Transportation Monitoring: Model-Based Vehicle Orientation Estimation

Ratnam Parikh
Arizona State University
rparikh4@asu.edu

Devshree Patel
Arizona State University
dpatel95@asu.edu

Vikram Reddy Sankepally
Arizona State University
vsankepa@asu.edu

Ashutosh Garg
Arizona State University
agarg69@asu.edu

Abstract

In this project, we improvise a vehicle localization and traffic scene reconstruction framework, CAROM, i.e., "CARs On the Map". We work on the vehicle orientation estimation (VOE) part of the tracking pipeline to improvise vehicle localization and 3D bounding box prediction. We introduce a model-based approach to achieve better performance relative to the current CAROM framework. The state-of-the-art model for VOE on KITTI benchmark is EgoNet model. Our objective is to integrate the EgoNet model with the CAROM framework. We face many challenges during the implementation of the project that will be discussed later in the report. We experiment various ways of integration and discuss the outcomes for the same.

1 Introduction

Traffic monitoring cameras and smart roadside units with vision-based sensors are becoming increasingly popular for traffic management purposes. Real-time videos from these cameras are used to evaluate driving safety, traffic congestions and rule violations. Interestingly, this functionality is a crucial component of "Autonomous vehicles" in the coming future. However, there are few problems pertaining to these monitoring cameras. Firstly, there is a significant cost associated with transmission and storage space of videos. Secondly, performing analysis on these videos is difficult as they contain unstructured data. Especially, it is difficult to obtain the 3D states of the vehicles from 2D images. Lastly, there are some privacy concerns associated with those videos which restricts accessbility. In order to resolve these issues, Duo et al. propose CAROM, a framework that can extract 3D information from the videos, generate a series of structured data records of vehicle states, and reconstruct traffic scenes on a 2D map or a 3D map.

Monocular object orientation estimation or estimating the 3D orientation of an object given a single 2D image of the object, is an important component of traditional computer vision problems like scene understanding and 3D reconstruction as well as modern vision challenges like autonomous driving and augmented reality, and robot manipulation. The main challenge of the object orientation estimation problem is that the task of estimating 3D orientation from a single 2D image is ill-posed. Recent developments propose the use of powerful deep learning features via Convolutional Neural Networks (CNNs) that learn appropriate features and models from the data to estimate the orientation in an end-to-end manner.

In this project, we aim on improving the vehicle tracking scene reconstruction pipeline proposed in the paper CAROM (Cars on the Map). CAROM estimates Vehicle Localization and Traffic Scene Reconstruction using monocular traffic cameras. It uses sparse optical flow to calculate the

optical vectors. These vectors are further used to calculate the vanishing points from which vehicle orientation is estimated. Heading direction fails when the detected instance is small or when the vehicle has a small motion in the image or when the RANSAC fails on optical vectors. To overcome this drawback in CAROM, we incorporate model-based orientation estimation in calculating the heading direction of the vehicle using EgoNet.

EgoNet uses a progressive approach to extract Intermediate Geometrical Representations to estimate egocentric vehicle orientation. This approach uses no prior information about the vehicle or the infrastructure of the surroundings. It uses the deep model to transform the image to IGRs(Intermediate Geometrical Representations) which are further lifted to the 3D representation of the object in the camera coordinate system. EgoNet is trained on the KITTI dataset and achieves better performance than traditional methods.

2 Implementation

2.1 Initial Approach

The first step after finalizing the EgoNet model as the VOE model to be integrated in CAROM pipeline, we setup both EgoNet and CAROM code over ASU Agave cluster. Our primary goal is to visualize vehicle orientation from EgoNet i.e. the orientation arrow. The initial roadblock was the dependency of EgoNet over a 2D/3D object detector. Like any other pose estimation model in Computer Vision, EgoNet too requires at least 2D bounding box predictions as an input to EgoNet. These bounding boxes are further cropped and used for 3D keypoints prediction. Along with this, for the output of orientation arrow, EgoNet requires ground truth 3D bounding boxes to estimate ground truth 3D keypoints that are used in calculation of the arrow. Original EgoNet paper uses output from D4LCN model (3D Detector model) as the ground truth 3D bounding boxes along with 2D bounding box as input to the model for orientation arrow prediction. This becomes an overhead if we want to directly integrate with CAROM.

To overcome the above mentioned issue, we tried various approaches for the same. Firstly, we tried to project combinations of 3D predicted keypoints in order to make an estimate of the heading direction. That will be further replaced as the heading direction in CAROM framework. After a few tries we were able to use points "1" and "5" as to draw a vector in heading direction. Another issue was that EgoNet doesn't encode object location i.e. centroid point when lifting 2D keypoints to 3D keypoints and thus we didn't have the center of object in 3D keypoints to start the heading vector. We used the 2D center and 2D 1st and 5th keypoints to estimate the heading vector. When integrated the heading direction was loaded from saved ".txt" files that were generated from prior run on EgoNet. This change in CAROM wasn't good on the output visualization and the 3D bboxes didn't get predicted as expected.

2.2 Second Approach

After few debugging sessions, we decided to use the 2D predicted keypoints directly to draw a 3D bounding box for the object and bypass various calculations in CAROM. We used 8 out of 33 2D keypoints, these are the endpoints that form the 3D box. We save the EgoNet output in the .txt files, which will be further loaded into CAROM pipeline. Now there are other issues as EgoNet requires 2D bounding boxes as an input we use the pretrained Mask-RCNN model as CAROM to get similar instances. But CAROM works on each instance but EgoNet predicts for a whole image. A pure integration wasn't possible and thus when we were loading the predicted txt files from EgoNet to CAROM, the instance tracking wasn't possible and thus we were missing on few instances and that resulted in an unstable final demo. We still tried to find a work-around for the issue of lack of instance tracking in EgoNet.

2.3 Final Approach

The Mask-RCNN model used by CAROM is retrained on the traffic data and also few post-processing methods are applied over those 2D predictions. This is the reason why a mismatch in number of instances occurred from EgoNet and CAROM. Thus we decided to use Mask-RCNN from

CAROM and also the post-processing methods to generate the input txt files for EgoNet. Further, after running EgoNet+CAROM on new data we get good results for predicted 3D bounding boxes. The flowchart below describes the workflow of the integrated pipeline.

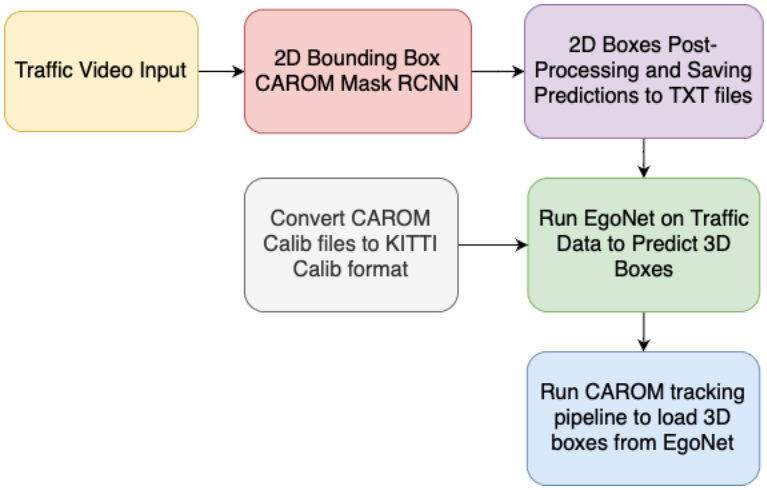

Figure 1: Flowchart

3 Experiments

3.1 Computing resources

For this project, we have used Agave super-computer provided by ASU research computing. Graphic configurations of our system was NVIDIA GeForce GTX 1080 Ti. CAROM was built using Intel distribution for python as it has all the binaries to perform data intensive computations.

3.2 Results

The results after first approach are very poor and the 3D bounding boxes are in a wrong direction. After the second approach, we get good results which are displayed below. But the final output is not stable due to the instance tracking problem discussed in implementation section. After the implementation of third approach we are able to get stable results as the instance tracking issue is solved.

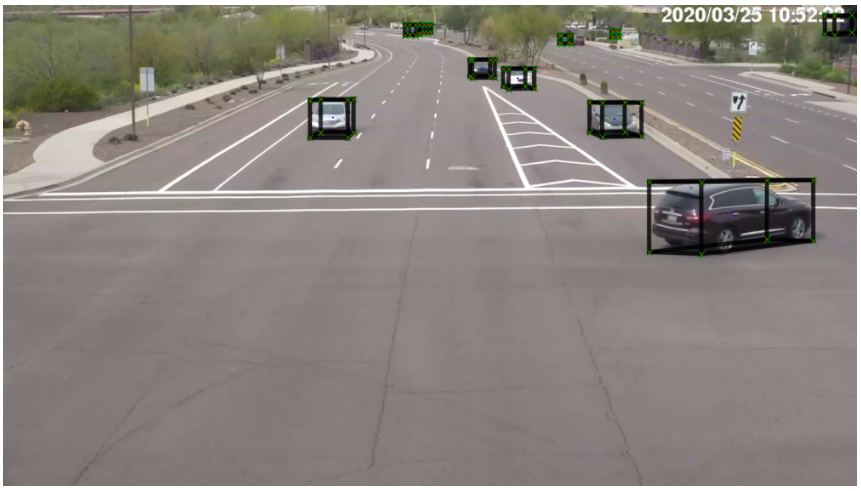

Figure 2: Output from EgoNet

Figure 3: Output CAROM+EgoNet-Second Approach

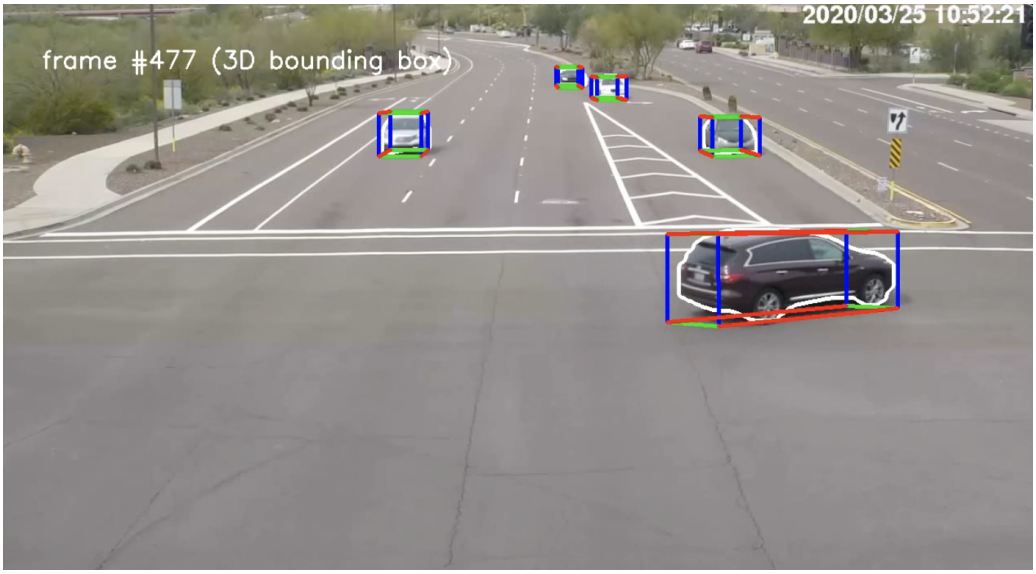

Figure 4: Output CAROM+EgoNet-Final Approach

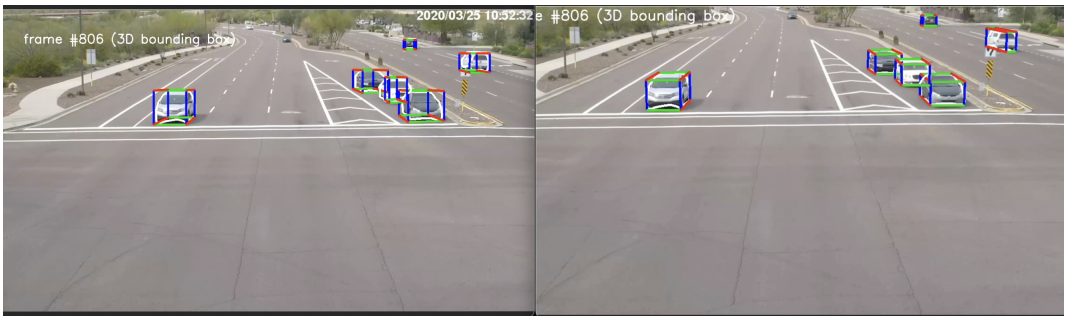

Figure 5: EgoNet+CAROM (Left) v/s CAROM (Right) Comparison

The Fig. 2 is the output of traffic data on EgoNet, where we can see that the 3D bounding boxes are generated with good quality. The Fig. 3 is output from CAROM after running EgoNet on the data

for the second approach mentioned above. Due to the instance tracking issue in both systems some cars' 3D bounding boxes don't appear. But, after solving the issue of instance tracking through same pre-processing techniques for EgoNet and CAROM, the 3D bounding boxes are getting visualized with better performance (Fig. 4) and the stability of output video increases. In Fig. 5 the comparison of original CAROM (right image) and EgoNet+CAROM (left image) is shown, where the car on the upper right side has errors for original CAROM 3D bounding box output whereas for the integrated system the output it is better.

4 Discussion and Conclusion

After various experiments and implementations we were able to achieve the goal of the project to successfully introduce model-based approach for VOE. We faced many challenges during the project from research of VOE, setup code repositories on ASU Agave, integrate both models and try to generate expected results through various implementations.The initial step in CAROM framework is to detect vehicles using Mask RCNN and doing post processing over them is missing out on few of the vehicles and thus remain undetected throughout the pipeline. This was observed during the second approach when we incorporated a different Mask-RCNN model before running EgoNet model. Thus, if the initial step is upgraded by incorporating better instance segmentation models and more robust post processing on detected vehicles, lesser vehicles will go undetected and increase the over performance of the integrated EgoNet+CAROM.

References

[1] Lu, Duo, et al. "CAROM-Vehicle Localization and Traffic Scene Reconstruction from Monocular Cameras on Road Infrastructures." 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021.

[2] Li, Shichao, et al. "Exploring intermediate representation for monocular vehicle pose estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. APA

[3] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.