase gradient $\phi_x / \|\phi_x\|$. Then, the normal flow $v_\perp(x_1, x_2)$ can be rewritten as:

$$v_\perp(x_1, x_2) = -\frac{\phi_t}{\|\phi_x\|}. \quad (4)$$

The above formulation assumes a single phase angle in the spatio-temporal image sequence. However, in practice, the image sequence is complex and includes different phase variations of the gray-level distributions. Thus, we would need to decompose the image sequence into different frequencies by applying a set of spatial filters at every frame. Here, we use quadrature Gabor filter pairs [14]. These filters are characterized by their center frequencies, (f_{x_1}, f_{x_2}). It should be noted that, all non-zero frequency components associated with the moving profile must lie on a line through the

center frequencies (f_{x_1}, f_{x_2}) in the frequency domain. Then, equation(4) can be rewritten as

$$v_{\perp}(x_1, x_2) = -\frac{\phi_t}{\|\phi_x\|} = -\frac{\phi_t}{2\pi(f_{x_1}^2 + f_{x_2}^2)}(f_{x_1}, f_{x_2}). \quad (5)$$

where the spatial phase gradient ϕ_x is substituted by the frequency vector $(2\pi/f_{x_1}, 2\pi/f_{x_2})$. The temporal phase gradient, ϕ_t is computed from the temporal sequence of its phase components by performing a least-squares linear regression on the (t, ϕ)-pairs. For more elaborate temporal phase gradient techniques, please refer to [12][14].

The proposed "sparse optical flow" $v_s(x_1, x_2)$ can be computed by keeping only the horizontal component of optical flow field:

$$v_s(x_1, x_2) = -\frac{\phi_t}{2\pi(f_{x_1}^2 + f_{x_2}^2)} f_{x_1}. \quad (6)$$

This reduces the effect of the vertical gradient error of the optical flow caused by vertical shocks of the camera.

A.2 Homogeneous sparse optical flow

Optical flow vectors with high similarity can be assigned into a homogeneous optical flow region. In order to extract the homogeneous regions, we need to establish a measure of similarity between optical flow vectors. Here, we use the angle between the x-axis and the sparse optical flow vectors. We call the regions composed by similar flow vectors *homogeneous sparse optical flow regions*.

Let $\psi(x_1, x_2)$ be the minimum angle difference between an optical flow vector $v_s(x_1, x_2)$ and its corresponding neighboring vectors $v_s(x_1 + k, x_2 + l)$, where k and l are the relative spatial positions of the neighboring sparse optical flow vectors:

$$\psi(x_1, x_2) = \min_{k,l} \arccos \frac{v_s(x_1, x_2) * v_s(x_1 + k, x_2 + l)}{\sqrt{v_s^2(x_1, x_2) + v_s^2(x_1 + k, x_2 + l)}} \quad (7)$$

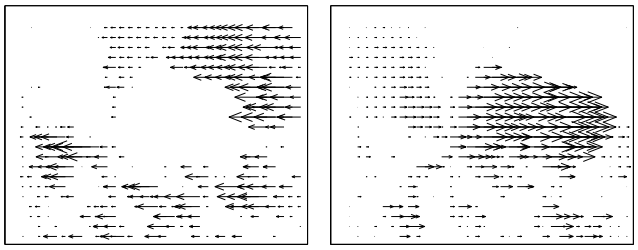
By considering the minimum angle difference among neighboring sparse optical flow vectors, we can capture a strong spatial correlation among the vectors in the boundary of foreground and moving target. This is because the angle differences of optical flow vectors between the moving target and the dynamic background is larger than those within a moving target. Therefore, the dynamical background $B(x_1, x_2)$ is extracted by

$$B(x_1, x_2) = \begin{cases} 1 & \text{dynamical background;} \quad \psi(x_1, x_2) = \pi; \\ 0 & \text{otherwise;} \quad \psi(x_1, x_2) = 0. \end{cases}$$

Figure 3 shows a comparison between traditional dense optical flow field and the proposed sparse optical flow field to model the dynamic background. Figure 3 (a) shows a sample scene. Figure 3 (b) shows the result of dynamic background subtraction. The dynamic background is marked by white color. The sparse optical flow vectors are shown in figure 3 (c) and (d). Figure 3 (c) shows clearly that the direction of sparse optical flow (i.e., right to left) describes the dynamic background accurately. Figure 3 (d) shows the regions having optical flow directions opposite to those modeling the dynamic background. These regions include both moving targets and quasi-static background. It should be mentioned that, it is difficult to directly distinguish these regions by simply modeling the motion of the moving targets since the two moving vehicles in this example have different optical flow fields. Moreover, the traditional optical flow approach is sensitive to noise while the flow vectors demonstrate a disordering as shown in Figures 3 (e) and (f). Therefore, the use of sparse optical flow allows representing the dynamic background more effectively and has lower computational requirements.

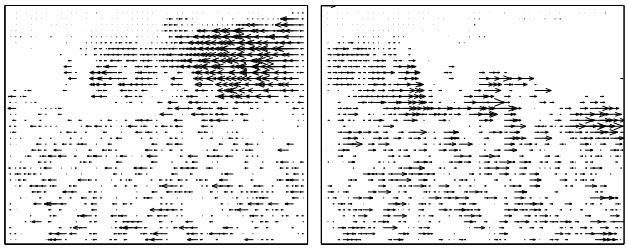
(a) Sample traffic scene.

(b) Dynamic background subtraction based on homogeneous sparse optical flow.



(c) Modeling the dynamic background using sparse optical flow having uniform direction from right to left.

(d) Sparse optical flow having uniform direction from left to right.



(e) Traditional phase-based optical flow with direction from right to left .

(f) Traditional phase-based optical flow with direction from left to right.

Fig. 3. Comparison of dynamic background modeling based on sparse optical flow and traditional dense optical flow.

B. Quasi-static background modeling

The quasi-static background model is used to represent special types of background that lack texture information and

have small gray-level variations over time. Small gray-level variations introduce significant instabilities in the computation of the spatial derivative errors which could increase the number of false positives. We have developed a block-based eigen-background prediction mechanism to address this issue. In this approach, the quasi-static background is statistically represented by a set of basis vectors computed from the n latest frames. The basis vectors retain the dominant information in the observed data. Changes in global and local illumination can be accounted by continuously updating the basis vectors over time.

B.1 Block-based eigen-background

Considering the strong correlation between neighboring pixels, we divide each image into a fixed set of blocks and represent the quasi-static background in each block using a set of eigenvectors. Such a representation enables capturing the global information of the gradually evolving background. The number of eigenvectors kept determine the amount of information preserved in each block. It should be mentioned that it is common in the literature to employ mixtures of Gaussians to model the background scene by estimating the variance of the observed data. In this study, eigenvectors are used to approximately estimate the dominant directions of variance.

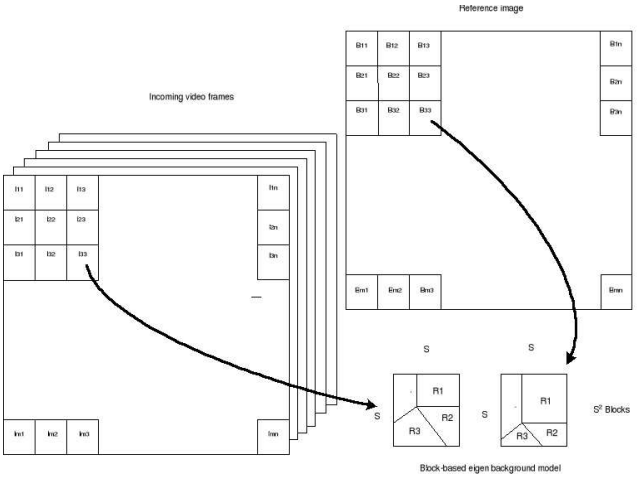


Fig. 4. The block diagram of the proposed quasi-static background model.

Figure 4 illustrates the main steps of the proposed quasi-static background model. Incoming video frames and reference frame are divided into blocks with index (x_1, x_2). Let $S_p(n_1, n_2, t)$ be the vector (i.e., gray or color) of the tth incoming video frame at position (n_1, n_2). Let $S_p(n_1, n_2)$ denote a pixel set at spatial location (n_1, n_2) from time $t-q-1$ to time t, where $S_p(n_1, n_2) = \{S_p(n_1, n_2, t-q), \cdots, S_p(n_1, n_2, t)\}$. The mean vector of $S_p(n_1, n_2)$ is computed by $\bar{S}_p(n_1, n_2) = \frac{1}{q} \sum_{r=t-q-1}^{t} S_p(n_1, n_2, r)$.

Each block in the reference image is modeled from the last q observed block scenes $S_b(x_1, x_2)$ with size $s \times s$. Formally, $S_b(x_1, x_2)$ is an $s \times s \times n$ array containing the vectors of each pixels in a fixed block from time $t-q-1$ to t, i.e., $S_b(x_1, x_2) = \{S_p(x_1, x_2, t-q-1), \cdots, S_p(x_1, x_2, t)\}$, where $S_b(x_1, x_2, t) = \{S_p(n_1, n_2, t), ..., S_p(n_1+k, n_2+l, t)\}$. The mean vector of these vectors in the each block is given by:

$$\bar{S}_b(x_1, x_2) = [\bar{S}_p(n_1, n_2), \cdots, \bar{S}_p(n_1+k, n_2+l)]^T. \quad (8)$$

Thus, the zero mean vectors of each block scene $S_b(x_1, x_2, t)$ are obtained by $\hat{S}_b(x_1, x_2, t) = S_b(x_1, x_2, t) - \bar{S}_b(x_1, x_2)$. The covariance over q observed block scenes is given by:

$$C = \frac{1}{q} \sum_{r=t-q-1}^{t} \hat{S}_b(x_1, x_2, t)[\hat{S}_b(x_1, x_2, t)]^T \quad (9)$$

Every quasi-static background scene at a fixed block (x_1, x_2), $Q(x_1, x_2, t)$, can be modeled by the eigenvectors $\{R^1(x_1, x_2), \cdots, R^n(x_1, x_2)\}$ of the covariance C and their corresponding coefficients $W(x_1, x_2, t) = \{\omega^1(x_1, x_2, t), \cdots, \omega^n(x_1, x_2, t)\}$:

$$Q(x_1, x_2, t) = \sum_{k=1}^{K} \omega^k(x_1, x_2, t) R^k(x_1, x_2). \quad (10)$$

where $K \ll n$ (i.e., in our experiments, we set K=3).

B.2 Updating quasi-static background model

To make the quasi-static background model robust, it is necessary to update the eigenvectors over time to account for global changes in the environment (i.e., illumination changes). Here, we keep information about each block in the image over a time interval T (e.g., 5 frames). The eigen-background model can be updated by removing the old frames and adding the new ones using standard Singular Value Decomposition (SVD) or incremental SVD [15].

IV. Overtaking vehicle detection, tracking, and velocity verification

Overtaking vehicle detection based on a mobile camera faces many challenges including that the motion caused by a moving target can be confused with the motion of the background due to camera motion. Several methods have been investigated to address this issue including modeling the motion distribution of the moving target and the mobile camera separately. These methods needs to accumulate the energy of the moving target over multi-frames while suppressing noise.

In general, it is relatively easier to represent the motion of the background, however, the motion distribution of moving target could be very difficult to capture in a real scenario. In this paper, we segment out the overtaking vehicles by subtracting the dynamic and quasi-static background from each frame, using a region-based hysteresis-thresholding strategy. In order to classify the status of overtaking vehicles, we estimate the dominant motion of each overtaking vehicle. Since it has been observed that pixel-based velocities are unreliable at the motion boundaries of the motion field [16], we search for the dominant motion in the homogeneous region of the sparse optical field associated with the vehicle.

A. Overtaking vehicle detection

To detect moving vehicles, we project each block in the incoming video frame onto the eigenvectors corresponding to that block location and we subtract the eigen-coefficients from the reference frame. Specifically, given a block (x_1, x_2) in the incoming frame, we define the discrepancy between the incoming block and the quasi-background as follows:

$$D(x_1, x_2, t) = \min_{r,k,l} \parallel W(x_1, x_2, t) - W(x_1 + k, x_2 + l, r) \parallel$$

where $W(x_1, x_2, t)$ corresponds to the eigen-coefficients computed by projecting the incoming scene onto the quasi-static background $E(x_1, x_2) = \{e^1(x_1, x_2), \cdots, e^K(x_1, x_2)\}$. $W(x_1 + k, x_2 + l, r)$ is the eigen-coefficients of the r quasi-static background scene at the spatial block $(x_1 + k, x_2 + l)$. This computation is performed between the incoming block (x_1, x_2) and each background blocks $(x_1 + k, x_2 + l)$ within a small window. Eventually, we pick the smallest difference to generate a binary map $M(x_1, x_2, t)$, indicating the presence of overtaking vehicles:

$$M(x_1, x_2, t) = \begin{cases} 0 & background \quad ; \quad D(x_1, x_2, t) < \alpha, \\ 1 & foreground \quad ; \quad otherwise. \end{cases} \tag{11}$$

Choosing a fixed threshold to decide whether there is some significant change within a block often leads to miss-detections and miss-classifications. Here, we have adopted a region-based hysteresis-thresholding strategy based on two thresholds [17]. Based on this strategy, a low and a high threshold, denoted by α_h and α_l, are used. Accordingly, two binary maps, $M_h(x_1, x_2, t)$ and $M_l(x_1, x_2, t)$, are generated using equation (11). A coarser resolution binary map $\hat{M}_h(x_1, x_2, t)$ is then obtained by dividing $M_h(x_1, x_2, t)$ into blocks and by counting all pixels labeled as '1' in each block. Then, a pixel in $\hat{M}_h(x_1, x_2, t)$ is labeled as '1' only if the majority of pixels in the corresponding block from $M_h(x_1, x_2, t)$ are labeled as '1'. The main purpose of the coarser resolution binary map is to filter out isolated regions due to noise in $M_h(x_1, x_2, t)$. A hypothesis is formed if a connected region of pixels in $M_l(x_1, x_2, t)$ corresponding to a connected region of pixels in $\hat{M}_h(x_1, x_2, t)$. The use of region-based hysteresis-thresholding enhances detection while suppressing false positives.

We have augmented the above strategy with a localized spatial segmentation step which is very useful when the color of quasi-background is similar to the color of a moving vehicle at the same spatial position. In this case, the local region-of-interest is segmented into homogeneous regions using a k-means clustering algorithm. Moving vehicles are identified by comparing their color similarity to the moving targets obtained in previous frames.

B. Overtaking vehicle tracking

The purpose of object association and tracking is to keep track of overtaking vehicles over time. Here, we track all detected vehicles using a simple correlation measure based on Euclidean distance. Specifically, let us assume that the location of a vehicle in frame t is (x_j^t,y_j^t). Once the vehicle has been detected, we compute its enclosing rectangle R_j^t to approximate its region. To determine whether a vehicle j, detected at frame t, is the same to vehicle i, detected at frame $t-1$, two conditions are required,

$$\begin{cases} d_{ij} = \min_s \{d_{sj} | s = 1, 2, \cdots, N\}; \\ R_j^t \cap R_i^{t-1} \neq \emptyset. \end{cases} \tag{12}$$

where $d_{ij} = \sqrt{(x_j^t - x_i^{t-1})^2 + (y_j^t - y_i^{t-1})^2}$ and N is the number of vehicles in the tracking list. If there is no overlapping region between several consecutive frames, then the vehicle is added to the tracking list as a new vehicle.

C. Velocity verification

The role of this step is to determine the status of the overtaking vehicles (i.e., overtaking vs passing) in order to issue appropriate warnings to the driver. When a vehicle is overtaking the host vehicle, different parts of the overtaking vehicle are associated with different velocities. For example, the distant parts of the overtaking vehicle seem to move slowly while its closer parts seem to move faster. Moreover, pixel velocities at the motion boundaries of the motion field of the detected vehicle are unreliable. To determine the status of overtaking vehicles, we need to extract their dominant motion. In this paper, we search for the dominant motion in the sparse optical flow field of the moving target. First, we establish a set of homogeneous regions corresponding to the motion of different parts of overtaking vehicles. This is performed by grouping together the sparse optical flow vectors into homogeneous regions using the amplitude of the optical flow vectors and k-mean clustering. The dominant motion is then determined by the motion of the largest area.

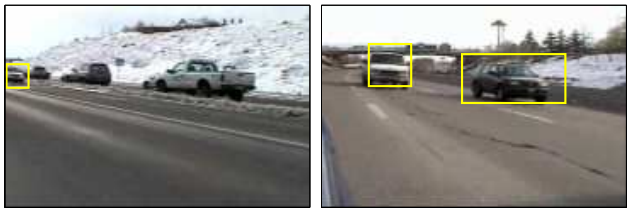
V. Experimental results

The proposed overtaking vehicle detection system has been tested under the highway scenarios in Reno, Nevada. In this section, we demonstrate the performance of the proposed overtaking vehicle detection system under several challenging highway traffic scenes. Our evaluation has been carried out both qualitatively and quantitatively.

A. Data Set

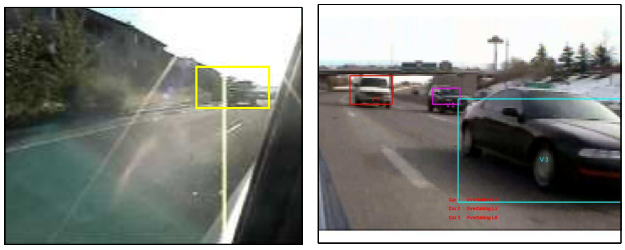
A video camera was held on the frame of the left window of the host vehicle to capture several rear-view image sequences. The overtaking vehicle video were captured on I80 in Reno, Nevada. To consider a variety of scenarios, we captured several different video sequences on different days and times, as well as under different weather conditions (e.g., sun and snow). The video was digitized using a sample rate of 15 frames per second. The size of each frame is $576 * 720$. The digitized image sequences contain various cases including partial occlusions, shadows, and multi-vehicles overtaking the host vehicle. To evaluate the performance of the proposed approach, we computed ground truth information by labeling

each sequence manually (i.e., enclosing overtaking vehicles by a rectangle).



(a) Partial occlusion. (b) Two overtaking vehicles.

(c) Partial occlusion of two overtaking vehicles. (d) Illumination changes due to passing through the overpass.



(e) Sun highlight. (f) Multiple vehicles simultaneously overtaking the host vehicle.

Fig. 5. Overtaking vehicle detection under various scenarios.

B. Qualitative evaluations

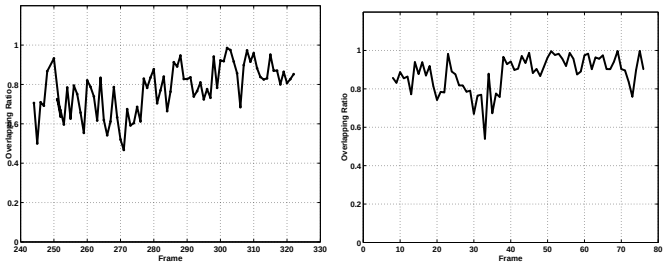
Figure 5 presents several examples of overtaking vehicle detection in a highway. Figure 5 (a) shows a case of detecting an overtaking vehicle is moving in from the left. The overtaking vehicle is partially occluded. Figure 5 (b) shows two different vehicles simultaneously overtaking the host vehicle. In Figure 5 (c), an overtaking vehicle is partially occluded by another overtaking vehicle. If the two overtaking vehicles have different speeds, our algorithm can separate them. Figure 5 (d) presents a case where the overtaking vehicle goes through an overpass, causing significant illumination changes. Figure 5 (e), illustrates a case where an overtaking vehicle is detected under sun highlight. Finally, Figure 5 (f) shows a case where multi-vehicles are simultaneously overtaking the host vehicle.

C. Quantitative evaluations

The proposed algorithm was also evaluated using an objective measure to quantify its detection performance in terms of the overlapping ratio between the rectangular area detected by the proposed method and the manually labeled rectangular area. The overlapping ratio r is defined as follows [18]:

$$r = \frac{2 * (A \cap B)}{A + B}, \tag{13}$$

where A is the ground truth area, and B is the area detected by our algorithm. Figure 6 shows a quantitative evaluation using several different video sequences. Figure 6(a) shows the overlapping ratio in a snow scene, recording an overtaking vehicle in the video sequence from frame 244 to frame 322. The overtaking vehicle was correctly detected as it is shown by the graph. Figure 6 (b) presents evaluation results in the presence of tree shadow, recording an overtaking vehicle in the video sequence from frame 8 to frame 76. The overtaking vehicle was detected correctly.



(a) Overlapping ratio curve in a snow scene. (b) Overlapping ratio curve assuming shadows.

Fig. 6. Performance results of overtaking vehicle detection under different video sequences. The x-axis denotes the frame number and y-axis denotes the overlapping ratio between the labeled rectangular area and the one detected by the proposed algorithm.

In Figure 7, the proposed algorithm shows consistent detection when operating in a snow scene for a long period, recording a total of 414 vehicles in a video sequence from frame 103 to frame 548. Five different vehicles overtook the host-vehicles in this example. In the same video sequence, three different vehicles overtook the host-vehicle simultaneously from frame 105 to frame 128. These overtaking vehicles were detected correctly as it is shown in the figure. The drop in performance from frame 150 to frame 160 for vehicle 2 is due to partial occlusions.

Table I shows the performance of the system in terms of true positives, false negatives and false positives under different scenarios. It should be noted that, most of the false positives were due to the relative motion between the host-vehicle and the manually held camera. We expect that fixing the camera on the host-vehicle would reduce the false positives.

VI. Conclusions and Future Work

In this paper, we proposed an overtaking vehicle detection algorithm by modeling the background of a traffic scene into

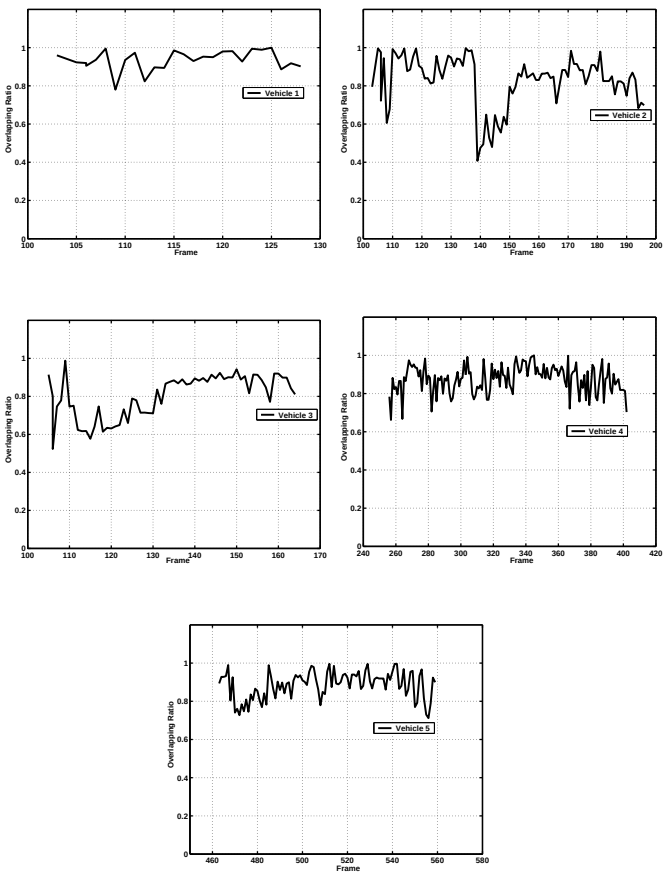


Fig. 7. Performance results of overtaking vehicle detection in a snow scene for a long period. Five different vehicles overtook the host-vehicles. The x-axis denotes the frame number and y-axis denotes the overlapping ratio between the labeled rectangular area and the one detected by the proposed algorithm.

Table I. Performance results of overtaking vehicle detection on typical scenes captured.(Legends: TP represents True Positives, FP represents False Positives and FN represents False Negatives.)

Video sequences	TP(%)	FP(%)	FN(%)
Sunny scene	96.2%	5.6%	3.8%
Vehicle go through the bridge	96.0%	3.1%	4.0%
Multi-overtaking vehicles	94.5%	4.6%	5.5%

dynamic and quasi-static regions. Homogeneous sparse optical flow was used to model the dynamic background due to camera motion. An eigenspace approach was used to model the quasi-static background, providing good robustness to illumination changes. A region-based hysteresis-thresholding approach, augmented by a localized segmentation method, was developed to make the proposed algorithm more robust to noise and reduce false positives. Our experimental result demonstrated the robustness of the proposed system under challenging traffic scenarios. The proposed technique will evidently fail when the overtaking vehicles are far from the host vehicle due to the velocity perspective effect and when overtaking vehicles are significantly occluded. There is also a significant relative deviation of motions when the direction of overtaking vehicles is not perpendicular to the image plane of the camera. For future work, we plan incorporate more sophisticated tracking (i.e., this would produce smoother overlapping ratio curves) and test our system more extensively.

ACKNOWLEDGMENTS

This research was supported by Ford Motor Company under grant No. 2001332R, and the University of Nevada, Reno (UNR) under an Applied Research Initiative (ARI) grant.

REFERENCES

[1] Z. Sun, G. Bebis and R. Miller, "On-road Vehicle Detection Using Optical Sensors: A Review," *IEEE Int. Conf. on Intelligent Transportation Systems*, 2004.

[2] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner and W. Seelen, "An image processing system for driver assistance," *Image and Vision Computing*, vol. 18, pp. 367-376.

[3] A. Kuehnel, "Symmetry-based recognition for vehicle rears," *Pattern Recognition Letters*, vol. 12, pp. 249-258, 1991.

[4] T. Zielke, M. Brauckmann and W. Seelen, "Intensity and edge-based symmetry detection with an application to car-following," *CVGIP: Image Understanding*, vol. 58, pp. 177-190, 1993.

[5] Z. Sun, G. Bebis and R. Miller, "On-road vehicle detection using Gabor filters and support vector machines," *IEEE International Conference on Digital Signal Processing*, Santorini, Greece, 2002.

[6] A. Giachetti, M. Campani and V. Torre, "The use of optical flow for road navigation, " *IEEE Tran. on Robotics and Automation*, vol. 14, pp 34-48,1998.

[7] W. Kruger, W. Enkelmann and S. Rossle, "Real-time estimation and tracking of optical flow vectors for obstacle detection," *IEEE Intelligent Vehicle Symposium*, pp 304-309, 1995.

[8] Y. Zhu, D. Comaniciu, M. Pellkofer and T. Koehler, "Passing vehicle detection from dynamic background using robust information fusion," *IEEE Intelligent Vehicle Symposium*, 2004.

[9] P. Batavia, D. Pomerleau and C. Thorpe, "Overtaking vehicle detection using implicit optical flow," *IEEE Int. Conf. on Transportation Systems*, 1997, pp. 729 - 734.

[10] J. Dlaz, E. Ros, S. Mota, G. Botella, A. Canas and S. Sabatini, "Optical flow for cars overtaking monitor: the rear mirror blind spot problem,"*10th. Int. Conf. on Vision in Vehicles*, 2003

[11] W. Gillner,"Motion based vehicle detection on motorways," *IEEE Intelligent Vehicles Symposium*, pp 25-26, 1995.

[12] D. Fleet and A. Jepson, "Computation of component image velocity from local phase information," *Int. Journal of Computer Vision*, vol. 5. no. 1, pp. 77-104, 1990.

[13] Z. Li and Z. Shen, *Dynamic Image Analysis*, National Defence Pulishing Firm, 1999.

[14] T. Gautama and M. Hulle, "Phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Trans. Neural Networks*, vol. 13. no. 5, pp. 1127-1136, 2002.

[15] M. Brand, "Incremental singular value decomposition of uncertain data with missing values," *European Conference of Computer Vision*, pp. 707-720, 2002.

[16] M. Nicolescu, G. Medioni, "Motion segmentation with accurate boundaries: a tensor voting approach," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 382-389, 2003.

[17] H. Eng, J. Wang, A.H Kam and W. Yau, "Novel region-based modeling for human detection within highly dynamic aquatic environment," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp.II-390-II-397, 2004.

[18] K. She, G. Bebis, H. Gu and R. Miller, "Vehicle tracking using on-line fusion of color and shape features," *IEEE Int. Conf. on Intelligent Transportation Systems*, 2004.