Integrating Motion and Appearance for Overtaking Vehicle Detection

Alfredo Ramirez, Eshed Ohn-Bar and Mohan Trivedi, Fellow, IEEE

Abstract—The dynamic appearance of vehicles as they enter and exit a scene makes vehicle detection a difficult and complicated problem. Appearance based detectors generally provide good results when vehicles are in clear view, but have trouble in the scenes edges due to changes in the vehicles aspect ratio and partial occlusions. To compensate for some of these deficiencies, we propose incorporating motion cues from the scene. In this work, we focus on a overtaking vehicle detection in a freeway setting with front and rear facing monocular cameras. Motion cues are extracted from the scene, and leveraging the epipolar geometry of the monocular setup, motion compensation is performed. Spectral clustering is used to group similar motion vectors together, and after post-processing, vehicle detections candidates are produced. Finally, these candidates are combined with an appearance detector to remove any false positives, outputting the detections as a vehicle travels through the scene.

I. INTRODUCTION

In the last few years, multi-modal sensors for driver assistance systems have become the norm for higher-end commercial vehicles. From side and rear radar for obstacle detection, to birds-eye-view and back-up cameras for parking assistance, modern vehicles have become more aware of their surroundings. To this end, this work focuses on a monocular computer vision system for detecting vehicles that overtake the ego-vehicle in freeway settings. Overtaking vehicles pose an inherent danger to the driver because of the high velocity needed to overtake on a freeway, and because vehicles must pass through a driver’s blind spot to perform an overtaking. We chose a monocular setup for this work because there is no costly or specialized equipment required (as opposed to a radar/lidar or a stereo camera rig). It is also possible to take advantage of monocular camera systems that are already integrated into the vehicle, like a forward facing lane camera, or a backup camera.

In this paper, we propose an algorithm for detecting overtaking vehicles using motion cues from the scene. Motion compensation of video data is performed using the optical flow of the scene and epipolar geometry. After post processing and outlier removal, overtaking vehicle candidates are produced. By combining these outputs with the output of an appearance based vehicle detector, we are able to detect vehicles earlier in the scene and increase the accuracy of the detections when compared to only using an appearance detector. To evaluate our algorithm, extensive video from front and rear cameras was collected from freeway driving scenarios under normal traffic and weather conditions. The performance of our algorithm and of the state-of-the-art appearance based detectors was compared using this data.

II. RELATED RESEARCH STUDIES

Vehicle detection and tracking from a stationary camera has been a long studied problem in computer vision [1] [2]. However, on road vehicle and overtaking detection can be a much more difficult computer vision problem due to the dynamic movements of the camera and high velocity of the surrounding vehicles, and can be approached in a variety of ways [3] [4]. Appearance based vehicle detectors are very popular due to their good performance after thorough training. These types of algorithms range from the deformable parts model (DPM) detector [5], which is a HOG feature and latent SVM based detector, to the algorithm shown in [6], which uses an extension of DPM to analyze 2D landmarks to generate a 3D analysis of a vehicle. These detectors can perform well when a vehicle in the scene moves relatively slowly; where the vehicle stays mostly visible within the scene and remains within the same perspective. However, appearance detectors will have trouble performing well when vehicles enter and leave the scene frequently, causing partial occlusions, or when they travel through the scene so that a different view of the vehicle is seen by the camera (e.g. front of vehicle, side of vehicle, etc.). An alternative approach to vehicle and overtaking detection is a motion-based algorithm that takes advantage of the movement within the scene. For example, in [7], optical flow is used to model and separate the background into regions. Then, the background regions are removed from the scene to detect the overtaking vehicle. More recently, Garcia et al. integrated information from a radar system with optical flow information in the overtaking region of the scene to perform the detections [8]. Although motion can be a useful tool for overtaking vehicle detection, it can be be difficult to calculate in areas with little texture (such as on roads), and is limited by the accuracy of the optical flow algorithm used. Certain groups have previously found success in fusing appearance based detectors with motion cues. In [9], optical flow in the edges of the image frame in combined with an appearance based active-learning vehicle detector in order to detect and track vehicles.

III. PROPOSED METHOD

A flow diagram of the proposed motion algorithm can be seen in Figure 1. We first start by calculating the optical flow from the scene. Three optical flow algorithms were considered for this work: Lucas-Kanade [10], Farnebäck [11], and Brox & Malik [12]. We found that the Farnebäck algorithm provided fairly accurate dense optical flow for our...
A. Pre-Processing

To eliminate incorrectly tracked points, a backwards-forwards consistency check is performed. Backwards and forwards dense optical flow is calculated between consecutive video frames and sampled at strong corners within the images to produce flow vectors. Strong corners are easier to track and are more likely to produce accurate flow vectors, since their appearance will remain relatively uniform as vehicles move through the scene. We then compare backwards vector originating from the endpoint of a forwards vector, and ensure they are consistent (i.e. sum to 0) Inconsistent flow vectors can occur in areas where the optical flow can be incorrect, such as at edges of the frame, or in areas with occlusions. Therefore, we can remove these from consideration to improve the overall accuracy of the optical flow. The consistency check used in this algorithm is implemented as:

$$||V_f - V_b||_2 \leq t_c$$  \hspace{1cm} (1)$$

Where $V_f$ and $V_b$ are the forward and backward flow vectors, respectively, and $t_c$ is the consistency threshold. An example of this procedure can be seen in Figure 2.

B. Vehicle Candidate Detection

Next, the corrected flow vectors are used with the moving camera setup to take advantage of the stereo (epipolar) geometry of the scene. The flow vectors give corresponding
point pairs for consecutive frames, and with the knowledge that the camera is moving (since the vehicle is moving), the fundamental matrix can be calculated for the frames. The fundamental matrix provides geometric constraints on a pair of stereo images, which can provide information about the scene, such as how the camera is moving. With the point pairs, the fundamental matrix is estimated the 8-point algorithm [13] and random sampling consensus (RANSAC) to remove any leftover incorrect matches. Finally, the epipole (i.e. the focus of expansion) location in the first image is extracted from the fundamental matrix.

The corrected flow vectors are grouped together by similarity in flow and location using the spectral clustering algorithm recommended in [14]. The output of the clustering algorithm can be seen in Figure 3. We found that the spectral clustering algorithm outperformed simple k-means clustering in outlining objects in the scene more closely.

With the motion vectors clustered, we must now isolate the clusters belonging to overtaking vehicles. For a forward moving vehicle, objects in the scene which are stationary or moving slower than the ego-vehicle will have motion vectors moving away from the epipole. Objects moving in parallel with the ego-vehicle but at a higher velocity (i.e. overtaking the ego-vehicle) will have motion vectors moving towards the epipole. To check if a vector is moving towards the epipole, we use the following equation:

\[ v \cdot l_v \leq t_{motion} \]  

Where \( v \) is the motion vector in question, \( l_v \) is a unit vector located at the origin at \( v \) and pointing towards the epipole, and \( t_{motion} \) is a threshold. The result of the above equation will be large if both vectors point in similar directions, and small if they point in opposite directions. \( t_{motion} \) controls how much the angle between the vectors can vary.

C. Post Processing

The spectral clustering algorithm has the tendency to overcluster the motion vectors, therefore oversegmenting objects.

in the scene. To reduce the number of clusters, we join clusters with similar average motion and in close proximity. Two clusters are merged if the following inequality is satisfied:

\[ ||c_i - c_j||_2 + ||v_i - v_j||_2 \leq t_{similar} \]  

Where \( c_i \) and \( c_j \) are the centroid locations of the \( i^{th} \) and \( j^{th} \) clusters, and \( v_i \) and \( v_j \) are the average flow vectors of the clusters. Here, we assume that clusters belonging to the same object in the scene will be near each other and have similar motion.

The final step in the motion algorithm is to clean the merged clusters in order to produce the detection candidates. The final clusters sometimes include outlier motion vectors that are a large distance away from the rest of the motion vectors in the cluster. In order to remove these outlier vectors, we calculate the sample standard deviation of the x and y coordinates for each cluster. Motion vectors that are more than two standard deviations away from the centroid in the x or y direction are removed from the cluster. Next, bounding boxes for all of the clusters are calculated. To remove false positives from the final output, clusters with few of motion vectors or too small a bounding box are removed from the final detections. This is done because overtaking vehicles that are near the ego-vehicle in the scene will appear large in the image. Therefore, small clusters or clusters with small bounding boxes are most likely a false positive detection, or vehicles that are far away in the scene, which do not need to be detected. Finally, the remaining vehicle candidates are

Fig. 3: The results of the spectral clustering on the sampled optical flow vectors. Each cluster is displayed with a distinct color. The larger, filled circles show the centroid of the cluster, and the black line indicates the average flow vector magnitude and direction of the cluster.

Fig. 4: Sample output detections of the motion-only algorithm. Note that the vehicles are being detected as they enter the scene, even though they are still partially occluded.
Fig. 5: The LISA-X testbed is equipped with multiple cameras, both inside and out, as well as radar systems and other sensors.

... and an fdr of .6409 on the same data. This performance is expected, as the motion cues persist and stay strong.

The stitched view contains artifacts from the stitching motion only algorithm using a stitched wide-angle view from [5].

D. Integration of Motion Algorithm and Appearance Detector

Integration occurs in two ways. The sets of vehicle boxes from the two algorithms are joined by taking their union and removing overlapping instances using non-maximal suppression. Secondly, boxes without sufficient flow after ego-motion compensation are removed. We found these two techniques significantly improve performance of object detectors. Additionally, early detections from the motion algorithm are added to the detection sequence.

IV. EXPERIMENTAL SETUP

The dataset used for the experiment was collected using the LISA-X vehicle testbed, which can be seen in Figure 5. Two Point Grey cameras with Theia wide-angle lenses are mounted on the outside of the testbed: one mounted on the windshield pointing forward, and one attached to the rear-roof of the vehicle pointing backwards. These cameras collect 896x672 pixel monochrome video at 20 fps. The dataset consists of video of naturalistic driving under free-way settings with medium traffic. Ground truth overtaking instances of the data was annotated with bounding boxes for 1327 frames containing 20 overtakings. The ground truth data is used to evaluate the tested algorithms by comparing the amount of overlap between bounding boxes. For evaluation, we found a great improvement when comparing to the standard Haar like-Viola Jones method [15] trained on front and rear vehicles, as vehicles would often be missed at the sides of the vehicle in close proximity with changing of the aspect ratio. Instead, we turned to a recent object detector that also allows for changing of the aspect ratio and object view. The algorithm used for this is the deformable part model in [5].

To test a more extreme case, we also evaluated the motion only algorithm using a stitched wide-angle view from [16]. The stitched view contains artifacts from the stitching process, such as paralax errors in the overlapping regions, and stretching around the edges of the image, that would cause problems for vehicle detectors. A sample frames of this data can be seen in Figure 6. We use this view to test non-ideal video inputs, and to simulate video effects from surround view equipment, including the Point Grey Ladybug camera system, or omnidirectional cameras, such as in [17]. The dataset consists of 22 minutes of video and was gathered under similar conditions as the previous data, but limited to the rear view of the ego-vehicle.

V. RESULTS

In this section we evaluate the approach for motion integration with the state-of-the-art vehicle detectors from [5], referred to as DPM. The DPM algorithm would produce detections only when the vehicle is entirely visible and have entered the scene. At these moments, the vehicle bounding box can be quite large, and although the motion-based algorithm is efficient at detecting the presence of a vehicle, it sometimes produces non-tight bounding boxes. Nonetheless, with a 0.2 overlap requirement for a true positive, Figure 7 shows a significant improvement of the method by motion integration. We evaluate two integration techniques, the motion compensation (referred to as MC) which allows for removal of false positives and the use of motion cues for vehicle box generation (referred to as MB) in addition to the boxes generated by the baseline appearance detector.

As previously mentioned, although the motion cues are strong in the proximity of the ego-vehicle and are reliable in identifying the presence of a vehicle, the bounding boxes generated are not always tight. We therefore evaluate the effect of changing the overlap (intersection over union) threshold needed to produce a true detection in Figure 8. The greatest improvement comes when relaxing the overlap. For instance, the immensely popular DPM scheme can be greatly improved by introducing motion cues. The strength of our approach is enhanced by the fact that no training is needed, and can therefore be generalized for improvement of any existing appearance-based object detectors. Off the shelf classification algorithms can be enhanced to produce a more continuous object detection trajectory. Nonetheless, even with the outlier removal and post-processing, further work needs to be done in order to guarantee tightness of motion-boxes. Even so, the method shows much promise, and the DPM+MC+MB solution is competitive with other techniques at less tight overlap requirements. The motion algorithm as a whole takes approximately 1 second per frame using a GPU implementation of the optical flow algorithm. This is faster than the baseline method used in evaluation in this paper, yet future speedups and optimization are needed.

In evaluating the performance of the motion-only algorithm on the stitched rear view, we found that even in non-ideal settings, the algorithm still performed moderately well. The motion only algorithm was able to achieve a true positive rate (tpr) of .4618 and a false detection rate (fdr) of .8231. In comparison, the DPM algorithm was able to achieve a tpr of .4218 and an fdr of .6409 on the same data. This performance is expected, as the motion cues persist and stay strong.
Fig. 6: Samples of the stitched view data with vehicle detections produced by the motion only algorithm. Observe the heavy distortion visible in the detected vehicles.

![Image of stitched view data with vehicle detections]

Fig. 7: Evaluation of the baseline vehicle detection schemes, the deformable parts model (DPM) from [5], with and without the proposed motion integration at a 0.2 overlap threshold. A significant advantage is shown by integration of the motion-based vehicle detection boxes (MB) and the false positive removal using motion compensation (MC).

![Graph of true positives rate vs false detections rate]

VI. CONCLUDING REMARKS AND FUTURE WORK

In this paper, we studied a method for improving an appearance-based vehicle detector using motion cues in the context of overtaking vehicle detection. Reliable motion cue extraction is a difficult task, and so a scheme of processing dense optical flow was used. We compared the performance to other vehicles detectors in order to judge its relative effectiveness, and found an improvement in vehicle detection performance. In the future, tighter bounding boxes may be generated through more robust outlier removal algorithms. Additionally, incorporating long term motion trajectories, as in [18], could improve the overall performance of the algorithm. Furthermore, it might prove useful to integrate our overtaking vehicle detector into other driver assistance systems, such as a system for merge recommendations [19].

![Graph of true positives rate at FDR = 0.1 vs overlap threshold]

Fig. 8: Varying the overlap threshold for a correct detection. Notice how motion boxes improve upon the baseline algorithm. The motion boxes are less accurate but provide a significant improvement at less strict overlap thresholds.
VII. ACKNOWLEDGMENT

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

REFERENCES