RBE/CS549: Computer Vision
Project 3 EinsteinVision

Shreya Bang
M.S. Robotics Engineering
Email: srbang@wpi.edu
Using 2 Late Days

Rajus Nagwekar
M.S. Robotics Engineering
Email: rmnagwekar@wpi.edu
Using 2 Late Days

***Abstract*—The project focuses on developing a completely vision-based, intuitive dashboard for assisting human drivers on the road. The challenge was to discover implemented algorithms for detecting vehicles, road signs, lane markings etc and to create a 3D render of the perceived world using Blender. Index: Object Detection, Depth Mapping, 6D Pose Detection**

A. *Overview*

Human-Robot interaction is a major problem in the robotics industry. It is very important for the car to have a good understanding of its surroundings using all the sensors available at its disposal, but it is also important for the car to relay this information to its human user so as to achieve a certain level of collaboration between the human and the machine. The 3D render developed during this endeavour aims to alleviate the gap between the robot's & the human's understanding of the world scene.

B. *Monocular Depth Detection*

Since it is assumed that the vehicle does not receive any kind of depth data at all, we use monocular depth detection to place the objects in the 3D scene. As seen in [1], the MiDas DPT-large model has been trained on 12 separate datasets. The dpt_beit_large_512 has almost 11 million parameters and an inference rate of 5.7fps.

Fig. 1: Monocular Depth

C. *Vehicle Detection*

Vehicle detection, vehicle subtype detection & vehicle orientation detection are of paramount importance for the vehicle to avoid collision and determine safe, drivable areas of the road. We used Resnet-101 trained on the COCO dataset to get upto 9 vehicle classes[2].

Fig. 2: Vehicle Detection & Subclassification

D. *Vehicle Pose Detection*

To appropriately place the vehicle in the scene, the 6D pose of the vehicle was calculated using YOLOv3 along with the COCO dataset, as described in [1].

Fig. 3: Vehicle Pose Detection

E. Lane Detection

Lanes are used by vehicles to identify the drivable area on the road and also to help the car drive within the marked region. We used a mobilenetv2 network trained on the CuLane dataset to detect these lanes [2].

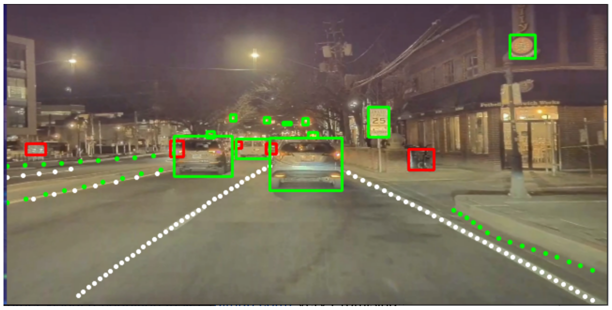

Fig. 4: Lane Detection

F. Object Detection

Objects commonly encountered on roads, such as pedestrians, traffic lights, street signs, dustbins, traffic cones, traffic lights etc were detected by using Resnet & Resnext101 models trained on the COCO & LVIS dataset respectively[4]. The traffic signal colour were detected using the RegNetY model [3].

Fig. 5: Object Detection

G. Taillight Detection

The LVIS model is capable of segmenting the taillight from a the image of the car. We used this obtained region of interest and used the HSV colour space of the RoI to determine whether the tailight is displaying the colour red.

H. Human Pose Detection

Human Pose Detection was implemented using a ResNext model trained on the COCO keypoints dataset. This dataset is capable of identifying 17 keypoints on the human body so as to estimate the current pose.

Fig. 6: Human Pose Detection

I. Speed Bump Detection

Whenever a speed bump is seen, there is a sudden change in the depth intensity map for the pixel which can be estimated to be on the road itself. We leveraged this to detect the speed bump on the road.

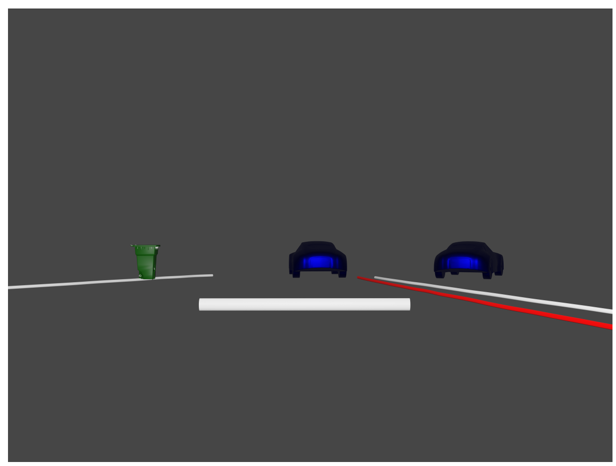

Fig. 7: Speed Bump Detection

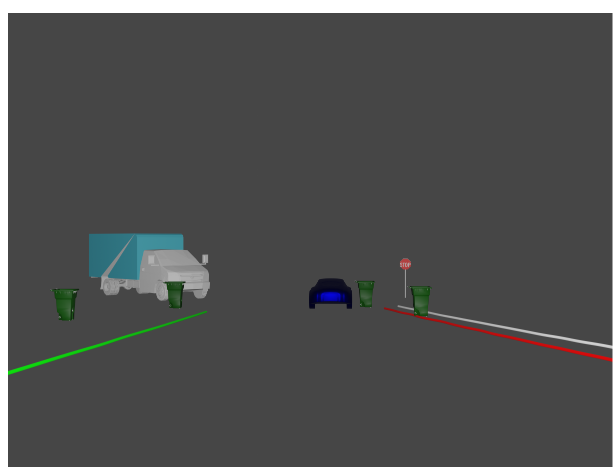

Fig. 8: Blender Output

REFERENCES

[1] John Flynn Jana Kosecka Arsalan Mousavian, Dragomir Anguelov. 3d bounding box estimation using deep learning and geometry. *CVPR*.

[2] Ignacio Sañudo Olmedo Marko Bertogna Carmelo Scribano, Giorgia Franchini. Cerberus: Simple and effective all-in-one automotive perception model with multi task learning. *IROS*.

[3] Amirhossein Kazerouni, Amirhossein Heydarian, Milad Soltany, Aida Mohammadshahi, Abbas Omidi, and Saeed Ebadollahi. An intelligent modular real-time vision-based system for environment perception.