Constructing a Map and a Human Driving Dataset from Birds-Eye View Video

Yesus Becerril, Zhiqian Qiao, John M. Dolan
Monterrey Tech, Carnegie Mellon University, Carnegie Mellon University

Introduction

Machine learning approaches can generate better autonomous driving models and behaviors, but they need data.

Problem:

- NGSIM dataset is the only suitable public dataset for this purpose. However its size/time period and scope are limited.
- Recent approaches to replicate this dataset need mounting fixed infrastructure, which requires permission, can be expensive, and is not portable.

Solution:

- Using a drone as our only infrastructure.
- Design a portable and easily repeatable flow work.
- Create or own flexible dataset.

Method

1.Extraction of the road

We use Semantic Segmentation, by using CNNs to achieve the mask of the image.

- The model: UNET, with VGG16 pre-trained on imagenet, as its encoder.
- Classes: 0 background, 1 Road

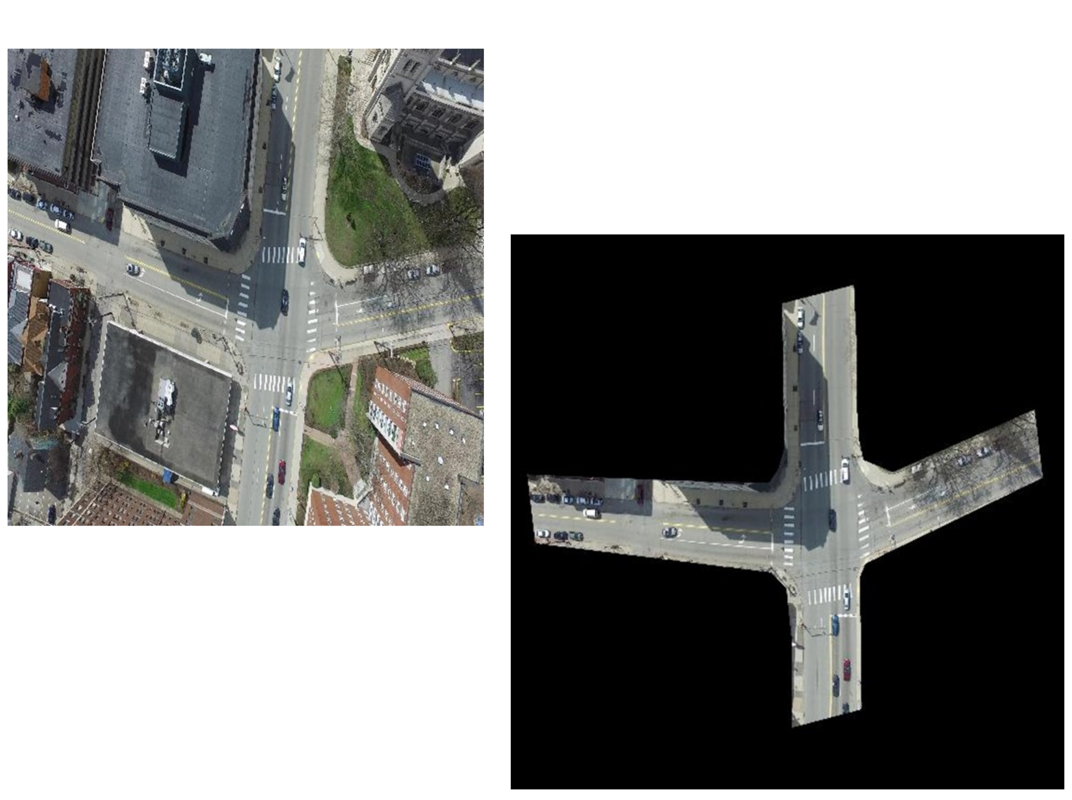

Fig 1. Given an aerial video of a intersection, our method gets the mask of the road.

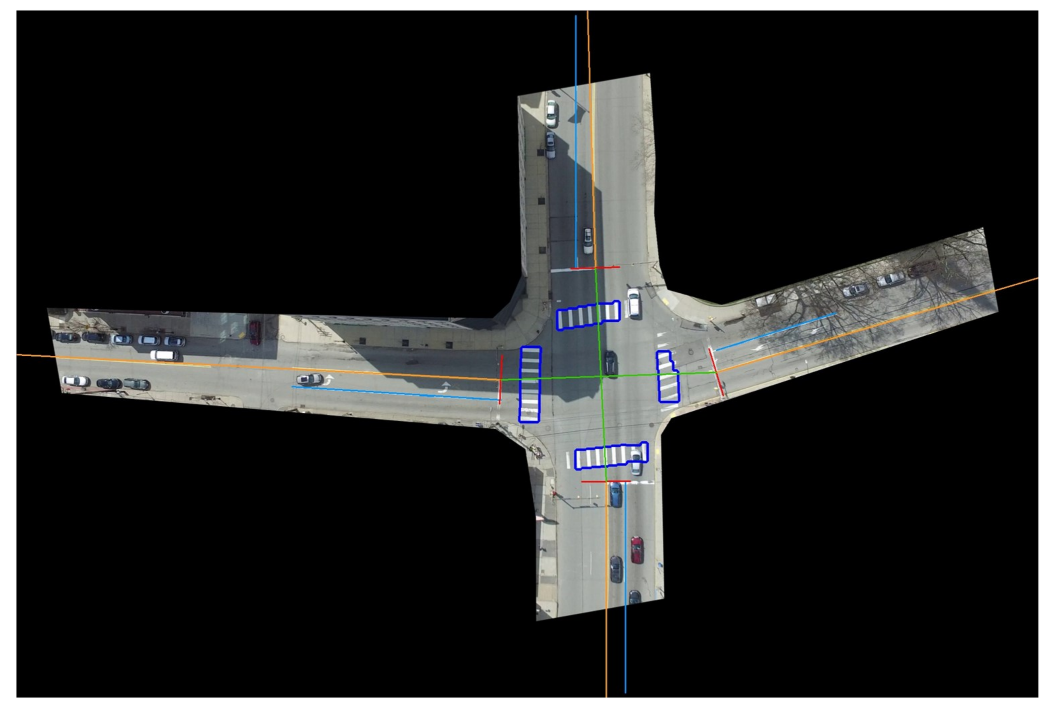

Fig 2. We were capable of clearly distinguish each of the lane-marks on the road. Yellow line, White lines, Zebra crossing and the stop line.

2. Global Coordinate Road Geometry

- Construction of the map, by identifying the lanemarks as reference points, to build the global coordinates.
- This was achieved by using diverse methods of computer vision, such as Canny Edge Detection, Color

3. Transition to local lane geometry

- Translation from pixel space to meters space.

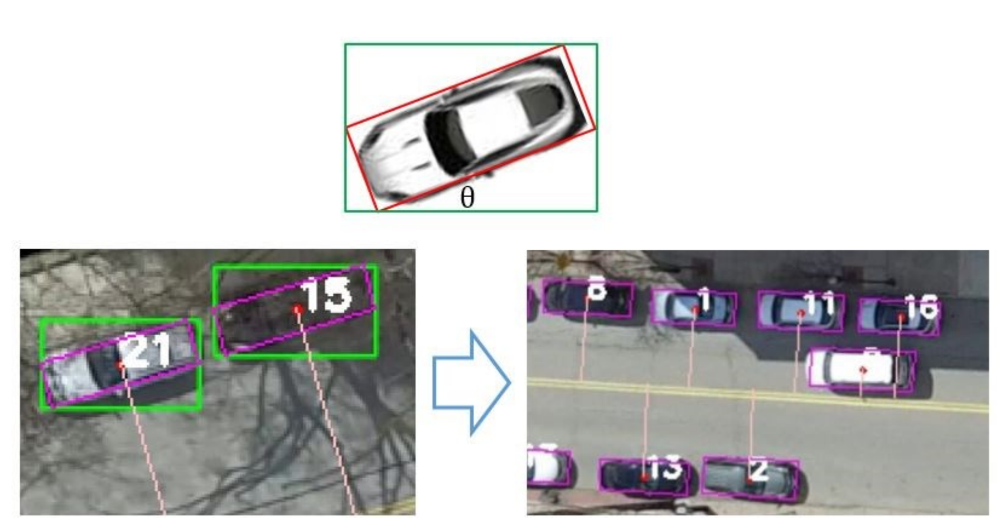

Fig 3. Using the angle of the street, we determine the head angle of the car. At the end, we have the height and with of its center position.

References

- O. Ronneberger, P. Fischer and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597, 2015.
- V. Iglovikov and A. Shvets, TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation, arXiv:1801.05746, 2018.
- H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.

Results

- We were able to represent a map and a dataset, that contains the data of:
  - Local X and Y
  - Vehicle Size
  - Section ID
  - Lane ID

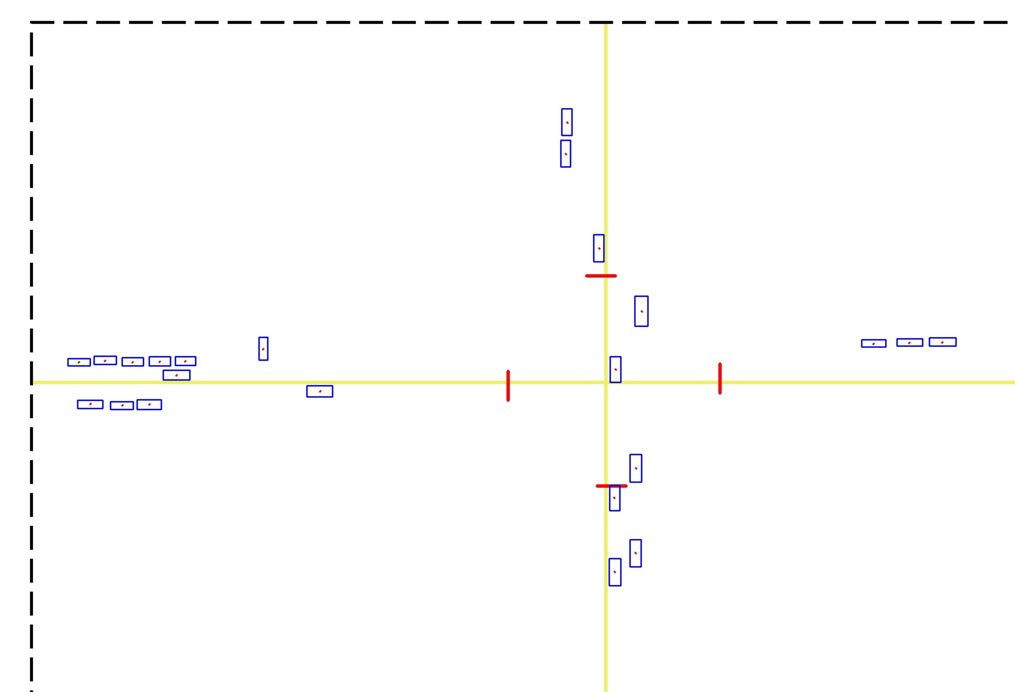

Fig 4. Given an aerial video of a intersection, our method gets the mask of the road. Next, it locates the lane marks. Finally, we export the driver location values about those lane marks in an x-y plane.

Conclusions

- We developed a process capable of building a map, to obtain diverse values of the location of the driver about the street, in an x-y plane.
- Scalable and easy to repeat.

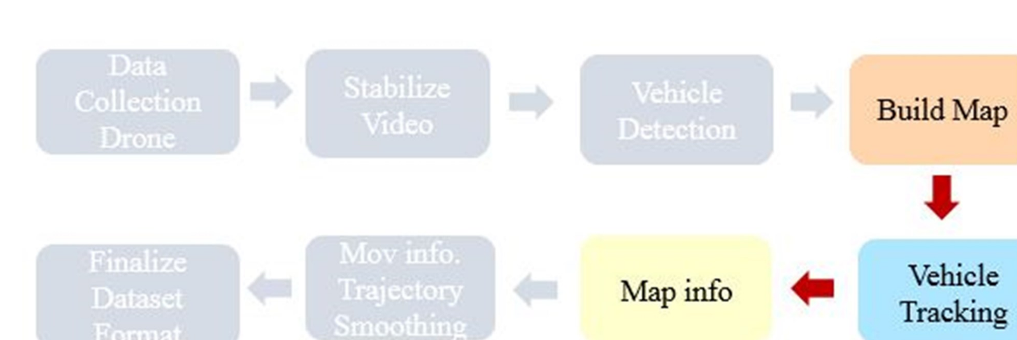

Fig 5. Entire work Flow of the method. In color, map section.

Acknowledgments

- Thanks to Dr. John Dolan and Zhiqian Qiao, for this opportunity and guidance.
- Thanks to Dr. Sergio Camacho and ITESM.
- An special thanks to Rachel Burcin, Ziqi Guo, Dr. John Dolan and all the entire cohort.