

Partially Occluded Vehicle Recognition and Tracking in 3D

Eshed Ohn-Bar, Sayanan Sivaraman *Student Member*, and Mohan Trivedi *Fellow, IEEE*

Abstract—Vehicle detection is a key problem in computer vision, with applications in driver assistance and active safety. A challenging aspect of the problem is the common occlusion of vehicles in the scene. In this paper, we present a vision-based system for vehicle localization and tracking for detecting partially visible vehicles. Consequently, vehicles are localized more reliably and tracked for longer periods of time. The proposed system detects vehicles using an active-learning based monocular vision approach and motion (optical flow) cues. A calibrated stereo rig is utilized to acquire a depth map, and consequently the real-world coordinates of each detected vehicle. Tracking is performed using a Kalman filter. The tracking is formulated to integrate stereo-monocular information. We demonstrate the effectiveness of the proposed system on a multilane highway dataset containing instances of vehicles with relative motion to the ego-vehicle.

I. INTRODUCTION

The number of on-road fatalities amounts to tens of thousands of fatalities each year in the United States alone [1]. Hence, there's a considerable interest in developing active safety systems. Fast and reliable monitoring of the vehicle's surround can provide assistance to save lives and potentially reduce accidents in severity and number. An automated vision system for detecting and tracking vehicles may be incorporated with other surround analysis and driver assistance or monitoring systems [2], [3], [4]. Such a system is an essential part of developing efficient intelligent driver assistance and autonomous vehicle systems.

The problem of vehicle detection poses several challenges for a vision-based system. Urban and freeway settings include shadowing, man-made structures, and ubiquitous visual clutter can introduce false positives. On top of the variation in the appearances of the objects to be detected, these objects are also often partially occluded. The vehicles to be detected are also encountered in a variety of orientations, including preceding, oncoming, and cross traffic. Although motion can be a useful cue under such settings, effects of ego motion may need to be compensated for [5].

This work aims to generalize an appearance-based vehicle detector to better detect partially visible vehicles. We term the system OVeRT (partially Occluded Vehicle Recognition and Tracking). The general overview of the framework can be seen in Fig. 1. Detections of vehicles are first made in the image plane using multiple cues, appearance-based and motion-based. The appearance-based detector is based on [6] an active-learning monocular vision approach. The features used are the Haar-like rectangular features. The system cannot reliably detect partially occluded vehicles.

The authors are with the Department of Electrical and Computer Engineering, University of California San Diego (UCSD), La Jolla, CA, 92092 USA (e-mail: eohnbar@ucsd.edu; ssivaram@ucsd.edu; mtrivedi@ucsd.edu).

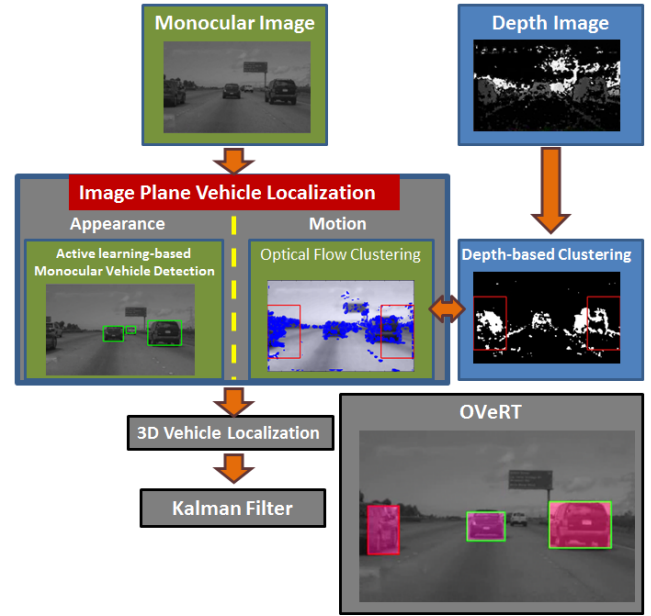


Fig. 1: Overview of the OVeRT (partially Occluded Vehicle Recognition and Tracking) approach proposed in this paper.

When a vehicle is partially visible, motion cues may provide information for detecting the vehicle. In the OVeRT system, optical flow is calculated in 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 with depth-based clustering. These image plane vehicle localizations can then be projected and tracked in 3D using a Kalman filter.

In addition to the novelty of improving a previously proposed vehicle detection system to be more robust to occlusions, we pursue an integrative approach between the monocular and stereo-vision domain in tracking of the vehicles. While sensor fusion has shown promise in combining complimentary modalities for increased performance [8], [9], [10], [11], [12], integrating cues from the domains of monocular and stereo-vision for vehicle detection and tracking is still an immature field.

Recently, Sivaraman and Trivedi [7] incorporated information from both the monocular and stereo-vision sources in order to track vehicles detected in the image plane. By formulating the tracking problem using both modalities, each vehicle's state is estimated in both image and 3D coordinates. The combined domain tracking performs in a higher overall detection true positive rate. The system runs at 46ms per frame. Since the initial detections rely on appearance-based

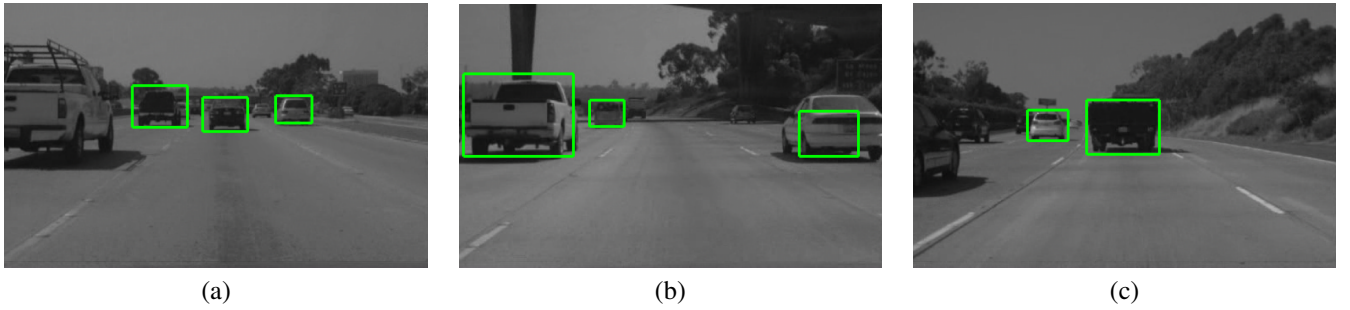


Fig. 2: Current limitations of the appearance-based baseline system in [7]. (a) Vehicle in left lane after 41 frames since entering the field of view is not fully visible, and is missed by the detector. (b) Example of a phenomenon of overtaking vehicles being detected with a partial bounding box, then detection is lost for several more frames before reliable detection and tracking is achieved. Vehicle on the right is partially covered by the detection box, 37 frames since entering the scene. Only after 7 more frames the box returns with a reliable and continuous detection. (c) On a multilane highway, some vehicles may be only visible while they enter the field of view in an overtaking motion, and later become occluded by either another vehicle, existing from where entered in a backward relative motion, or simply become too distant to detect, thereby never being detected at all throughout their trajectory. The vehicle on the far left lane is at 51 frames since entering the scene with a higher relative speed to the ego-vehicle, never to be detected by the appearance-based detection process.

Haar-like features, the detections are currently limited vehicles that are entirely visible (Fig. 2). We therefore build upon the work in [7] to integrate additional cues from the two domains into the OVeRT system.

We use optical flow-based clustering to identify additional vehicles who would be missed with an appearance-based only detector. These vehicles are at the adjacent lanes exhibiting distinct motion (in opposite direction to most of the rest of the scene) while overtaking the ego-vehicle. The experiments were taken place using data collected on multilane highways. Vehicles with different relative motion in nearby adjacent lanes, or the lanes next to those, were analyzed. Since the appearance-based method for vehicle detection poorly detects a vehicle at the periphery of the view, with inaccurate detections (i.e. bounding boxes not fully covering the visible portion of the vehicle - Fig. 2), motion cues from the two modalities will complement the systems in [7], [6].

II. RELATED RESEARCH

This work is concerned with object detection from monocular and stereo-vision, as well as occlusion handling and overtaking vehicle detection techniques. Hence, we review state-of-the-art vehicle detectors using each of these two modalities, and then turn to relevant appearance and motion-based features for incorporating the domains.

Vehicle detection based on monocular input generally employs techniques relying on an appearance-based feature set. A review can be found in [13]. Common methods draw upon a Haar-like feature set and a cascade structure for the selection of features [14]. Another common appearance-based feature set follows Dalal and Triggs [15]. In order to increase the invariance of these methods to partial visibility and different orientations, recent efforts include the work of Felzenszwalb *et al.* [16], which was utilized in [17] for

vehicle detection using the deformable part-based model and a Latent SVM.

Stereo-vision techniques for vehicle detection may incorporate 3D information to detect and track vehicles using geometric models and temporal filtering. Different segmentation methods are usually employed. Barrois *et al.* [18] used clustering of optical flow for a foreground separation based on motion. Next, the pose of the vehicles proposed in the first step by the motion segmentation is estimated by fitting a cuboid geometry to the vehicle using a modified Iterative Closest Point algorithm. Hermes *et al.* [19] achieved stereo-based vehicle detection and tracking using a two-stage mean-shift algorithm. A particle motion pattern is utilized for learning trajectory patterns. Pantilie *et al.* [20] used depth maps with optical flow to acquire the 3D position, velocity, and orientation information of vehicles and pedestrians, providing a general methodology for using stereo-vision input for separating moving objects from static ones. Lefebvre and Ambellouis [21] used Mean-shift tracking of 3D point clouds, forming objects using sparse stereo-matching. Perrollaz *et al.* [22] used optical flow and a spatio-temporally smoothed occupancy grid. Erbs *et al.* [23] also incorporate motion-based features by tracking stixels and fitting a cuboid model for vehicle detection. In general, due to the movement of the camera, successfully incorporating motion features into vehicle detection schemes is challenging, as the camera motion needs to be accounted for. This may be done using ego-motion compensation [5], [24] or long-term analysis of point trajectories [25].

As part of the OVeRT framework, cues from both monocular and stereo-vision domains are incorporated in order to track vehicles. We review previously suggested techniques for such cue integration. The use of both monocular and stereo-vision cues typically manifests itself in the use of

monocular vision for detection, and stereo-vision for 3D localization and tracking. Stein *et al.* [9] showed that monocular vision can detect objects that are missed in stereo-vision approaches when objects lie close in 3D space. Toulminet *et al.* [26] combined input from the two domains to detect and track the preceding vehicles in the ego lane. The depth image is used for segmentation and edge extraction. These are combined with monocular-based features, such as symmetry operators, for the final detection. Tracking is performed by employing cross-correlation over the frames. Sivaraman and Trivedi [7] showed that formulating the tracking of the monocular appearance-based vehicle detections using an approach fusing both monocular and stereo domains was superior to tracking in the image plane alone. Depth-based features can be used to reduce occlusion-related errors. HOG applied to the monocular image, optical flow image, and depth image, were combined as an input to a SVM in [8] for pedestrian detection showing improvement in the performance on occluded objects.

We extend the vehicle detection and 3D tracking system using stereo-monocular fusion proposed in [7] to include detection and tracking of non-rear view objects, such as in the case of overtaking vehicles. Cues from both domains are integrated to achieve this goal. Since appearance-based features are prone to occlusions and varying appearance, they tend to produce high false positive rates in complex driving environments. Furthermore, a trained appearance-based classifier on vehicles' rear is not sufficient for detection of vehicles at different orientation views, such as in the case of overtaking vehicles. In such cases, motion-based features have been shown to work well. Recently, Garcia *et al.* [27] combined radar and optical flow derived from monocular images for overtaking vehicle detection. Based on the candidate objects given by the radar modality, optical flow is used in a particular forward direction and threshold to determine an overtaking vehicle. The optical flow is calculated in a window around the radar detection. The motion features were shown to be useful during night time as well.

III. PARTIALLY OCCLUDED VEHICLE RECOGNITION AND TRACKING (OVERT) IN 3D

The novelty of this work lies in further generalizing an object detection system to handle partially occluded vehicles. In particular, we would like to leverage the distinct motion cues of overtaking vehicles since they enter the field of view and until they become fully visible. We first detail the algorithm for fully-visible vehicles, and then the OVERT development.

A. Fully-Visible Vehicle Detection and Tracking using Stereo-Monocular Cues

A vehicle detector from monocular input was trained using an Adaboost cascade of Haar-like rectangular features, as mentioned in [6], [14]. An active-learning framework 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. Active-learning has shown to significantly increase the performance of a classifier for vehicle detection [28].

The left monocular image and the depth image were calibrated to have the same coordinate system. Because the depth map contains significant noise, spatial filtering provides an improved estimation. For a given bounding box, parametrized by its top left corner coordinates, width, and height, as described in Equation 1, we calculate the median of the window to define the longitudinal distance of the vehicle (Equation 2).

$$\begin{aligned} v_k &= [i_k \ j_k \ w_k \ h_k]^T \\ Z_k &= \text{median}(D_{v_k}), \\ D_{v_k} &= \{\text{depth pixels in } v_k\} \end{aligned} \quad (1)$$

Hence we can derive the real-world coordinates of the vehicle by projecting the centroid of the box, using Z_k , to get the other two 3D coordinates, X_k and Y_k . This provides the full state vector, given in Equation 2.

$$V_k = [i_k \ j_k \ w_k \ h_k \ X_k \ Y_k \ Z_k \ \Delta X_k \ \Delta Y_k \ \Delta Z_k]^T \quad (2)$$

$$\begin{aligned} V_{k+1} &= \begin{pmatrix} \mathbf{I}_{4 \times 4} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{3 \times 3} & \Delta_t \mathbf{I}_{3 \times 3} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{3 \times 3} \end{pmatrix} V_k + \eta_k \\ M_k &= (\mathbf{I}_{7 \times 7} \ \mathbf{0}) V_k + \xi_k \end{aligned} \quad (3)$$

Each vehicle is tracked using a single Kalman filter, with the full state-space system given in Equation 3, where η_k and ξ_k are the plant and observation noise, respectively. The state transition matrix and the observation matrix are given in 3 as well.

B. Motion-based Vehicle Detection of Partially Visible Vehicles

In order to localize vehicles in the two search windows (shown in Fig. 1), we combine bounding box proposals, both from an optical flow-based clustering and depth-based segmentation. Only nearby objects are looked upon for overtaking motion, hence we threshold the depth image to produce a set of vehicle bounding boxes proposals, v_d^j , $j = 1, 2$ (for the left and right search windows). The resulting image undergoes morphological opening with a disk-shaped structuring element of radius 4. We search for the largest connected component in each of the left and right search windows.

Next, the optical flow is computed in the search windows and filtered so that only optical flow in the orientation of the overtaking motion is kept and used for clustering of pixels with similar motion. Optical flow is a well known computer vision technique for motion estimation. We use a parallel implementation of the coarse-to-fine Lucas-Kanade [29] optical flow algorithm.

Optical flow is derived based on the constraint of a constant brightness profile. The approximation for the solution produces the velocity components (V_x, V_y) for an image $I(x, y)$. These can be used to determine a magnitude and an orientation, (Δ, θ) .

Flow Attributes 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 $i = (x, y)$ in the $m \times n$ discrete image signal I , with optical flow orientation within a certain range. For instance,

$$\Theta_I = \{i \in I | \theta(i) \in [0, \frac{\pi}{2}]\}$$

for detecting left overtaking motion. Alternatively, for the right search window we use the range $[\frac{\pi}{2}, \pi]$. Next the optical flow vectors that fit the appropriate range above are clustered together into one, inclusive bounding box, v_{OF}^j , for $j = 1, 2$. Angular restriction limits the flexibility of the system-as not all overtaking scenarios will produce such optical flow. Nonetheless, such an approach is experimentally validated to be much superior to the baseline. Following a similar approach for undertaking and overtaking vehicles with relative motion to the ego-vehicle was also successful.

Given such set of motion vectors, we can perform a classification of overtaking occurrences by choosing a function $\varphi(\Delta_I) : \mathbb{R}^{|\Theta_I|} \rightarrow \mathbb{R}$, and thresholding it. We compare three choices: Entropy, L1-norm, and L2-norm. The frequency of vector occurrences in a certain direction, $|\Theta_I|$ (the Cardinality) is also of interest.

A detection event must last longer than two consecutive frames in order to be inputted to the tracker described in Section III-A. The detection bounding box is given by the intersection of the bounding box from the optical flow clustering and the depth-based clustering, $v_{OT} = v_{OF}^j \cap v_d^j$.

Finally, once the vehicle becomes fully visible and is detected by the appearance-based detector in Section III-A, we associate the tracks using the minimum euclidean distance in 3D of the centroids of the bounding boxes:

$$k^{t+1} = \underset{k^t}{\operatorname{argmin}} \|c_{OT}^t - c_k^{t+1}\|_2 \quad (4)$$

where v_k^{t+1} are all the detections bounding boxes at time $t + 1$, with centroids $c_k^{t+1} \in \mathbb{R}^3$, and $c_{OT}^t \in \mathbb{R}^3$ is the centroid of the bounding box of v_{OT} , the bounding box of the overtaking vehicle given by the optical flow and depth-based clustering.

IV. EXPERIMENTAL EVALUATION AND DISCUSSION

A. Experimental Setup in LISA Testbeds

Data has been captured using a stereo rig, looking forward. Video of the left and right monocular images, and the depth image which is aligned with the left view is captured at a resolution of 500×312 at 25 fps. Stereo matching is implemented using [30].

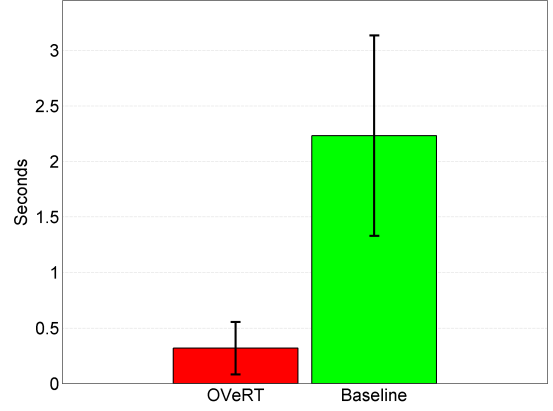


Fig. 3: Vehicle localization of partially occluded vehicles as they enter the scene in a relatively higher speed. OVeRT's tracking is initialized, on average, within 0.32 seconds or 8 frames of the vehicle entering the field of view (standard deviation = 0.23 seconds). On the other hand the baseline appearance-based only detection system [7] detects such vehicles on average within 2.32 seconds (55.8 frames) of the vehicle entering the scene (standard deviation = 0.90 seconds).

The proposed system was evaluated on multi-lane highway driving settings, captured using the calibrated vehicle-mounted stereo-rig. A total of 60 overtakings were annotated from 4 video sequences, providing about 4800 frames of annotated positive examples. Furthermore, 5000 frames were annotated with negative examples, of either no vehicle in the motion-based detector search windows or when the vehicles there were being overtaken by the ego-vehicle. Overtaking was defined as a vehicle at a higher speed entering the scene from the left or the right nearby lanes, as well as the lanes next to those. The instance a part of the vehicle was visible in the video to a human was defined as the beginning of the overtaking. This allows us to evaluate how long into the overtaking each detector fires up, OVeRT or the baseline detector and tracker from [7], and compare the two. Usually, the vehicle becomes fully-visible and is detected by the appearance-based detector. Nonetheless, in some instances the appearance-based detector never fires up, from the time the vehicle enters the field of view and until it disappears. For instance, a vehicle may begin an overtake but then becomes steady or slows down (or the ego-vehicle speeds up) resulting in a backward motion and existing from the scene without being detected by the system in [7] at all.

The improvement of OVeRT on the baseline in [7] is depicted in Fig. 3, with a specific example of the detection and tracking in Fig. 4. Fig. 5 evaluated the performance of the motion-based detection process of OVeRT on overtaking vehicles in nearby lanes to the ego-lane. Vehicles in adjacent lanes produce more optical flow and are easier to detect than vehicles in the lanes next to those and are further. In Fig. 3, we note that on average, OVeRT produce reliable detections within about 8 frames of the vehicle entering the field of

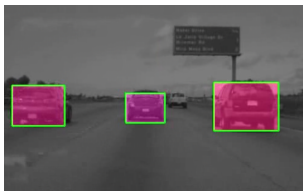
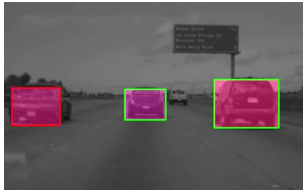
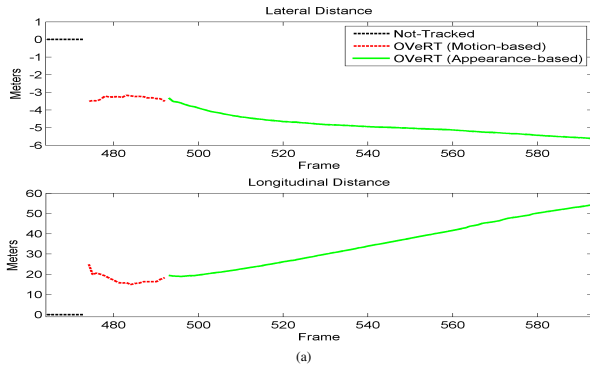


Fig. 4: Evaluation of the performance of the OVerT system. (a) Tracking results of an overtaking vehicle’s trajectory using OVerT. The baseline system maintains detection only along the plotted green line. Sequence of images of the tracked vehicle: (b) A vehicle on the left (black box) with higher relative motion to the ego-vehicle enters the scene, image shown is at 4 frames since vehicle entered field of view. (c) OVerT tracking begins after two frames using motion cues, 9 frames since visibility. (d) The scene after 37 frames, before handing the tracking to the Appearance-based detector, still only motion-based cues. (e) 38 frames since entering the scene, the vehicle is detected using appearance-based features only (this is where the baseline method [7] would have detected the vehicle).

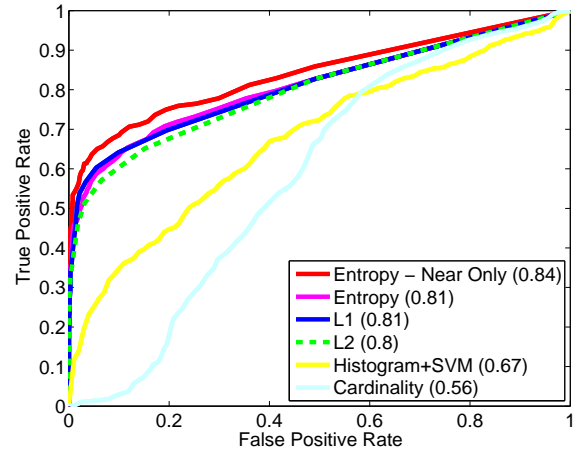


Fig. 5: Evaluation of the detections provided by OVerT using different motion attributes. In parenthesis is the AUC measure. ‘Near only’: performance on instances of vehicles in adjacent lanes to the ego-lane only. See Section III-B for more detail on the methods.

view, or about 2 seconds before the baseline system in [7] detects and tracks them. As the vehicles produce a strong and distinct motion cue very early on, this is expected.

Fig. 3 shows only vehicles in the dataset that were eventually detected by one of the detectors or both. It should be noted, that out of the 60 annotated instances, some vehicles which overtake are not detected at all throughout their visible trajectory by both OVerT and the baseline system. Most of the missing detections occur due to the challenging choice of some of the overtaking instances-where a vehicle may begin an overtake and then become occluded by another vehicle (as happens with vehicles in the farther lanes). This occurred in 13 out of the 60 annotated samples with OVerT, and 30 out of the 60 with the baseline system.

V. CONCLUDING REMARKS AND FUTURE WORK

Vision-based vehicle detection must be robust to the common partial visibility of the objects. We have proposed a system which improves upon a baseline for detecting and tracking vehicles in 3D with different relative motion to the ego-vehicle. We have explored our approach using a dataset of overtaking vehicles, where an appearance-based detector is prone to errors and missed detections. Future work should involve studying and addressing possible false detections as a result of the optical-flow motion cue detector. Track handling, as described in [10] can be used. Additionally, the threshold method should be further tested and compared with other possible approaches under more intricate urban settings. Clustering based on long-term point trajectories in the image can provide more reliable motion saliency cues, as well as allow for motion-cue integration in the entire scene, as opposed to just in the periphery.

VI. ACKNOWLEDGEMENTS

We acknowledge the support of the UC Discovery Program, associated industry partners, and our UCSD LISA colleagues. We thank the reviewers for their constructive comments.

REFERENCES

- [1] "Traffic safety facts," National Highway Traffic Safety Administration, Washington, D.C., Tech. Rep. DOT HS 811 402, 2011. [Online]. Available: <http://www-nrd.nhtsa.dot.gov>
- [2] A. Doshi, S. Y. Cheng, and M. M. Trivedi, "A novel active heads-up display for driver assistance," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 39, no. 1, pp. 85–93, 2009.
- [3] B. T. Morris and M. M. Trivedi, "Vehicle iconic surround observer: Visualization platform for intelligent driver support applications," in *IEEE Conf. Intell. Veh. Symp.*, 2010.
- [4] T. Gandhi and M. M. Trivedi, "Vehicle surround capture: Survey of techniques and a novel omni video based approach for dynamic panoramic surround maps," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 293–308, Sep. 2006.
- [5] —, "Parametric ego-motion estimation for vehicle surround analysis using an omnidirectional camera," *Machine Vision and Applications*, vol. 16, no. 2, pp. 85–95, 2005.
- [6] 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.
- [7] S. Sivaraman and M. M. Trivedi, "Combining monocular and stereo-vision for real-time vehicle ranging and tracking on multilane highways," in *IEEE Conf. Intell. Transp. Syst.*, 2011.
- [8] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. M. Gavrila, "Multi-cue pedestrian classification with partial occlusion handling," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
- [9] G. Stein, Y. Gdalyahu, and A. Shashua, "Stereo-assist: Top-down stereo for driver assistance systems," in *IEEE Conf. Intell. Veh. Symp.*, Jun. 2010, pp. 723–730.
- [10] Y.-C. Lim, C.-H. Lee, S. Kwon, and J. Kim, "Event-driven track management method for robust multi-vehicle tracking," in *IEEE Conf. Intell. Veh. Symp.*, 2011.
- [11] T. Kowsari, S. Beauchemin, and J. Cho, "Real-time vehicle detection and tracking using stereo vision and multi-view adaboost," in *IEEE Conf. Intell. Trans. Syst.*, 2011.
- [12] Q. Baig, O. Aycard, T. D. Vu, and T. Fraichard, "Fusion between laser and stereo vision data for moving objects tracking in intersection like scenario," in *IEEE Conf. Intell. Veh. Symp.*, 2011.
- [13] S. Sivaraman and M. M. Trivedi, "A review of recent developments in vision-based vehicle detection," in *IEEE Conf. Intell. Veh. Symp.*, 2013.
- [14] P. Viola and M. Jones, "Robust real-time face detection," *IEEE Intl. Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2005.
- [16] P. Felzenszwalb, R. Girshick, and D. McAllester, "Cascade object detection with de-formable part models," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
- [17] A. Takeuchi, S. Mita, and D. McAllester, "On-road vehicle tracking using deformable object model and particle filter with integrated likelihoods," in *IEEE Conf. Intell. Veh. Symp.*, Jun. 2010.
- [18] B. Barrois, S. Hristova, C. Wohler, F. Kummert, and C. Hermes, "3D pose estimation of vehicles using a stereo camera," in *IEEE Conf. Intell. Veh. Symp.*, 2009.
- [19] C. Hermes, J. Einhaus, M. Hahn, C. Wohler, and F. Kummert, "Vehicle tracking and motion prediction in complex urban scenarios," in *IEEE Conf. Intell. Veh. Symp.*, 2010.
- [20] C. Pantilie and S. Nedevschi, "Real-time obstacle detection in complex scenarios using dense stereo vision and optical flow," in *IEEE Conf. Intell. Trans. Syst.*, 2010.
- [21] S. Lefebvre and S. Ambellouis, "Vehicle detection and tracking using mean shift segmentation on semi-dense disparity maps," in *IEEE Conf. Intell. Veh. Symp.*, 2012.
- [22] M. Perrollaz, J.-D. Yoder, A. N. andgre, A. Spalanzani, and C. Laugier, "A visibility-based approach for occupancy grid computation in disparity space," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1383–1393, 2012.
- [23] F. Erbs, A. Barth, and U. Franke, "Moving vehicle detection by optimal segmentation of the dynamic stixel world," in *IEEE Conf. Intell. Veh. Symp.*, 2011.
- [24] K. S. Huang, M. M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," in *IEEE Conf. Intell. Veh. Symp.*, 2003.
- [25] T. Brox and M. Jitendra, "Object segmentation by long term analysis of point trajectories," in *European Conf. of Computer Vision*, 2010.
- [26] G. Toulminet, M. Bertozzi, S. Mousset, A. Bensrhair, and A. Broggi, "Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis," *IEEE Trans. Image. Proc.*, Aug. 2006.
- [27] F. Garcia, P. Cerri, A. Broggi, A. de la Escalera, and J. Armingol, "Data fusion for overtaking vehicle detection based on radar and optical flow," in *IEEE Proc. Intell. Veh. Symp.*, 2012.
- [28] S. Sivaraman and M. M. Trivedi, "Active learning for on-road vehicle detection: A comparative study," *Machine Vision and Applications*, 2011.
- [29] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of Imaging Unertstanding Workshop*, 1981.
- [30] (2011, Jan. 27) Tyzx dense stereo. [Online]. Available: <http://www.tyzx.com>.