RGB-D Images for Object Segmentation,Localization and Recognition

Akash Verma

Supervisor: Dr. Sunil Kumar

ABV-IIITM, Gwalior

August 6, 2022

विश्वजीवनामृतं ज्ञानम्

Table of Contents

Introduction

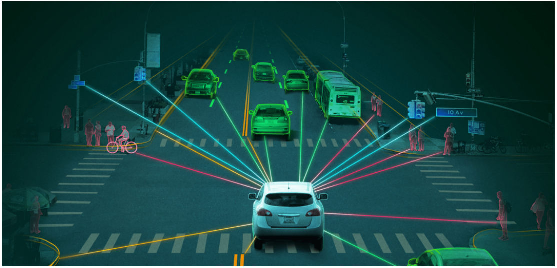

Figure 1: Autonomous Driving

One of the most critical components of autonomous driving is the detection of 3D objects: a self-driving car must accurately detect and locate objects such as cars pedestrians and even occluded objects in order to plan the route safely and avoid collisions.

Lidar System

- To this end, existing algorithms primarily rely on LiDAR (Light Detection and Ranging) as the input signal, which provides precise 3D point clouds of the surrounding environment.

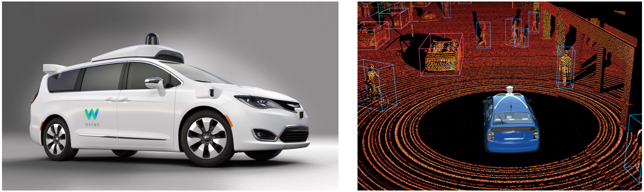

Figure 2: Working Principle of Stereo Camera

- LiDAR, however, is very expensive. Another problem of Lidar is its inability to measure distance through heavy rain, snow, and fog.

Stereo System

To this end, existing algorithms primarily rely on LiDAR (Light Detection and Ranging) as the input signal, which provides precise 3D point clouds of the surrounding environment.

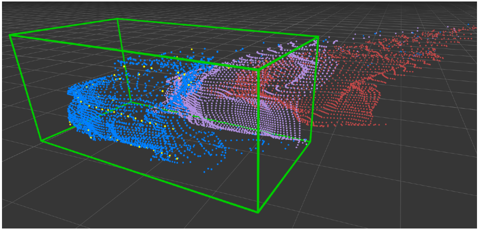

Figure 3: Stereo System

The green box is the ground truth location of the car in the KITTI dataset. The red points are obtained with a stereo disparity network. Purple points, obtained with our stereo depth network (SDN), are much closer to the truth. After depth propagation (blue points) with a few (yellow) LiDAR measurements the car is squarely inside the green box.

Working of Stereo Camera

- A stereo camera is a type of camera with two or more lenses with a separate image sensor or film frame for each lens. This allows the camera to simulate human binocular vision, and therefore gives it the ability to capture three-dimensional images

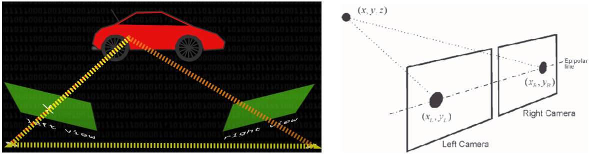

Figure 4: Working Principle of Stereo Camera

Pseudo-LiDAR

Pseudo-LiDAR first utilizes an image-based depth estimation model to obtain predicted depth pseudo-LiDAR requires two systems: a depth estimator, typically trained on a generic depth estimation (stereo) image corpus, and an object detector trained on the point cloud data

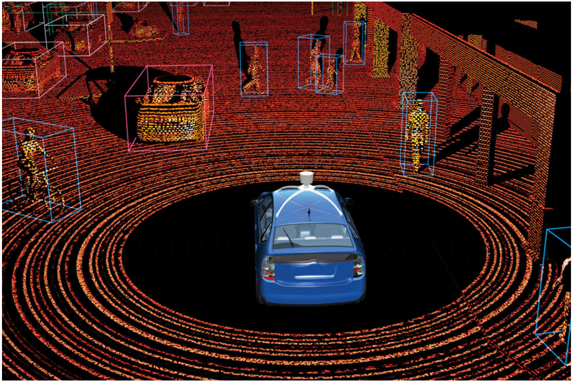

Figure 5: Psuedo Lidar

Pseudo-LiDAR and Point Cloud

Pseudo-LiDAR first utilizes an image-based depth estimation model to obtain predicted depth Z(u; v) of each image pixel (u; v). The resulting depth Z(u; v) is then projected to a "pseudo-LiDAR" point (x; y; z) in 3D by

$$z = Z(u; v)$$

$$x = \frac{(u - c_u)z}{f_u}$$

$$y = \frac{(u - c_v)z}{f_v}$$

where $(c_u; c_v)$ is the camera center and f_U and f_V are the horizontal and vertical focal lengths. The "pseudo-LiDAR" points are then treated as if they were LiDAR signals, over which any LiDAR-based 3D object detector can be applied.

Working of our Model

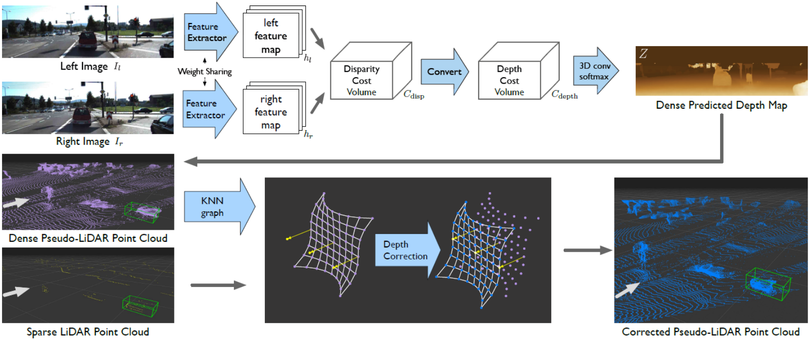

Figure 6: Model

Stereo Depth Network

We propose changes to adapt stereo networks for direct depth estimation.
We learn stereo networks to directly optimize the depth loss.

$$z \propto \frac{1}{D} \Rightarrow \delta Z \propto \frac{1}{D^2}\delta D \Rightarrow \delta Z \propto Z^2 \delta D$$

$$\sum_{u,v \in A} l(Z(u,v) - Z^*(u,v))$$

Objectives and Research Gap

- Depth Loss
- Depth Correction
- Depth Map

References I

[1] Y. You, Y. Wang, W. Chao, D. Garg, G. Pleiss, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-lidar++: Accurate depth for 3d object detection in autonomous driving," *CoRR*, vol. abs/1906.06310, 2019.

[2] Y. Wang, W. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving," *CoRR*, vol. abs/1812.07179, 2018.

[3] R. Qian, D. Garg, Y. Wang, Y. You, S. J. Belongie, B. Hariharan, M. Campbell, K. Q. Weinberger, and W. Chao, "End-to-end pseudo-lidar for image-based 3d object detection," *CoRR*, vol. abs/2004.03080, 2020.

Thank You