MUSA 611 Final Project Proposal

Xingxing Shu

Problem

Understanding and promoting urban tree cover has never been more important. So, I tried to create an application that shows the greenery index of communities in Philadelphia, especially from the perspective of pedestrian walking on roads instead of looking down from the sky. This application measures the canopy cover in different communities of Philly. Rather than count the individual number of trees or calculate NDVI of road samples, I've developed a scaleable and universally applicable method by analyzing the amount of green perceived while walking down the street. The visualization maps street-level perception only.

Dataset

The data of this application comes from mainly two part: Google Street View and Twitter.

Google Street Map Part: Downloading Google street images of the sample points of the street networks of research region using Google API and extracted greenery pixels from these images and calculate the average Green View Index of each street sample points. Then to find out the difference of greenery index between NDVI and GSV, I used LandSAT 8 image to calculated NDVI and spatial joined the value to those sample points and performed correlation analysis on it.

Twitter part: Due to limitation of time and most because of the gigantic amount of street images if I chose to download all the images of streets in Philadelphia which is impossible for me to process on my laptop. So I decided to choose two typical communities in Philly as my research object. One with high NDVI and the other with relatively low NDVI. Then I collected Twitter near these two communities and tried to find out how urban greenery affect people's sentiment and their polarity and sensitivity of their tweets. The data of this part is collected from Twitter API using tweepy package in Python.

Method

With the street network and boundary shapefile of Philadelphia as input, a shapefile containing points every 50m was created to be fed into the Google API to retrieve Google Street View Images;

With the shapefile as input, metadata containing the panoID, panoDate, latitude, longitude and tilt specifications for the image will be stored in text files to be later used to calculate the Green View Index;

Traditionally we need near infrared band and red band to detect vegetation because vegetation shows high reflectance at near infrared band but shows high absorption at red band. However, GSV images do not have both near infrared band and red band but only red, green and blue bands. So we developed and used a simple automatically unsupervised classification method to extract green vegetation from GSV images.

GVI was defined as the ratio of the total green area from four pictures taken at a street intersection to the total area of the four pictures, calculated using the following equation:

$$\text{Green View} = \frac{\sum_{i=1}^{6}\sum_{j=1}^{3} Area_{g\_ij}}{\sum_{i=1}^{6}\sum_{j=1}^{3} Area_{t\_ij}} \times 100\%$$

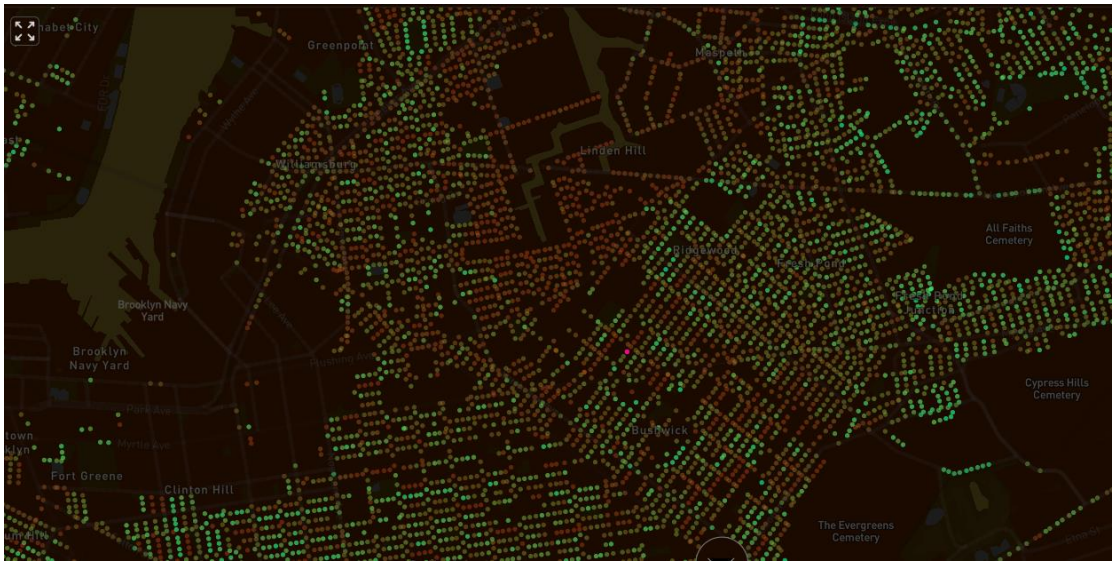
where $Area_{g\text{-}ij}$ is the number of green pixels in one of these images captured in six directions with three vertical view angles (−45°, 0°,45°) for each sample site, and $Area_{t\text{-}ij}$ is the total pixel number in one of the 18 GSV images.

We restrict our work to using tweets that are explicitly geo-tagged as such tweets allow us to determine where they are posted from. These steps include the following: 1) Filtering tweets that are explicitly geo-tagged with latitude and longitude coordinates and within 1km of Philly. 2) Selecting tweets that are written in English, based on the "language" field provided by the Twitter API. 3) Tokenizing each tweet into individual words based on separation by spaces. 4) Converting all tweets and tokenized words into lower-case.

Design

Planners are the primary user of this application as well as residents of Philly. We believe in open data and mobilizing science into the public realm. This application aims to raise a proactive awareness of urban vegetation improvement, using computer vision techniques based on Google Street View images. Our focus is on street trees: it doesn't map parks, as GSV doesn't venture into them as it does on average streets.

Of course parks are an essential component of urban vegetation. But have you ever wondered how sustainable your street or neighborhood is? Does your city need more efforts to make the streets greener? Did you know that you can also contribute by joining a bottom-up approach instead of waiting for things to be done by politicians and planners?



Anticipated Difficulties

Tweets with latitude and longitude information count for very small part of all tweets I collected and I can't find a way to collect tweets within specific region but to collect them based on 'key words' which makes tweets that are useful for analysis is very limited. If I couldn't find a better way to solve this, maybe I'll just show the GSV and NDVI maps of the two communities I chose.

Missing pieces

Hard to tell, JS is so new and sort of difficult to me, I feel like the final application is going to deviate from what I expected ☹