Brady Richardson

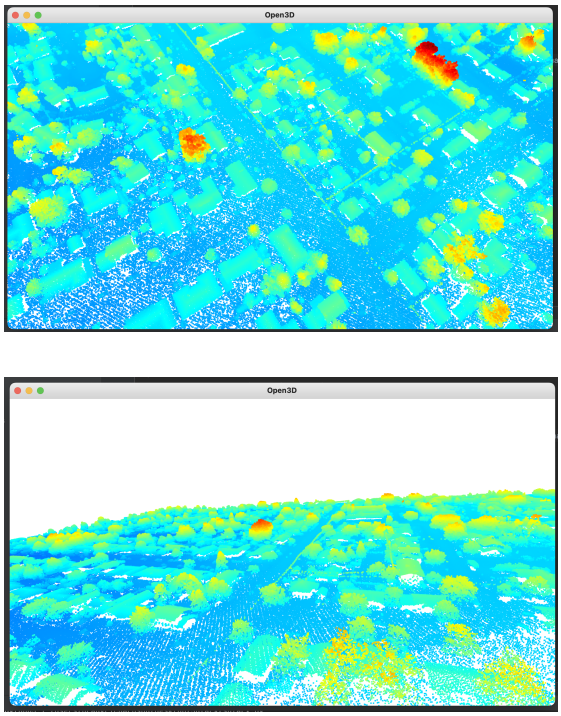
CS 270

Desire to participate: 10/10

10/03/23

Project Proposal: Labeling 3D Data

Description: LiDAR data is very detailed and often can include millions (if not billions) of points in a single dataset. One of the challenges of working with these datasets and the pointclouds that are created from them is the difficulty in annotating the points themselves. Some datasets come labeled, however, there are many that lack classification. This can be difficult when trying to filter out objects or items of interest in a pointcloud. One example of this problem could be a pointcloud of a neighborhood where there are no labels for buildings—you would have to manually derive a way to single out the buildings in the point cloud, which could be a complex process. The purpose of this project is to create a ML tool that can accurately classify these data points using a clustering algorithm such as KNN.



Example of pointcloud without classification/filtering

Features: The features necessary to accomplish this can be found in the LAS files that are used to hold LiDAR data. The relevant features within that data would things like XYZ coordinates, the classifications (for labeled data), intensity (with reference to the reflection of the laser), and the scan angle. All of these features can be found within the LiDAR files, the only caveat is that not all LAS files contain all of this data. The other challenge that I think is appropriate for this class is the need to process the data after it's been downloaded - it comes in a file format that must be parsed and organized using libraries such as

numpy and open3d. These features will need to be assessed and combined in a unique way to determine how to classify each of the pointcloud objects.

X - X coordinate of point in pointcloud

Y - Y coordinate of point in pointcloud

Z - Z coordinate of point in pointcloud

Classification - often unlabeled, 31 possibilities listed here:

https://desktop.arcgis.com/en/arcmap/latest/manage-data/las-dataset/lidar-point-classification.htm

Intensity - intensity of the returned pulse, ranging from -32k to 32k

Scan_Angle - ranging from -90 degrees to 90 degrees (0 degrees is directly below aircraft scanner)

X	Y	Z	Classification	Intensity	Scan_Angle
1745.01	1245.47	114.00	Building	21130.02	20.20
continuous	continuous	continuous	discrete	continuous	continuous

Data Collection: Luckily, there is a lot of publicly-accessible LiDAR data available at https://opentopography.org/. There are many datasets available for everyone, and even more for those who have academic credentials. The data is also customizable in terms of size—you can select different regions within the dataset that you want to analyze. This is good because it can help keep the size of the data down in the event that you have to do batched processing. The only requirement to access this data is to make a free account.

Seeing as it is a unique opportunity to both create a useful tool and get exposure to 3D data visualization/modeling, I think this project should be considered for the class. If classifying all 31 classes would be too difficult for the scope of this project, I suggest focusing in on a few possible object types (buildings, ground, etc.).