Large-Scale 3D Point Cloud Representations via Graph Inception Networks with Applications to Autonomous Driving

Chen, Siheng

TR2019-039 June 28, 2019

Abstract

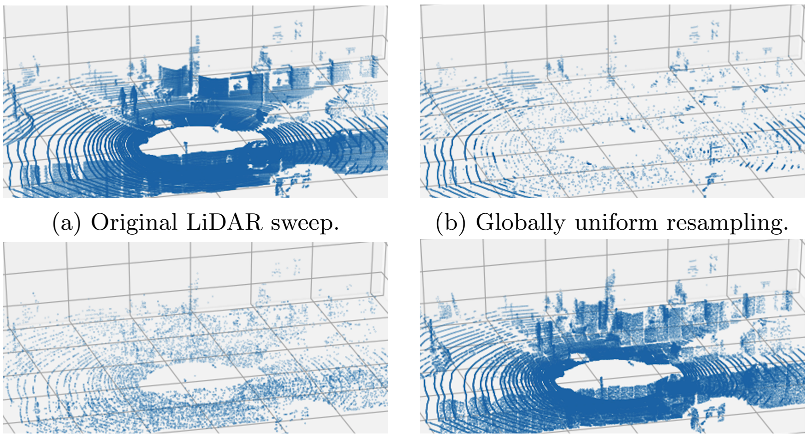
With the growth of 3D sensing technologies, one can now use a large number of 3D points to precisely represent objects' surfaces and surrounding environments. We call those 3D points a 3D point cloud; it has a growing impact on various applications, including autonomous driving, drones, robotics, virtual reality and preservation of historical artifacts [4]. For example, a self-driving car could use multiple sensors to observe the world, such as LiDARs, cameras and RADARs [2]. Among those, LiDARs produce two types of 3D point clouds: real-time LiDAR sweeps and high-definition maps. Both types of 3D point clouds provides accurate range information for self-driving cars, which are critical to localization and perception systems. We consider these point clouds large-scale point clouds because they contain a large number of 3D points and record outdoor, open areas.

Graph Signal Processing Workshop (GSP)

Large-Scale 3D Point Cloud Representations via Graph Inception Networks with Applications to Autonomous Driving

Siheng Chen
Mitsubishi Electric Research Laboratories (MERL)

With the growth of 3D sensing technologies, one can now use a large number of 3D points to precisely represent objects' surfaces and surrounding environments. We call those 3D points a *3D point cloud*; it has a growing impact on various applications, including autonomous driving, drones, robotics, virtual reality and preservation of historical artifacts [4]. For example, a self-driving car could use multiple sensors to observe the world, such as LiDARs, cameras and RADARs [2]. Among those, LiDARs produce two types of 3D point clouds: real-time LiDAR sweeps and high-definition maps. Both types of 3D point clouds provides accurate range information for self-driving cars, which are critical to localization and perception systems. We consider these point clouds *large-scale point clouds* because they contain a large number of 3D points and record outdoor, open areas.



(a) Original LiDAR sweep. (b) Globally uniform resampling.

(c) Octree. (d) PCT: GIN.

Figure 1: Comparison between an original LiDAR sweep and its reconstructions. All reconstructions use 2.78% of original data.

Motivations. Based on the corresponding sensing mechanisms, 1D time series are usually associated with regular-spaced time stamps and 2D images are usually associated with regular-spaced 2D lattices. They are naturally discretized and supported on a regular structures; however, 3D point clouds are irregularly and sparsely scattered in the 3D space and the classical lattice-based methods cannot be directly used to 3D point clouds. To fully exploit 3D point clouds, we need advanced processing and learning techniques to handle a series of challenges, including compression, denoising recognition, semantic segmentation, registration and object detection of 3D point clouds. A fundamental task behind all those challenges is *3D point cloud representations*; that is, representing a 3D point cloud in a compact format, such that it is easy to conduct subsequent processing procedures. As a classical representation tool, Octree partitions the 3D space adaptively and has been an effective representation tool [5]; however, it represents a 3D point cloud only in the 3D spatial domain and does not fully exploit geometry structures formed by 3D points. In classical signal processing, based on the underlying data structures, representation techniques are designed to transform data from one space to another, leading to easy and compact representations. For example, we use the Fourier transform for 1D time-series and the discrete cosine transform for 2D images [6]. A specific aim of this work is to consider such basic representations for large-scale 3D point clouds.

Solutions. There are two main challenges to represent large-scale 3D point clouds: huge variations in a large scene and unknown transform to capture complex geometry structures formed by irregular 3D points. To solve these challenges, we propose a novel graph-neural-network-based system, called the *point cloud neural transform* (PCT). The proposed PCT includes three stages: voxelization and the voxel-level representations. In the voxelization stage, we consider two views to discretize the 3D space into a series of voxels, including the bird-eye view and the range view. The bird-eye view regularly discretizes the 3D

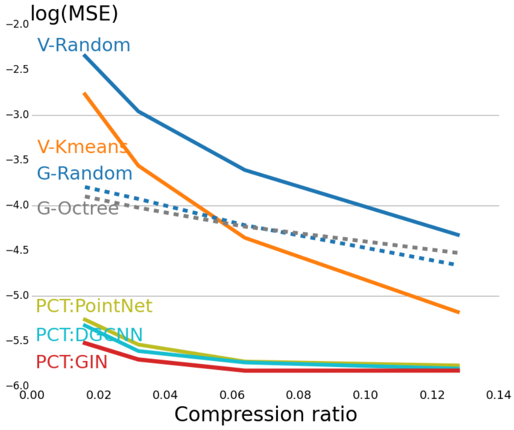


Figure 2: Mean square errors as a function of the compression ratio in the dataset of KITTI. The proposed PCT significantly outperforms its competitors.

space along each of the x, y, z axis, which is general and works for arbitrary 3D point clouds; the range view discretizes the 3D space along the elevate angle and azimuth angle, which reflects the sensing order and works for organized 3D point clouds. In the voxel-level representations, we propose novel *graph inception neural networks* (GIN) to produce a low-dimensional feature vector, which represents 3D points inside each voxel. The proposed GIN have two key features: i) Using an inception structure to handle nonuniform point distributions (ii) Satisfying permutation and translation invariant and promoting scale and rotational equivalence. Because of the two-stage design, the proposed PCT can be considered as the 3D counterpart of the windowed Fourier transform for 1D time series and the blocked discrete cosine transform for 2D images.

Related works. [1] proposed a deep autoencoder that directly handles unorganized 3D point clouds; [7, 3] introduced a 2D lattice to help decoding. The learning-based approach has strong representation power, but requires a huge amount of training data and is hard to be scalable. In this work, we combine the voxelization and learning-based approaches, leading to a system that avoids discretization error and handles large-scale scenarios.

Impacts. As a general representation tool, the PCT can be potentially used to compression, denoising, recognition, semantic segmentation, registration and object detection of 3D point clouds. Here we apply the PCT to represent both real-time LiDAR sweeps and high-definition maps collected by self-driving cars. Figures 1 and 2 shows that the proposed PCT significantly outperforms its competitors.

In summary, the main contributions of this work are as follows:

• We propose the point cloud neural transform (PCT) to represent large-scale 3D point clouds;
• We propose novel graph inception networks (GIN) to implement the voxel-level encoding;
• The proposed PCT is applied to represent real-time LiDAR sweeps produced by self-driving cars and outperforms its competitors.

References

[1] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas. Representation learning and adversarial generation of 3d point clouds. arXiv:1707.02392., June 2017.

[2] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, Providence, RI, June 2012.

[3] T. Groueix, M. Fisher, V. Kim, B. Russell, and M. Aubry. Atlasnet: A papier-mâché approach to learning 3d surface generation. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, June 2018.

[4] R. B. Rusu and S. Cousins. 3D is here: Point cloud library (PCL). In *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, May 2011.

[5] R. Schnabel and R. Klein. Octree-based point-cloud compression. In *SPBG'06 Proceedings of the 3rd Eurographics / IEEE VGTC conference on Point-Based Graphics*, Boston, MA, July 2006.

[6] M. Vetterli, J. Kovačević, and V. K. Goyal. *Foundations of Signal Processing*. Cambridge University Press, Cambridge, 2014. http://foundationsofsignalprocessing.org.

[7] Y. Yang, C. Feng, Y. Shen, and D. Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, June 2018.