2022 Web Mapping and Spatial Data Visualization Class Final Project Proposal

Xin Yan, 3/29/2023

What data can help us estimate building heights at high resolution?

Abstract: Building height is an important indicator for estimating energy consumption, material stock allocation, greenhouse gas emissions, urban heat island effects, distribution of population, and urban development plan. However, openly accessible building height information over large areas at high spatial resolution is still missing. To estimate the building height at a high accuracy, satellite data is the best choice to do that. Besides that, the building morphology also plays a big role. Different satellite data has various band setting, observation characteristics and research focus, the most important work is finding a set of stable features that help us build a model to estimate the building height. After having these features, we can easily predict the building height!

Overview: Building height estimation is a complex and challenging task, and there are several limitations that researchers and practitioners must consider. Limited availability and high cost of high-quality aerial and satellite imagery and lidar data, especially in remote or developing regions, can hinder the accuracy and precision of building height estimates. Using multiple data sources, such as lidar, InSAR, and optical imagery improves building height estimates' accuracy and precision. Sentinel data have a high spatial and temporary resolution, and it shares free in charge, which helps us conduct researches in many areas. I proposed a building height estimation model based on Sentinel 1/2 optical and SAR images and building footprint data. I chose 3 large, medium, and small cities in the United States as training area. The reference building height of each building can be calculated from the LiDAR and DEM data, which has high accuracy and can be reliable, the government also published the building height data in their GIS websites. The width, length, orientation, and building density of each building was calculated in ArcGIS pro using minimum boundary geometry tool, I also extracted the satellite image value of each pixel within the building geometry. Then I will get a corresponding relationship between building height and features. I combined several feature selection methods to select 13 of the 160 features that contributed the most and apply them in the random forest regression model. The methods include Shapley Additive Explanations, Permutation feature importance, and Random forest variable importance ranking, which can also give us the feature importance. And I put these features into a boosted Random Forest Model, which will give me a prediction based on the selected features. The height estimation results were compared and evaluated with Lidar data in training and other cities. The realization of this model can help to obtain large-scale (national or even global) high-precision (10m resolution) building height information and provide data support for other studies and applications. In order to do the further analysis, I will show the building height distributions in different states, for example, the average building height and the number of buildings.

Technology:
Minimum boundary geometry tool in ArcGIS pro;
Feature selection methods in Python;
Machine learning algorithms in R;
Tableau for web mapping;
Shiny for data publish.

Data:
Sentinel 1: bands 2, 3, 4, 8;
Sentinel 2: VV and VH polarization;
LiDAR data;
DEM data;
Building footprint data.

Inspiration: https://public.tableau.com/app/profile/julia.donovan/viz/BuildingsHeights/Dashboard1;
https://github.com/microsoft/USBuildingFootprints

Potential Challenges:
Image displacement issue happens due to the building material variety when the resolution is high;
Data process consuming lots of time and energy;
Raster image mapping

Timeline:
data collection: 3.31 – 4.20
data cleaning: 4.20 – 4.27
data mapping: 4.27 – 5.7