Dynamic 3D Reconstruction of Vehicles for Safer Intersections

Project ID: 30

Srinivasa Narasimhan (PI), Associate Professor
Robotics Institute, Carnegie Mellon University
https://orcid.org/0000-0003-0389-1921

Robert Tamburo (Co-PI), Senior Project Scientist
Robotics Institute, Carnegie Mellon University
https://orcid.org/0000-0002-5636-9443

Dinesh Reddy, Graduate Student
Robotics Institute, Carnegie Mellon University

Technologies for Safe & Efficient Transportation

THE NATIONAL USDOT UNIVERSITY TRANSPORTATION CENTER FOR SAFETY

Carnegie Mellon University | UNIVERSITY of PENNSYLVANIA

Contract No. DTRT12GUTG11

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

1 Problem

Over the past decade, there have been concerted nationwide efforts to improve roadway safety. Contributions include the implementation of new educational programs, deployment of new design strategies, and adaptation of new technologies. These efforts, in part, have resulted in historically low fatalities rates (fatalities per miles traveled) [2]. However, the fatality rate and number of fatalities have increased over the past couple of years from reckless behaviors such as speeding, alcohol impairment, and not wearing seat belts [1]. Intersections are particularly dangerous sections of the roadway system due to being points of conflict between vehicles, pedestrians, and bicyclists. According to the Federal Highway Administration, more than 25% of all fatal crashes and 50% of crashes occur at intersections resulting in over 3 million accidents at a cost of over $100 billion per year.

Multiple video cameras are becoming increasingly common at urban traffic intersections. This provides us a strong opportunity to reconstruct moving vehicles crossing those intersections. The shapes (even sparse) and motions of the vehicles can be invaluable to traffic analysis, including vehicle type, speed, density, trajectory and frequency of events such as near-accidents. Infrastructure-to-Vehicle (I2V) communication systems can provide such analysis to other (semi-)autonomous vehicles approaching the intersection. That said, reconstructing moving vehicles in a busy intersection is hard because of severe occlusions. Furthermore, the cameras are often unsynchronized, provide wide baseline views with little overlap in fields of view and need to be calibrated each frame as they are often not rigidly attached and sway because of wind or vibrations.

2 Approach and Methodology

We present a comprehensive framework that fuses (a) incomplete and imprecise structured points (part detections) across multiple views with (b) precise but sparse single-view tracks of unstructured points, to reconstruct moving vehicles even in severe occluded scenarios. This framework consists of three main stages: (1) a novel object-centric (as opposed to feature-centric) RANSAC approach to provide a good initialization of the 3D geometry of the structured points of the vehicle, (2) a novel algorithm that fully exploits the complementary strength of the structured and unstructured points via rigidity constraints, and (3) closing-the-loop by reprojecting the reconstructed structured points to all views to retrain the part detectors. We implemented a full end-to-end system that also includes a pre-processing stage to self-calibrate and synchronize the cameras by adapting recent prior works [3]. A detailed overview of our system is illustrated in Fig. 1.

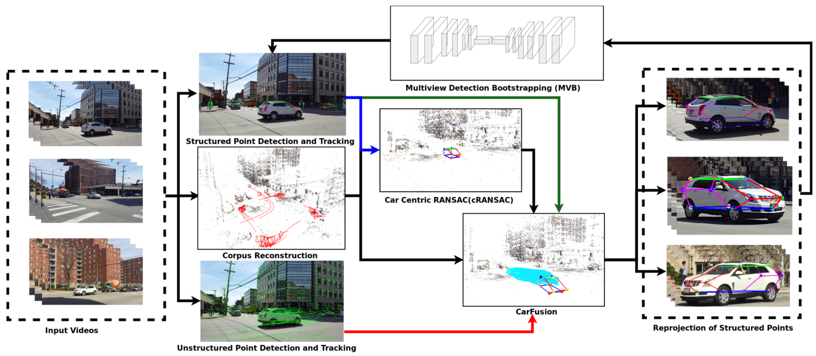

Figure 1: Our overall pipeline for dynamic 3D reconstruction of multiple cars from uncalibrated and unsynchronized video cameras. We fuse the structured points (detected vehicle parts) and the tracks of the unstructured feature points to obtain precise reconstruction of the moving vehicle. The reconstructions are reprojected into all the views and are used to bootstrap and improve the detectors.

3 Findings

We demonstrated reconstruction of vehicles at a busy intersection shown in Fig. 2. About 62 vehicles were detected, tracked and reconstructed within a 3-minute duration captured from 21 handheld cameras that are uncalibrated and unsynchronized and were panning to cover wider fields of view. A subset of vehicle structured point trajectories are augmented within the Google Earth image of the intersection. They include cars of different types (sedans, SUVs, hatch-backs, jeeps, etc.) making left and right turns, going straight-ahead as well as changing lanes. Several views of two specific cars in various occluded scenarios are shown with the reprojections of the structured points.

We evaluated our framework on a traffic scene captured with six Samsung Galaxy 6, ten iPhone 6, and six Gopro Hero 3 cameras at 60 fps in a busy intersection for 3 minutes. In total the algorithm is run on nearly 210000 frames. These videos were captured by 13 people, some of whom carried two cameras. The sequence is challenging as there are no constraints on the camera motion or the vehicle motion in the scene. We manually annotated the 2D locations of the structured points for every visible cameras for

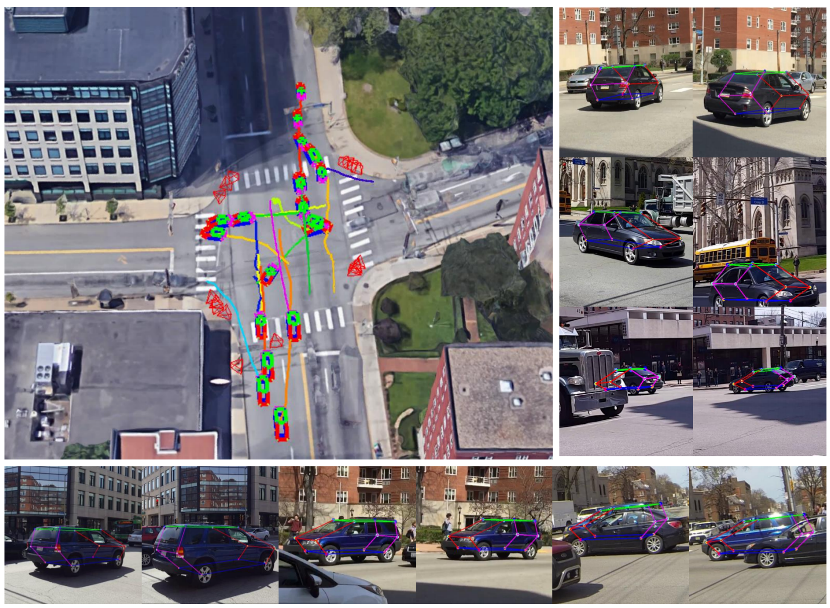

Figure 2: Reconstruction of vehicles crossing a busy intersection, making turns, going straight and changing lanes. A subset of vehicle skeletons (3D detector locations) and their 3D trajectories are augmented within the Google Earth view of the intersection. The reconstructions are reprojected into multiple views of two cars (a sedan and an SUV) demonstrating good performance under partial occlusions.

2793 frames from different viewing angles in the Intersection dataset. These data are available for research purposes at http://www.cs.cmu.edu/∼ILIM/projects/IM/CarFusion.

The accuracy of reconstructed structured points with respect to the ground truth annotations according to the tracking length and the number of unstructured points is shown in Fig. 3. As expected, the increase in visibility (track length) of structured points better stabilize the structured points which leads to higher quality reconstruction. We also found that the larger number of unstructured points improve the quality of the structured points due stronger rigidity constraints and the improvement is more evident for stricter threshold.

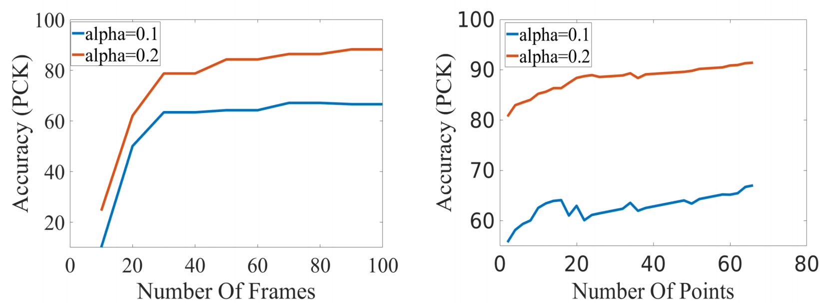

Figure 3: Analysis of accuracy with respect to increase in number of frames (left) and increase in number of unstructured points (right).

Our approaches are designed to handle partial occlusions but fail when a vehicle is mostly occluded at all times. The estimated 3D vehicle tracks are accurate but slightly wobbly and will benefit from additional domain specific priors. In Fig. 4, we illustrate the complete 3D reconstruction of trajectories of structured points on moving cars using our method and the 2D projection to inlier views for several cars. As can be seen from the results we are able to accurately reconstruct the trajectories of the cars over time captured from unsynchronized videos.

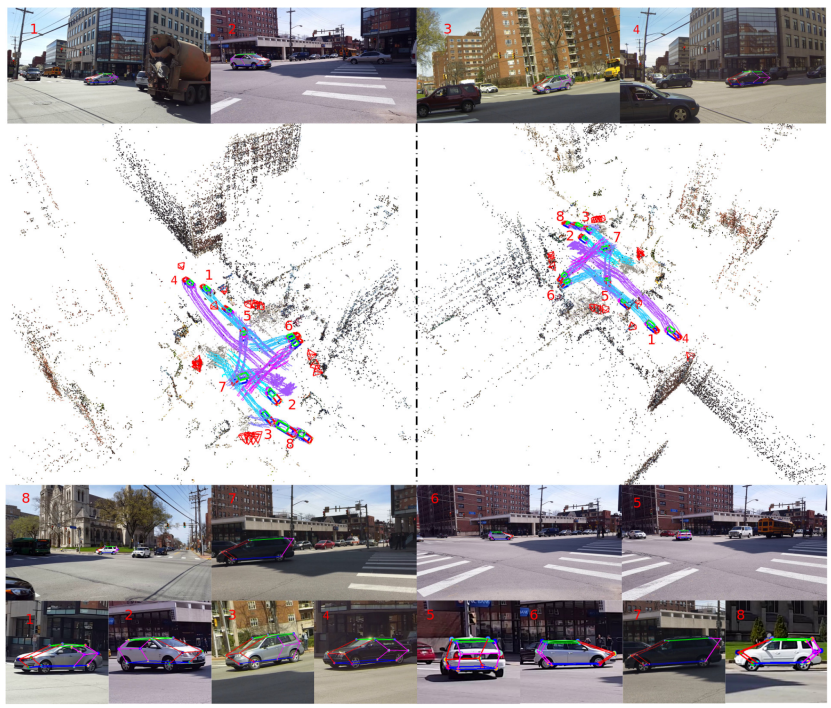

Figure 4: Visualization of the reconstructed cars using our method. We show the 2D re-projection of the reconstructions onto sample frame containing those cars. All the re-projected points fit the cars well.

4 Conclusions

We have presented a method to fuse imprecise and incomplete part detections of vehicles across multiple views and the more precise feature tracks within a single view to obtain better detection, localization, tracking and reconstruction of vehicles. This approach works well even in the presence of strong occlusions. We have quantified improvements due to the different stages of the end-to-end pipeline that only uses videos from multiple uncalibrated and unsynchronized cameras as input. We believe this approach can be useful for stronger traffic analytics at urban intersections.

5 Recommendations

The future of transportation will include semi- and fully autonomous vehicles that communicate with each other, as well as the infrastructure. Sensors on vehicles have a limited, often occluded view of the environment. Cameras mounted on poles in the intersection will provide a bird's eye view of the intersection. Coupled with computing at the edge and algorithms such as 3D reconstruction can be wirelessly transmitted to connected vehicles for collision avoidance, path planning, and assisted navigation.

References

[1] National Highway Safety Administration, “National Fatality Statistics,” https://www-fars.nhtsa.dot.gov/Main/index.aspx, Accessed: 24-Sept-2018.

[2] National Highway Traffic Safety Administration, “USDOT Releases 2016 Fatal Traffic Crash Data,? https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data. 2016, Accessed: 24-Sept-2018.

[3] M. Vo, S. G. Narasimhan, and Y. Sheikh, “Spatiotemporal bundle adjustment for dynamic 3d reconstruction,” In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.