Automated Condition Analysis of Roads

Joel Stansbury, Claire Chan, and Anna Shonia
CX 4240: Introduction to Computational Data Analysis
Spring 2019

Introduction

What are we doing?

- Detecting and Classifying different types of road damage (cracking, potholes, etc.)

Why are we doing it?

- GDOT needs this information in order to determine which roads to work on. The way they obtain it is very inefficient:
  - Performed manually; Inconsistent; expensive; requires road closure

Why is it important?

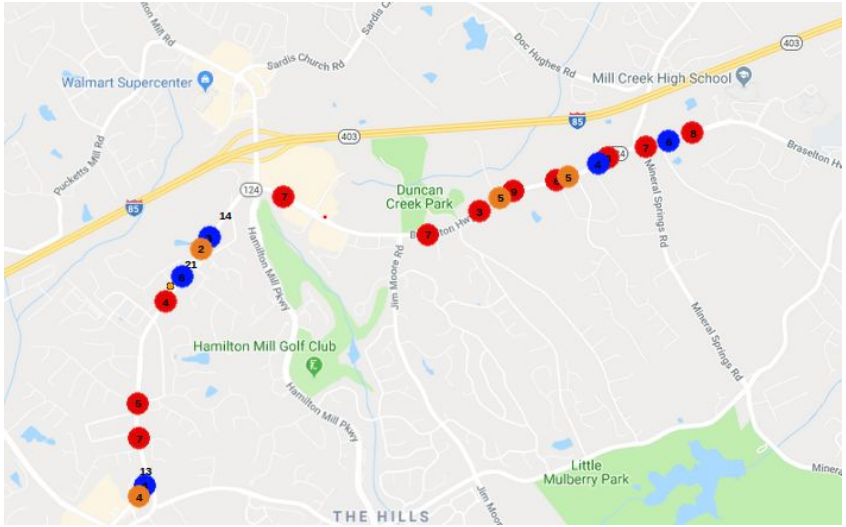
- More efficient resource allocation
  - Make it easier for GDOT to know which roads need to be worked on; help minimize cost of fixing these roads

Data Collection

Made a smartphone app to record

- Audio
- Accelerometer
- GPS

Recorded data as we drove back and forth along the path 5 times



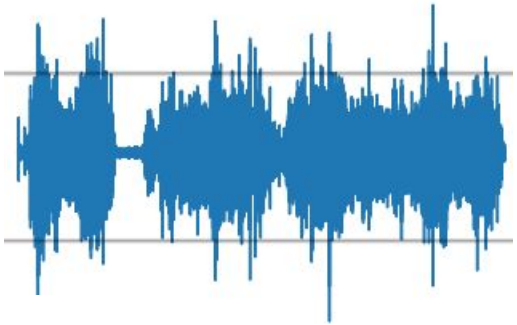
Results of clustering algorithm over the section of road we analysed

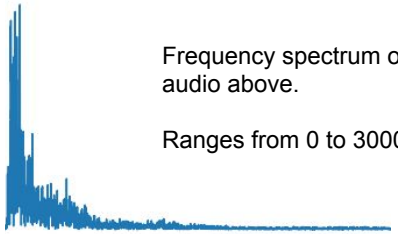
Extracting Features

Look closer at all of the points in time where the volume exceeds a user-specified threshold

At all of those points find the frequency spectrum using FFT.

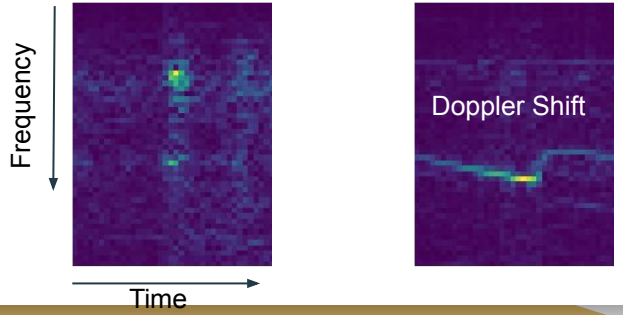
We can place these frequency spectrums side by side to see how they change in time





Frequency spectrum obtained from a 1 second clip of the audio above.

Ranges from 0 to 3000Hz



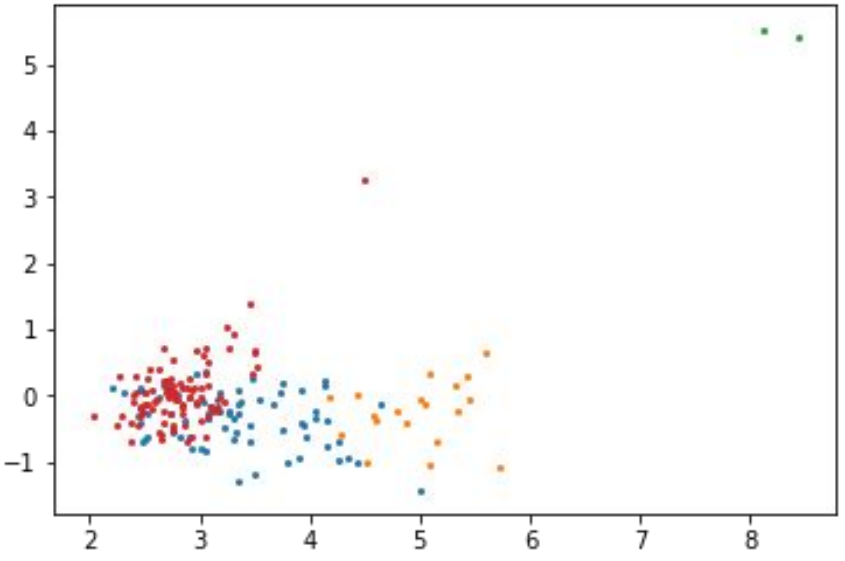
Algorithm

What is our approach to the problem?

- First, **SVD** to perform dimensionality reduction on the feature matrix
- Then, **GMM** algorithm for anomaly classification

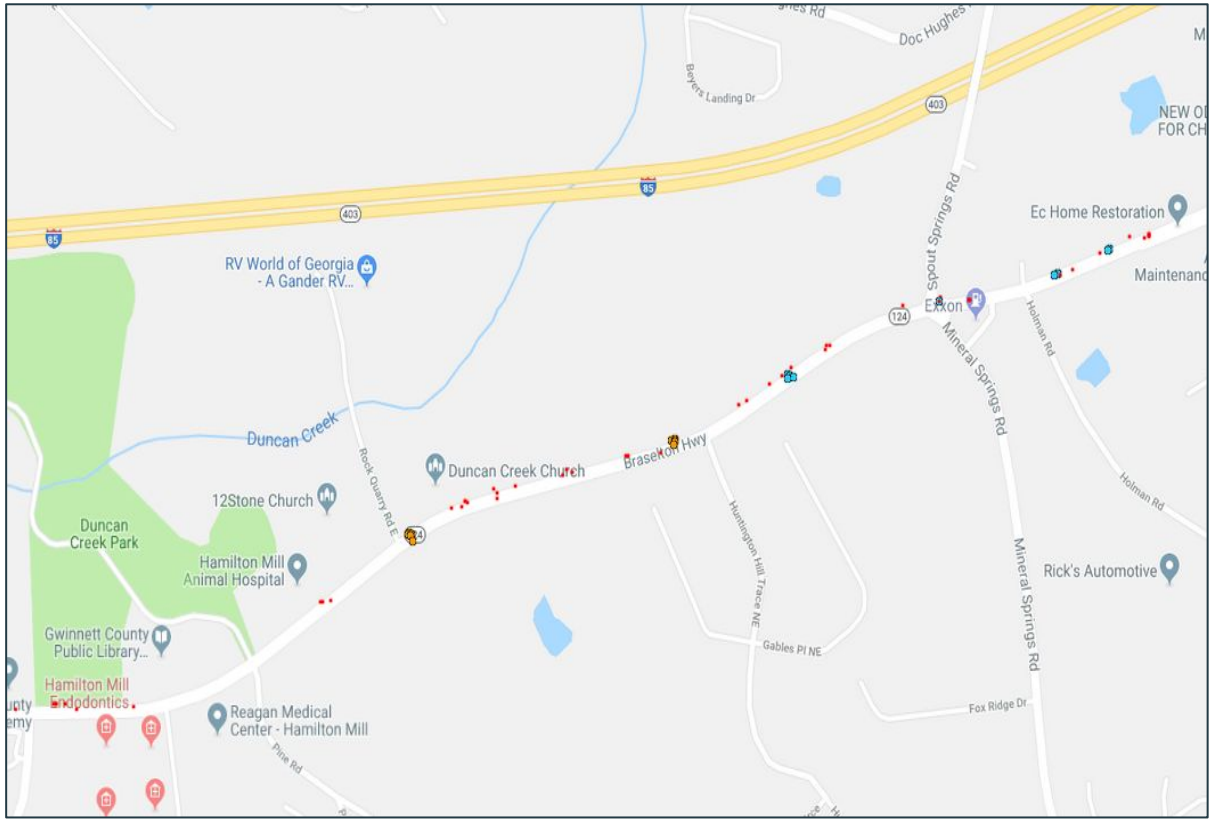
What is new about this approach?

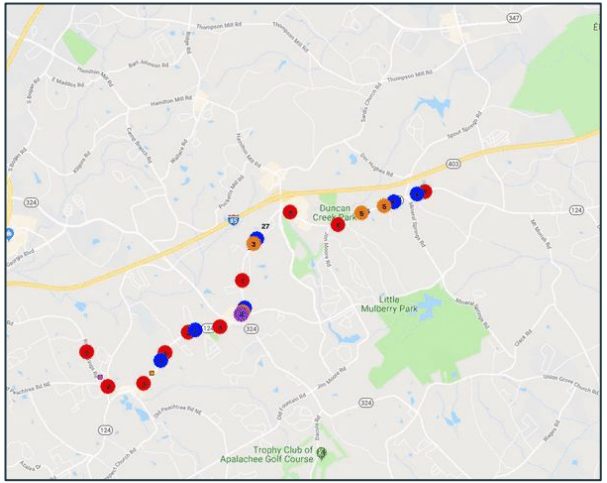
- Automates analysis of road conditions
- Eliminates variation between operators
- Mobile app so no implementation cost



2 principle axes

Visualization





Results

What methods did we use to evaluate our algorithm and what were the results?

- Davies Bouldin score: 1.6004179081836094
  - Ratio b/w intra-cluster distance and inter-cluster distance/ avg similarity b/w each cluster
- Silhouette score: 0.37762391442145105
  - Measure of how similar object is (avg distance) to its own cluster compared to other clusters
- DBSCAN for:
  - Verifying that an anomaly detected is actually an anomaly
  - Differentiating between anomalies on opposite sides of the road

How do these methods compare to others?

- Was this an efficient measure of road conditions?