Partially Occluded Vehicle Recognition and Tracking in 3D

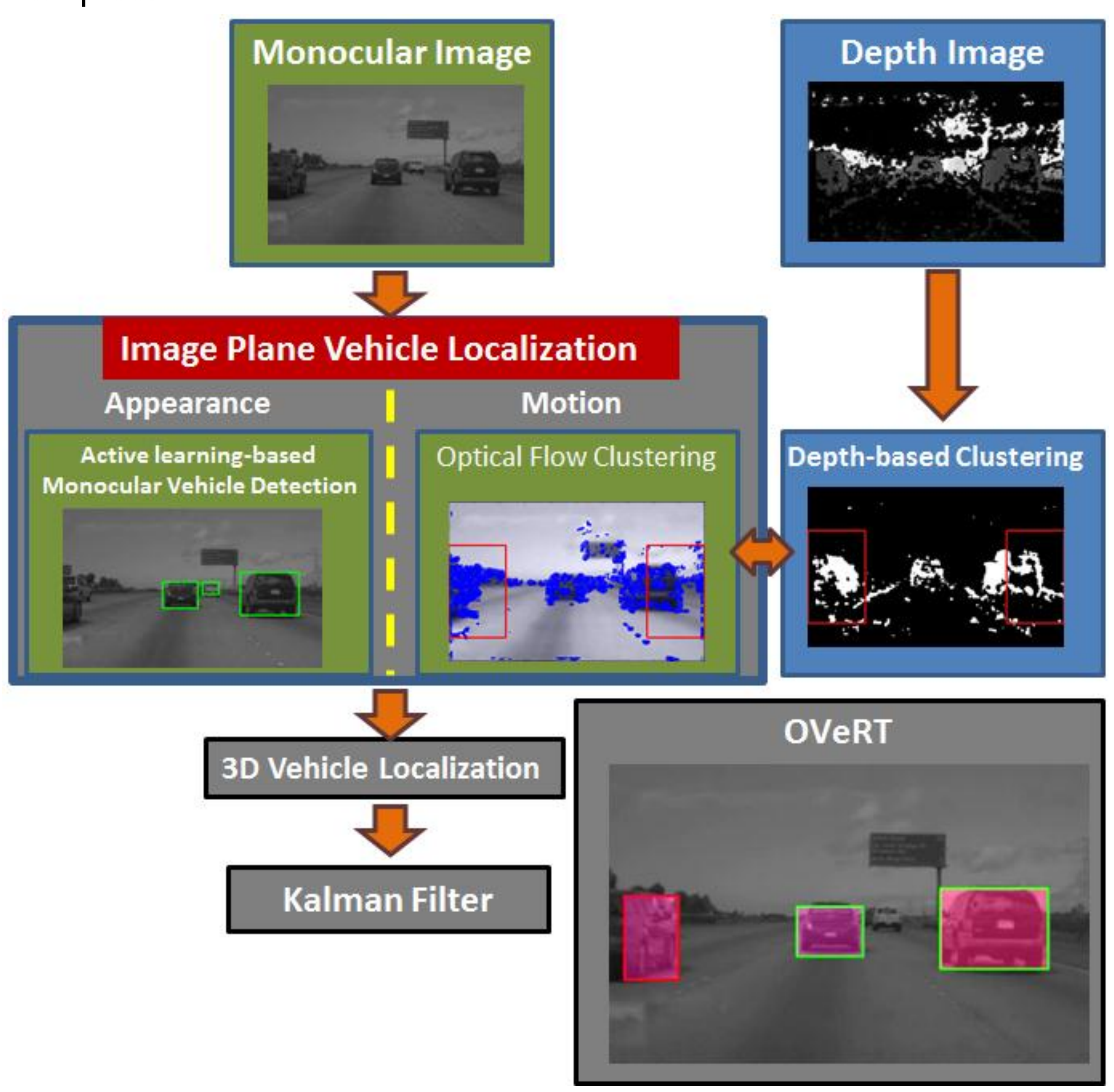
Eshed Ohn-Bar, Sayanan Sivaraman, and Mohan M. Trivedi
Laboratory for Intelligent and Safe Automobiles
University of California, San Diego

LISA: Laboratory for Intelligent & Safe Automobiles

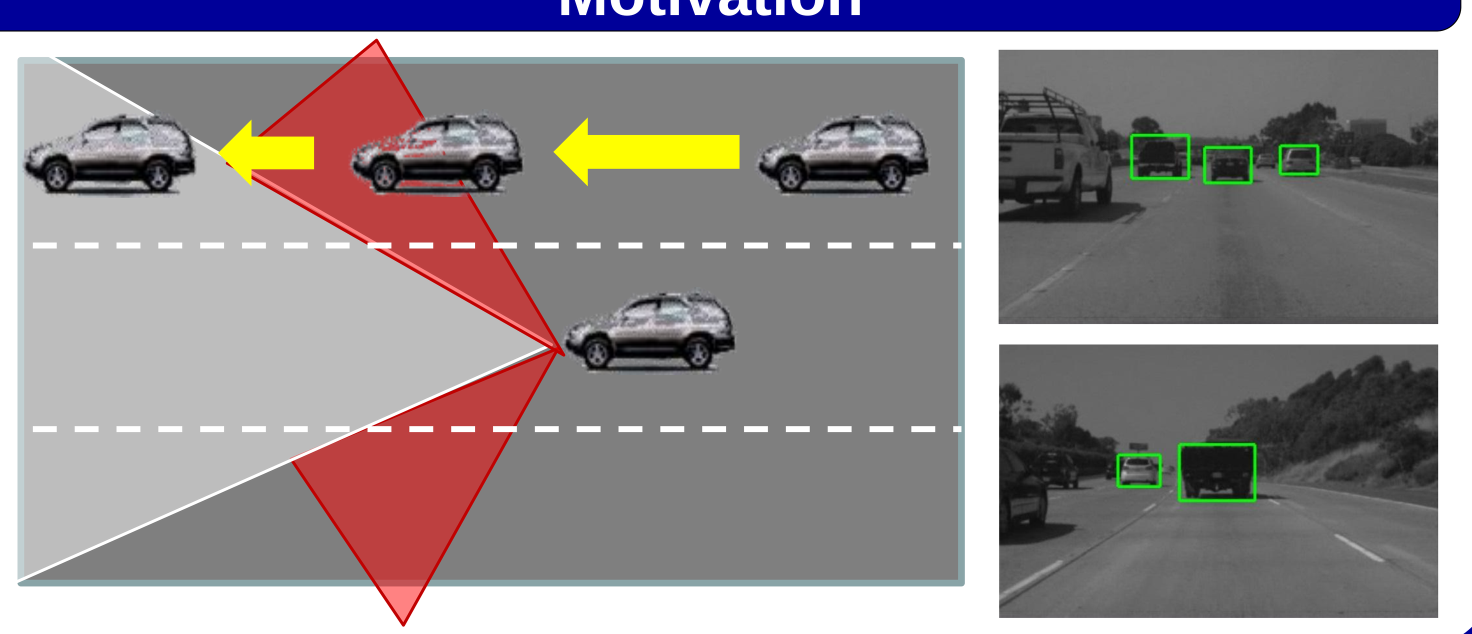
CVRR UCSD

Abstract

The proposed vision system detects partially visible vehicles using an active-learning based monocular vision approach [1,2] and motion (optical flow) cues. The appearance-based technique can't reliably detect partially visible vehicles, yet motion cues can provide useful information for detecting the vehicle. In the OVeRT (partially **O**ccluded **Ve**hicle **R**ecognition and **T**racking) system, optical flow is calculated in two peripheral search windows to identify distinctive motion patterns of a vehicle with relatively different speed than the ego-vehicle. The optical flow is clustered to produce detections, which are refined using stereo. Furthermore, we use the depth map to obtain the real-world coordinates of each detected vehicle. Tracking is performed using a Kalman filter, which is formulated to integrate stereo-monocular information. We demonstrate the effectiveness of the proposed system on a multilane highway dataset containing 60 overtaking instances, 4800 frames of positive examples and 5000 frames of negative examples.



Motivation



Fully-Visible Static-Cue Vehicle Detection

A vehicle detector from monocular input was trained using an Adaboost cascade of Haar-like rectangular features. An active-learning framework [1] was employed using two stages. First, an initialization of the classifier is performed in a supervised manner using a set of positive target class and negative class. Next, in the query and retraining stage, the classifier was evaluated on an independent dataset and retrained to include missing and false positive detections.

Motion-based Vehicle Detection of Partially Visible Vehicles

As critical, near range vehicles that perform overtaking are of interest - we threshold the depth image to produce a set of vehicle bounding boxes proposals, v_d^j , $j = 1,2$ (for the left and the right search window).

Flow Attributes: Given a flow field with magnitude and orientation (Δ, θ), several methods for transforming the flow to a descriptor will be explored. As a baseline a histogram of flow vectors and linear SVM scheme is used. Alternatively, within each frame we may restrict to pixels with optical flow with the range

$$\Theta = \{i \mid \theta(i) \in [0, \frac{\pi}{2}]\}$$

for detecting left overtaking motion (for the right search window we use the range $[0, \frac{\pi}{2}]$). Next the optical flow vectors that fit the appropriate range above are clustered together into one, inclusive bounding box v_{OF}^j. Given such set of motion vectors, we can perform a classification of overtaking presence by a score produced by applying a function to the flow magnitude, such as the **entropy**, **L1-norm**, and **L2-norm**, as well as the frequency of occurrences of flow vectors (**cardinality** of the set) in the overtaking direction. The detection bounding box is given by the intersection of the bounding box from the optical flow clustering and the depth clustering $v_{OT}^j = v_{OF}^j \cap v_d^j$.

Tracking Both in Image Plane and 3D Coordinates with a Single Kalman Filter

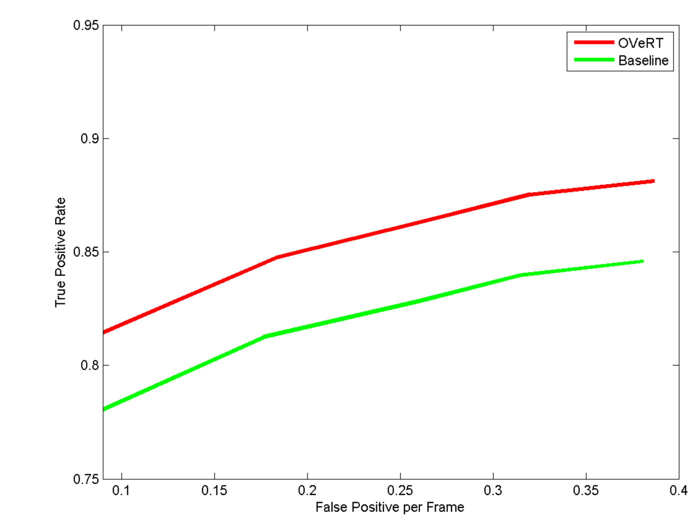
$$v_k = [i_k \;\; j_k \;\; w_k \;\; h_k]^T$$
$$V_k = [i_k \;\; j_k \;\; w_k \;\; h_k \;\; X_k \;\; Y_K \;\; Z_k \;\; \Delta X_k \;\; \Delta Y_k \;\; \Delta Z_k]^T$$

$$V_{k+1} = \begin{pmatrix} \boldsymbol{I}_{4\times4} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{3\times3} & \boldsymbol{\Delta_t I_{3\times3}} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I_{3\times3}} \end{pmatrix} V_k + \eta_k$$

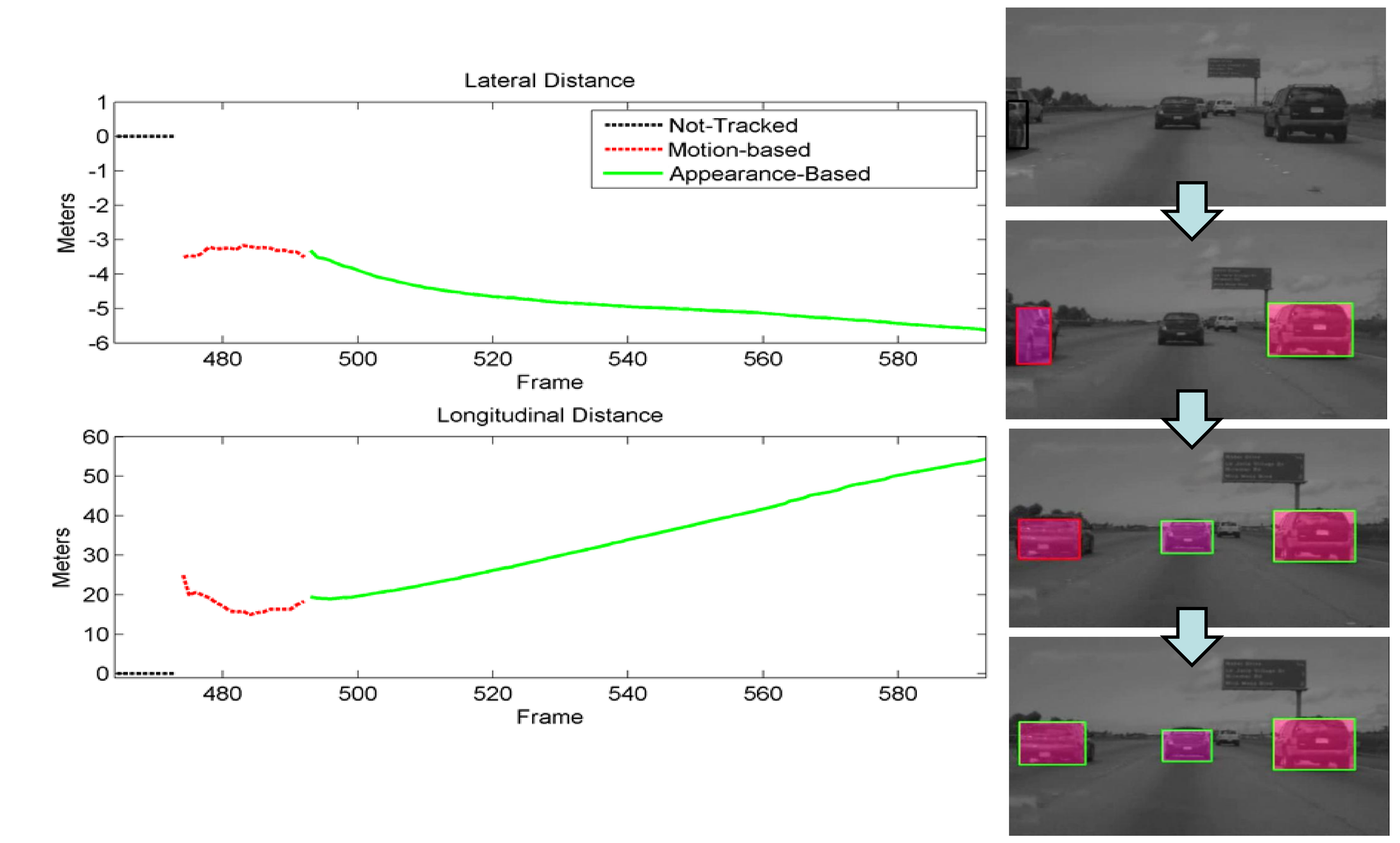
$$M_k = (\boldsymbol{I}_{7\times7} \;\; \boldsymbol{0})\, V_k + \xi_k$$

For a given bounding box, parameterized by its top left corner coordinates, width, and height, the depth of the vehicle is given by the median value in the bounding box, Z_k, and project the centroid to obtain the other two 3D coordinates, X_k and Y_k. Each vehicle is tracked using a single Kalman filter.

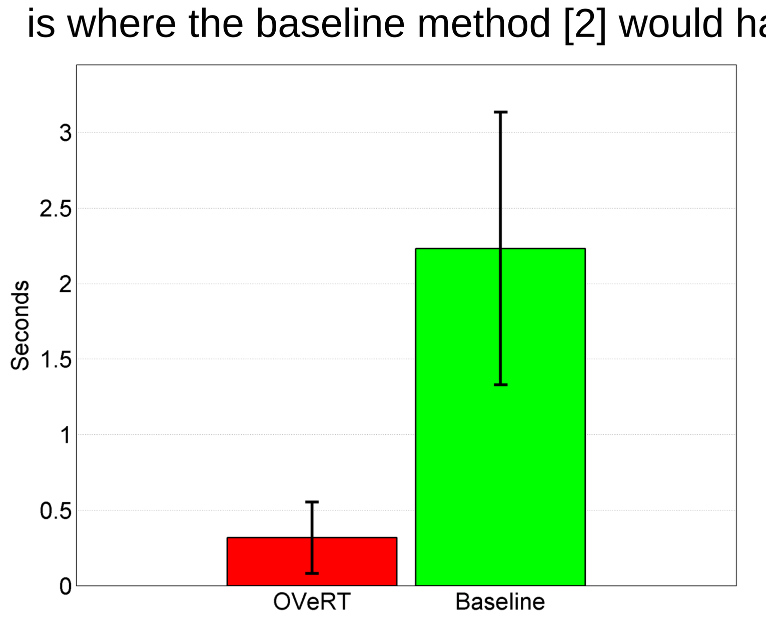
Experimental Evaluation



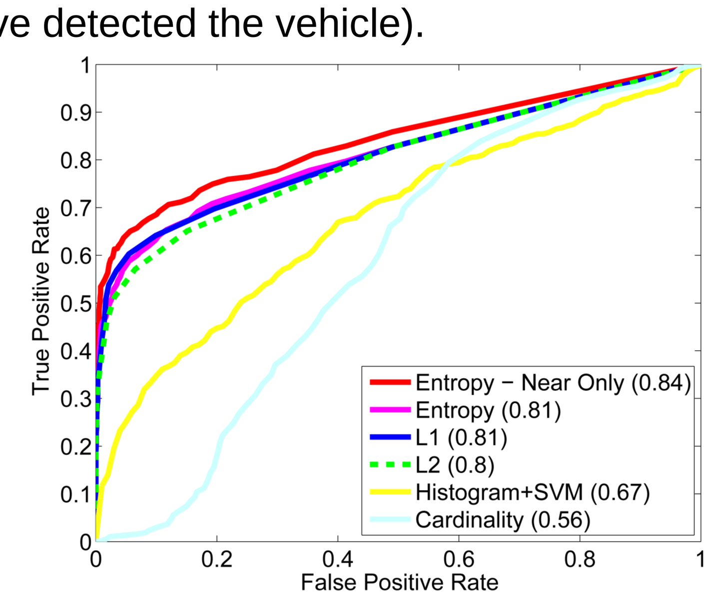
Evaluation of the system with motion only and without the stereo-based box proposals, compared to the non-motion baseline in [2]. Figure on left was produced using a subset of the complete 60 overtaking dataset, with 6667 vehicle boxes and 7 overtaking instances.



Example of the performance of the OVeRT system. The baseline system maintains detection only along the plotted green line. Sequence of images of the tracked vehicle on the right: A vehicle with higher relative motion enters the scene. OVeRT tracking begins after 9 frames using motion cues. Next, after 38 frames since entering the scene, the vehicle is detected using appearance-based features (this is where the baseline method [2] would have detected the vehicle).



Vehicle localization of partially visible vehicles as they enter the scene in a relatively higher speed. OVeRT's tracking is initialized, on average, within **0.32 seconds** of the vehicle entering the field of view (standard deviation = 0.23 seconds). On the other hand, the baseline appearance-based only detection system [2] detects such vehicles within **2.32 seconds** (standard deviation = 0.90 seconds).



Evaluation of the detections provided by OVeRT using different motion attributes: we threshold a score functions $\phi(\Delta): \mathbb{R}^{|\Theta|} \to \mathbb{R}$ applied to the magnitude field. In parenthesis is the AUC. 'Near only': performance on instances of vehicles in adjacent lanes to the ego-lane only.

References

[1] **Sayanan Sivaraman and Mohan M. Trivedi,** "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Transactions on Intelligent Transportation Systems*, 2010.
[2] **Sayanan Sivaraman and Mohan M. Trivedi,** "Combining monocular and stereovision for real-time vehicle ranging and tracking on multilane highways," *IEEE International Conference on Intelligent Transportation Systems, 2011.*