Deep Neural Networks and stereo vision for Enhanced Distance Estimation and Object Detection

Askil Folgerø, Jonas Hedegaard

February 2024

Advisor: Christian Møller Dahl, Syddansk Universitet
Company contact: Carl Johan Danbjørg, SCAPE Technologies

1 Motivation

Robots used in current industrial setting are getting automated to solve complex and changing tasks. Their ability to problem-solve is dependent on their ability to see, traditionally solved by 3D cameras and CAD models. This Master's Thesis project proposes a application of neural networks and stereo 2D image processing to address the limitations of traditional point cloud imaging and CAD based models. By leveraging deep neural networks and traditional 2D image data, used in a stereo configuration, the project aims to develop a more efficient and robust solution for robotic handling systems used at Scape Technologies.

2 Problem statement

Can stereo imaging, in conjunction with deep neural networks and object detection, be used for accurate distance estimation, and serve as a robust and efficient replacement for traditional point clouds and CAD models used by robots in a production environment?

3 Project description

The primary objective of this project is development of a deep neural network, capable of accurately estimating distance from object to sensor, and a system to detect a fixed location on objects from where a robot is safe to grab, using two 2D images of object/plane. This project seeks to overcome the challenges different materials such as reflective surfaces, close proximity positioned objects and different light environments pose to current technologies and solutions. Thereby enabling a more robust, and reliable automation of robotics.

Theory and methodology: The methodology encompasses collecting and prepossessing stereo image data from real production robotics environments, designing a stereo input deep neural network to process data, distance estimation, disparity mapping, object detection. Implementing a series of experiments to train, test, and refine the model. The project will focus on the model's distance depiction accuracy and robustness to ensure it can be seamlessly integrated into production workflows. To attain our goal we will use and be working with stereo image data, machine learning, deep neural networks, computer vision, generate disparity image data.

4 Timeline and milestones

- **Prototyping and data collections**. Exploring small scale physical solutions, point cloud label generation, and synthetic generated data using open source 3D software, for image and labels.
- **Model testing**, and development of small scale model on data, proof of concept, testing, validating.
- **Object detection**: Generating a object detection system for hierarchically choosing object which are safe to pick up by robot.
- **Pipeline for final data collection**. Collecting stereo image data from production environment.
- **Prepossessing**. Converting video feed to image date, and generating ground truth labels, used for large scale model development.
- **Model generation**. Using deep learning libraries such as, but not limited to PyTorch to generate small/large scale deep neural network models, for testing and validation.
- **Performance analysis**. Assessing the systems prediction against performance of already existing solutions.
- **Finalize model**. Tuning and assesing performance.
- **Documenting and reporting**. Document the methodology, analysis, and findings in a master's thesis.

5 Literature

- Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., & Bry, A. (2017). End-to-End Learning of Geometry and Context for Deep Stereo Regression. In Proceedings of the IEEE International Conference on Computer Vision (Vol. 2017-October).

- Luo, W., Schwing, A. G., & Urtasun, R. (2016). Efficient Deep Learning for Stereo Matching. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Vol. 2016-December).
- Žbontar, J., & Le Cun, Y. (2015). Computing the Stereo Matching Cost with a Convolutional Neural Network. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Vol. 07-12-June-2015).