Homework 2: Sexy Semantic Mapping

Alex Chiluisa
Used two late day

I. Introduction

This paper reports on the process to build a map from a point cloud generated from LIDAR raw data and then transferring the result semantic labels from the camera image onto the LIDAR point cloud.

II. Building The Map

For this section, we used Iterative Closest Point(ICP) algorithm. ICP handles the principle of point cloud registration to define the Euler transformation relationship between the 3D point cloud data obtained from several images from different angles, therefore, the different views can be unified into the same coordinate system. In other words, if we have two sets of point cloud data that came from the same target from different angles, the purpose of the registration is to transform those cloud data by translation and rotation transformations [1].

Assuming we have A and B (two set of point cloud data); for every point in a_i in A, the nearest point on B is searched as the corresponding point b_i, then the rigid body transformation is solved so the objective function (eq. 1) obtains the minimum value [2].

$$min \sum \|b_i - a_i\|^2 \tag{1}$$

ICP gets accurate registration results of any shape of point cloud data. Additionally, it does not require the segmentation of point clouds or any feature extraction process. It will guarantee coverage by using better initial values. However, the computational cost to find the corresponding point is large and sometimes the algorithm assumes that the nearest Euclidean distance of a point is the corresponding point in the other cloud set, which could be the cause of wrong corresponding points [1]. This process will help us to build a high-definition map, in our case based on the LIDAR information.

III. Semantic Point Painting The Map

In this section, we utilized RGB images to generate semantic information. The approach described in [3], [4] generates external image labels to point cloud by applying two operations: 1) Search based Superpixel Labeling and 2) 3D Contextual Refinement. For the first step, it performs a 2-dimensional in superpixels units by over-segmentation. For each superpixel labeled in the reference images, it tries to find the most similar superpixels in the external labeling pool with ground truth label superpixels. This means finding the k nearest neighbors (*k*NN) in the labeling pool by defining an objective function to minimize the measure the distance among superpixels by determining the euclidean distance in the feature space [3]. However, the euclidian measurement may not capture the right intrinsic visual between superpixels, and the label classifier hardly relies on the training data, which could increase the error on the propagation of external image labels to point clouds. One way to mitigate this problem is by applying Exemplar Support Vector Machines (ESVM) to build robust metrics. In this case, for each superpixel withdrawn from the pool, a linear SVM is trained to find visually similar superpixels. ESVM for each superpixel is trained to optimize the margin of the classification boundaries, therefore each superpixel is translated and rotated to expand to more positive examples for training. In the second stage, a 3D Contextual Refinement approach is implemented. Here, the point cloud is labeled given the **k**NN of each superpixel in the reference image without any labeled 3D training data. The construction of the graphical model is shown in Fig. 1. It constructs the nodes and edges from the point cloud to optimize the model.

Each node endorsed its label as the variables on the graphical model, to set the potential function to implement intra and inter-images. Finally, by minimizing those potential functions, the semantic labels for 3D segments are deducted. By introducing the intra-image smoothing term, it is possible that the neighbor superpixels have related labels to encode the intra-image consistency in the reference images. While the inter-image smoothing term helps to make the 3D labeling outcomes consistent among reference images. By integrating both the intra-image and inter-image in and among the reference images, the graphic model combines individual superpixel labels onto 3D points.

This approach helps to mitigate the lack of enough 3D training data for semantic labeling by exploring the existing 2D data.

For this report, we use the definitions and concepts described in [1]–[4] to implement the semantic mapping described in [5]

A. Chiluisa is with the Department of Robotics Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA (e-mail: ajchiluisa@wpi.edu)

References

[1] X. Shi, T. Liu, and X. Han, "Improved iterative closest point (icp) 3d point cloud registration algorithm based on point cloud filtering and adaptive fireworks for coarse registration," *International Journal of Remote Sensing*, vol. 41, no. 8, pp. 3197–3220, 2020.

[2] B. Hexsel, H. Vhavle, and Y. Chen, "Dicp: Doppler iterative closest point algorithm," *arXiv preprint arXiv:2201.11944*, 2022.

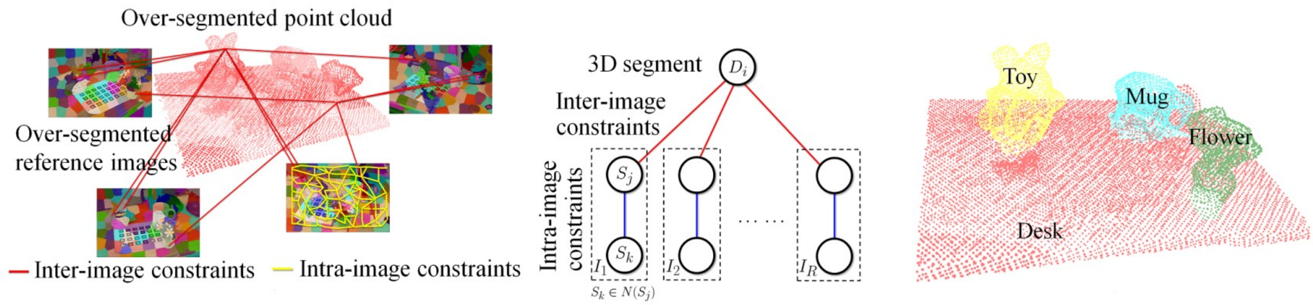

Fig. 1. (Right) Nodes and edge construction from the point cloud, (center) abstract representation of the graphical mode. (left) output after optimizing on the model

[3] Y. Wang, R. Ji, and S.-F. Chang, "Label propagation from imagenet to 3d point clouds," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3135–3142.

[4] A. Boulch, J. Guerry, B. Le Saux, and N. Audebert, "Snapnet: 3d point cloud semantic labeling with 2d deep segmentation networks," *Computers & Graphics*, vol. 71, pp. 189–198, 2018.

[5] https://github.com/AmrElsersy/PointPainting.