Towards Semantic Understanding of Surrounding Vehicular Maneuvers: A Panoramic Vision-Based Framework for Real-World Highway Studies

Miklas S. Kristoffersen[1,2], Jacob V. Dueholm[1,2], Ravi K. Satzoda[2],
Mohan M. Trivedi[2], Andreas Møgelmose[1,2], and Thomas B. Moeslund[1]
[2] University of California, San Diego, USA. rsatzoda@eng.ucsd.edu,mtrivedi@ucsd.edu
[1] Aalborg University, Denmark. {mskr11,jdueho11}@student.aau.dk {am,tbm}@create.aau.dk

Abstract

This paper proposes the use of multiple low-cost visual sensors to obtain a surround view of the ego-vehicle for semantic understanding. A multi-perspective view will assist the analysis of naturalistic driving studies (NDS), by automating the task of data reduction of the observed sequences into events. A user-centric vision-based framework is presented using a vehicle detector and tracker in each separate perspective. Multi-perspective trajectories are estimated and analyzed to extract 14 different events, including potential dangerous behaviors such as overtakes and cut-ins. The system is tested on ten sequences of real-world data collected on U.S. highways. The results show the potential use of multiple low-cost visual sensors for semantic understanding around the ego-vehicle.

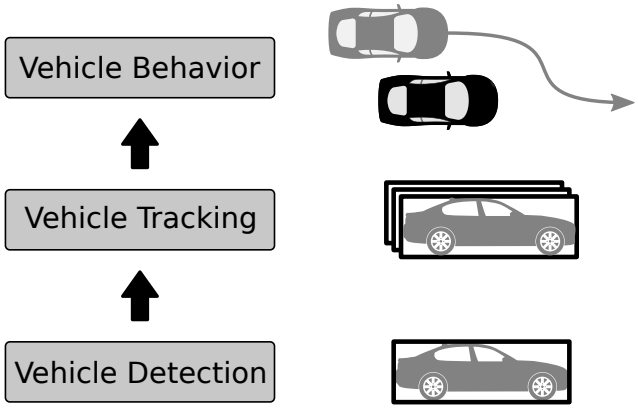


Figure 1. The ascending levels of vehicle interpretation in a vision-based application. At the lowest level is vehicle detection, which locates visible vehicles on a single-camera and single-frame basis. One level up, detections are associated between frames and views, in order to track vehicles on a multiple-camera and multiple-frame basis. At the highest level, the spatio-temporal trajectories are used to classify behaviors of vehicles.

1. Introduction

Trajectories of surrounding vehicles are essential to the extraction of higher-level semantics. Recent scientific progress in visual vehicle detection and tracking allows for robust trajectories [19] that enables us to automate exploration of vehicle behaviors, which has previously been a time-consuming manual hand-labeling process. However, until now, visual cameras have not been used to cover full surroundings of a vehicle with the purpose of estimating trajectories of surrounding vehicles and analyzing maneuvers. In this study we show how existing methods for monocular vehicle detection and tracking adapts to a multi-perspective framework with the purpose of reaching a higher level understanding of surrounding vehicle maneuvers and behaviors as shown in Fig. 1. If successful, these trajectories contain information, which is valuable to naturalistic driving studies (NDS) that seek to answer how drivers behave and why, in order to understand circumstances of crashes and near-crashes. By learning how surrounding trajectories develop over time, it is possible to predict which route the vehicles will probably follow in the near future. The prediction of trajectories is an integral part of path-planning in advanced driver assistance systems (ADAS).

The leading technologies in terms of sensing vehicular surroundings are LiDAR and radar. A lot of research has been conducted in the field using three-dimensional point clouds, consequently enabling autonomous vehicles to successfully drive public roads without causing accidents. However, by introducing low-cost passive visual cameras it is possible to add a level on top of the already existing solutions that rely purely on spatiotemporal positions and shapes. The visual modality contains appearance cues that can help improve the performance, e.g. by detecting brake lights, estimating orientation of vehicles, and recognizing traffic signs and signals. Thus, by using multi-perspective visual cameras together with existing ADAS, it is possible to achieve rich information of surroundings.

The main contributions of this paper can be summarized

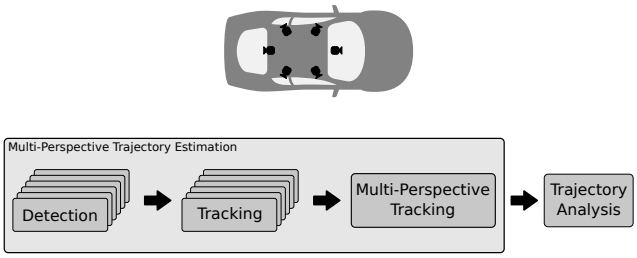


Figure 2. The top image displays the placement of the six synchronised cameras. The bottom image shows the flow of the system from the input of six video sequences to the output of a trajectory analysis.

(a) Front left (b) Front (c) Front right

(d) Rear left (e) Rear (f) Rear right

Figure 3. Sample images captured from the synchronized multi-perspetive setup. Note the challenges of e.g. glare, shadows, and distortion.

as follows: (1) Using six cameras, we develop a framework for estimating vision-based multi-perspective trajectories on a moving platform. The method has three steps: Vehicle detection in six different perspectives, vehicle tracking between frames in the six perspectives, and multi-perspective tracking that connects the trajectories across perspectives; (2) The multi-perspective trajectories are analyzed for semantics of surrounding vehicular events. We show how the combination of six perspectives, a top-down visualization of trajectories, and a list of events that have occurred, can be used as a powerful tool to interpret higher-level semantics of the surrounding vehicular maneuvers; (3) A vehicle equipped with six cameras is used to capture several hours of free-flow highway driving. We show a real-world study of 10 sequences chosen to prove the potential of the system.

2. Related Work

High-level semantics have previously been analyzed, identifying maneuvers as overtakes, lane-changes, cut-ins, cut-outs, or simply staying in lane. Early examples [9] use simulated data, while recently, real data are used in a front view of a moving platform [17, 12, 21], classifying up to 27 maneuvers regarding lane-changes. In [18], both a mono and a stereo camera are used to obtain trajectories in front of the ego-vehicle. The behaviors of the obtained trajectories are then learned using an unsupervised learning approach. A similar approach is seen in [16] with vehicles behind the ego-vehicle. Trajectories are furthermore used to infer traffic patterns in intersections using stereo vision [24, 7]. Estimating trajectories from vision-based sensors can be divided into classic computer vision disciplines as detection and tracking of vehicles. These are well researched fields with public available databases with common benchmarks. Multi-target vehicle tracking is mainly found in KITTI [8] and DETRAC [22], where multi-perspective tracking is mainly found for pedestrian tracking, as seen in Pets2009 [5] with overlapping views, MOT Challenge [14], and MCT Challenge [1] with non-overlapping views. In comparison to trajectories observed from pedestrians with static cameras [13], vehicle trajectories discovered with a multi-camera setup on a moving platform are subject to additional difficulties [18], such as effects of relative motion. Non-overlapping perspectives require the use of re-identification, which is traditionally used in surveillance applications [10]. In the application of tracking surround vehicles, the re-identification problem between perspectives is considerably simplified, since only a limited number of candidates exist, depending on the traffic density.

Previous studies have detected and tracked vehicles using multi-camera setups. An early example is seen in [6], where an omnidirectional camera together with a pan-tilt-zoom camera are used to detect and classify vehicles. In [3] surrounding vehicles and pedestrians are detected and tracked in a simple low-velocity parking environment. In [20] vehicles are detected around the ego-vehicle in a highway scenario using a method based on the deformable parts model (DPM) [4]. These studies focus on the low-level aspects of detecting and tracking in surround view applications, whereas we in this work furthermore show the potential use of the resulting trajectories as a tool for analyzing the behaviors of surrounding vehicles.

The challenge of associating trajectories between perspectives is studied in [15], where four cameras are used with partial overlap. Trajectories are extracted from each individual camera and projected to a common plane, where trajectories are associated. A similar approach is seen in [2], finding local trajectories and projecting to a common plane and linked if both the spatio-temporal features match.

3. System Overview

The synchronized data used in this work are collected on U.S. highways in California. The vehicle used for data collection is equipped with six Point Grey cameras and a

GPS tracker. Furthermore, data are logged from the controller area network (CAN) bus. The six cameras are placed strategically around the vehicle, as shown in Fig. 2, in order to achieve a full surround view as seen in Fig. 3. The front and rear cameras are considered the most important in the process of estimating the multi-perspective trajectories, for which reason they are capturing with a resolution of 1280×960. The two cameras use low-distortion lenses with horizontal field of view of 70° and 80°, respectively. The four side view cameras are captured at a lower resolution of 640×480, to achieve a frame rate of 15 frames per second (FPS) for the synchronized data collection. The side view cameras are mounted with wide angle lenses with a horizontal field of view of 135°, to ensure a full surround coverage with overlapping views, at the cost of a higher degree of distortion.

A flow diagram of the system is shown in Fig. 2. Vehicle detection is performed for each of the six inputs of the cameras. The detections in each perspective are used by the vehicle tracker, which associates the detections between frames for each of the six perspectives. The trajectories are connected between perspectives, and finally an analysis of the multi-perspective trajectories is performed.

4. Multi-Perspective Trajectory Estimation

In the following section we present the methods designed for estimating trajectories of vehicles present in surroundings of the ego-vehicle using six different visual perspectives.

4.1. Vehicle Detection

Visual vehicle detection is a well researched topic that has seen recent scientific progress, but is not yet considered a solved problem. In this work we use six different perspectives from the same location on a moving platform, and are thus subject to variances in capturing such as the viewpoint of vehicles, lighting, shadows, and glare. An example of these challenges is shown in Fig. 3. The side views are especially challenging with lower resolutions and severe distortion caused by the wide angle lenses. The multi-perspective challenges require the vehicle detection to be either one versatile detector, or to use a separate detector optimized for each perspective.

In this work we use the model-based Deformable Parts Model (DPM) detector [4] in a two-stage implementation presented in [24, 7]. The implementation includes a pretrained vehicle model trained on the KITTI dataset [8], which is used for all six perspectives. The first stage is a regular DPM detector, while the second stage detects vehicles in an upscaled version of the image in an area around the horizon. The horizon is specified for each of the perspectives. Detections for both stages are combined in a non-maxima suppression.

We have a set of captured video sequences in the time interval T, which is $\boldsymbol{V}^T = \{\boldsymbol{V}_1^T, \boldsymbol{V}_2^T, \cdots, \boldsymbol{V}_K^T\}$ for K cameras. A video sequence for one camera is a subset, $\boldsymbol{V}_k^T \subset \boldsymbol{V}^T$. Each video sequence has F images, thus $\boldsymbol{V}_k^T = \{I_1, I_2, \cdots, I_F\}$. We use the two-stage DPM detector to find a set of detections $\boldsymbol{D}^T = \{\boldsymbol{D}_1^T, \boldsymbol{D}_2^T, \cdots, \boldsymbol{D}_K^T\}$ for K cameras in the time interval T. Furthermore, the set of detections in camera k over time T has a length of N and is $\boldsymbol{D}_k^T = \{d_1, d_2, \cdots, d_N\}$. Each detection is $d_n = [t, x_1, y_1, x_2, y_2, s]$ where t is the time index/frame number, x_1 is the horizontal coordinate of the top left corner of the bounding box with respect to the top left corner of the input image, y_1 is the vertical coordinate of the top left corner, x_2 and y_2 are the bottom right corner of the bounding box, and s is a score.

4.2. Vehicle Tracking

Just like visual vehicle detection, the topic of visual vehicle tracking has received a lot of attention in scientific research. The challenge of tracking vehicles in six different perspectives over longer time periods is mainly difficult due to three things; sudden changes in capturing conditions, similar appearance of vehicles, and inter-vehicle occlusions. Despite these challenges, the visual vehicle tracking methods have reached an accuracy that allows for higher-level understanding of trajectories in a scene.

We use the online tracking method presented in [23] in a tracking-by-detection manner for each perspective in order to track vehicles between frames. It uses Markov decision processes (MDP) in combination with the widely used Tracking-Learning-Detection (TLD) tracker [11].

The tracker is originally designed for tracking pedestrians, for which reason, it is optimized for tracking vehicles in this study. The first change is the aspect ratio of the template, which is chosen based on typical vehicle aspect ratios in the annotations of the KITTI dataset [8] as shown in Fig. 4. Note that the aspect ratio of vehicles varies with the orientation at which they are observed. From this follows that vehicles observed in the side views will have a larger aspect ratio than vehicles observed in the rear and front views. We use an aspect ratio of 1.5, which is the mean of the annotated bounding box aspect ratios of the KITTI dataset. The second change is the state transition parameters of the MDP, which has been trained for vehicles. The MDP is trained on a sequence from the KITTI dataset [8] using available ground-truth annotations and detections computed by the DPM detector.

We find a set of associations of detections between frames $\boldsymbol{A}^T = \{\boldsymbol{A}_1^T, \boldsymbol{A}_2^T, \cdots, \boldsymbol{A}_K^T\}$ for K cameras in the time interval T. The kth set of associations has a length of M and is $\boldsymbol{A}_k^T = \{a_1^k, a_2^k, \cdots, a_M^k\}$. Each association is $a_m^k = [ID, d_n]$ where ID is a unique vehicle identification number.

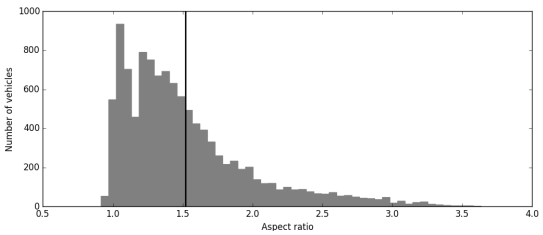


Figure 4. Histogram of annotated vehicle aspect ratios in the KITTI dataset [8]. The mean is shown with the vertical line at approximately 1.5.

4.3. Multi-Perspective Tracking

The final step of the multi-perspective trajectory generation is the connection of trajectories between cameras. Stationary setups have shown reliable performance, but in this study we have six perspectives on a moving platform, which makes the challenge of correctly associating trajectories non-trivial.

The trajectories are associated between perspectives, by assigning the same identification number to trajectories belonging to the same vehicle across perspectives. The association is done directly in the image planes, where stationary multi-perspective setups often perform the trajectory association in a common ground-plane. Since the camera views are known to overlap, predefined overlap regions are determined for each view denoted $\Omega^k = [\Omega^k_L, \Omega^k_R]$. Each trajectory is only evaluated once, in the first frame it appears. The bounding box of the new trajectory is firstly examined to be positioned in either the left or right overlapping region. Secondly, the corresponding adjacent view is examined for possible candidates to be associated with. Associated trajectories between cameras are described as $\boldsymbol{B}^T = \{\boldsymbol{B}^T_{k,k\pm1}\}$ for $k \in [1, 2, \cdots, K]$ in the time interval T, with K being the number of cameras. Note that k wraps around, such that k_1 and k_K are adjacent perspectives. Each set of associations between two cameras k and $k \pm 1$ is $\boldsymbol{B}^T_{k,k\pm1} = \{b_1, b_2, \cdots, b_L\}$ where $b_l = [a^k_m, a^{k\pm1}_{m'}]$ is the lth association.

$$b_l = \begin{cases} [a^k_m, a^{k-1}_{m'}] & \text{if } \Omega^{k-1}_R < a^{k-1}_{m'}(x_2) \text{ and } a^k_m(x_1) < \Omega^k_L \\ [a^k_m, a^{k+1}_{m'}] & \text{if } \Omega^k_R < a^k_m(x_2) \text{ and } a^{k+1}_{m'}(x_1) < \Omega^{k+1}_L \end{cases}$$

As an example, see Fig. 7(b), where the leftmost car just appeared, and is being associated with the rightmost car in Fig. 7(a). A similar association is made between Fig. 7(f) and Fig. 7(e). In the case with multiple possible matches in the adjacent view, a constraint is added, where an ID only can exists once in each view, or else the closest match is chosen.

This association scheme is seen to fail at high density scenes, or at late detections, when the vehicle has already passed the overlapping part of the image, resulting in two trajectories not being associated. The simple association method is found sufficient in free-flow highway scenarios.

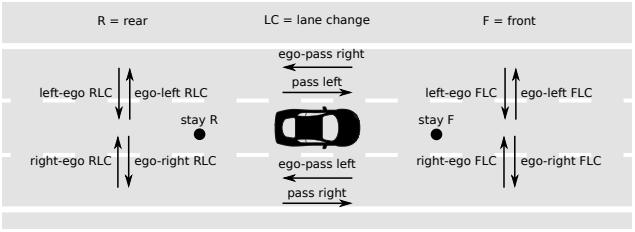


Figure 5. The 14 events detected in the trajectory analysis.

One advantage of using multiple views, is the ability to remove short-lived faulty trajectories, since the trajectories of interest are considered as long tracks in order to describe an event. All trajectories with a length less than a certain threshold measured in frames are removed. The threshold has been determined experimentally to 75 frames, corresponding to 5 seconds with 15 FPS, for the results presented in this work.

5. Trajectory Analysis

A map or a list of the dynamics and behaviors of surrounding vehicles is an integral part of understanding what is happening around the ego-vehicle, and why something is happening. In this section we present how the multi-perspective trajectories are transformed to a common framework and analyzed for events and certain behaviors. The system output is thus two-fold; a visualization of trajectories in the road surface enabling an in-depth analysis and a list with events that allows for fast interpretation.

5.1. Visualization of Trajectories

The visualization enables NDS to describe why events at certain time instances are happening. Combined with the actual video feeds, this is a powerful tool for studying on-road vehicle behaviors in a way that has not been presented previously.

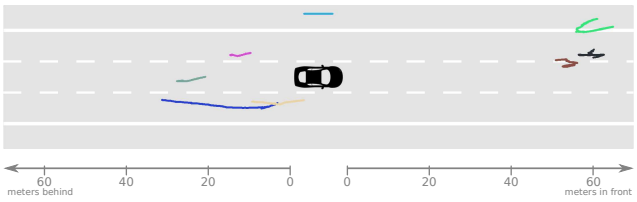
The multi-perspective trajectories are mapped to a common framework being the road surface. This is achieved by inverse perspective mapping (IPM) the front and rear perspectives, and using the middle of the bottom of the bounding box as a position of tracked vehicles. The trajectories are filtered using the average of the last n positions in order to achieve smooth tracks. The side views are used as discrete positions for rear left, rear right, front left, and front right. Furthermore, a simple lane estimator is used to show in which lane vehicles are positioned when they are on the side of the ego-vehicle. As the road might have a curve or slope, the IPM can not be expected to be accurate at larger distances. Vehicles are therefore tracked up to a distance of approximately 70 meters behind and in front of the ego-vehicle.

5.2. Data Reduction

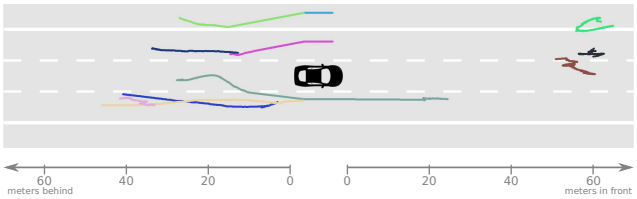
Visualizations are valuable for analyzing vehicle dynamics, but they contain a lot of data that are not easily interpretable. This problem can be solved by reducing the amount of information presented to the end-user. Furthermore, It allows for NDS to be automated.

The top-view trajectories are used to compute which events are occurring. We detect 14 different events as shown in Fig. 5. The method is currently limited to detecting lane changes in the front and rear perspectives, and only for adjacent lanes. Passing vehicles are found for all available lanes. For example, if a vehicle moves from a rear left to a front left position it is passing the ego-vehicle on the left. Likewise, if a vehicle moves from a front left to a rear left position the ego-vehicle is passing it on the right.

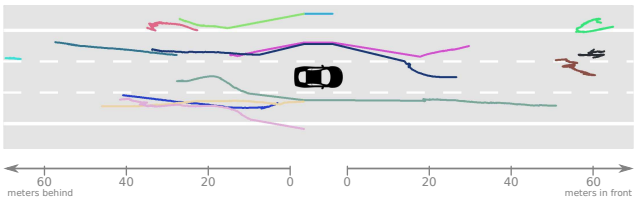
A combination of events can be grouped into semantics allowing for a higher-level understanding of vehicular maneuvers. For example, if a vehicle stays in front of the ego-vehicle within a certain distance over a time period, it can be concluded that the ego-vehicle is tailgating the vehicle in front. Another example is a vehicle that changes from ego-lane to left lane to pass the ego-vehicle on the left. This is defined as an overtake. If a passing vehicle changes lane to the ego-lane close to the front of the ego-vehicle, it is called a cut-in. A behavior that is potentially dangerous.



(a) Frame 100



(b) Frame 350



(c) Frame 600

Figure 6. Top-view trajectories of Seq2 at three time instances. As seen over time, the ego vehicle is being overtaken multiple times, where one vehicle furthermore makes a potential dangerous cut-in.

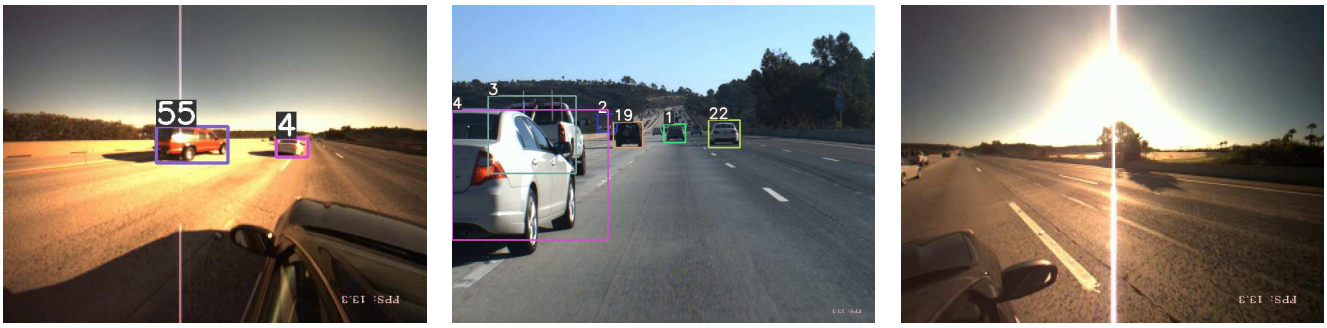
6. Experimental Evaluation

In this section we evaluate the performance of the system based on ten highway sequences ranging from 10 seconds to 40 seconds. The sequences are chosen from several hours of captured data in free-flow traffic, where interesting events are observed, to prove the potential of the system. These events include overtaking, tailgating, cut-ins, and cut-outs. In order to gain further insight in the performance, we show a detailed evaluation of one of the sequences.

It would be time consuming for NDS to analyze the events from six different perspectives. Our visualization allows for a top-down view of the scene, helping to get an overview of the different events. Fig. 6 shows the visualized trajectories at three time instances of a 40 seconds sequence (Seq2). In this way it is possible to see what is happening in the sequence over time. At the first time instance Fig. 6(a), the ego-vehicle has two receding vehicles in the rear right lane, one approaching vehicle in the rear ego-lane, one approaching vehicle in the rear left lane, one vehicle on the left side, and three vehicles in the lanes in front. At the second time instance Fig. 6(b), one of the vehicles has chosen to overtake on the right side of the ego-vehicle, which is probably caused by the vehicle overtaking on the left that has a lower velocity. Also, a new vehicle is approaching in the rear left lane. Note that this vehicle has a higher velocity than the vehicle currently overtaking on the left. This might be the reason why the very same vehicle at the third time instance Fig. 6(c), cuts into the ego-lane after overtaking the ego-vehicle on the left and starts to overtake the slower moving vehicle on the right. Fig. 7 displays the six perspectives at the time instance where the vehicle cuts into the ego-lane.

The list of events for Seq2 shown in Table 1 reduces the information further. With three detected left side passes and one right side pass, it can be concluded that the ego-vehicle drives slower than the surrounding traffic. However, as a vehicle stays in front of the ego-vehicle, it is likely that the ego-vehicle drives with a velocity similar to that vehicle. The combination of the visualization and the list of events is a powerful tool that allows for fast interpretation of behaviors occurring in a scene.

Table 1 summarizes the number of occurrences of each event for all ten sequences compared to the ground-truth obtained by manual inspection of each sequence. An overview of the ten sequences is shown in Fig. 8 along with all the trajectories from all ten sequences plotted in Fig. 8(k). This demonstrates the variety in the sequences of vehicles overtaking on both left and right, lane changes, and a few potential dangerous cut-ins. The system shows approximately the same tendencies as the ground-truth throughout all the ten sequences. This is also confirmed by the precision $TP/(TP + FP)$ and recall $TP/(TP + FN)$, where TP



(a) Front left (b) Front (c) Front right

(d) Rear left (e) Rear (f) Rear right

Figure 7. The six perspectives of Seq2 at frame 506. The multi-perspective tracked vehicles are shown by their latest detection in colored bounding boxes with corresponding identification number. Note some vehicles can be seen in multiple perspectives due to overlap, thus assigned the same identification number.

is true positives, FP is false positives, and FN is false negatives. The most frequent event is found to be vehicles passing the ego-vehicle on the left, while there was no one going from the ego-lane to the right-lane in front of the ego-vehicle. This indicates a passive driver, not forcing any of the cars in front to make a lane change. Also noteworthy is the event of a vehicle changing from ego-lane to left lane in front of the ego-vehicle, having a precision and recall of zero. This is partly explained by the false positives caused by the inaccuracy at far distances as seen in Fig.8(a). The inaccuracy is mainly caused by a road surface that is not completely flat or curved, which will make the IPM inaccurate, or the fact that only a small number of pixels are available the further away the vehicle is. The two false negatives seen in sequence three and seven respectively, may be caused by the filtering of trajectories, resulting in the trajectories coming up short, as the trajectories direction indicate a lane change, according to Fig.8(c) and Fig. 8(g).

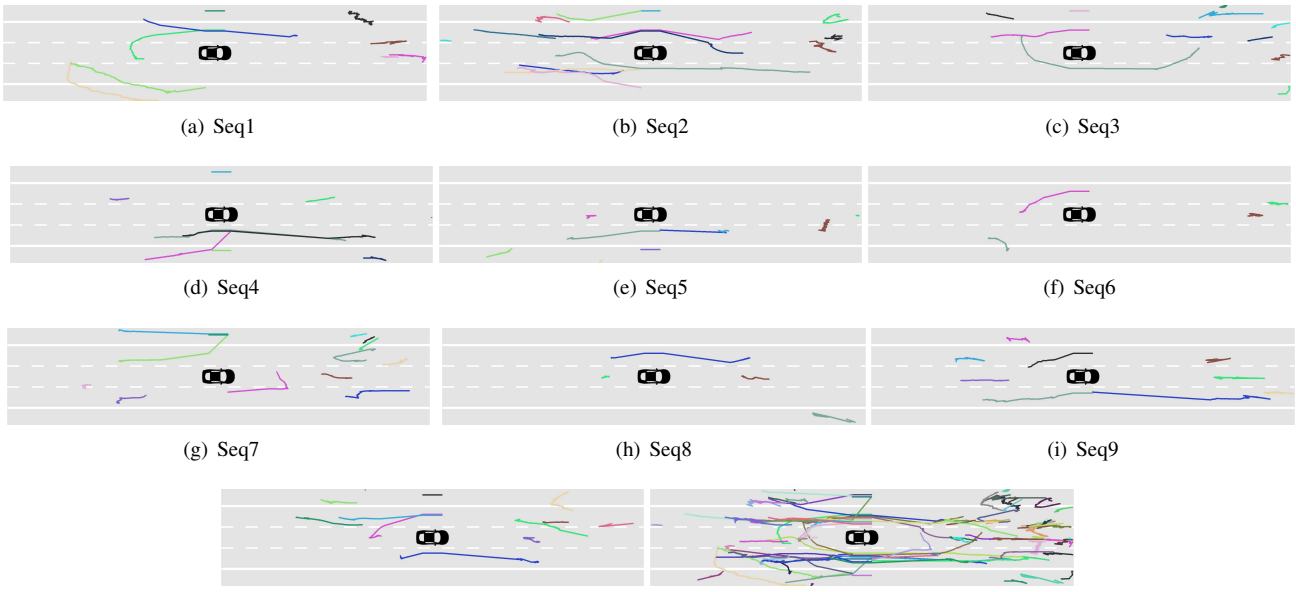
As seen in Fig. 8(k), the system is primarily tracking vehicles in the ego-lane and adjacent lanes. This is primarily due to frequent occlusions of vehicles in other lanes, but also the fact that they need a bigger distance to the ego-vehicle before appearing in the front and rear perspectives. The result is that vehicles in outer lanes have a higher probability of causing false negatives, which also reflects in the result for left passes in Table 1. Also, the association between views has difficulties if two vehicles pass on the same side simultaneously. Including more features than position may solve this problem, e.g. by using appearance cues. Furthermore, instead of using the overlap restriction, vehicles can be associated between views by allowing them to appear in other views within a certain time frame. This is however more a task of vehicle re-identification than overlap-association.

7. Concluding Remarks

This work developed a multi-perspective framework for analyzing on-road vehicle behavior in real-world highway data. The usage of multiple overlapping cameras proves useful for estimating persistent trajectories in full surrounding of the ego-vehicle. The multi-perspective framework successfully enables in-depth analysis despite the challenges introduced in the visible domain such as variances in point of view, glare from the sun, shadows of different sizes and shapes, and distortion (see Fig. 7 and Fig. 9 for examples), and is efficiently removing short-lived false trajectories. Furthermore, by using low-cost passive sensors in the visible spectrum the system allows for an interface that is easily understandable by humans, which is an important property in terms of human-computer-interaction

Table 1. Events detected by the system for all ten sequences compared to ground-truth (GT) [System/GT].

Event	Seq1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	Seq8	Seq9	Seq10	Precision	Recall
Stay front	2/1	0/1	1/1	1/0	0/1	1/1	1/1	1/1	1/1	1/1	0.78	0.78
Stay rear	0/0	0/0	0/0	0/0	1/1	0/0	0/0	1/1	1/1	0/0	1.00	1.00
Pass on left	3/4	3/4	1/2	1/1	0/0	1/1	3/3	1/1	0/0	3/5	1.00	0.76
Pass on right	0/0	1/1	1/1	0/0	1/1	0/0	0/0	0/0	0/0	1/1	1.00	1.00
Ego-pass on left	0/0	0/1	0/1	4/4	1/1	0/0	0/0	0/0	1/1	0/0	1.00	0.75
Ego-pass on right	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	1/1	0/0	1.00	1.00
In front, left to ego-lane	1/0	2/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0.50	1.00
In front, right to ego-lane	0/0	0/0	1/1	0/1	1/1	0/0	1/1	0/0	0/0	0/0	1.00	0.75
In front, ego-lane to left	1/0	0/0	0/1	0/0	0/0	0/0	0/1	0/0	0/0	0/0	0.00	0.00
In front, ego-lane to right	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1.00	1.00
In rear, left to ego-lane	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/1	0/0	1.00	0.50
In rear, right to ego-lane	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1.00	1.00
In rear, ego-lane to left	1/1	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0	1/1	1.00	1.00
In rear, ego-lane to right	0/0	1/1	0/1	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0.50	0.50
Precision	0.7	0.88	1.0	0.83	1.0	1.0	0.83	1.0	1.0	1.0		
Recall	0.88	0.7	0.6	0.83	0.8	1.0	0.83	1.0	0.8	0.78		



(a) Seq1 (b) Seq2 (c) Seq3

(d) Seq4 (e) Seq5 (f) Seq6

(g) Seq7 (h) Seq8 (i) Seq9

(j) Seq10 (k) Total

Figure 8. Visualization of the ten sequences along with all the trajectories in total. Evaluated in Table 1.

(HCI). This makes the system an attractive addition to the sensor suite of intelligent vehicles.

The potential of the system is not limited to highway driving. More complex scenarios are a logical next step for example in urban areas as shown in Fig 9. In this specific scenario, the vehicle is stopped at an intersection with vehicles coming from the front right, and going through multiple perspectives, before disappearing in the rear left perspective. This is only one scenario among many. Applications able to model scenes by utilizing the surround view allow for sophisticated understanding of events and behavior. The obtained information can be used for both NDS and ADAS, ultimately answering questions such as: Why did this vehicle make a cut-in? Is it safe to make a left turn now?

A more comprehensive study of semantics from the detected events would include classification of e.g. safe and aggressive lane changes. Thus, a movement towards understanding high-risk semantics that need the attention of the driver or the ADAS. Also, by using a data-driven learning approach instead of the heuristic rule-based event classification, it will be possible to model typical trajectories allowing for future predictions of dynamics and behaviors in the scene.

(a) Front left (b) Front (c) Front right

(d) Rear left (e) Rear (f) Rear right

Figure 9. Six perspectives at an intersection in an urban scenario.

Acknowledgment

The authors would like to thank their colleagues at the Laboratory for Intelligent and Safe Automobile (LISA), University of California, San Diego, for assisting with the data gathering and their invaluable discussions and comments.

References

[1] Multi-Camera Object Tracking (MCT) Challenge. [Online]. Available: http://mct.idealtest.org.

[2] N. Anjum and A. Cavallaro. Trajectory Association and Fusion across Partially Overlapping Cameras. In *IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009.*

[3] M. Bertozzi, L. Castangia, S. Cattani, A. Prioletti, and P. Versari. 360 Detection and tracking algorithm of both pedestrian and vehicle using fisheye images. In *IEEE Intelligent Vehicles Symposium, 2015*.

[4] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 2010.

[5] J. Ferryman and A. Shahrokni. Pets2009: Dataset and challenge. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009*.

[6] T. Gandhi and M. Trivedi. Video Based Surround Vehicle Detection, Classification and Logging from Moving Platforms: Issues and Approaches. In *IEEE Intelligent Vehicles Symposium, 2007*.

[7] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun. 3D Traffic Scene Understanding From Movable Platforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5), 2014.

[8] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[9] T. Gindele, S. Brechtel, and R. Dillmann. A Probabilistic Model for Estimating Driver Behaviors and Vehicle Trajectories in Traffic Environments. In *IEEE Conference on Intelligent Transportation Systems, 2010*.

[10] T. Huang and S. Russell. Object Identification in a Bayesian Context. In *IJCAI*, volume 97, 1997.

[11] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-Learning-Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7), 2012.

[12] D. Kasper, G. Weidl, T. Dang, G. Breuel, A. Tamke, A. Wedel, and W. Rosenstiel. Object-Oriented Bayesian Networks for Detection of Lane Change Maneuvers. *IEEE Intelligent Transportation Systems Magazine*, 4(3), 2012.

[13] M. S. Kristoffersen, J. V. Dueholm, R. Gade, and T. B. Moeslund. Pedestrian Counting with Occlusion Handling Using Stereo Thermal Cameras. *Sensors*, 16(1):62, 2016.

[14] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. *arXiv:1504.01942 [cs]*, 2015.

[15] M. J. Mirza and N. Anjum. Association of moving objects across visual sensor networks. *Journal of Multimedia*, 7(1), 2012.

[16] B. T. Morris and M. M. Trivedi. Unsupervised Learning of Motion Patterns of Rear Surrounding Vehicles. In *IEEE International Conference on Vehicular Electronics and Safety, 2009*.

[17] R. K. Satzoda and M. M. Trivedi. Drive Analysis Using Vehicle Dynamics and Vision-Based Lane Semantics. *IEEE Transactions on Intelligent Transportation Systems*, 16(1), 2015.

[18] S. Sivaraman, B. Morris, and M. Trivedi. Learning Multi-Lane Trajectories using Vehicle-Based Vision. In *IEEE International Conference on Computer Vision Workshops, 2011*.

[19] S. Sivaraman and M. Trivedi. Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis. *IEEE Transactions on Intelligent Transportation Systems*, 14(4), 2013.

[20] C. Wang, Y. Fang, H. Zhao, C. Guo, S. Mita, and H. Zha. Probabilistic Inference for Occluded and Multiview On-road Vehicle Detection. *IEEE Transactions on Intelligent Transportation Systems*, 17(1), 2016.

[21] G. Weidl, A. Madsen, D. Kasper, and G. Breuel. Optimizing Bayesian Networks for Recognition of Driving Maneuvers to Meet the Automotive Requirements. In *IEEE International Symposium on Intelligent Control, 2014*.

[22] L. Wen, D. Du, Z. Cai, Z. Lei, M. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu. DETRAC: A New Benchmark and Protocol for Multi-Object Tracking. *CoRR*, abs/1511.04136, 2015.

[23] Y. Xiang, A. Alahi, and S. Savarese. Learning to Track: Online Multi-Object Tracking by Decision Making. In *IEEE International Conference on Computer Vision, 2015*.

[24] H. Zhang, A. Geiger, and R. Urtasun. Understanding High-Level Semantics by Modeling Traffic Patterns. In *International Conference on Computer Vision, 2013*.