2D Object detection with monocular depth estimation

Abrar Naim Shahiruddin Bin Shahbudin*, Che Wab Ar-Rayyan Bin Che Wan Shamsiruddin†,
Muhammad Ammar Bin Mohd Hazlan‡, Muhammad Tareq Adam Bin Ellias§,
Muhammad Zahirul Isyraf Bin Mohamed Aidi Shahriz¶

***Abstract*—The advancement of robotic arm systems has increasingly relied on object detection algorithms for efficient pick-and-place tasks. While 3D object detection models such as YOLOv8 have demonstrated high performance in these scenarios, their application is limited in low-cost robotic systems due to hardware constraints. To address these limitations, this project develops an integrated framework that combines 2D object detection with monocular depth estimation, enabling robust object localization and depth analysis using cost-effective hardware. The proposed system utilizes YOLOv8 for precise 2D object identification and classification. Simultaneously, depth estimation is achieved through a monocular camera setup, offering a streamlined approach to determining object distances without the need for expensive stereo or LiDAR systems.**

***Index Terms*—2D Object Detection, Monocular Depth Estimation, YOLOv8, Robotics, Computer Vision**

I. Introduction

In this section, [1]you will introduce the topic of your paper, providing background information on the problem you're addressing. You should also state the primary objectives of your research and the significance of your work.

Briefly describe the structure of the paper, outlining the contents of each section. The introduction should capture the reader's attention and explain why the research is important.

II. Related Work

This section discusses previous work in the field that is relevant to your study. Provide an overview of key studies and methods that have been used in the past to solve similar problems. Highlight the strengths and weaknesses of these approaches, and explain how your work builds upon or differs from them.

III. Methodology

The methodology for the 2D object detection and monocular depth estimation system leverages the integration of YOLOv8 for object detection, a robotic arm for manipulation, and monocular depth estimation using a pinhole camera model. The approach is designed to detect an object from a monocular camera feed, estimate its depth, and position a robotic arm to interact with the object. The main components of the system are described below. Throughout the system, safety protocols are incorporated to ensure safe interaction with objects and the environment. The system's joint angles are clamped to a safe range to avoid damaging the robotic arm. Additionally, the camera calibration and depth estimation parameters are periodically validated to maintain system accuracy.

IV. Experiments and Results

This section should include the experimental setup, evaluation metrics, and results obtained from running your experiments. Use tables, figures, and charts to present the results. You should also provide an analysis of the results, comparing them to other approaches or baselines if applicable.

V. Conclusion

The conclusion summarizes the main findings of the paper, discusses their implications, and suggests potential future work. You may briefly mention any limitations of your study and how these could be addressed in future research.

References

[1] A. Smith and C. Jones, "Title of the paper," *Journal Name*, vol. 12, pp. 1–10, 2020.