

CHARACTERIZING ROAD CONDITIONS VIA SMART MOBILE CROWD SOURCING

A Thesis
IN
Computer Science

Presented to the Faculty of the University
of Missouri-Kansas City in partial fulfillment of
the requirements for the degree

MASTER OF SCIENCE

by
DUSTIN SHORTER

B.S., Kansas State University

Kansas City, Missouri

2019

© 2019

DUSTIN SHORTER

ALL RIGHTS RESERVED

CHARACTERIZING ROAD CONDITIONS VIA SMART MOBILE CROWD SOURCING

Dustin Shorter, Candidate for the Master of Science Degree

University of Missouri-Kansas City, 2019

ABSTRACT

Bad road conditions can cause vehicle damage and create hazardous driving conditions. Current reporting methods for bad road conditions is a burden for the reporter, so much so that these conditions might not be reported. With current smartphone technology this reporting method can be greatly improved. Smartphones come with an accelerometer and GPS positioning on-board. Utilizing these sensors, once the application is installed and run, a smartphone user can gather bad road conditions automatically while driving. Then the user can upload the data to the server. These road conditions are then classified and optionally displayed on a map while the user is driving. The potential for finding bad road conditions is greatly increases with crowd sourcing. When the road condition data is analyzed an algorithm is used to determine the size of the bad road condition and how many samples are needed. Even though the smartphone might not have the sampling rate of a dedicated accelerometer it's potential for gather large amounts of data using crowd sourcing and ease of use outweighs this deficiency.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Computing and Engineering, have examined a thesis titled “Characterizing Road Conditions via Smart Mobile Crowd Sourcing,” presented by Dustin Shorter, candidate for the Master of Science degree, and certify that in their opinion it is worth of acceptance.

Supervisory Committee

Sejun Song, Ph.D., Committee Chair
Department of Computer Science & Electrical Engineering

Baek-Young Choi, Ph.D.
Department of Computer Science & Electrical Engineering

Sanghak Sung, Ph.D.
Department of Computer Science & Electrical Engineering

CONTENTS

| | |
|--|-----|
| ABSTRACT..... | iii |
| ILLUSTRATIONS | vii |
| ACKNOWLEDGMENTS | ix |
| CHAPTER 1. INTRODUCTION | 1 |
| CHAPTER 2. BACKGROUND | 3 |
| CHAPTER 3. IMPLEMENTATION | 4 |
| 3.1 Application..... | 4 |
| 3.1.1 GPS Sensor..... | 4 |
| 3.1.2 Accelerometer Sensor..... | 4 |
| 3.1.3 Save and Upload Data | 5 |
| 3.1.4 Send Current Location to Server | 7 |
| 3.1.5 Display Road Condition Data on Map | 7 |
| 3.2 Server | 7 |
| 3.2.1 Download Data from Cloud | 8 |
| 3.2.2 Create Zones..... | 8 |
| 3.2.3 Classify Road Conditions..... | 12 |
| 3.2.4 Good Road Condition..... | 13 |
| 3.2.5 Poor Road Condition | 14 |
| 3.2.6 Bad Road Condition | 16 |
| 3.2.7 Summary of Good, Poor, and Bad Road Conditions..... | 16 |
| 3.2.8 Receive GPS Location from Smartphone and Send Data | 16 |
| CHAPTER 4. EXPERIMENTATION | 18 |

| | |
|---|----|
| 4.1 Setup | 18 |
| 4.2 Dependencies | 20 |
| 4.2.1 X, Y versus Z Axis Dependency | 21 |
| 4.2.2 Speed Dependency | 24 |
| 4.2.3 Elevation Change Dependency..... | 27 |
| 4.3 Classification of Road Conditions | 28 |
| 4.3.1 Count and Magnitude of Road Conditions..... | 30 |
| 4.4 Road Conditions..... | 33 |
| 4.4.1 Intentional Road Condition Dip | 34 |
| 4.4.2 Road Condition Enter Bridge | 35 |
| 4.4.3 Road Condition Exit Bridge | 36 |
| 4.4.4 Road Condition Fixed Potholes..... | 37 |
| 4.4.5 Road Condition 3 Potholes | 38 |
| 4.4.6 Road Condition Pothole and Low Spot | 39 |
| 4.4.7 Road Condition Bad Area Lone Elm Road | 40 |
| 4.4.8 Road Condition Raised Crosswalk | 41 |
| 4.4.9 Intentional Road Condition Speed Bump | 41 |
| CHAPTER 5. CONCLUSION | 43 |
| REFERENCE LIST | 44 |
| VITA..... | 46 |

ILLUSTRATIONS

| | |
|--|----|
| Figure 1: Architecture | 2 |
| Figure 2: Sample GPS and Accelerometer data | 6 |
| Figure 3: Example of filenames | 6 |
| Figure 4: Zone size..... | 9 |
| Figure 5: Create Zone Diagram..... | 10 |
| Figure 6: Good road condition section..... | 14 |
| Figure 7: Poor road condition section..... | 15 |
| Figure 8: Road condition data on map while driving | 17 |
| Figure 9: Smartphone in open center console..... | 19 |
| Figure 10: RoadConditionApp..... | 20 |
| Figure 11: Axes x, y, z Comparison | 22 |
| Figure 12: x vs. z and y vs. z | 23 |
| Figure 13: ASD changes with vehicle speed | 24 |
| Figure 14: Speed and z average..... | 25 |
| Figure 15: Speed vs. z average..... | 26 |
| Figure 16: Speed variation test..... | 27 |
| Figure 17: Elevation and z axis accelerometer data | 28 |
| Figure 18: Road Classification Figure | 29 |
| Figure 19: Route For Count and Magnitude Consideration | 31 |
| Figure 20: Good, Poor, and Bad Road Conditions..... | 32 |
| Figure 21: Intentional Road Condition Dip..... | 34 |
| Figure 22: Road Condition Enter Bridge..... | 35 |

| | |
|---|----|
| Figure 23: Road Condition Exit Bridge | 36 |
| Figure 24: Road Condition Fixed Potholes | 37 |
| Figure 25: Road Condition 3 Potholes | 38 |
| Figure 26: Pothole and Low Spot | 39 |
| Figure 27: Road Condition Bad Area Lone Elm Road | 40 |
| Figure 28: Intentional Road Condition Raised Crosswalk..... | 41 |
| Figure 29: Intentional Road Condition Speed Bump | 41 |

ACKNOWLEDGMENTS

First and foremost, I would like to thank my academic advisor Dr. Sejun Song for allowing me to work with him on my master's thesis. He has provided guidance and support on my thesis and in classes. Without his support and guidance, I doubt I would have been able to complete the thesis on time. I would also like thank Dr. Baek-Young Choi and Dr. Sanghak Sung for their feedback and time.

Secondly, I would like to thank the University of Missouri-Kansas City for providing me the opportunity to pursue and master's degree and to grow as an engineer.

Finally, I would like to thank my wife, daughter and parents for their continued support of my pursuit of a master's degree. They are the support that allowed me to pursue my passion for science and engineering.

CHAPTER 1. INTRODUCTION

Potholes and rough road conditions are a large problem that can cause vehicle damage or hazardous driving conditions. The current reporting standard for a bad road condition or pothole is to report via phone call to local area department of transportation or online web browser or email. In these cases, additional information must also be provided, such as: pothole location, county, highway number, nearest local cross street or interchange, direction (i.e. northbound), lanes impacted (i.e. right lane, center lane, left lane, shoulder), and per reporting phone number (for after-hours calls only, in-case of questions) [1]. This standard of reporting places a great amount of burden upon the reporter. Also, there is not any information about the road condition's size or shape seen in [2].

The thesis topic I will be addressing will use a smartphone's capabilities to capture and display road conditions using onboard GPS and accelerometer sensors with mobile data. The GPS and accelerometer data logged and then sent to the cloud. The server processes the data from the cloud and communicates with the user's smartphone device and transmits road condition data based on the current location of the device using mobile data. When the device receives the data, it's communicated to the user optionally via a map using markers of the road conditions as seen in Figure 1. This data may also be used to report pothole and road conditions to local public works departments.

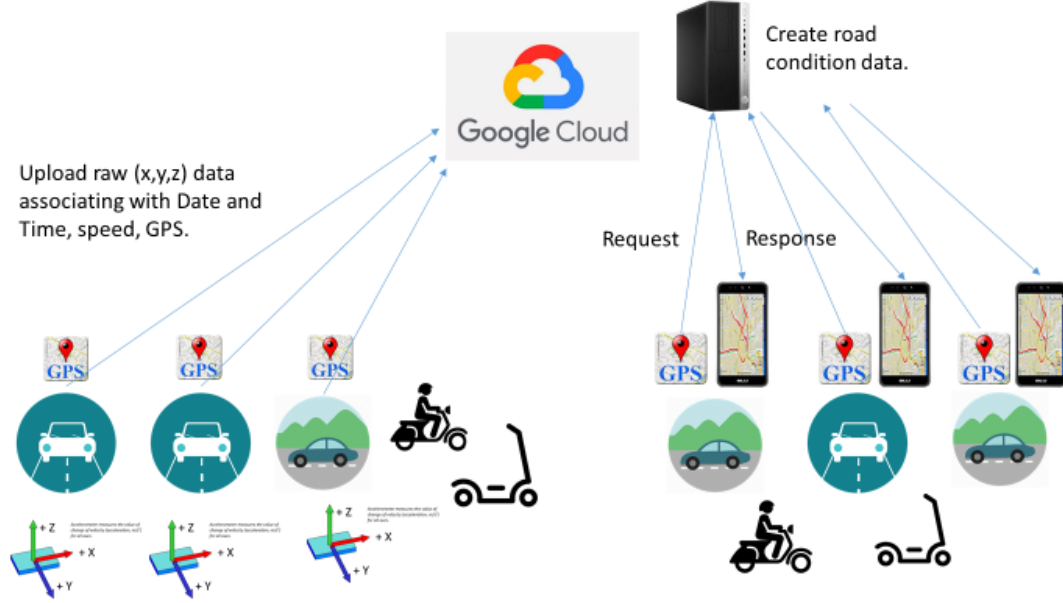


Figure 1: Architecture

Throughout the paper I will mention other works like this topic regarding detecting potholes with smartphone and accelerometers in the Background section. In the Implementation section I will mention about the application and the server in more details on how they work together and their main functions. In the Experimentation section I will briefly mention about any dependencies that we need to consider when capturing the accelerometer data and other dependencies that we can ignore. Also, in this section I will display multiple road conditions and the associated data with an image of the road condition. I will then classify the road condition using a formula to decide the impact and size of the road condition.

CHAPTER 2. BACKGROUND

Using a smartphone to detect road conditions is a relatively new concept but is covered by multiple papers as seen in [3], [4], [5]. In [3] they use the Android smartphone as an online petition system to gather data from users regarding the road conditions and then use an algorithm to process the data and with image processing to determine the damage. In [4] they can detect potholes in real time using the Android accelerometers. They use and evaluate data with certain processing algorithms and can get a positive rate as high as 90 percent based on actual data from the actual use case data. In [5] they use an Android smartphones and machine learning to evaluate the road conditions in real-time. They use the smartphone to do automatic evaluations and provide the data through a web browser. With their system they can obtain a 90 percent performance using their designated road classifications.

Crowd sourcing is also present a paper as seen in [3] as stated above they use crowd sourcing to gather their data from users. They use the crowd sourcing for data collection, selection and assessment regarding the road conditions that are processed.

Road classification is mentioned in [5], [6]. In [5] they use four categories to classify their road conditions and count the amount of times the road condition occurs. In [6] they use machine learning to develop and impact detection strategy that is in accordance to ISO 8608.

CHAPTER 3. IMPLEMENTATION

I have created an Android application to capture GPS and accelerometer data. When capturing data, I used multiple Android devices: Google Pixel 2 XL, Samsung Galaxy S8, Samsung Galaxy Note 10.1, and Motorola Moto G6. When analyzing the data that will be shown in this paper, I used the Google Pixel 2 XL to reduce variables.

3.1 Application

The application's main functions are capture data, sending the data to cloud storage, sending current location to server, and displaying road conditions to users while driving after receiving the classified data from the server. Capturing the data includes getting the current position of the smartphone, capturing accelerometer data for x, y, z axis, and getting the current date and time. This is done while the user is driving in their car. Currently, the GPS is logging every one second for the Google Pixel 2 XL and the accelerometer data is captured approximate 50 samples per GPS data point.

3.1.1 GPS Sensor

The GPS sensor was the sensor currently on the smartphone. The accuracy of the GPS location is generally about 16 feet for a smartphone according to [7]. When capturing GPS data and plotting on a map you will notice that sometime the GPS coordinate is off the road. This shows that sometimes it's imprecise. For example, if the user is driving at 30 mph the vehicle drives about 44 ft. per second, so having an accuracy of about 16 feet is acceptable.

3.1.2 Accelerometer Sensor

The accelerometer I chose to use is TYPE_LINEAR_ACCELERATION [8]. This type of accelerometer sensor removes the force of gravity. This was chosen to avoid the need

remove the force of gravity from gathered accelerometer data. The sensor rate that currently being used is `SENSOR_DELAY_FASTEST` [9] as this captures data as fast as possible for the current smartphone device. I originally was using `SENSOR_STATUS_ACCURACY_HIGH` [9] but this was only generating 5 accelerometer samples per 1 second. I originally thought that accuracy on the output of the sensor was the most important, but soon realized that more data samples was better.

If we were to use a dedicated accelerometer, we could obtain a much higher sampling depending on the accelerometer. Some accelerometers can sample at a rate of approximately 400 to 500 samples per second. This sampling rate is approximately 10 times that of the smartphone I used. The units of measurement for the accelerometer are in $\frac{m}{s^2}$ which is a unit of measurement for acceleration.

3.1.3 Save and Upload Data

When the data is captured it is saved to a CSV file on the smartphone. While the data is being captured the date and time is also logged for both GPS and accelerometer data. Figure 2 is an example of the data output.

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|------------------------------|-----|--------|-----|--------|-------|--------|--------|---|--------|---|--------|---|
| 1 | Tue Mar 19 07:38:53 CDT 2019 | Lat | 38.878 | Lon | -94.86 | Speed | 12.01 | | | | | | |
| 2 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.371 | y | -0.511 | z | 0.0804 | |
| 3 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.239 | y | -0.546 | z | -0.075 | |
| 4 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.403 | y | 0.0583 | z | -0.235 | |
| 5 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.338 | y | 0.7149 | z | -0.382 | |
| 6 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.455 | y | 0.673 | z | -0.163 | |
| 7 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.305 | y | 0.8374 | z | -0.354 | |
| 8 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.398 | y | 1.0897 | z | -0.233 | |
| 9 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.339 | y | 0.9614 | z | 0.2032 | |
| 10 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.367 | y | 0.4164 | z | -0.131 | |
| 11 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.282 | y | 0.0666 | z | 0.1773 | |
| 12 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.224 | y | -0.446 | z | 0.0545 | |
| 13 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.169 | y | -0.922 | z | 0.1072 | |
| 14 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.126 | y | -1.154 | z | -0.135 | |
| 15 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.156 | y | -1.155 | z | 0.0019 | |
| 16 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.115 | y | -1.072 | z | -0.133 | |
| 17 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.121 | y | -0.845 | z | -0.262 | |
| 18 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.09 | y | -0.644 | z | -0.288 | |
| 19 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.09 | y | -0.375 | z | -0.428 | |
| 20 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.129 | y | -0.221 | z | -0.216 | |
| 21 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.176 | y | -0.194 | z | -0.242 | |
| 22 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.315 | y | -0.037 | z | -0.236 | |
| 23 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.303 | y | 0.0328 | z | -0.232 | |
| 24 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.32 | y | -0.202 | z | -0.295 | |
| 25 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.419 | y | -0.382 | z | -0.233 | |
| 26 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.239 | y | -0.223 | z | -0.128 | |
| 27 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.476 | y | -0.019 | z | -0.031 | |
| 28 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.441 | y | -0.08 | z | -0.162 | |
| 29 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.354 | y | -0.238 | z | -0.026 | |
| 30 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.427 | y | -0.294 | z | -0.127 | |
| 31 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.383 | y | -0.375 | z | 0.0427 | |
| 32 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.456 | y | -0.527 | z | 0.1165 | |
| 33 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.426 | y | -0.817 | z | 0.2553 | |
| 34 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.345 | y | -0.513 | z | 0.3017 | |
| 35 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.251 | y | -0.522 | z | -0.005 | |
| 36 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.219 | y | -0.611 | z | -0.063 | |
| 37 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.248 | y | -0.34 | z | -0.086 | |
| 38 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.142 | y | -0.759 | z | -0.02 | |
| 39 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.23 | y | -0.391 | z | -0.381 | |
| 40 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.074 | y | -1.033 | z | -0.005 | |
| 41 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.27 | y | -1.105 | z | -0.395 | |
| 42 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.155 | y | -0.983 | z | -0.244 | |
| 43 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.228 | y | -0.957 | z | -0.185 | |
| 44 | Tue Mar 19 07:38:53 CDT 2019 | | | | | | x | -0.277 | y | -1.191 | z | -0.353 | |
| 45 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.251 | y | -1.074 | z | -0.232 | |
| 46 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.277 | y | -0.974 | z | -0.127 | |
| 47 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.13 | y | -1.132 | z | -0.504 | |
| 48 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.126 | y | -1.237 | z | -0.243 | |
| 49 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.443 | y | -1.183 | z | -0.041 | |
| 50 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.314 | y | -1.195 | z | -0.308 | |
| 51 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.281 | y | -1.268 | z | -0.119 | |
| 52 | Tue Mar 19 07:39:00 CDT 2019 | | | | | | x | -0.22 | y | -1.316 | z | -0.021 | |
| 53 | Tue Mar 19 07:39:01 CDT 2019 | Lat | 38.878 | Lon | -94.86 | Speed | 15.846 | | | | | | |
| 54 | Tue Mar 19 07:39:01 CDT 2019 | | | | | | x | -0.285 | y | -0.993 | z | -0.234 | |
| 55 | Tue Mar 19 07:39:01 CDT 2019 | | | | | | x | -0.254 | y | -0.814 | z | 0.2844 | |

Figure 2: Sample GPS and Accelerometer data

When the application is closed it creates another file. The filenames contain the date they are captured along with a Universal Unique ID (UUID). Example filename shown in Figure 3.






 2019-03-19_17cb4c3f-94b9-4d68-a7fd-7f3737cbde5c_log.csv
 2019-03-19_ff3cac53-b6ea-4c30-9995-7f56ed98e320_log.csv
 2019-03-20_b4ab7c9a-f8be-430b-93c7-37573376b3b0_log.csv
 2019-03-21_42d84cb6-0ef3-4d32-bdbb-d5f8a77bf4a6_log.csv
 2019-03-22_61be0da1-aa11-473a-aa1a-0d7e1d7f2cca_log.csv

Figure 3: Example of filenames

After the files are stored on the smartphone the user can upload the data by pressing the upload button on the screen. The file will only upload when the smartphone is within Wi-Fi range. I chose to include the Wi-Fi check to avoid accidental upload over mobile data. This saves the user from inadvertently using their mobile data for helping the research. User manual upload

was implemented because I want the user to know they are uploading freely, and there was no readily available method to pause uploading if the signal is lost. The file is then uploaded to a Google Cloud Storage bucket. After the file is successfully uploaded it is deleted off the smartphone.

3.1.4 Send Current Location to Server

The third function of the application is to send its current GPS location to the server when the application detects a change in location. This GPS location is used by the server to determine what road condition data should be sent to the smartphone. This will be further discussed in the server section.

3.1.5 Display Road Condition Data on Map

The final function of the application is to display the data sent from the server to be displayed on the map. In the later sections I will discuss the server in more detail but for now I will explain what the application does after it receives the data. When the application receives the data, it contains information regarding the GPS location of the road condition and the marker for the road condition. In later sections I explain the classification for the road conditions further. These are the rankings given for road conditions as displayed on the map. The road condition information is received in the form either orange, or red and contains a value. The orange represents poor and red bad road conditions. I chose not to use any marker for good road conditions because if it's not marked as poor (orange) or bad (red), I can be assumed it's good. This also aids user ability to easily see marked road conditions.

3.2 Server

The server's main functions are to download the captured data from the cloud storage, creating zones with the data, receiving current location from the smartphone, and sending the

classified zone data to the smartphone to be displayed. These functions allow the smartphone to focus on getting the data and displaying the data.

3.2.1 Download Data from Cloud

The server downloads the captured data from the cloud storage using the Google Cloud API. This API allows the server to get the files and then store them locally so they can be processed.

3.2.2 Create Zones

After the files are downloaded locally on the server they are analyzed and then sorted into zones. These zones are used to consolidate data that is representing the same part of the road. Since the overall idea is to use crowd sourcing to capture data for the roads, I want to combine this data. When creating zones, the two things we need to determine are the size of the zone and how much data each zone should hold. The idea is to keep the most recent data available and not use older data, so the road data is representing the current road condition.

The size of the zone we should is width a lane on the road. Then to determine what other data points should be place within the zone we can use the radius of the lane. Using the radius, we will calculate if the data point is less than the radius of the zone center. If the data point is less than it will be added to the zone otherwise it will create a new zone or determine if it belongs in another zone. Examples of the zone size is shown in Figure 4.

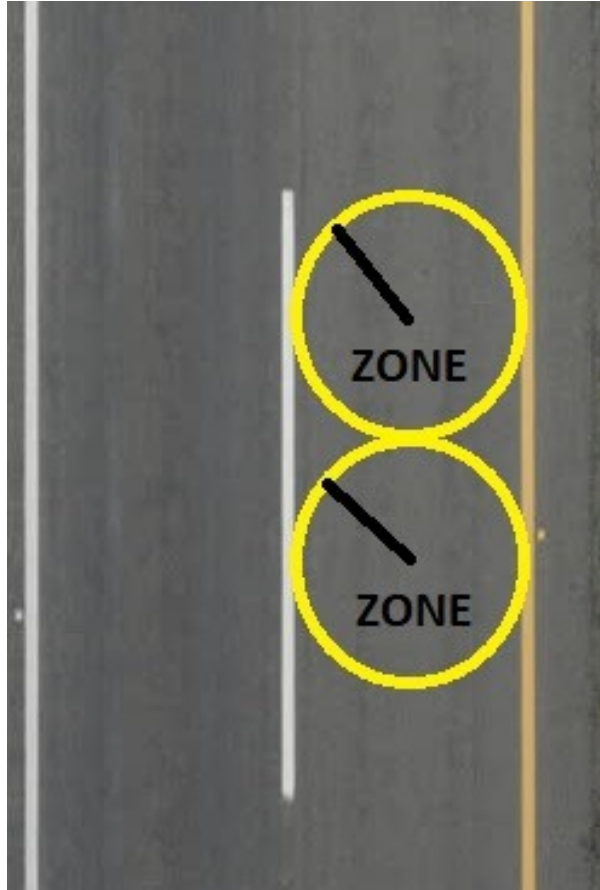


Figure 4: Zone size

To determine the amount of data to hold in the zone we will use Figure 5 to determine an approximate number of samples we should use to calculate the road condition status.

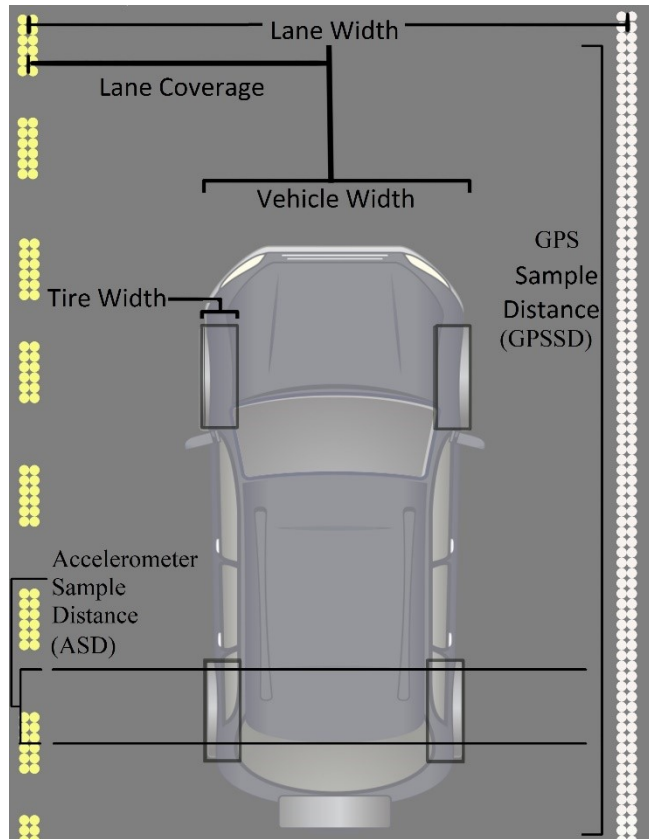


Figure 5: Create Zone Diagram

To determine how much data, we calculate the area of the zone. In Figure 5 there are multiple variables that are shown. The first thing to note is we do not need to use the full lane width for calculating the zone area. This is because the vehicle is a static object and when it moves left or right the entire object moves together. This means that when left wheels are going over a certain portion of the road the right wheels are going over a different portion on the right side at the same time. Due to the vehicle being a static object that moves together we need to only worry about the half of the lane width which we will call lane coverage (LC). The next variable we need to consider is the GPS sample distance (GPSSD). This is how far the vehicle will travel between each GPS interval that is captured. To determine the distance the vehicle will travel we need to know how fast the vehicle is going and for what time interval. Now that we

know the area of the zone, we need a way to determine how many samples we can get in that zone.

To determine how many samples in that zone we need, we need to know how many tire width's per ASD are in the zone area. To determine this, we use the tire width and the ASD. The ASD is calculated by taking the GPSSD and dividing by amount of accelerometer data per GPS iteration as seen in Figure 5. Also, since we will have two tires in the area, we need to account for both. Now that we know the necessary information to determine how many samples we want in the zone; we can create a formula.

$$Samples = \frac{GPSSD * LC}{2 * ASD * Tire Width}$$

The formula above tells us basically how many ASD by tire width samples are in the zone area. But if we want more samples to improve our estimate of the zone, we need a way to determine how many samples we need to use that new ASD value. The new formula is the same as the previous except we just enter the new ASD as a value in inches. For example, we want our new ASD to be 1 inch, so instead of calculating ASD we just enter the value we want.

$$Samples = \frac{GPSSD * LC}{2 * NewASD * Tire Width}$$

Now we have two formulas, but they are not easy to use because they require many variables that are not easily attainable, and if we want to use crowd sourcing this might be a burden to calculate this for every zone to determine the how many samples we need. So, we simplify these formulas into something that is more accessible. To do this we need to make some assumptions. The first assumptions we will make are regarding the tire width. Regarding the tire width I looked up some information on common tire widths at [10] and took the average of the smallest and largest tires that are listed. The value I came up with is 9.84 inches. The

next variable to simplify is the GPSSD. The GPSSD multiplies the speed of the vehicle by the GPS sample rate of the device. The current speed of the vehicle is in mph but we would like the result to be measured in inches per second so we will multiply the speed by 17.6. Also, the GPS sample rate of the device is set to 1 second in the application.

$$GPSSD = Speed(mph) * 17.6 \left(\frac{in}{sec} \right) * GPS Sample Time(sec)$$

To compute the current ASD we just divide the GPSSD by the accelerometer sample rate. Currently, the Google Pixel 2 XL generates an average of 50 samples per GPS sample rate. Then we just divide GPSSD by 50. The final variable we need to determine is the lane coverage LC. This is half the lane width. When determining lane width, I will be referencing [11]. The website shows that 12 feet seems to be common in all the types of roads and is the width of the Freeway. Now that we have provided some values for the variables, we can create some new formulas.

$$Samples Current ASD = 183$$

$$Samples for New ASD = \frac{64.4 * speed(mph)}{New ASD}$$

The formulas above represent using the current ASD and new ASD respectively. The first formula works out to a constant value. This is because GPSSD and current ASD are related and ASD is a fraction of GPSSD so when the calculations are finished the speed gets cancelled out and the value that is obtained is 183.

3.2.3 Classify Road Conditions

Determining road conditions that should be captured and reported can be somewhat difficult due to what is considered a bad road condition can be subjective. We can make an

educated guess using some data and video footage to determine what a good, poor, and bad road is on average and provide a tolerance to allow some tolerance.

3.2.4 Good Road Condition

To find what a good road is we gather the z axis data and then calculate the average, standard deviation, and add them together to get a baseline for a good road condition and set a tolerance. I then calculated the average and standard deviation for five different day's and times. Figure 6 shows the good road section of the route. To calculate the z value for the GPS data point, I take the absolute value of the z axis values for the GPS data point and then calculate the average.

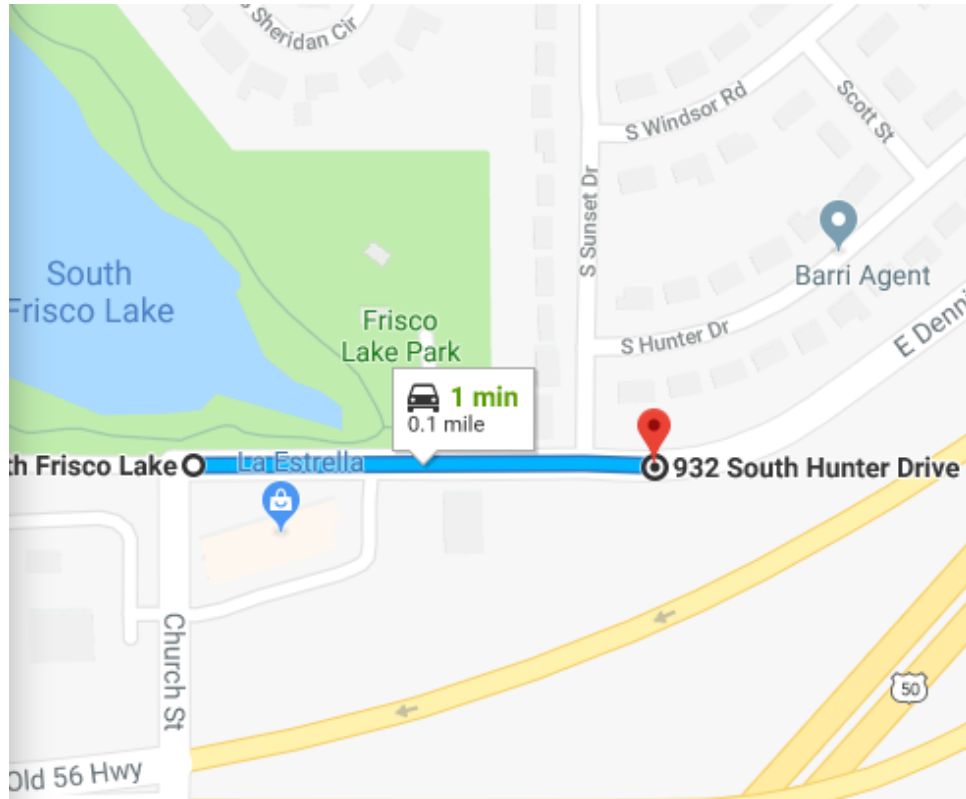


Figure 6: Good road condition section

The average of the road section was 0.31 and the standard deviation was 0.07. Add these two values and it equals 0.38. To keep the numbers simple, I chose to round up to 0.40. This value can be used to filter out good road conditions.

3.2.5 Poor Road Condition

Using this value as a filter, for the good road conditions, is a good start but it would be nice if we could categorize unwanted road conditions into two categories. The first category as poor road conditions and the next can be bad road conditions. If the lower bound for the poor road conditions was greater than 0.40 then there needs to be an upper bound. To determine the poor road condition, I will repeat the process for the good road condition. Figure 7 shows the road for the poor condition section.

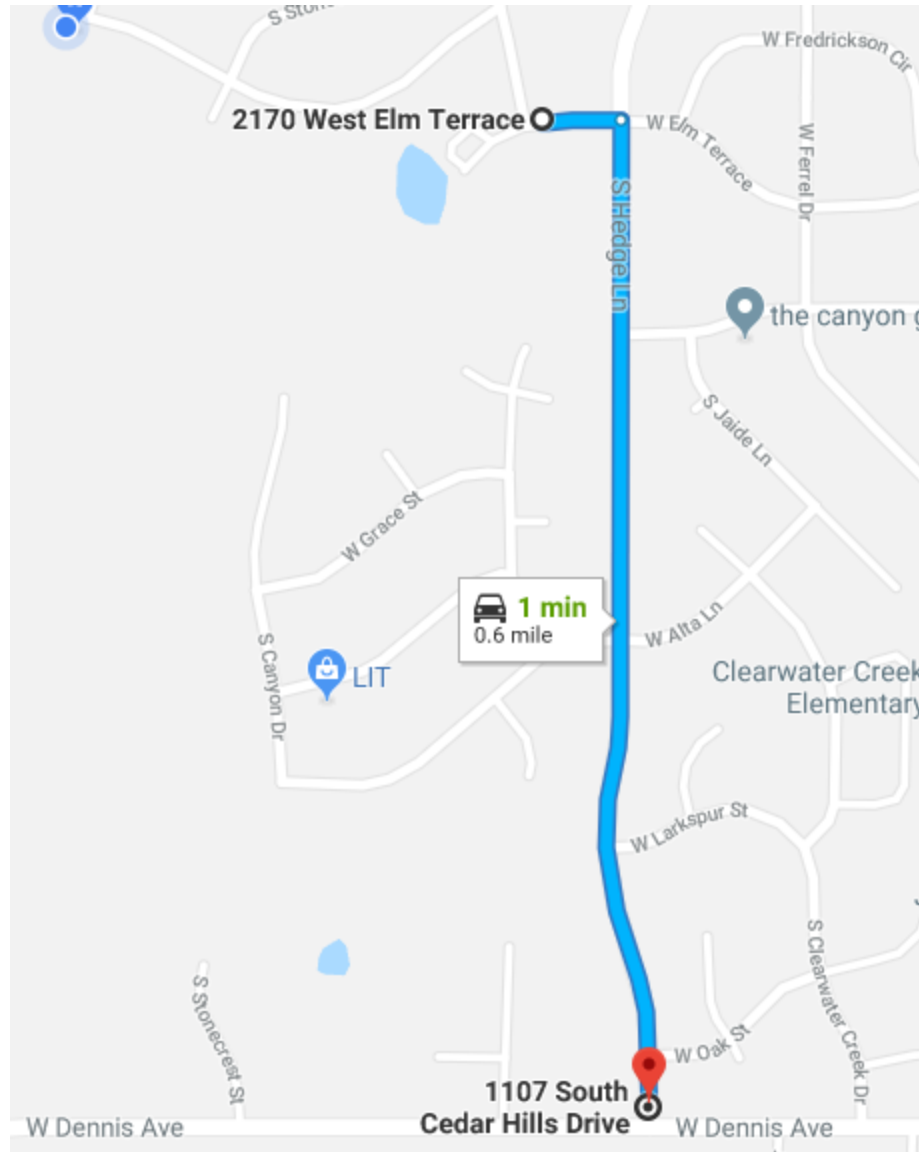


Figure 7: Poor road condition section

This section of the road has multiple cracks and bumps. It seems like a decent choice to determine the upper bound condition for the poor road condition. To determine the poor road condition, I will calculate the average of the poor sections on the route. The average for the poor sections of the route is 0.67 and the standard deviation is 0.11. Sum these two together and the value is 0.78. Round this up and the upper bound for the poor road condition is 0.80.

3.2.6 Bad Road Condition

The final road condition would be a bad road condition. Since this value captures the values above the poor road condition we just need to check if the road condition is greater than the upper bound of 0.80.

3.2.7 Summary of Good, Poor, and Bad Road Conditions

Road conditions are classified as good, poor or bad. Good road conditions are GPS z axis values less than or equal to 0.40. The poor road conditions are z axis values greater than 0.40 and less than or equal to 0.80. The bad road conditions are z axis values greater than 0.80.

3.2.8 Receive GPS Location from Smartphone and Send Data

The last two functions of the server are to receive the GPS location from the smartphone and send data to the smartphone. The current setup for the communication between the server and the smartphone is mobile data. This was chosen because many new smartphones come with a mobile data plan that uses mobile data and many plans are unlimited. The smartphone creates a socket connection with the server. Then sends the current GPS location and then waits for the server to reply with the zone data that is within a specified range determined by the server. This range can be set to the server administrator or user's preference. I currently use 500 feet to keep the data sent to a minimum. Finally, closes the connection and repeats. Figure 8 shows the data being sent to the device from the server while driving around my neighborhood.

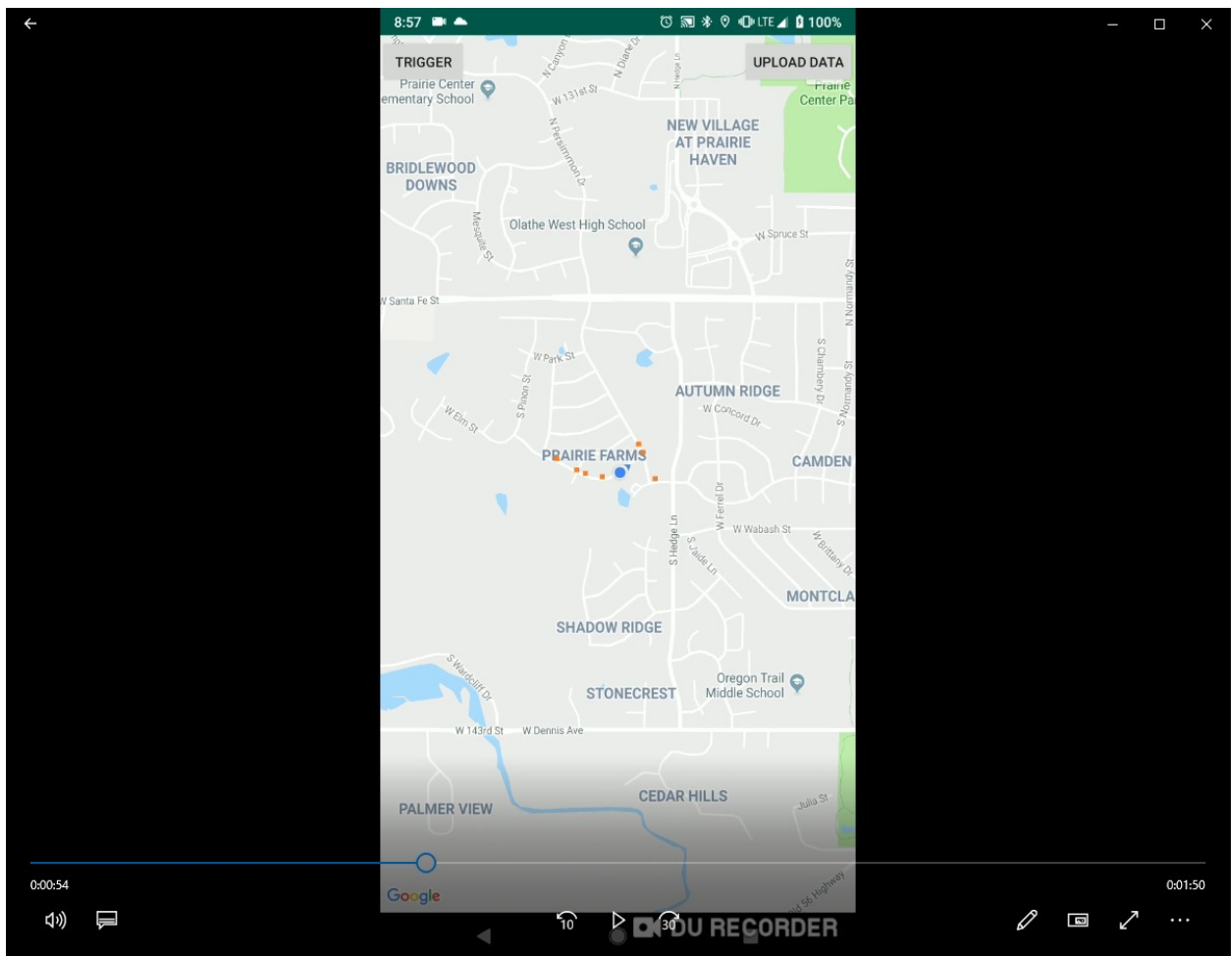


Figure 8: Road condition data on map while driving

CHAPTER 4. EXPERIMENTATION

For this chapter I will discuss my setup for the experimentation section, any dependencies that need to be considered and ignored, and show some road condition examples with the data and images of the road conditions. I will also calculate the bad road conditions affected area of the zone.

4.1 Setup

For the experimentation portion of the paper I chose to minimize the number of variables that could affect the results. The variables that could be controlled were smartphone used, smartphone placement, smartphone orientation, route driven, vehicle, dash camera, and the application used to collect the data. Some of the variables which could not be controlled were weather, traffic on the route, changes in road conditions, and syncing time of route. The smartphone used is a Google Pixel 2 XL with Android Version 9. The smartphone placement seen in Figure 9 and orientation are in open center console, screen up, forward facing. The vehicle's make and model are 2016 Kia Soul. The dash camera is the Garmin Speak Plus with Amazon Alexa.

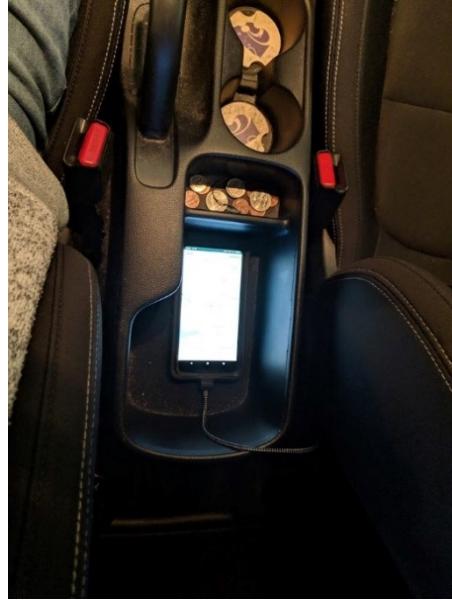


Figure 9: Smartphone in open center console

I chose to place the smartphone in the open center console of my Kia Soul because I could see the application run while driving and to prevent the smartphone from sliding due to the smartphone case and the console being rubber.

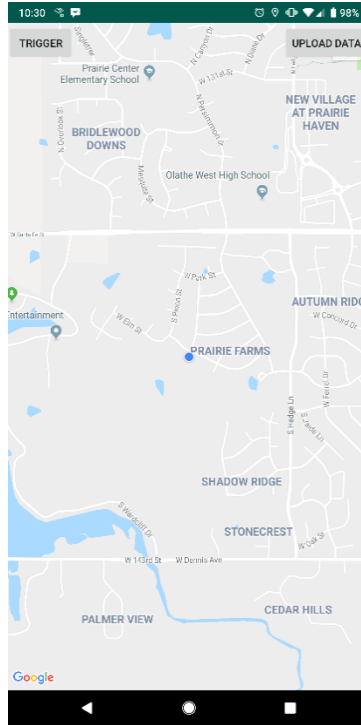


Figure 10: RoadConditionApp

The application shows the current location of the user and allows the user to upload the data when connected to Wi-Fi. The application was designed in this fashion to avoid accidental use of mobile data to upload the data as I do not have an unlimited data plan. The user can also place a trigger (flag) in the data if they want to note something that needs further review later.

4.2 Dependencies

If we want to use the z axis only for detecting potholes we need to check if there are other dependencies that might affect our results. This will not be an exhaustive test but more of a brief analysis of these dependencies. The dependencies we will be analyzing are the x, y versus z axis of the accelerometer. The speed of the vehicle which include if the ASD is affected due to speed being included if we want a new ASD, if driving speed changes

accelerometer data, and driving over the same road condition varying the speed affects the accelerometer data.

4.2.1 X, Y versus Z Axis Dependency

After collecting the GPS and accelerometer data I will determine if there are any dependencies between x and y versus z. The first part I will show x, y, and z axis' together in one figure. Then I will chart x versus z and y versus z in a scatter plot and calculate the correlation coefficient for each of the figures.

Below in Figure 11 the x, y and z axis are shown together.

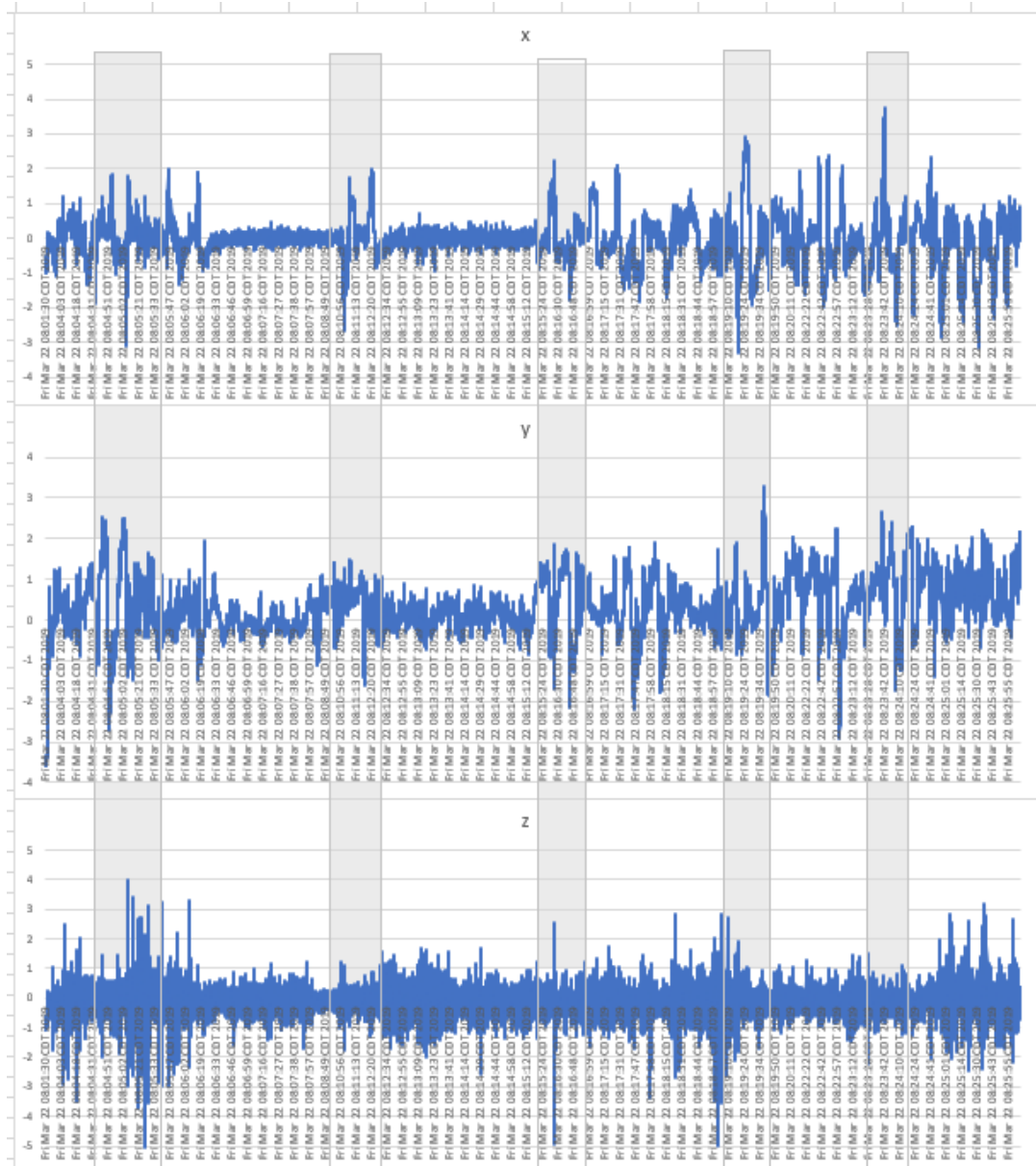


Figure 11: Axes x, y, z Comparison

In comparing the three graphs (z axis vs. the x and y axis) you will notice that there are inconsistencies between each axis. In the first grey slicing, all three axes increased and decreased but, the z axis spiked after the x and y axis were stabilizing. In the second grey slicing, the x and y axis spike but, the z axis seems unaffected. In the third grey slicing, the x axis increased and then decreases; the y axis increased, then decreased, then repeat; the z axis

seemed stable and then rapidly increased and decreased in a short time period. The fourth grey slicing, the x axis spiked in the middle, where the y axis had a small increase in the beginning and a large spike at the end, with the z axis increased and decreased and then stabilized. In the fifth grey slicing, the x axis had a very large spike in the middle then a couple smaller negative spikes, the y axis increased then spike up and down, with the z axis being unaffected by what x and y are reported. Based on these slices of data and graph comparisons it appears that x, y and z axis are relatively independent of each other.

The next thing we will do is chart the x versus z and y versus z and calculate the correlation coefficient. In Figure 12 for both charts it doesn't appear there are any relationships between the x and y axis versus z axis. When we calculate the correlation coefficient for x versus z, we get 0.05 and for y versus z, we get 0.08. These values are low and close to 0 so we can assume there is a small correlation between the x and y axis versus the z axis.

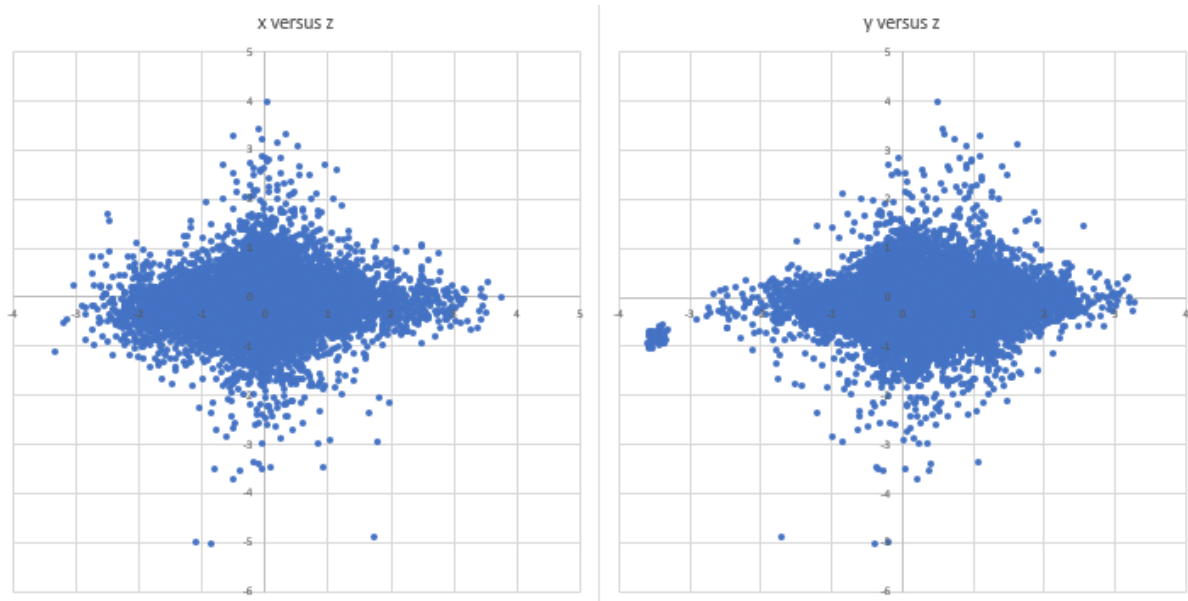


Figure 12: x vs. z and y vs. z

Using these two methods we can conclude that the x and y axis do not affect the z axis, so we do not need to use these two axes when determine road conditions. We can focus on using the z axis alone.

4.2.2 Speed Dependency

The next dependency to look at is speed. When determine if speed is a dependency, we need to look at three scenarios. First, in the Create Zones section we see one formula seems to depend on the speed of the vehicle. Second, does driving faster or slower cause the z axis to generate high or lower accelerometer values. Third, does driving over the same road condition with varied speed affect the accelerometer data.

To determine if speed affects GPSSD or ASD we look at the two portions of the formulas that use speed. Using the formula to calculate ASD with a sample rate of 1 second we can see that the ASD increases when the speed increases in Figure 13. According to Figure 13 the speed does change the ASD.

$$ASD = \frac{GPSSD}{Accel\ Samples} = \frac{17.6 * Speed(mph) * Sample\ Rate}{50}$$

Speed 10 MPH

» 3.5 inches

Speed 30 MPH

» 10.6 inches

Speed 50 MPH

» 17.6 inches

Speed 70 MPH

» 24.6 inches

Figure 13: ASD changes with vehicle speed

To determine if driving faster or slower can cause the z axis to generate high or lower accelerometer values I will use two methods. Method 1 is shown in Figure 14, which charts

speed vs z average data. Method 2 uses a scatter plot shown in Figure 15 with speed versus z average and then I will calculate the correlation coefficient between speed and z average.

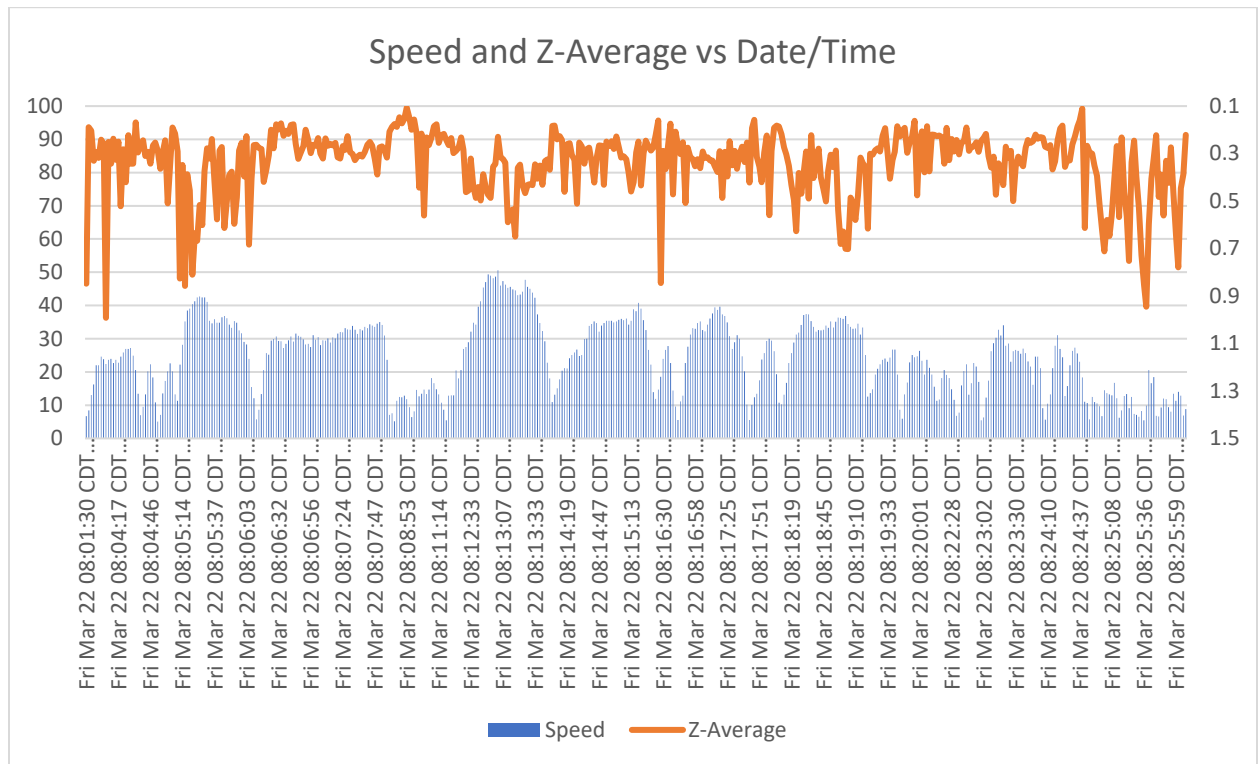


Figure 14: Speed and z average

This chart aligns the z axis average data for the GPS coordinate data and then charts the speed of the vehicle vs Date and Time. Looking at the chart, there are times when the road condition spikes when the speed increases, as shown at 08:05:14. However, there are other times when the road condition data is spiking, and the speed is low at 08:04:17. At the end of the route the speed is slower, while the accelerometer data is spiking often. This seems that the road condition data doesn't depend on the speed of the vehicle.

The next test will use a scatter plot and the correlation coefficient to determine a relationship between speed and z average. In Figure 15 it appears that there might be a very small linear correlation. So, when we calculate the correlation coefficient, we get 0.17. This

value is low but there seems to be a small relationship between speed and z average. Since this is low, we can safely not consider it. But more testing might be needed.

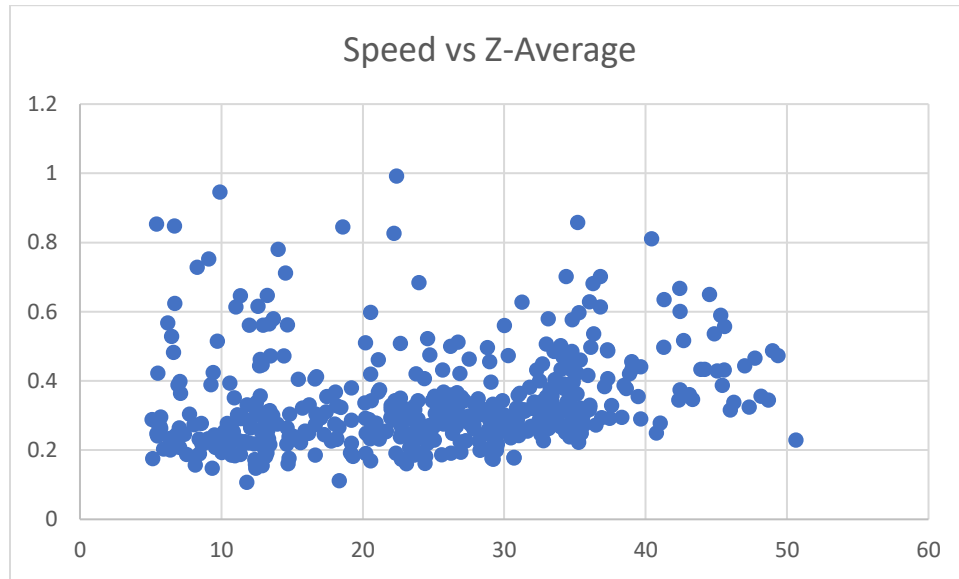


Figure 15: Speed vs. z average

Finally, does driving over the same road condition with varied speed affect the accelerometer data. In Figure 16 the data looks different for each time we drove over the road condition but the maximum value for each of the different speeds never exceeded much over the absolute value of 4. This seems to indicate the small variation of speed that was tested, the accelerometer data didn't increase or decrease based on the speed. But as seen before, when we increase our speed our accelerometer sample distance increases so the values, may not be completely showing the full picture of the road condition. So, with this small test we cannot conclude that varying the speed over a road condition does doesn't affect the z axis accelerometer data. More information and testing on this will be needed for a conclusion.

One thing to also consider is that with this test it isn't feasible to vary the speed by much due to speed limit restrictions. So, we cannot drive over the same road condition going 5 MPH and then 70 MPH for safely or legally.

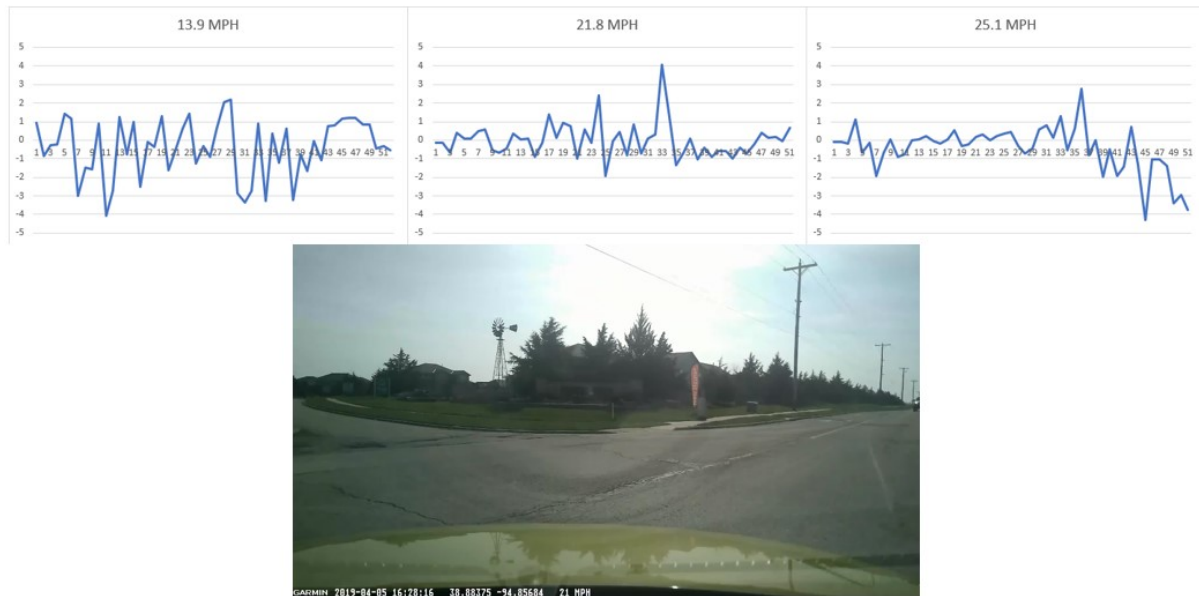


Figure 16: Speed variation test

4.2.3 Elevation Change Dependency

The final dependency evaluated is elevation of the route vs the z axis. See Figure 17, with the z axis data and the elevation of the route.

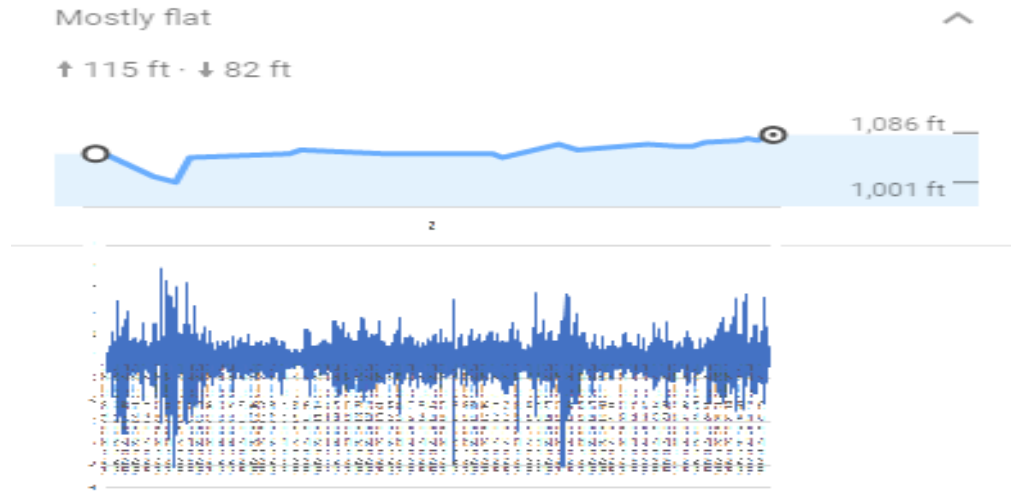


Figure 17: Elevation and z axis accelerometer data

At the beginning of the route there are many spikes in the z axis data both positive and negative while the route is decreasing in elevation. Also, when the route elevation is flat there are some spikes in data which seem to be elevation independent. A little over half way the route increases in elevation and around that time more z axis spikes are both negative and positive. Also, notice that when there are elevation changes there does seem to be spikes around those locations.

With this small test we cannot conclude elevation change don't affect z axis accelerometer data. More testing on this will be needed to make a better conclusion.

4.3 Classification of Road Conditions

To classify the road conditions, we need to know how we can accomplish this task. For visual reference we will use Figure 18.

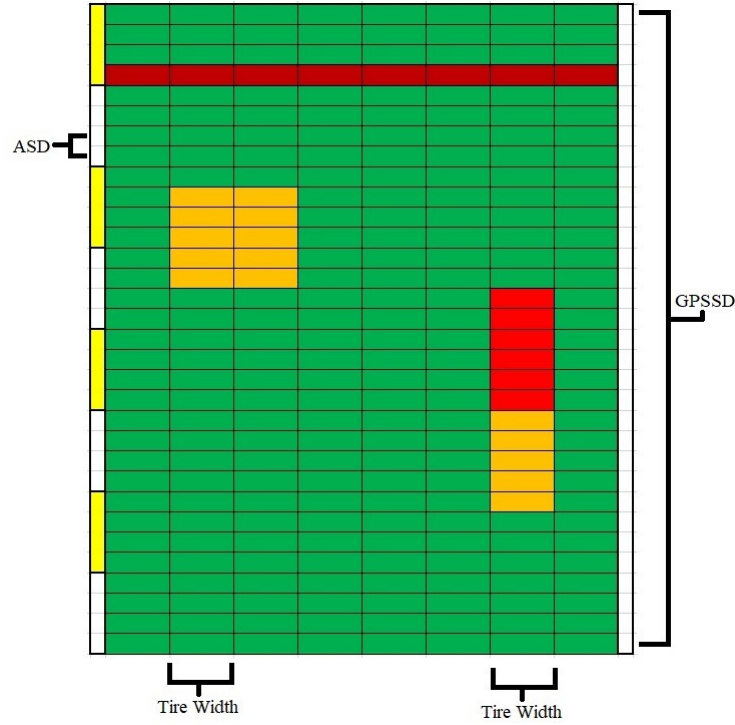


Figure 18: Road Classification Figure

This figure is an example of a zone and the colors in the grid spaces represent good, poor, and bad road conditions. In the figure above we see that we have the GPSSD the ASD and tire widths. The GPSSD is the GPS sample distance and the ASD is the accelerometer sample distance, and the tire width is the average width of a vehicle tire. When the vehicle drives over the section of road the tires can hit two different types of road conditions at the same time with the tires. Since the tires and the vehicle will react to driving over these road conditions, we will use the road condition that is worse for consideration. When we drive over the road condition, we will count how many of each type we encounter. We also need to know how many total ASD samples we have after driving over the zone.

In the example above, we have 32 ASD samples, 15 good, 10 poor, and 7 bad. At this point we can calculate a percentage of each classification. For the good portion of this zone we have $\frac{15}{32}$, for poor $\frac{10}{32}$, and for bad $\frac{7}{32}$. Now we can take the overall area of the zone and

multiply by the percentage to get a rough estimate of the zone classification. To calculate the zone area, we use the formula below.

$$\text{Zone Area} = \text{Classification \%} * \text{GPSSD} * \text{LC}$$

Using this idea to count the number of each road classification in the zone that we encounter in the data we can use this to give an overall picture of a route with regards to the number of good, poor, and bad road conditions are in that specific zone. In the next section we will show the route driven and display the same route with regards to the number of good, poor, and bad sections of that route.

4.3.1 Count and Magnitude of Road Conditions

In the previous section, we demonstrate that we can count the number of good, poor, and bad samples in the zone and determine a percentage of each classification. Figure 19 shows the route driven and Figure 20 shows this route but categorized as good, poor and bad respectively with the associated count and magnitude values. To compute the values for the figures, I have taken the absolute value of the z axis data point. Then I count the number of values that meet the range state above. Finally, I average the z axis values that are in that specific range.

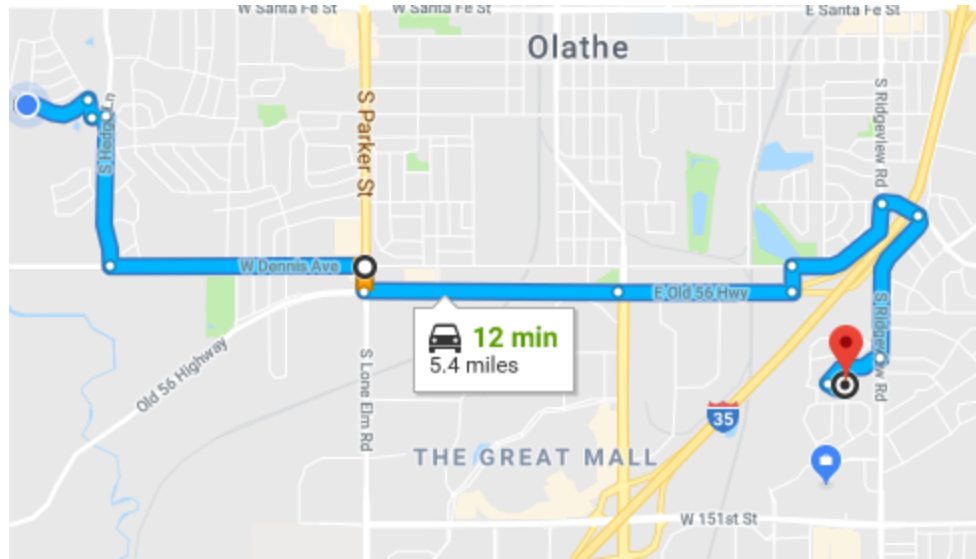


Figure 19: Route For Count and Magnitude Consideration

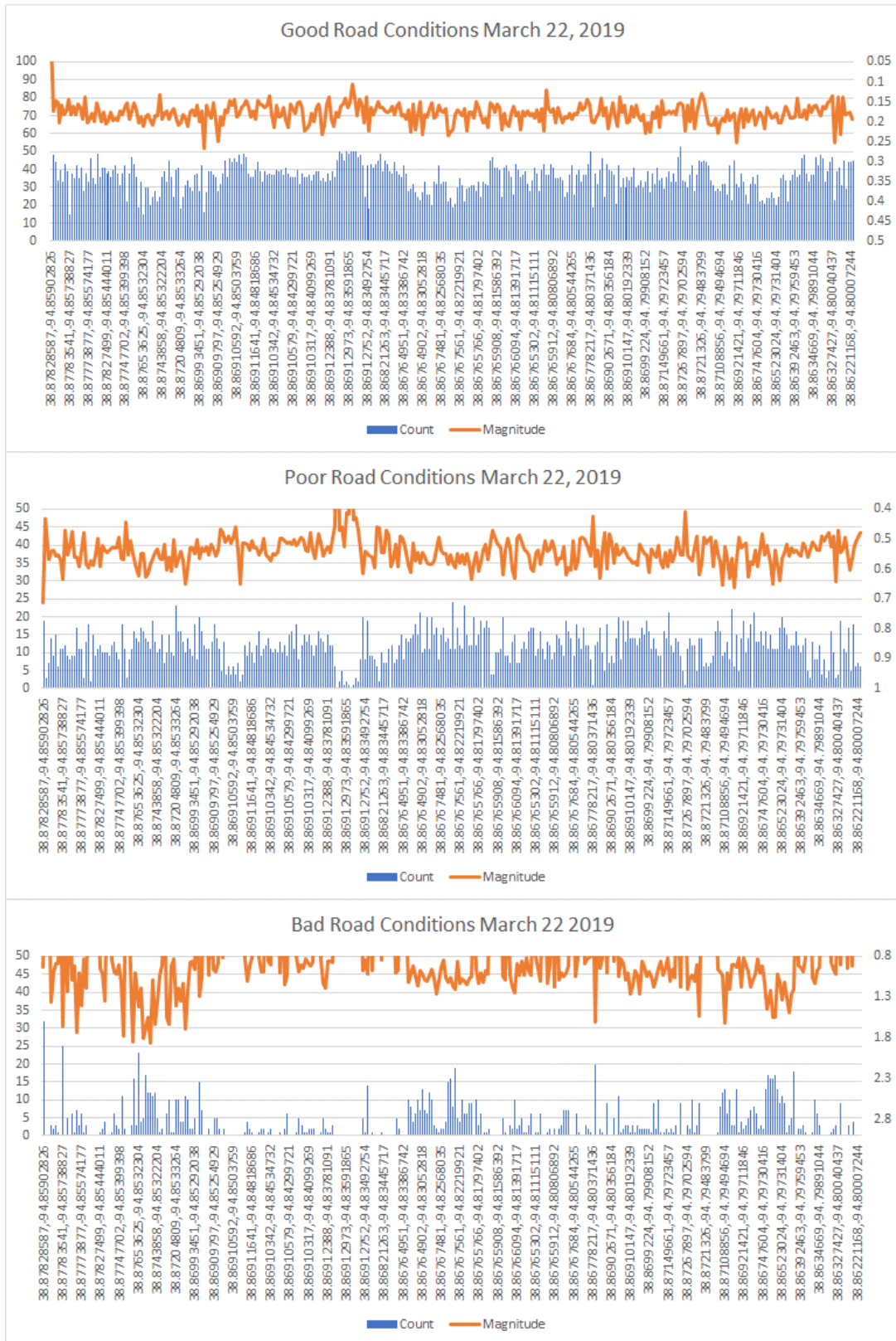


Figure 20: Good, Poor, and Bad Road Conditions

Figure 20 shows the road conditions on the route. The good road condition shows the magnitude of the good road conditions is between 0 and approximately 0.25 on average and the count that are considered a good road condition are between approximately 15 and 50. As discussed previously in the paper, we are sampling approximately 50 samples per second. This tells us that most of the road conditions are considered good. This makes sense because the entire road or route isn't bad just small portions. The poor road condition shows the magnitude is between 0.4 and 0.7 on average. The count that are considered a poor road condition are between 0 and 24. Finally, the bad road condition's magnitude is approximately between 0.8 and 1.90. The count that are considered a bad road condition are between approximately 0 and 33. The majority of the count for the bad road conditions are below 10 but a few are above. This indicates the problem areas on the route since there are many values that are in the bad road condition range.

4.4 Road Conditions

The types of road conditions that will be shown in the next sections will include intentional and unintentional road conditions. Intentional road conditions are road conditions that are present but are on purpose like speed bumps and dips. Unintentional road conditions are road conditions like potholes and cracks. In some of the sections below I will display routes that I have driven over multiple times and some only once. Some of the sections will also include consecutive zones due to multiple road conditions next to each other. The same smartphone, vehicle, and smartphone placement were used in all the sections below.

With all the road conditions below I will show the average speed and use the calculation discussed in the Create Zone section to determine the number of samples needed for the basic ASD. I will show the estimated area affected by the bad road condition using the area of the

zone and the ratio of bad road conditions detected over the total collected samples. Also, if there are multiple samples taken, I will display the average magnitude in $\frac{m}{s^2}$ and $\frac{ft}{s^2}$ for the magnitude of the bad road condition.

The figures in the sections below are using the same graph to show two different types of data. The count value is shown as a bar chart and the values associated are on the left-hand side, the magnitude values are the orange line and the values associated on the right-hand side of the figure.

4.4.1 Intentional Road Condition Dip

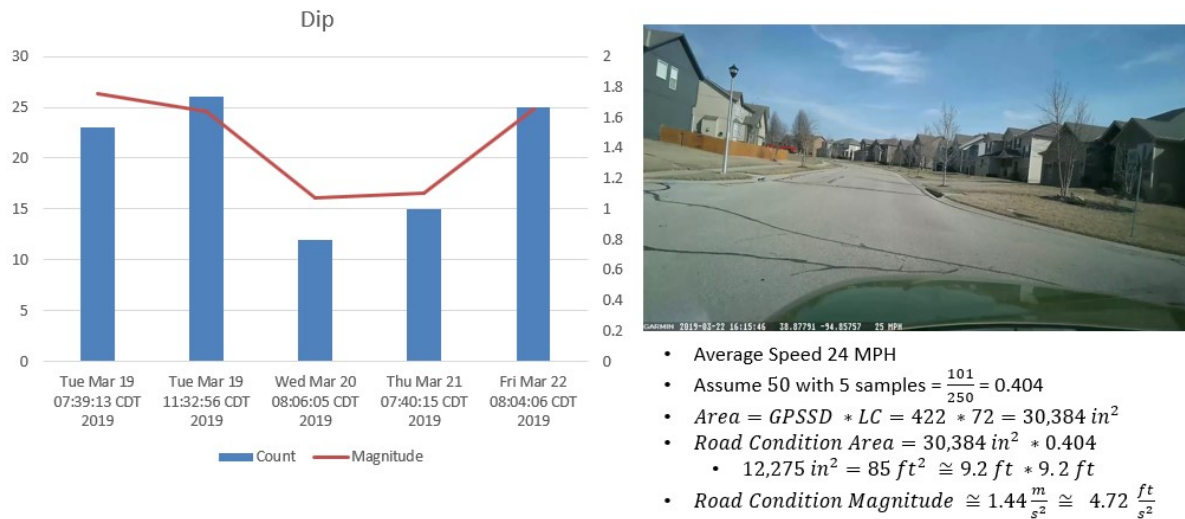


Figure 21: Intentional Road Condition Dip

The dip road condition above was driven over five times. Two times was driven on the same day at different times and the other three times was three different days. The bad road condition affected area is approximately $85 ft^2$ with a magnitude of $1.44 \frac{m}{s^2}$ or $4.72 \frac{ft}{s^2}$.

4.4.2 Road Condition Enter Bridge

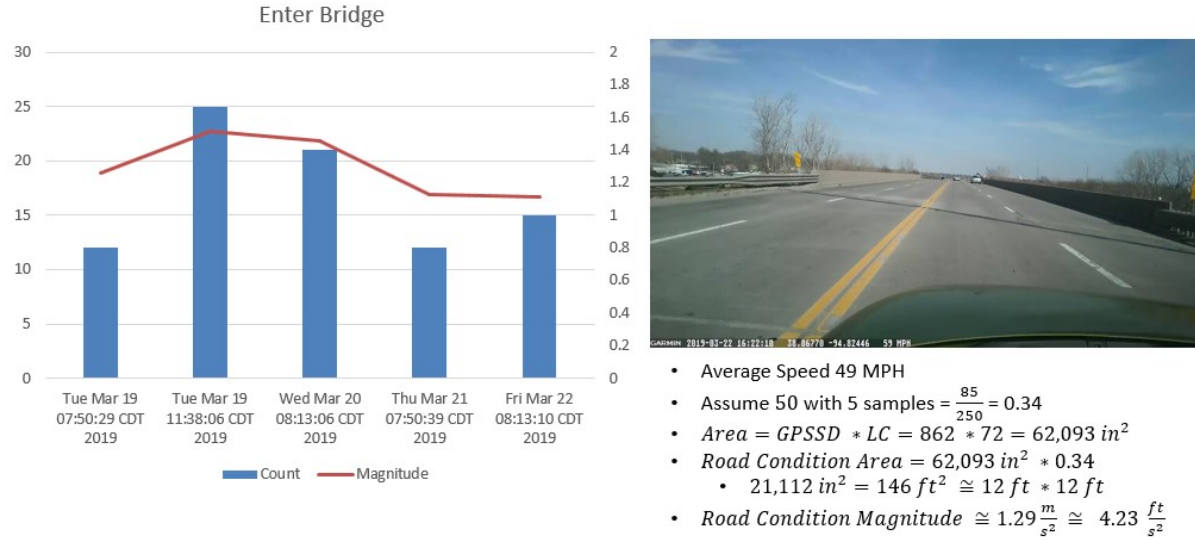


Figure 22: Road Condition Enter Bridge

The enter bridge road condition above was driven over five times. Two times was driven on the same day at different times and the other three times was three different days. The bad road condition affected area is approximately 146 ft^2 with a magnitude of $1.29 \frac{m}{s^2}$ or $4.23 \frac{ft}{s^2}$.

4.4.3 Road Condition Exit Bridge



Figure 23: Road Condition Exit Bridge

The exit bridge road condition above was driven over five times. Two times was driven on the same day at different times and the other three times was three different days. The bad road condition affected area is approximately 129 ft^2 with a magnitude of $1.22 \frac{m}{s^2}$ or $4.00 \frac{ft}{s^2}$.

4.4.4 Road Condition Fixed Potholes



Figure 24: Road Condition Fixed Potholes

The fixed potholes road condition above was driven over five times. Two times was driven on the same day at different times and the other three times was three different days. The bad road condition affected area is approximately 30 ft^2 with a magnitude of $1.28 \frac{m}{s^2}$ or $4.20 \frac{ft}{s^2}$.

4.4.5 Road Condition 3 Potholes



Figure 25: Road Condition 3 Potholes

The three potholes road condition above was driven over once. The first pothole is shown in the figure in front of my car. The second and third potholes are around the location of the car directly in front of me on the road. The bad road condition affected area for pothole 1 is approximately 38 ft^2 with a magnitude of $1.29 \frac{m}{s^2}$ or $4.23 \frac{ft}{s^2}$. For pothole 2 and 3 the affected area is approximately 133 ft^2 with a magnitude of $1.68 \frac{m}{s^2}$ or $5.51 \frac{ft}{s^2}$. The area for potholes 2 and 3 is combined as one since it happened to be captured in a single zone.

4.4.6 Road Condition Pothole and Low Spot

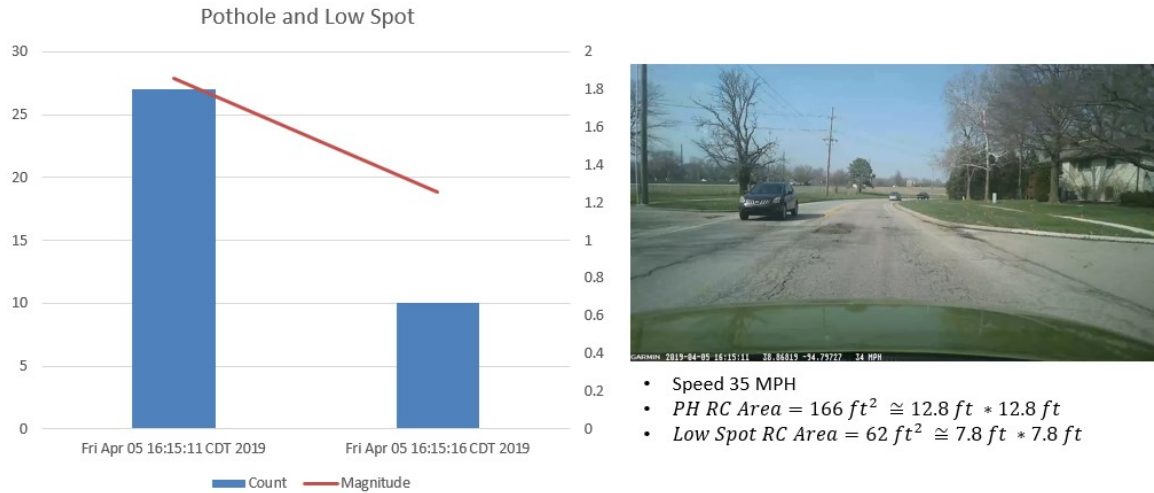


Figure 26: Pothole and Low Spot

The pothole and low spot road condition above was driven over once. The pothole is surrounded by an unsmooth surface that accounts for some of the area that will be detected. The low spot road condition is located near the car in front of me right before the turn ahead. The bad road condition affected area for the pothole is approximately $166\ ft^2$ with a magnitude of $1.86\ \frac{m}{s^2}$ or $6.10\ \frac{ft}{s^2}$. For low spot road condition, the affected area is approximately $62\ ft^2$ with a magnitude of $1.26\ \frac{m}{s^2}$ or $4.13\ \frac{ft}{s^2}$.

4.4.7 Road Condition Bad Area Lone Elm Road

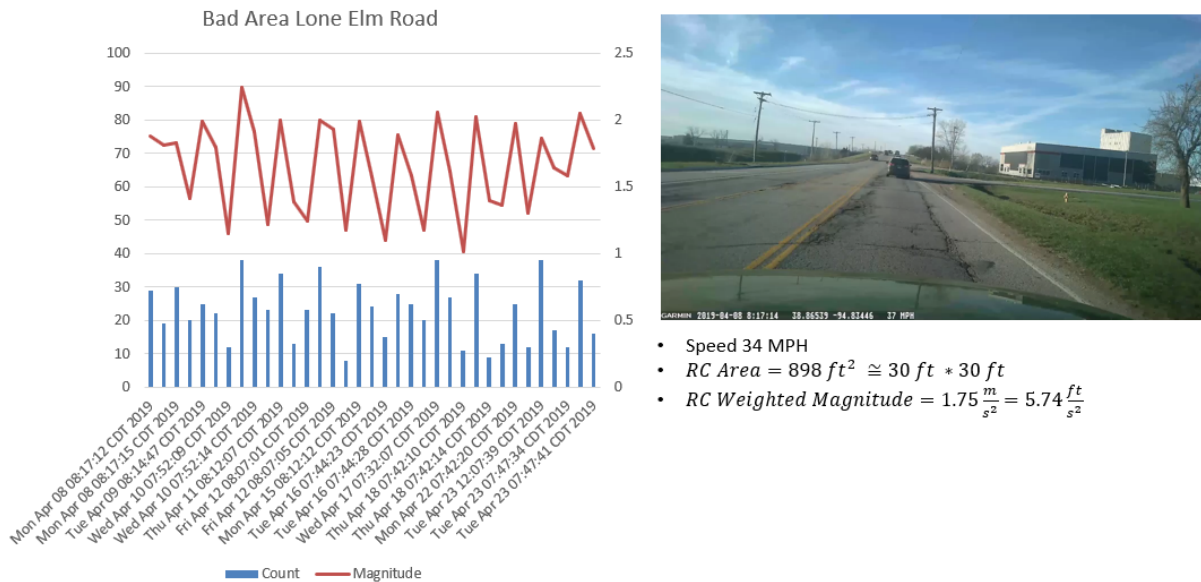


Figure 27: Road Condition Bad Area Lone Elm Road

The road condition bad area on Lone Elm Road above was driven over eleven times. The bad area spans approximately three zones so the area will be large. The road condition contains multiple potholes, cracks and low spots in the road. The bad road condition affected area is approximately $898\ ft^2$ with an average weighted magnitude of $1.75\ \frac{m}{s^2}$ or $5.74\ \frac{ft}{s^2}$.

4.4.8 Road Condition Raised Crosswalk



Figure 28: Intentional Road Condition Raised Crosswalk

The intentional road condition raised crosswalk above was driven over once. The bad area spans two zones because the crosswalk is large. The intentional road condition affected area is approximately $114\ ft^2$ with an average weighted magnitude of $2.33\ \frac{m}{s^2}$ or $7.64\ \frac{ft}{s^2}$.

4.4.9 Intentional Road Condition Speed Bump

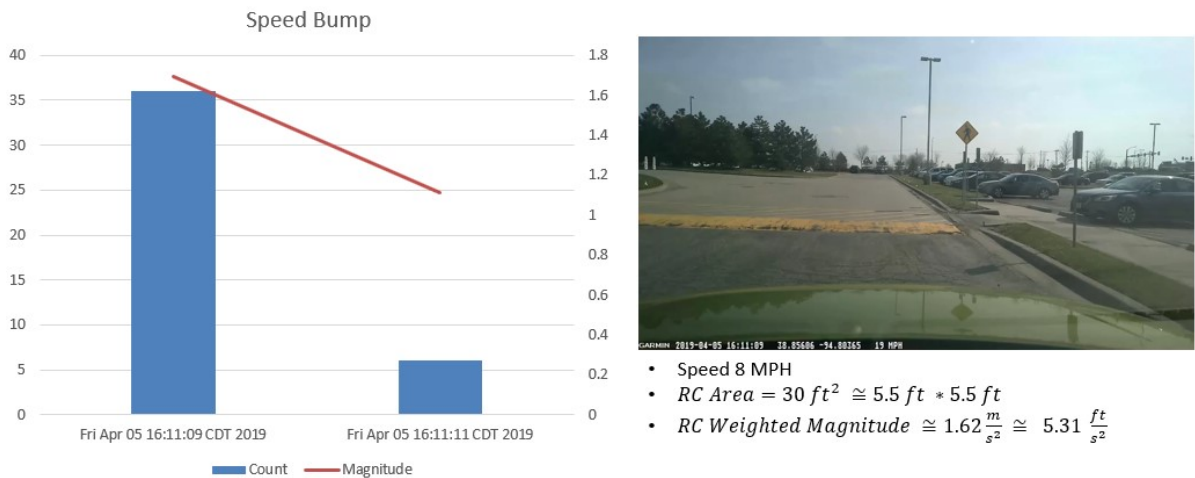


Figure 29: Intentional Road Condition Speed Bump

The intentional road condition speed bump above was driven over once. The bad area spans two zones because of the slow speed needed to drive over the speed bump with cause

more than a second to elapse. The intentional road condition affected area is approximately 30 ft^2 with an average weighted magnitude of $1.62 \frac{m}{s^2}$ or $5.31 \frac{ft}{s^2}$.

CHAPTER 5. CONCLUSION

The current method of reporting bad road conditions is a burden to the reporter. If the burden of reporting is too great, multiple instances will not be recorded or at the least delayed. This paper has shown that using a smartphone can be used to capture and display road conditions. After these road conditions are captured, they can be classified into good, poor, and bad road conditions. Along with classification, we can estimate the affected area of these road conditions. While a single user does not have the capacity to capture all the road conditions, the ability to use crowd sourcing dramatically increases the amount of data that can be gathered. Since the smartphone is used by a large portion of the population the ability to gather a lot of data is incredible. I have also shown that the smartphone might not have the sampling rate of a dedicated accelerometer. However, what it lacks in sample rate it makes up for in usability and ease of use. If the reporting burden is left solely to the user, it is likely to delay reporting which will have a trickle-down effect leading to delay of effective budgeting and road improvements.

REFERENCE LIST

- [1] "LibAnswers," 26 October 2017. [Online]. Available: <https://answers.jocolibrary.org/faq/209511>. [Accessed 24 3 2019].
- [2] "OPENDATA KC," [Online]. Available: <https://data.kcmo.org/311/311-Call-Center-Service-Requests/7at3-sxhp/data>. [Accessed 24 4 2019].
- [3] V. F. Enigo, T. V. Kumar, S. Vijay and K. G. Prabu, "CrowdSourcing Based Online Petitioning System for Pothole Detection Using Android Platform," *In Fourth International Conference on Recent Trends in Computer Science & Engineering*, vol. 87, no. Procedia Computer Science, pp. 316-321, 2016.
- [4] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs and L. Selavo, "Real time pothole detection using Android smartphones with accelerometers," in *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS) Distributed Computing in Sensor Systems and Workshops*, Barcelona, Spain, Spain, 2011.
- [5] V. M. Souza, R. Giusti and A. J. Batista, "Asfalt: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning," *In Pervasive and Mobile Computing*, vol. 51, pp. 121-137, 2018.
- [6] C. Gorges, K. Öztürk and R. Liebich, "Impact detection using a machine learning approach and experimental road roughness classification," *Mechanical Systems and Signal Processing*, vol. 117, pp. 738-756, 2019.

- [7] "GPS Accuracy," GPS.gov, 5 12 2017. [Online]. Available: <https://www.gps.gov/systems/gps/performance/accuracy/>. [Accessed 24 3 2019].
- [8] "Sensor," Android Developers, [Online]. Available: <https://developer.android.com/reference/android/hardware/Sensor.html>. [Accessed 24 3 2019].
- [9] "SensorManager," Android Developers, [Online]. Available: <https://developer.android.com/reference/android/hardware/SensorManager.html>. [Accessed 24 3 2019].
- [10] "Common Tire Sizes," [Online]. Available: <http://www.thetirestore.com/common-tire-sizes.html>. [Accessed 24 4 2019].
- [11] "Lane Width," U.S. Department of Transportation, [Online]. Available: https://safety.fhwa.dot.gov/geometric/pubs/mitigationstrategies/chapter3/3_lanewidth.cfm. [Accessed 24 4 2019].

VITA

Dustin Shorter completed his bachelor's degree in industrial engineering from Kansas State University in May 2007. He started his master's in computer science at the University of Missouri-Kansas City (UMKC) in January 2017, with an emphasis in Networking and Telecommunications and graduates in May 2019. While studying at UMKC, he works as a software engineering at Garmin. Upon completion of this requirements for the master's program, he will continue to work at Garmin as a software engineer.