Passing Vehicle Detection from Dynamic Background Using Robust Information Fusion

Ying Zhu Dorin Comaniciu
Siemens Corporate Research, Inc.
Princeton, NJ, USA
{Ying.Zhu, Dorin.Comaniciu}@scr.siemens.com

Martin Pellkofer Thorsten Koehler
Siemens VDO Automotive AG
Regensburg, Germany
{Martin.Pellkofer; Thorsten.Koehler}@siemens.com

***Abstract*— This paper presents a robust method of passing vehicle detection. Obstacle detection algorithms that rely on motion estimation tend to be sensitive to image outliers caused by structured noise and shadows. To achieve a reliable vision system, we have developed two important techniques, motion estimation with robust information fusion and dynamic scene modeling. By exploiting the uncertainty of flow estimates, our information fusion scheme gives robust estimation of image motion. In addition, we also model the background and foreground dynamics of road scenes and impose coherency constraints to eliminate outliers. The proposed detection scheme is used by a single-camera vision system developed for driver assistance. Our test results have shown superior performance achieved by the new detection method.**

I. INTRODUCTION

In a monocular vision system designed for driver assistance, a single camera is mounted inside the ego-vehicle to capture image sequence of forward road scenes. Various vehicle detection methods have been developed to detect vehicles in the central field of the view [7], [14]. In this paper, we study the problem of passing vehicle detection, i.e. detecting vehicles that are passing the ego-vehicle upon the left or right and entering the field of view at a higher speed. Figure 1 shows such an example. Passing vehicle detection has a substantial role in understanding the driving environment. Because of the potentially unsafe driving situation that an overtaking vehicle could create, it is important to monitor and detect vehicles passing by.

Since passing vehicles need to be detected earlier on while they are entering the view and only partially visible, we can not completely rely on appearance information. Instead, characteristic optical flows are generated by a vehicle passing by. Hence, motion information becomes an important cue in detecting passing vehicles. Several obstacle detection methods using optical flow have been reported in literature [7], [8], [9], [11], [12], [15]. The main idea is to compare a predicted flow field calculated from camera parameters and vehicle velocity with the actual image flows calculated from motion estimation. An obstacle is declared if the actual flows do not match the predicted flows. These methods work well if neither strong noise nor illumination change is present. However, structured noise and strong illumination change happen quite often in practical situations. They can cause spurious image features and unreliable flow estimates. To address this problem, we have developed a highly reliable method to detect events of vehicle entering and trigger warnings in real time. In particular, a robust motion estimation scheme using *variable bandwidth density fusion* is proposed. It enables highly reliable analysis on scene dynamics and leads to superior detection performance.

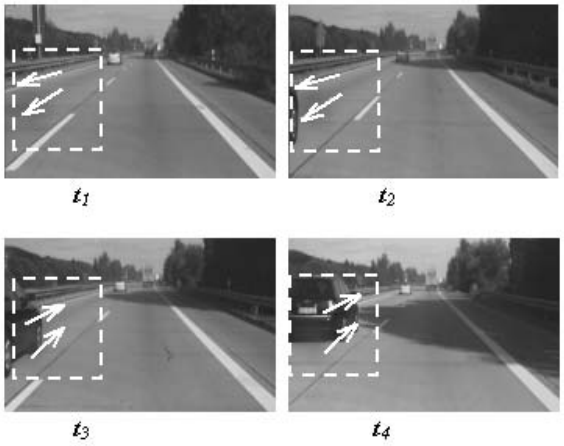

Fig. 1. Examples of vehicle passing by.

The rest of the paper is organized as follows. An overview of the proposed detection method is provided in section II. In Section III and IV, we introduce the dynamic models of road scenes and derive the corresponding hypothesis testing method. In Section V, we present variable bandwidth density fusion for robust motion estimation. Experimental results are presented in Section VI and conclusions are drawn in Section VII.

II. PASSING VEHICLE DETECTION

Vehicle passing is a sporadic event that changes the scene configuration from time to time. As Figure 1 shows, when a vehicle enters the field of view ($t_2 \sim t_3$), it forms a local foreground layer that temporarily blocks the road scene. Both image appearance and image motion around the entry point deviate from road scene dynamics. Thus, the problem is formulated as detecting changes in the scene dynamics around entry points. To solve the problem, we need to address three issues: modeling the dynamics of

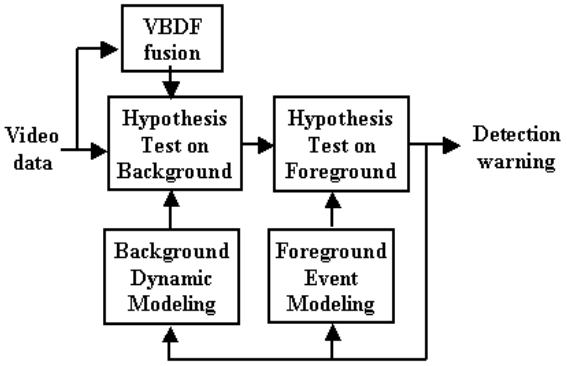

Fig. 2. Passing vehicle detection module.

road scene and vehicle passing, deriving a decision rule for passing vehicle detection, and estimating relevant features and statistical quantities involved in hypothesis testing. The proposed solution is illustrated in Figure 2. We characterize the dynamics of road scenes and passing vehicles through modeling image intensity and image motion around entry points. For the event of vehicle passing, the temporal coherency in vehicle movement is imposed. The decision rule is implemented as a decision tree. We monitor the image intensity and image motion inside analysis windows placed around entry points, and detect any change of scene dynamics. Relevant parameters used in hypothesis testing are updated over time in response of detection results. An important contribution of our work is the robust motion estimation using *variable bandwidth density fusion*(VBDF)[1].

III. Background and Foreground Dynamics

In the absence of passing vehicles, the visible road scene, i.e. the background, is moving consistently in the field of view as the camera is moving along with the ego-vehicle. Given the vehicle velocity and camera calibration, the image motion and image intensity of the background scene is predictable over time. In other words, the background scene follows a dynamic model defined by camera parameters and camera motion. Denote the image intensity at time instance t by $I(\mathbf{x}, t)$ and the motion vector by $\mathbf{v}(\mathbf{x}, t)$, where $\mathbf{x}$ is the spatial coordinate of an image pixel. The hypothesis of the dynamic background is described as

$$\mathrm{H}_{\mathrm{road}}: \quad \begin{cases} I(\mathbf{x}+\mathbf{v}(\mathbf{x},t)\cdot\delta t, t-\delta t) = I(\mathbf{x},t) + n_I \\ \mathbf{v}(\mathbf{x},t) = h(\mathbf{x}, V_0(t), \Theta) \end{cases} \tag{1}$$

The true image motion $\mathbf{v}(\mathbf{x}, t)$ is decided by the vehicle speed $V_0(t)$ and camera parameters Θ. Under the brightness constancy condition, image intensity is predictable from previous frames given the true motion. Nevertheless, brightness constancy is frequently violated in practice due to changing illumination. In addition, intensity is also affected by various image noise. Therefore, a noise term n_I is adopted to account for the perturbation on intensity. These

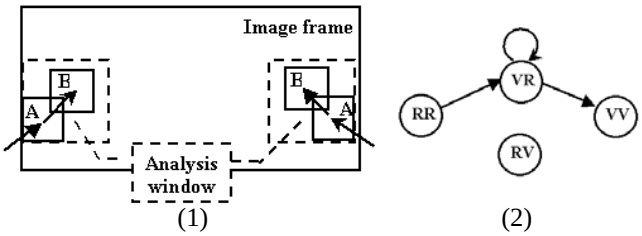

Fig. 3. Dynamics of passing vehicles. (1)Two subwindow A and B lie on the trajectory of passing vehicles. (2) Admissible paths of state transitions.

hypotheses on scene dynamics impose useful domain-specific constraints.

When a passing vehicle enters the view, the dynamics of the background scene is violated. From (1), violations of background dynamics can be detected through hypothesis testing on image intensity and image motion. However, a violation can happen under conditions other than vehicle passing, such as strong illumination changes and structured noise. To validate that a violation is indeed caused by a passing vehicle, it is necessary to exploit the domain-specific constraints introduced by passing vehicles as well.

Considering the diversity of vehicle appearance and velocity, we characterize the dynamics of passing vehicles by underlining the coherency present in vehicle motion. As Figure 3-1 illustrates, to describe the motion coherency, we take two subwindows A and B along the trajectory of a passing vehicle and study the motion pattern. As a passing vehicle enters the field of view, it arrives at A and B in an orderly fashion. For a vehicle to arrive B, it has to arrive A first. Thus violation of background dynamics should happen in A no later than in B. In contrast, such coherency is lacking in the case where the violation of scene dynamics is a consequence of irregular causes such as sudden illumination changes, structured noise and shadows. Therefore, the hypothesis made on passing vehicles helps to further distinguish events with coherent motion from irregular causes known as outliers. Denote S_A and S_B as the state variable of subwindow A and B respectively. Use R to represent the state where motion and intensity comply with road dynamics, and V for the state where the road dynamics is violated. The event of vehicle passing is described as a series of state transitions of $S_A S_B$ starting with RR and ending with VV. As Figure 3-2 shows, coherent events are distinguished by a set of admissible paths of state transitions

$$\mathrm{H}_{\mathrm{vehicle}}: \mathcal{P} = \{RR \rightarrow VR \rightarrow \cdots \rightarrow VV\}. \tag{2}$$

The preceding formulation of passing vehicle dynamics has been demonstrated to work well in our tests. However, it is necessary to point out that a more general framework of hypothesizing domain-specific constraints is through extensive learning [6].

IV. Decision tree

In solving the problem of passing vehicle detection, we encounter different contexts in the analysis windows, e.g.

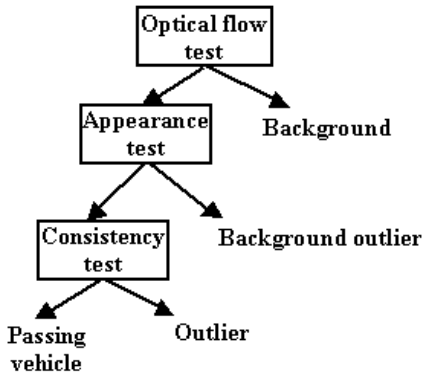

Fig. 4. Decision tree.

road scenes, outliers and vehicles. Decision trees classify these contexts by sorting them through a series of hypothesis testing represented in a tree form. The decision tree adopted in our algorithm is shown in Figure 4. Image motion and image intensity are tested against the dynamic model of the road scene (1). Coherency test is performed on the contexts that violate the scene dynamics. The decision rule for passing vehicles is summarized as follows.

$$\begin{array}{c}(\text{background dynamics is violated})\\ \wedge\ (\text{conherency is satisfied})\end{array} \tag{3}$$

A. Hypothesis testing on background dynamics

The true motion $\mathbf{v}(\mathbf{x},t)$ of the road scene is given in (1). Assume the observed image motion $\hat{\mathbf{v}}(\mathbf{x},t)$ can be estimated. Then the hypothesis testing on background dynamics is expressed as

$$\begin{array}{l}\text{violation of background dynamics if}\\ (||\hat{\mathbf{v}}(\mathbf{x},t)-\mathbf{v}(\mathbf{x},t)|| \geq \tau_{motion}) \vee (||R(\mathbf{x},t)|| \geq \tau_{residual})\end{array} \tag{4}$$

where $R(\mathbf{x},t) = I(\mathbf{x}+\mathbf{v}(\mathbf{x},t)\cdot\delta t, t-\delta t) - I(\mathbf{x},t)$ is the residual from motion compensation and reflects the mismatch between the predicted image and the actual image. By testing motion and residual, we classify all the instances into two groups, instances complying with background dynamics and instances violating background dynamics.

Although further testing is performed to classify instances of violations, it is important to have reliable motion estimation $\hat{\mathbf{v}}(\mathbf{x},t)$ that faithfully reflects the context for an accurate initial classification. We have developed a robust motion estimation algorithm using *variable bandwidth density fusion* and spatial-temporal filtering. The next section is devoted to detailed discussions on our approach towards robust motion estimation.

When motion estimation is not reliable, the residual test helps to identify background scenes. For instance, the aperture problem [2] has been encountered very often in our experiments with real videos captured on highways. In such scenarios, it is very difficult to obtain accurate motion estimation. Thus, the presence of background can not be identified by motion but can be easily identified by testing the image residual.

The thresholds τ_{motion}, $\tau_{residual}$ as well as the admissible state transitions $\mathcal{P}$ are part of the decision tree solution. There are generally two ways to solve them, offline learning and online learning. Offline decision tree learning has been studied previously [6]. Online learning enables system adaptation to the gradual change of scene dynamics. Take $\tau_{residual}$ as an example, online learning can be achieved by modeling the residual data $\{R(\mathbf{x},T), R(\mathbf{x},T-1), R(\mathbf{x},T-2), \cdots\}$ computed online. Our nonparametric density estimation and mode finding techniques [1] can be used to cluster the data, obtain a Gaussian mixture model and update the model over time. The mixture model learned online is then used to predict the context from new observations $R(\mathbf{x},T+1)$.

B. Hypothesis testing on passing vehicles

The coherency test is performed on instances where background dynamics is violated. The purpose of this test is to further rule out outliers caused by structured noise and sudden illumination changes. From the hypothesis formulated on passing vehicles (2), the decision rule is expressed as

$$\begin{array}{l}\text{passing vehicle :}\\ \{\cdots S_A(t-2)S_B(t-2), S_A(t-1)S_B(t-1),\\ S_A(t)S_B(t)\} \in \mathcal{P}\\ \text{outlier :}\\ \{\cdots S_A(t-2)S_B(t-2), S_A(t-1)S_B(t-1),\\ S_A(t)S_B(t)\} \notin \mathcal{P}\end{array} \tag{5}$$

V. Motion estimation

In this section, we introduce the technique of *variable bandwidth density fusion* and an robust motion estimation algorithm.

A. Initial estimates of image motion

Following the assumption of brightness constancy, the motion vector for a given image location is computed by solving the linear equation

$$\bigtriangledown_{\mathbf{x}} I(\mathbf{x},t)\cdot\mathbf{v} = -\bigtriangledown_t I(\mathbf{x},t) \tag{6}$$

The biased least squares solution is given by[1], [3]

$$\hat{\mathbf{v}} = (A^T A + \beta\mathbf{I})^{-1} A^T \mathbf{b} \tag{7}$$

where A is a matrix defined by the spatial image gradients $\bigtriangledown_{\mathbf{x}} I$ in a local region, and $\mathbf{b}$ is vector composed of temporal

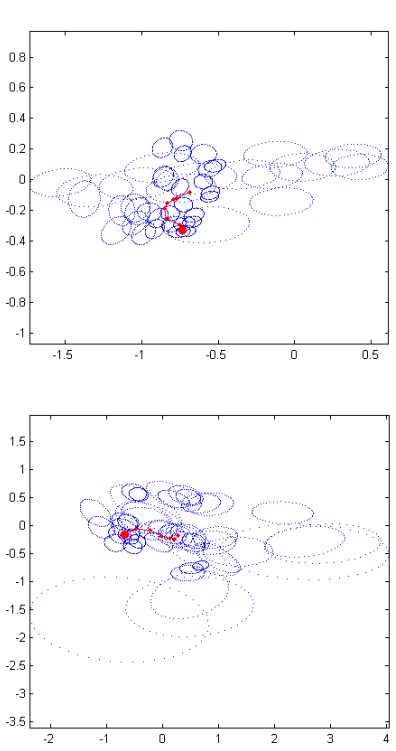

Fig. 5. Variable bandwidth density fusion. 45 initial motion estimates with covariance are plotted as blue ellipses. The trajectory of mode propagation through scales is shown in red. The convergent point is shown as a large solid circle in red.

image gradients $\bigtriangledown_t I$. To describe the uncertainty of the motion estimation, we also define its covariance [4] as

$$C = \hat{\sigma}^2 (A^T A + \beta \mathbf{I})^{-1}$$
$$\hat{\sigma}^2 = \frac{1}{N-3} \sum_{i=1}^{N} (\bigtriangledown_{\mathbf{x}} I(\mathbf{x}_i, t) \cdot \hat{\mathbf{v}} + \bigtriangledown_t I(\mathbf{x}_i, t))^2. \quad (8)$$

where N is the number of pixels in the local region and $\hat{\sigma}^2$ is the estimated variance of image noise. Unreliable flow estimates are associated with covariance matrices with a large trace. This information is very important for robust fusion.

A multiscale hierarchical framework of computing $\hat{\mathbf{v}}$ and its covariance C is discussed in [1]. For every frame, the initial motion vector is estimated at different spatial locations inside the analysis window. Thus, we get a sequence of motion estimates with covariances $\{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}\}$ in space and in time.

B. Variable bandwidth density fusion

The initial motion estimates are sensitive to structured noise and illumination changes which introduce outliers in motion estimates. To overcome these outliers, joint spatial-temporal filtering is performed on the initial motion estimates through a technique called variable bandwidth density fusion (VBDF). VBDF is a fusion technique first introduced in [1]. In contrast to conventional BLUE and covariance intersection fusion, with the presence of multiple motions, VBDF is able to locate the most significant mode of the data and thus is robust again outliers. Given the initial motion estimates $\mathbf{v}_{\mathbf{x},t}$ and covariance $C_{\mathbf{x},t}$ across multiple spatial and temporal locations $\mathbf{x} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$, $t = \{T, T-1, \cdots, T-M\}$, we apply VBDF to obtain the dominant motion in the analysis window of the T-th frame.

VBDF is implemented through the following mean shift procedure [5]. First, a pointwise density estimator is defined by a mixture function:

$$f(\mathbf{v}; \{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}\}) = \sum_{\mathbf{x},t} a_{\mathbf{x},t} K(\mathbf{v}; \mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t})$$
$$K(\mathbf{v}; \mathbf{v}_i, C_i) = \frac{\exp(-\frac{1}{2}(\mathbf{v}-\mathbf{v}_i)^T C_i^{-1} (\mathbf{v}-\mathbf{v}_i))}{(2\pi)^{d/2} |C_i|^{1/2}} \quad (9)$$
$$\sum_{\mathbf{x},t} a_{\mathbf{x},t} = 1$$

Here, $a_{\mathbf{x},t}$ defines a weighting scheme on the data set, and $K(\mathbf{v}; \mathbf{v}_i, C_i)$ is the Gaussian kernel with center $\mathbf{v}_i$ and bandwidth C_i. The variable bandwidth mean shift vector at location $\mathbf{v}$ is given by

$$\mathbf{m}(\mathbf{v}) = H(\mathbf{v}) \sum_{\mathbf{x},t} w_{\mathbf{x},t} C_{\mathbf{x},t}^{-1} \mathbf{v}_{\mathbf{x},t} - \mathbf{v}$$
$$H(\mathbf{v}) = (\sum_{\mathbf{x},t} w_{\mathbf{x},t} C_{\mathbf{x},t})^{-1} \quad (10)$$
$$w_{\mathbf{x},t}(\mathbf{v}) = \frac{\frac{a_{\mathbf{x},t}}{|C_{\mathbf{x},t}|^{1/2}} \exp(-\frac{1}{2}(\mathbf{v}-\mathbf{v}_{\mathbf{x},t})^T C_{\mathbf{x},t}^{-1} (\mathbf{v}-\mathbf{v}_{\mathbf{x},t}))}{\sum_{\mathbf{x},t} \frac{a_{\mathbf{x},t}}{|C_{\mathbf{x},t}|^{1/2}} \exp(-\frac{1}{2}(\mathbf{v}-\mathbf{v}_{\mathbf{x},t})^T C_{\mathbf{x},t}^{-1} (\mathbf{v}-\mathbf{v}_{\mathbf{x},t}))}$$

The iterative computation of the mean shift vector recovers a trajectory starting from $\mathbf{v}$ and converging to a local maximum, i.e. a mode of the density estimate $f(\mathbf{v}; \{\mathbf{v}_{\mathbf{x},t}\})$.

$$\mathbf{v}_0 = \mathbf{v}$$
$$\mathbf{v}_{j+1} = \mathbf{v}_j + m(\mathbf{v}_j) \quad (j \geq 0) \quad (11)$$
$$\mathbf{v}_j \rightarrow \text{mode}(\mathbf{v}; \{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}\}) \text{ as } j \rightarrow \infty$$

To treat $f(\mathbf{v}; \{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}\})$ with multiple modes, we introduce a series of analysis bandwidths $C_{\mathbf{x},t}^l = C_{\mathbf{x},t} + \alpha_l \mathbf{I}$ ($\alpha_0 > \alpha_1 > \cdots > 0$) which leads to multiple smoothed density estimates $f(\mathbf{v}; \{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}^l\})$. The number of modes in the density estimate decreases as larger analysis bandwidths are adopted. At the initial scale, α_0 is set large such that the density $f(\mathbf{v}; \{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}^0\})$ has only one mode $\text{mode}_0 = \text{mode}(\mathbf{v}; \{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}^0\})$ which is invariant to the starting point $\mathbf{v}$ in VBDF. The mode point is then propagated across scales. At each scale, VBDF uses the mode point found from the last scale as the initial point to locate the mode for the current scale.

$$\text{mode}_l = \text{mode}(\text{mode}_{l-1}; \{\mathbf{v}_{\mathbf{x},t}, C_{\mathbf{x},t}^l\}) (l = 1, 2, \cdots) \quad (12)$$

The mode point will converge to the most significant sample estimate as α_j decreases. The convergent point defines

the dominant motion $\hat{\mathbf{v}}_t$ inside the analysis window of frame T. Figure 5 shows two examples of using VBDF to obtain the dominant motion inside the analysis window. For every frame, initial motion estimates are computed at 9 equally spaced locations in the analysis window. The initial estimates from 5 consecutive frames are used as the input to VBDF. Exponential forgetting is employed to weight these initial motion estimates temporally. The results shown in Figure 5 demonstrates the robustness of the fusion algorithm against outliers.

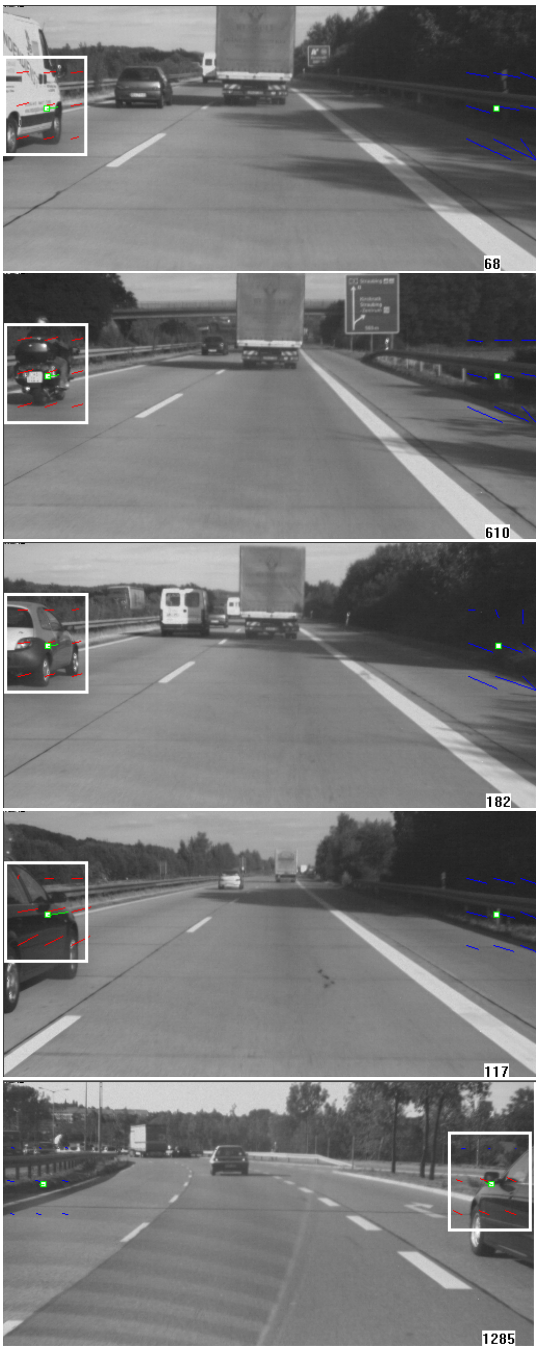

Fig. 6. Passing vehicle detection results.

VI. Experiments

To evaluate the performance of our passing vehicle detection algorithm, we tested 59 real videos captured on highways as well as city streets. The system is running in real time. Ego-vehicle velocity is provided in 10 videos. Among these 10 videos, there are 53 passing vehicles. Our vision system correctly detected all of them and made no false detections. For the remaining 49 videos containing 41 passing vehicles, the ego-vehicle velocity is unknown. In such cases, we assumed forward motion of the ego-vehicle in our test. The system missed 1 passing vehicle due to the persistant dark lighting under a bridge. In Figure 6, a few examples of detection results along with the estimated motion vectors are shown. Vehicles of different shapes are correctly detected. And more importantly, the visible illumination changes, shadows and structured noise did not trigger any false warning. Figure 7 and 8 shows the detection results on a video heavily contaminated by noise caused by glares on the windshield. The structured noise causes large mismatch of the actual flow and image from their predictions. Our system was able to correctly detect all three passing vehicles without triggering any false alarm. These results demonstrate the robustness of the proposed detection algorithm.

VII. Conclusions

We have proposed a highly reliable method for passing vehicle detection. Robust information fusion and dynamic scene modeling are important techniques that lead to the success of our vision system. Through variable bandwidth data fusion, we are able to locate the major mode of a data set with uncertainty and disregard outliers in the data. In addition, the hypotheses on the scene dynamics impose useful constraints to correctly identify passing vehicles. The superior performance achieved in our experiments including difficult cases demonstrates the power of the new techniques. These techniques also apply to effective integration of different modules of a vision system.

References

[1] D. Comaniciu, "Nonparametric information fusion for motion estimation", *IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 59-66, 2003.

[2] E. Trucco and A. Verri, *Introductory techniques for 3-D computer vision*, Prentice Hall, 1998.

[3] D. W. Marquardt, "Generalized inverse, ridge regression, biased linear estimation, and nonlinear estimation", *Technometrics*, 12(3):591-612, 1970.

[4] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer-Verlag, 2001.

[5] D. Comaniciu, V. Ramesh, P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift", *IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 142-149, 2000.

[6] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.

[7] W. Enkelmann, "Obstacle Detection by Evaluation of Optical Flow Fields from Image Sequences", *Image and Vision Computing*, 9(3):160-168, 1991.

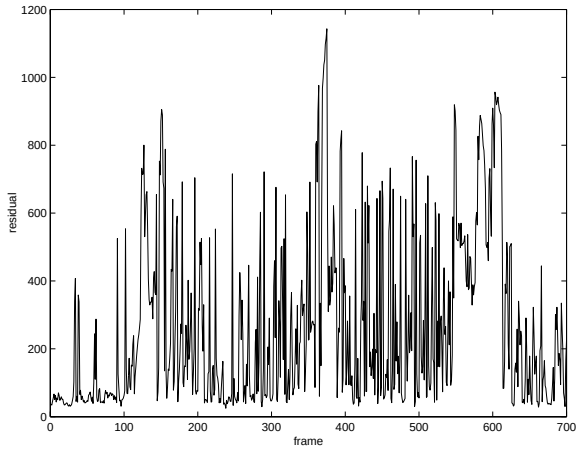

Fig. 7. Image mismatch error and flow estimates.

[8] R. C. Nelson and J. Aloimonos, "Obstacle Avoidance Using Flow Field Divergence", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(10):1102-1106, 1989.

[9] W. Krüger, "Robust Real-Time Ground Plane Motion Compensation From A Moving Vehicle", *Machine Vision and Applications*, 11:203-212, 1999.

[10] W. Enkelmann, "Video-Based Driver Assistance–From Basic Functions to Applications", *Int. J. Computer Vision*, 45(3):201-221, 2001.

[11] S. Carlsson and J. Eklundh, "Object Detection using Model Based Prediction and Motion Parallax", *European Conference on Computer Vision*, pp. 297-306, 1990.

[12] M. Betke, E. Haritaoglu and L. S. Davis, "Real-Time Multiple Vehicle Detection and Tracking from A Moving Vehicle", *Machine Vision and Applications*, 12:69-83, 2000.

[13] M. Haag and H.-H. Nagel, "Combination of Edge Element and Optical Flow Estimates for 3D-Model-Based Vehicle Tracking in Traffic Image Sequences", *Int. J. Computer Vision*, 35(3):295-319, 1999.

[14] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner and W. von Seelen, "Computer Vision for Driver Assistance Systems", *Proc. SPIE*, 3364:136-147, 1998.

[15] P. Batavia, D. Pomerleau, and C. Thorpe, "Overtaking Vehicle Detection Using Implicit Optical Flow", *Proc. IEEE Transportation Systems Conf.*, pp. 729-734, 1997.

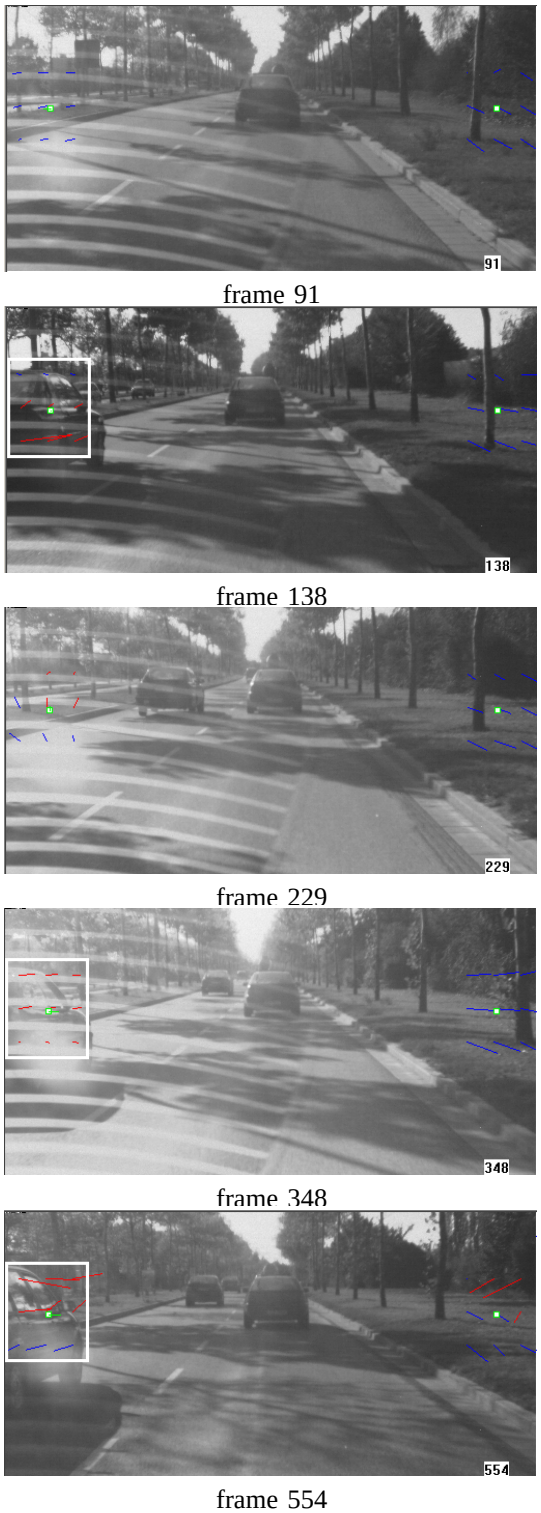

Fig. 8. Passing vehicle detection results on data affected by structured noise.