IEEE International Conference on Pattern Recognition 2014

Go with the flow: Improving Multi-View Vehicle Detection with Motion Cues

Alfredo Ramirez, Eshed Ohn-Bar, and Mohan M. Trivedi
LISA: Laboratory for Intelligent and Safe Automobiles
University of California San Diego
La Jolla, CA, 92092 USA
alr034@eng.ucsd.edu, eohnbar@ucsd.edu, mtrivedi@ucsd.edu

***Abstract*—As vehicles travel through a scene, changes in aspect ratio and appearance as observed from a camera (or an array of cameras) make vehicle detection a difficult computer vision problem. Rather than relying solely on appearance cues, we propose a framework for detecting vehicles and eliminating false positives by utilizing the motion cues in the scene in addition to the appearance cues. As a case study, we focus on overtaking vehicle detection in a freeway setting from forward and rear views of the ego-vehicle. The proposed integration occurs in two steps. First, motion-based vehicle detection is performed using optical flow. Taking advantage of epipolar constraints, salient motion vectors are extracted and clustered using spectral clustering to form bounding boxes of vehicle candidates. Post-processing and outlier removal further refine the detections. Second, the motion-based detections are then combined with the output of an appearance-based vehicle detector to reduce false positives and produce the final vehicle detections.**

I. Introduction

Robust object detection is an active branch in the pattern recognition and computer vision community, particularly in object detection from multiple views. Recent progress is mostly centered around integration of multiple models. For instance, a different model can be learned for a particular view and represented as a component in a mixture model [1], [2], [3], [4]. In this work, we use motion cues in order to augment the detections produced by such methods. This integration shows improved performance, both in terms of increased true positive rate and decreased false positive rate.

In this work, we focus on vehicle detection from two wide-angle monocular cameras capturing the forward and backward views of a vehicle, which is an important application of multi-view object detection. The last decade has seen an emergence of multiple sensors and intelligent systems in vehicles for driver assistance and autonomous driving. In particular, computer vision and radar system have become increasingly common in modern vehicles to sense surroundings and obstacles. Monocular computer vision systems, which are cheap and easy to install, face many challenges. For instance, images of vehicles vary in aspect ratio, scale, and shape as they travel throughout a scene. This is especially true for the most critical of vehicles: those that are in close proximity to the ego-vehicle. On the opposite side of the spectrum are distant vehicles, which are easier to model and detect because they appear relatively static in the scene, and are a popular subject of study in much of the vehicle detection work.

Overtaking vehicles are inherently risky and critical to the driver due to the high relative velocity needed for an overtake, and because vehicles become occluded or obscured within blind spots as they come from behind to pass. The monocular vision based system is advantageous because it can be integrated with other camera systems on the vehicle, such as a rear-facing backup camera, or a forward-facing lane detection camera. In this paper, we extensively evaluate the proposed algorithm on video data of real world driving in a freeway setting. We subsequently compare the performance of our algorithm to state-of-the-art appearance-based detectors.

Fig. 1: Motion cues can be integrated with an appearance-based detector to handle partial visibility, occlusion, and remove false positive detections. Left: detections from an appearance based detector (deformable part model from [1]). Right: detections from appearance based detector with proposed motion cues integrated. Detections are visualized as a yellow box, and motion flow vectors are visualized in blue.

This work includes two main contributions: given a trained appearance-based vehicle detector, we remove false positive detections using motion information from the scene. Furthermore, clustering moving objects in the scene provides additional proposals, which are integrated to produce a significant improvement in overall vehicle detection. Motion cues have the distinct advantage of being available even when a vehicle is partially visible, such as at the edge of the field of view, or when occluded by other objects or vehicles. Thus, the vehicle detection can provide a more complete and continuous

Research Study	Features	Description
Lenz *et al.* [5]	Stereo and Motion	Stereo disparity and optical-flow are clustered and tracked over time.
Garcia *et al.* [6]	Radar and Motion	Raw radar data is projected to the camera image and associated with optical flow.
Jazayeri *et al.* [7]	Motion	Scene motion is probabilistically modeled and an HMM separates vehicles from the background.
Wang, Bebis, and Miller [8]	Motion	Homogeneous sparse optical-flow and block-based eigenspaces are used to model and seperate the foreground and background regions of the scene.
Ohn-Bar, Sivaraman, and Trivedi [9]	Stereo, Appearance and Motion	Vehicles are localized using active-learning based appearance detector and optical-flow at edges of the scene, then tracked with stereo disparity and a Kalman filter.
This study	Appearance and Motion	Optical flow and the epipolar geometry of the scene are used to find initial vehicle candidates and then compared to appearance-based detections to refine final detections and remove false positives.

TABLE I: Selection of recent vehicle detection methods which incorporate motion cues.

trajectory, especially when the detected vehicle is near the ego-vehicle. This is highly useful for vision-based scene analysis for driver assistance and criticality assessment. Additionally, motion cues can still be extracted when there is visual distortion in the scene, such as from surround view systems (e.g. from a Point Grey Ladybug camera system, or from a panoramic system [10]). Since appearance-based detectors can be negatively affected in this situation, motion cues can be a very useful addition, as is shown in [11].

II. Related Research Studies

Object detection in computer vision is a long studied area, with appearance based detectors being quiet popular. There are a wide range of appearance based detectors (a survey can be found in [12]). Recent algorithms, such as the deformable parts model (DPM) detector [1], specifically reason over the appearance variation that results from multiple views of object categories. Nonetheless, we found that these methods are still limited in terms of occlusion handling. They also rely on extensive training on large datasets, which affects their performance. On the other hand, motion cues from objects changing in orientation, various lighting conditions, and with partial occlusions, can still be extracted from a video.

Recent studies have found success in motion-based vehicle detectors, a list of which can be seen in Table I. For example, Jazayeri *et al.* [7] use a probabilistic model of the motion cues in the scene, and a hidden Markov model to detect the moving vehicles. In addition, we found motion cues to be important when considering two wide angle views: the front and rear of the vehicle. This motivated us to incorporate motion-based cues, as these are more robust to changes in appearance, with appearance-based cues to improve performance of the proposed vehicle detector. Recently, in [9], an active-learning vehicle detector was combined with the motion-cues from the edges of the scene to accurately detect vehicles earlier in the video sequence.

III. Robust Motion Cue Extraction For Vehicle Detection

The proposed vehicle detector begins with the optical flow calculation of the video sequence. We investigated three optical flow algorithms: Lucas-Kanade [13], Brox and Malik [14], and Farnebäck [15]. We found that a GPU accelerated version of the Farnebäck optical flow algorithm provided relatively accurate and quick results, and thus, this was used for generating the results in this paper. Nevertheless, we found that all three optical-flow algorithms produced incorrect flow at certain pixels (an example of the Farnebäck dense optical-flow is seen in Fig. 3(a)).

Consistency Check Dense optical flow is produced both forwards and backwards at every frame in the video sequence in order to address issues with wrongly tracked points. The dense optical flow is sampled at corners in order to produce accurate flow vectors. A consistency check of the sampled flow vectors is performed by checking whether the forward and backwards vectors are inverses of each other. Flow vectors that are incorrectly calculated, like at edges of frames, or at occlusions, will fail the consistency check, improving the average accuracy of the flow vectors. The consistency check

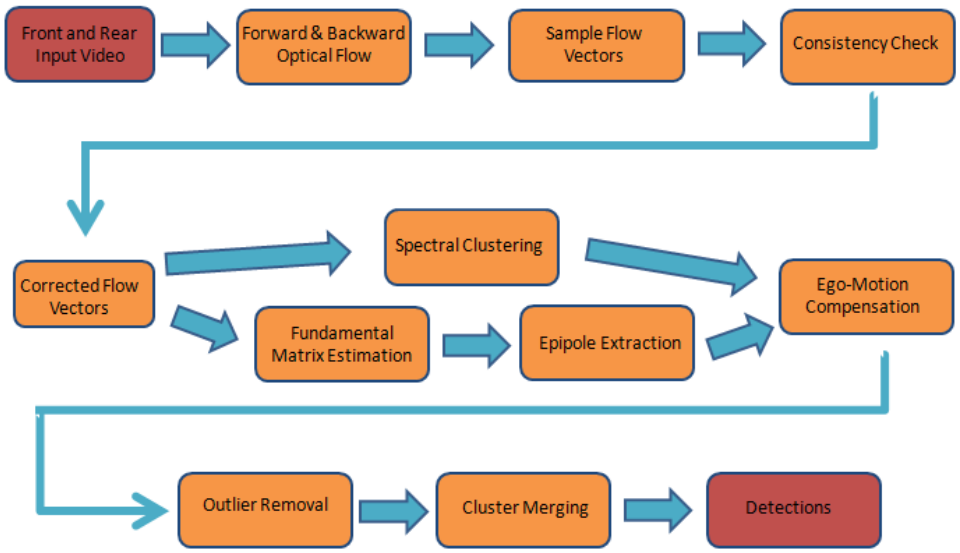

Fig. 2: Flowchart illustrating the proposed algorithm for the overtaking vehicle detector using motion cues.

(a) (b) (c) (d)

Fig. 3: The consistency check for sampled optical-flow vectors can remove any incorrectly calculated motion vectors from consideration, particularly at the edges of the field of view, or near occlusions. (a) The forward dense optical flow for a freeway scene. (b) The forward optical-flow vectors sampled at strong corners. (c) The sampled backwards optical-flow vectors. (d) The corrected optical-flow vectors, with many texture-less areas and boundary errors resolved.

used is:

$$||V_f - V_b||_2 \leq t_{consistent} \quad (1)$$

Where V_f is the forwards flow vector, and V_b is the backwards vector, and $t_{consistent}$ is an accuracy threshold. The results of the consistency check can be seen in Fig. 3.

Epipole Extraction The corrected flow vectors are next used to estimate the fundamental matrix for the current pair of frames. This is done to leverage the epipolar geometry of the moving camera setup. The fundamental matrix is estimated using RANSAC and the 8-point algorithm [16], with the flow vectors used as corresponding point pairs. Next, the focus of expansion of the initial frame is estimated by extracting the epipole coordinates from the fundamental matrix.

Spectral Clustering The flow vectors are clustered together using an iterative spectral clustering algorithm recommended in [17]. The clustered flow vectors for a sample frame can be seen in Fig. 4. The flow vectors are clustered with respect to their similarity in flow and spatial proximity in the frame. The spectral clustering algorithm tightly grouped foreground objects when compared to simple k-means clustering, which would often extend out from the foreground objects and include background pixels in the cluster.

Ego-motion Compensation In order to separate the motion clusters belonging to overtaking vehicles from the background, ego-motion compensation must be performed. Flow vectors moving in contradiction to the focus of expansion are not following the epipolar-geometry model, and are thus assumed to belong to an independently moving vehicle, possibly performing an overtaking. The amount of motion is characterized by the deviation from the vector emitted from a point in the previous frame to the epipole, l_v, and the estimated optical flow vector, v.

$$v \cdot l_v \leq t_{moving} \quad (2)$$

t_{moving} is a threshold for handling estimation errors in the epipole location. For forward motion, the dot product is expected to be positive, as the direction of motion should roughly follow the direction of l_v. For the rear we use the inverse of the condition, as we expect the two vectors to point in opposite directions.

Cluster Merging The spectral clustering provides a slight over-segmentation of the flow, so objects in the scene may

(a)

(b)

Fig. 4: The spectral clustering algorithm tightly groups optical-flow vectors for the objects in the scene. (a) The output of the spectral clustering step on the sampled optical flow vectors. The clustered motion vectors are shown, where each cluster is visualized with a unique color. The centroid and average flow-vector of the clusters are visualized as large dots, and thick black lines, respectively. (b)Motions vectors after ego-motion compensation. Note that only vectors remaining belong to the vehicle entering the frame.

be divided into multiple clusters and must be merged back together. We assume two clusters belong to the same object if they have a similar average flow vector, and their centroids are in spatial proximity to each other. This is implemented as the following inequality:

$$||c_1 - c_2||_2 + ||v_1 - v_2||_2 \leq t_{similar} \quad (3)$$

Where c_1 and c_2 are the centroid coordinates of a pair of clusters , and v_1 and v_1 are the average flow vectors of the pairs of clusters. Here, we assume that clusters belonging to the same object or vehicle will be in close proximity to each other in the image, and will exhibit a similar optical flow, expressed in the average optical flow term v_i. $t_{similiar}$ is a user set threshold for the level of similarity required for the merge.

Post Processing With the clusters merged, some final post-processing is done to clean the detections, as well as remove any false positives affecting the results. The final clusters

Fig. 5: Sample output detections for the motion-only algorithm from a forward and backward view. Note how vehicles are still detected, even as vehicles enter and exit the scene.

sometimes include outlier motion vectors that will affect the accuracy of the final detection. We assumed a Gaussian distribution in the origin location of the optical flow vectors, and calculate the standard deviation of the vector origins for each cluster. Optical-flow vectors that exceed two standard deviations away from the centroid in either x or y direction are eliminated from the cluster. To further increase the accuracy of the final detections, clusters with very small bounding boxes, or with a small number of flow vectors are removed. Since vehicles near the ego-vehicle will appear large in the scene, small clusters are likely to belong to distance vehicles, or false positives, and can thus be removed from consideration. The final vehicle detections are presented as bounding boxes calculated from the flow vector clusters. Sample outputs of the motion-only detector can be seen in Fig. 5.

Motion Cue Integration With Appearance-Based Detector With motion and appearance-based vehicle bounding boxes calculated, the two sets of boxes are merged by taking their union and removing overlapping instances using non-maximal suppression. Bounding boxes found to have too little motion after ego-motion compensation are removed from consideration. This integration technique was found to significantly improve performance of the appearance-based object detectors.

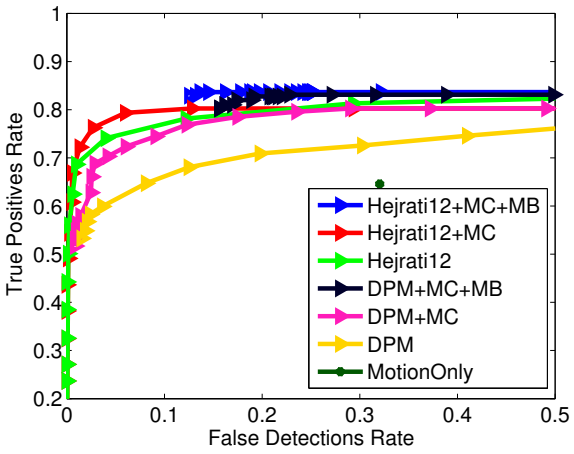

Fig. 6: Two appearance-based detectors were used in our experiments: the deformable parts model (DPM) from [1] and Hejrati and Ramanan [2]. The proposed motion integration with a 20% overlap threshold was also evaluated. Integration of the motion-based detection boxes (**MB**), along with the false positive removal scheme using motion compensation (**MC**) show significantly improved results over the state-of-the-art methods.

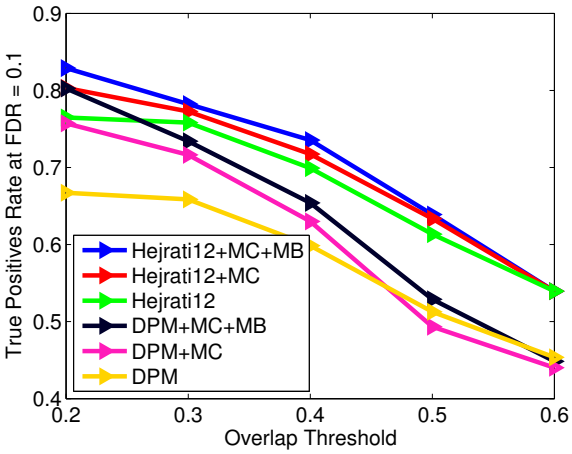

Fig. 7: Measuring the localization accuracy of the algorithms. Note how motion boxes produce higher true positive rates at lower overlap threshold requirements. Because the motion detections are not tight, performance deteriorate as the localization requirement increase.

IV. Experimental Setup

The video data used throughout this work was acquired using two monochrome Point Grey Flea3 cameras fitted with Theia wide-angle lenses, located on the windshield of the vehicle, pointing forwards, and on the roof of the vehicle, pointing backwards. The data was collected in a freeway setting, during the day, with light to moderate traffic. Ground truth data consists of 1327 vehicle instances containing 20 overtakings. Tested algorithms are evaluated by comparing the overlapping areas between the ground truth data, and the detected bounding boxes. We found that our proposed algorithm greatly improves on the standard Viola Jones method [18] trained on images of the front and rear of vehicles, as the changing aspect ratio caused many misses. To better evaluate our proposed algorithm, two state-of-the-art detectors were tested: the deformable part model in [1] and an extension of the DPM from Hejrati and Ramanan [2].

V. Experimental Evaluation

Generally, Hejrati12 performed significantly better than the DPM on our dataset, as it is designed to detect occluded and partially visible vehicles. Furthermore, the DPM produces detections only when a vehicle has completely entered the scene. Additionally, although the motion-only algorithm is proficient at detecting vehicles as they come into the scene, non-tight vehicle bounding boxes are sometimes outputted. Nevertheless, as shown in Fig. 6, with a 0.2 overlap requirement for a true positive, a significant improvement is shown with both methods of motion integration. Two motion integration techniques are evaluated in this work: motion based detections (referred to as **MB**), and the motion compensation scheme (referred to as **MC**), which removes many false positive detections.

Although the motion cues in the scene are strong and realiable in proximity to the ego-vehicle, the detections can lack tightness around overtaking vehicle. Therefore, we also evaluate how changing the overlap threshold between the ground truth bounding box and the detected bounding box affects true positive rate, seen in Fig. 7. We note that there is a great improvement in all tested algorithms when the overlap threshold is relaxed. For example, the DPM algorhtm (which is significantly faster than the Hejrati12 algorithm) is significantly improved with the motion integration. The advantage of our algorithm is that no training is needed for the motion step, and can thus be generalized for any existing appearance-based vehicle detector to produce more continuous detections. However, even with the post-processing steps, the tightness of the detected motion-based bounding boxes must be improved. Nonetheless, our proposed motion integration method is promising, and the **DPM+MC+MB** scheme is comparable to the occlusion-reasoning Hejrati12 algorithm with relaxed overlap requirements. The motion-only algorithm runs at approximately one fps using GPU acceleration for calculating dense optical flow. At this speed, the algorithm is much faster than the two state-of-the-art methods evaluated in this work, but further speedups will be required for this algorithm to be used in real-time applications.

VI. Concluding Remarks and Future Work

We presented a method for integrating motion cues with an appearance-based vehicle detector for multi-view vehicle detection. Evaluation was done by improving detection rates of state-of-the-art vehicle detection algorithms. Future work should include additional outlier removal schemes and post processing for tightening the output bounding boxes of our algorithm. Additionally, more sophisticated approaches involving a combination of the appearance-based and motion-based detections into the spectral clustering as proposed in [19] can be pursued. Also, the proposed method provides a more complete vehicle trajectory, which can be used as part of a traffic pattern analysis system as in [20].

ACKNOWLEDGMENT

The authors would like to thank associated industry partners, the reviewers for their constructive notes, and the members of the Laboratory for Intelligent and Safe Automobiles (LISA) for their assistance.

REFERENCES

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," in *PAMI*, vol. 32, no. 9, Sep 2010.

[2] M. Hejrati and D. Ramanan, "Analyzing 3d objects in cluttered mages," in *NIPS*, 2012, pp. 602–610.

[3] C.-H. Kuo and R. Nevatia, "Robust multi-view car detection using unsupervised sub-categorization," in *WACV*, Dec 2009, pp. 1–8.

[4] E. Ohn-Bar and M. M. Trivedi, "Fast and robust object detection using visual subcategories," in *CVPRW*, 2014.

[5] P. Lenz, J. Ziegler, A. Geiger, and M. Roser, "Sparse scene flow segmentation for moving object detection in urban environments," in *Intelligent Vehicles Symposium (IV)*, 2011.

[6] 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 *IV*, 2012.

[7] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," *TITS*, vol. 12, no. 2, pp. 583–595, June 2011.

[8] J. Wang, G. Bebis, and R. Miller, "Overtaking vehicle detection using dynamic and quasi-static background modeling," in *CVPR*, 2005.

[9] E. Ohn-Bar, S. Sivaraman, and M. M. Trivedi, "Partially occluded vehicle recognition and tracking in 3d," in *IV*, June 2013.

[10] A. Ramirez, E. Ohn-Bar, and M. M. Trivedi, "Panoramic stitching for driver assistance and applications to motion saliency-based risk analysis," in *ITSC*, Oct. 2013.

[11] ——, "Integrating motion and appearance for overtaking vehicle detection," in *IV*, June 2014.

[12] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *TITS*, vol. 14, no. 4, pp. 1773–1795, Dec 2013.

[13] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of Imaging Unertstanding Workshop*, 1981.

[14] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *PAMI*, vol. 33, no. 2, pp. 500–513, 2011.

[15] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proceedings of the 13th Scandinavian Conference on Image Analysis*, ser. LNCS 2749, Gothenburg, Sweden, June-July 2003, pp. 363–370.

[16] R. Hartley, "In defense of the eight-point algorithm," *PAMI*, 1997.

[17] T. Sakai and A. Imiya, "Practical algorithms of spectral clustering: Toward large-scale vision-based motion analysis," in *Machine Learning for Vision-Based Motion Analysis*, ser. Advances in Pattern Recognition. Springer London, 2011, pp. 3–26.

[18] P. Viola and M. Jones, "Robust real-time object detection," in *IJCV*, 2001.

[19] K. Fragkiadaki, H. Hu, and J. Shi, "Pose from flow and flow from pose," in *CVPR*, June 2013, pp. 2059–2066.

[20] B. Morris and M. M. Trivedi, "Understanding vehicular traffic behavior from video: a survey of unsupervised approaches," in *JEI*, vol. 22, no. 4, 2013.