ABSTRACT

A software system is developed using the YOLO object detection model and the Midas depth estimation algorithm. The system is designed to detect and localize objects in real-time and estimate their depth from a single monocular image or video stream. The YOLO model is used for object detection, while the Midas depth estimation algorithm is used to estimate the depth of the detected objects.

The system is built using Python 3.6 and PyTorch 1.7, and it utilizes the CUDA Toolkit and cuDNN library for GPU acceleration. The COCO dataset is used for training and testing the YOLO model, while the Midas dataset is used to train and test the depth estimation algorithm. The software system is tested on various image and video datasets, and its performance is evaluated in terms of accuracy, speed, and resource utilization.

The results show that the YOLO model with Midas depth estimation achieves high accuracy in detecting and localizing objects, while also providing reliable depth estimates. The system is capable of processing images and videos in real-time, making it suitable for applications such as autonomous vehicles, robotics, and surveillance systems. The software system is also designed with a user-friendly interface, allowing users to easily input images or videos and obtain the results in real-time.

An alternative to this project is Faster R-CNN with Depth Estimation. Faster R-CNN is another popular object detection model that can be used for detecting and localizing objects in real-time. By integrating it with a depth estimation algorithm, such as the StereoBM algorithm, depth information can be estimated for the detected objects. But this object detection algorithms use regions to localize the object within the image. The network does not look at the complete image. Instead, parts of the image which have high probabilities of containing the object. YOLO object detection algorithm on the other hand is much different from the region based algorithm. In YOLO a single convolutional network predicts the bounding boxes and the class probabilities for these boxes and also is orders of magnitude faster(45 frames per second) than other object detection algorithms.

Overall, this project demonstrates the potential of combining the YOLO object detection model with the Midas depth estimation algorithm to develop an efficient and accurate object detection and depth estimation system. The software system developed in this project can be further optimized and extended to support additional features and functionalities.