Anthony Galczak
Eric Chan
Anthony Li
10-2-19

Robust 3D Object Tracking in Autonomous Vehicles

Project Logistics

This final project will be a mixing of the topics from CS230 (Deep Learning) and CS238 (Decision Making under Uncertainty). Anthony Galczak and Eric Chan are taking CS238 for 4 units, so we will need to prepare a paper at the end. Anthony Galczak and Eric Chan are also in CS230, but Anthony Li is not.

Introduction/Description of General Problem

A requirement for safe autonomous vehicles is object detection in 3D space. By detecting potential obstacles, such as other cars, an autonomous vehicle can plan a route and avoid collisions. However, tracking 3D objects and rotations is a notoriously difficult problem for even neural networks to solve. In 3D detection, we must infer the 3D shape of objects from a single viewpoint with imperfect sensors. Because there are often multiple plausible real-world interpretations for a given image, the bounding boxes generated by most methods tend to jump around between frames, resulting in an unstable detection. Additionally, most pure computer vision approaches are vulnerable to occlusions--they lose tracking when objects pass behind trees, buildings, or other obstacles. In the case of autonomous vehicles, occlusions are common, and tracking loss can be disastrous.

Our goal is to generate stable and accurate 3D bounding boxes around vehicles in a self-driving car dataset. We aim to provide a vehicle tracking system that is robust under short-term occlusions and is more stable frame-to-frame than a pure neural network object detection approach. We plan to use the *Lyft 3D Object Detection for Autonomous Vehicles* dataset on Kaggle[1] and the *KITTI Vision Benchmark Suite*[2].

Discussion of Uncertainties

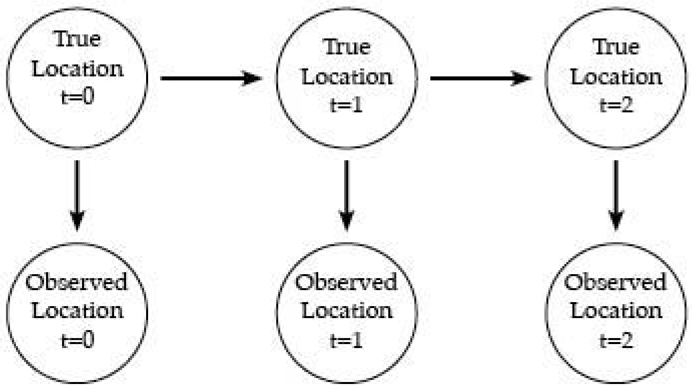
Our goal is to decide the best possible bounding box position for each detected vehicle for each frame of camera/LIDAR data. While computer vision and neural network techniques can provide us with approximate bounding boxes, these results are often noisy, jumping from frame to frame. Additionally, detecting object rotation typically poses a significant challenge for computer vision techniques. Finally, occlusions can cause computer vision algorithms to lose track of the vehicle entirely. We hypothesize that applying decision making techniques will result in more accurate and stable bounding box tracking, even with noisy data and momentary occlusions.

Possible Approach

In order to track vehicles smoothly and robustly, we will approach the problem from two angles. Using a neural network object detection approach, we will use single frames to generate "directly observed" bounding boxes. This direct observation method should give us rough estimates of the real-world positions of vehicles in the scene.

To include information about previous frames, we will introduce a transition model. Our hypothesis is that if we know the position and velocity of a vehicle at frame *t,* we should be able to predict the position of the vehicle at frame *t+1*. By using a transition model, we can predict "interpolated" bounding boxes based on previous frames.

We would like to combine the accuracy of the direct observation model with the smoothness and robustness of our transition model. To do so, we believe that we can represent the problem of object tracking as a hidden Markov model. In this case, the real-world positions of the cars in our scene are hidden states, since they are not directly observable. Our observations are the outputs from our neural network detection model. The transitions between hidden states are defined by our transition model. Using a filtering technique such as a Kalman filter, we believe we can fuse the "directly observed" bounding box prediction with the "interpolated" bounding box prediction, ideally giving a more stable and robust prediction of true location.



References

[1] https://www.kaggle.com/c/3d-object-detection-for-autonomous-vehicles/overview
[2] http://www.cvlibs.net/datasets/kitti/