A Point Cloud Semantic Segmentation Framework for Embedded Systems in Agricultural Robots

Gary Storey
Department of Computer Science
Loughborough University
Loughborough, UK
G.Storey@lboro.ac.uk

Lei Jiang
Department of Computer Science
Loughborough University
Loughborough, UK
L.Jiang2@lboro.ac.uk

Qinggang Meng
Department of Computer Science
Loughborough University
Loughborough, UK
Q.Meng@lboro.ac.uk

***Abstract*—In this paper the task of point cloud semantic segmentation from RGB-D data in outdoor agricultural environments is addressed specifically for the challenges of detecting trees, obstacles and the safe ground to traverse. A multi-step framework is proposed to enable real-time processing on an embedded system for use in wheeled agricultural robots. The initial step uses a MobileNetV2 based deep learning model to perform semantic segmentation on an RGB image. A point cloud is created from the segmentation and depth images which is then down-sampled, finally RANSAC plane segmentation refines the final segmentation output. Initial findings show the method performs well at labelling the desired targets while also running at up to 9fps on the embedded system.**

I. Introduction

Within computer vision and robotics, point cloud semantic segmentation (PCSS) has in recent years became an activate area of research [1]. The aim of PCSS is to provide semantic labelling to each point in the cloud, this information can then be used to identify specific objects/areas within an environment for example pedestrians, trees and roads. This can provide a robot with contextual understanding of the environment and therefore can plan and act accordingly.

A variety of different approaches have been applied for the task of PCSS from a range of different data sources including RGB-D and Lidar. Methods such as PointNet [2] and PointSeg [3], perform segmentation in the point cloud based upon 3D shape matching and edge features, without the consideration of RGB data. These methods are reliant on available ground truth data for model training, while available data sets like Semantic3D.net [4] are limited to urban and indoor environments. In comparison the RGB domain contains much richer data sets like the ADE2K [5] which also contains agricultural scenes.

In this paper a novel multi-step framework for PCSS from RGB-D data is presented. The application domain for the framework is wheeled field robots, specifically those operating in outdoor agricultural environments. The targeted hardware is an embedded system the Nvidia Jetson TX2, with data from an RGB-D stereo depth camera the Intel Realsense D435. The PCSS framework proposed aims to perform segmentation of trees, obstacles and the ground. This task provides three distinct challenges firstly unlike their urban counterparts which benefit from distinct roads with relatively flat surfaces, an agricultural wheeled robot may be required to drive on multiple ground types, i.e gravel, mud, grass some of which have deeply uneven surfaces. Therefore not only is it important to identify the ground but also to detect ground areas which may unsuitable to traverse, to allow safe path planning. Secondly there needs to be robustness to unknown or incorrectly labelled obstacles as segmentation models can be prone to error on unseen object types. Finally computational efficiency is required to run on the embedded hardware platform. To tackle these challenges the proposed framework initially employs a deep learning model to provide semantic segmentation on a RGB image to exploit the rich data sets in this domain, this is then refined within the point cloud initially applying intelligent down-sampling followed by RANSAC plane segmentation [6].

This research was funded through Innovate UK (Project 104016)

II. Proposed Method

The proposed multi-step framework for PCSS initially captures aligned RGB-D frame data from the Realsense D435 camera before performing the three main steps as detailed in this section. An overview of this framework is shown in Figure 1. The Open3d [8] Python library was used for creating and working with point cloud data.

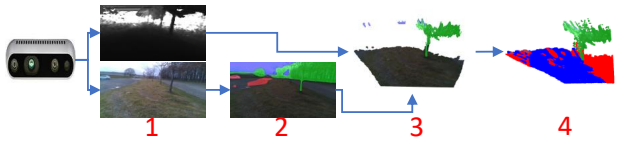

Fig. 1. Framework Steps Overview. (1) Captured RGB and Depth Data, (2) RGB Semantic Segmentation, (3) Point Cloud Creation, (4) Refined Semantic Segmentation.

A. RGB Semantic Segmentation

Semantic segmentation is the process of predicting a class label for each pixel in the input image. The proposed framework applies a lightweight model to enable fast performance on the Nvidia Jetson TX2. A MobileNetv2 [7] encoder with dilated convolutions [9] in tandem with a single layer decoder is used and implemented in PyTorch. The model was trained using the ADE2K training data set [5]. Given an input RGB

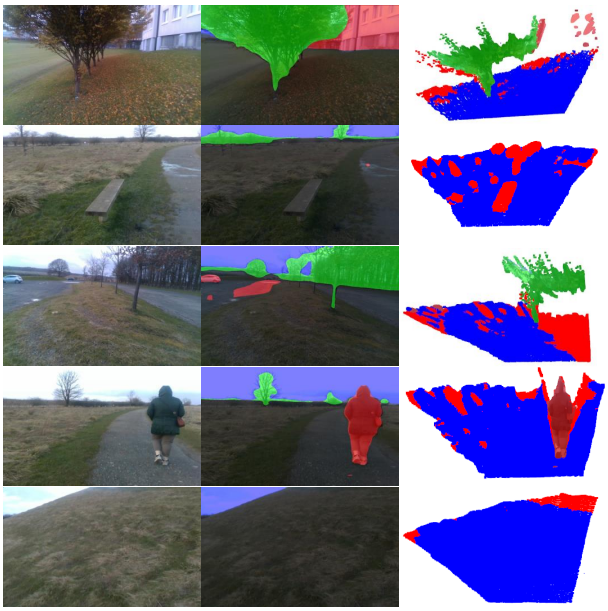

Fig. 2. Point Cloud Semantic Segmentation Results. (Left) RGB Image, (Centre) RGB Segmentation Output, (Right) Final output, blue denotes safe ground, red for obstacles and green for trees.

image of $n \times m \times 3$, the output of the RGB segmentation model S is $n \times m \times 3$ where each predicted class correlates to a specific RGB value.

B. Point Cloud Creation

Given the RGB segmentation output of S, the corresponding depth image D, a depth scale λ and the intrinsic camera parameters $C = \{c_x, c_y, f_x, f_y\}$, a point cloud $PC = \{pc_1, pc_2, ..., pc_n\}$ with n points is then generated as:

$$x_i = (u_j - c_x) \times z / f_x \tag{1}$$

$$y_i = (v_j - c_y) \times z / f_y \tag{2}$$

$$z_i = d / \lambda \tag{3}$$

where u_j and v_j are the jth location in D. While c_x, c_y, f_x and f_y are the camera intrinsic parameters. Each point pc_i is a vector given by $\{x_i, y_i, z\}$. An associated RGB colour matrix $C = \{c_1, c_2, ..., c_n\}$ also exists where $c_i = \{r_i, g_i, b_i\}$.

C. Point Cloud Down-Sampling and Segmentation

Processing point cloud data can be computationally expensive depending upon the number of points, to reduce this point removal via down-sampling is used. To retain sufficient detail the following down-sampling method is proposed. Firstly points corresponding to classes of high confidence are made exempt from calculation. Confidence is established via validation of the RGB segmentation model with ADE2K validation data set [5]. Specific points can be established by the RGB value of the class e.g. trees in C and removed from PC. Uniform nearest neighbour down-sampling is then applied. The PCSS is refined using RANSAC plane segmentation [6]. Final RGB class values are assigned to the relevant points in C where blue denotes flat terrain and red obstacles.

III. Experiments

To produce an initial evaluation two experiments were performed. One to assess PCSS accuracy and a second to evaluate computational efficiency on the target hardware. The experimental parameters were as follows, RGB and depth image size was 640 by 480. Nearest neighbour down-sampling was set to 10. RANSAC applied a distance threshold of 0.1, initial points of 100 and 50 iterations.

A. Segmentation Accuracy

Five rural images were evaluated, varying in ground and objects types. Figure 2 shows the images and outputs. The RGB segmentation proves adept at classifying trees, sky, ground and some obstacles classes such as people. Where deficiencies in RGB segmentation exist, like a missed bench object in one image the point cloud based method successfully detects the obstacle. The PCSS method can also distinguish flat from uneven ground surfaces, though depth data sparsity has an impact on some predictions further into the point cloud.

B. Computational Efficiency

Point clouds were tested at depths of 1, 5 and 10 metres. The five images where each processed 10 times and final mean values are given in Table I. When depth is 1 meter the performance is 9fps at 10 meters this is reduced to 5-6fps. Prior to down-sampling the average points per cloud where 69788 and 220722 for 1 and 10 meters respectively.

TABLE I
Mean Computational Efficiency Breakdown in Seconds

Step	1 Meter	5 Meter	10 Meter
RGB Segmentation	0.064s	0.064s	0.064s
Point Cloud Creation	0.017s	0.034s	0.038s
Point Cloud Down-Sampling	0.018s	0.049s	0.046s
RANSAC	0.008s	0.018s	0.021s
Total	0.11s	0.16s	0.17s

IV. Conclusion

In this paper a framework for PCSS is proposed, the initial finding are promising providing a foundation for future research. Using both RGB and point cloud semantic segmentation methods leverages the current strengths of each domain and aid robustness, for example in detecting missed obstacles from the RGB domain. The ability to distinguish flat from uneven terrain is extremely relevant for field robots especially as this framework runs on a embedded system and RGB-D camera costing less than £600. There are limitations due to the qualitative nature of the results, generating ground truth data regarding ground contours is an area for future research so more substantial quantitative results can be generated, while opening up the possibly of applying direct point cloud segmentation techniques like PointSeg in fusion with RGB segmentation.

REFERENCES

[1] J. Zhang, X. Zhao, Z. Chen and Z. Lu, "A Review of Deep Learning-Based Semantic Segmentation for Point Cloud," in IEEE Access, vol. 7, pp. 179118-179133, 2019.

[2] R. Q. Charles, H. Su, M. Kaichun and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 77-85.

[3] C. Li, X. Wei, H. Yu, J. Guo, X. Tang and Y. Zhang, "An Enhanced SqueezeNet Based Network for Real-Time Road-Object Segmentation," 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, 2019, pp. 1214-1218.

[4] H. Timo, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler and M. Pollefeys. "Semantic3D.net: A new Large-scale Point Cloud Classification Benchmark." ArXiv abs/1704.03847, 2017.

[5] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso and A. Torralba, "Scene Parsing through ADE20K Dataset," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5122-5130.

[6] M. A. Fischler, R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Comm. of the ACM, Vol 24, 1981, pp 381-395.

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 4510-4520.

[8] Q-Y. Zhou, J. Park and V. Koltun, "Open3D: A Modern Library for 3D Data Processing," arXiv:1801.09847, 2018.

[9] F. Yu and V Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv preprint arXiv:1511.07122, 2015.