Cost-effective Location of Road Anomalies Using a Microcontroller and Three-axis Sensors

Alice Huston

Toms River High School South

s.huston001118@gmail.com

1. Introduction

Roads in the United States, according to the 2017 Infrastructure Report Card, are "often crowded, frequently in poor condition, and becoming more dangerous," yet $420 billion is budgeted for repairing existing highways alone out of the $941 billion provided by the US government [4]. The severity of the impact on citizens also means that it is more important than ever to decide which roads need to be repaired the most. Previous projects have been high in cost and lack the portability required for larger scale projects. The *Pothole Patrol* (P^2) used a Soekris 4801 embedded computer, a WiFi card, a Sprint EVDO Rev A network card, an external GPS, and a 3-axis accelerometer to make surface anomaly measurements [2]. Meanwhile the RABIT robot is designed to be efficient and effective in its measurements, but requires shutting down a lane of the road in order to measure road data while not damaging the robot, traffic cannot continue to use the road alongside the RABIT [3]. There is a need for a system for detecting road surface anomalies that can be both precise and accurate while maintaining a low cost. The main goal of this system primarily relies on an Arduino Uno R3 microcontroller, an Adafruit 9-DOF breakout board, and a Sparkfun GPS Shield, where the breakout would be used to collect the sudden changes in accelerometer and gyroscope data that would signify a jolt and possibly the existence of a road anomaly, the GPS shield would be used to track the location of possible road anomalies, and the microcontroller would act as the hub for all of the data and write the data to the appropriate files.

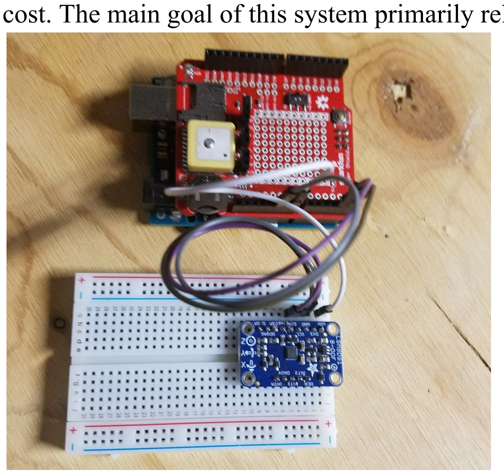

Figure 1: The system used in all trials, a GPS shield stacked on an Arduino Uno at the top and a breakout board at the bottom.

2. Equipment Preparation

In this section, descriptions of the assembly process for the system and any pre-experiment setup are provided. The method for gathering sample data is also provided.

2.1 System Assembly

Using the names given from the P^2 project, newer equivalents of the components used in the P^2 were found and their costs calculated as shown in Table 1 [2]. As the value of many of the items was unknown, N/A was used as a placeholder. The sensors used here are designed to match projects like P^2 in effectiveness while decreasing the overall cost. Even with partial estimates for P^2, the difference shown is still $64.97 cheaper than P^2. The data from this experiment were gathered using the breakout, GPS shield, and Arduino Uno R3 microcontroller previously discussed, with both the breakout and the shield connecting to the microcontroller. The shield was placed onto the microcontroller using header pins and from there the breakout was connected to the shield from which data were fed to the microcontroller. This method of breakout to shield to microcontroller wiring was repeated on another board in order to accommodate collecting two sets of data

Part Name		Total Cost	P^2 Equivalent Cost	P^2 Part Name	P^2 Equivalent Part Name
SparkFun GPS Logger Shield x2	$44.95	$89.90	N/A	Model Not Stated	Model Not Stated
Arduino Uno R3 (Atmega328 - assembled) x2	$24.95	$49.90	$315.23	Soekris 4801 embedded computer	Soekris net6501-30 embedded computer
Adafruit ATWINC1500 WiFi Breakout with uFL Connector - fw 19.4.4 x2	$24.95	$49.90	$25.99	Sprint EVDO Rev A network card	KuWFI 3G unlocked hotspot USB dongle
Adafruit 9-DOF Accel/Mag/Gyro+Temp Breakout Board LSM9DS1 x2	$14.95	$29.90	N/A	Model Not Stated	Model Not Stated
12V DC 1000mA (1A) regulated switching power adapter - UL listed x2	$8.95	$17.90	N/A	N/A	N/A
Premium Male/Male JumperWires - 40 x 6" (150mm) x3	$3.95	$11.85	N/A	N/A	N/A
Half Size Breadboard x2	$5.00	$10.00	N/A	N/A	N/A
9V battery holder with switch & 5.5mm/2.1mm plug x2	$3.95	$7.90	N/A	N/A	N/A
2.4GHz Mini Flexible WiFi Antenna with uFL connector -100mm x2	$2.50	$5.00	N/A	N/A	N/A
Arduino Stackable Header - 8 Pin x4	$0.50	$2.00	N/A	N/A	N/A
Arduino Stackable Header - 10 Pin x2	$0.50	$1.00	N/A	N/A	N/A
Arduino Stackable Header - 6 Pin x2	$0.50	$1.00	N/A	N/A	N/A
Total Cost		$276.25	$341.22		

Table 1: A comparison between the total price of the system used in this experiment and the system used in the Pothole Patrol experiment

simultaneously. Following its physical construction, the system was connected via USB to a computer and tested programmatically to ensure that data were accurate from both systems. The program itself was created in the Arduino IDE v1.8.3 using the C/C++ programming languages and several libraries provided by Arduino to communicate with the breakout and shield from the microcontroller, the TinyGPS++ library to parse the NMEA GPS strings provided by the shield, the Adafruit_LSM9DS1 library to read data from the breakout, and the Adafruit_Sensor library to normalize breakout sensor readings [1]. The shield and breakout returned all data to the microcontroller, but of the data only the parsed GPS strings, accelerometer, and gyroscope data were saved as individual inputs to a csv file on a laptop connected to the microcontroller. This data were the only data taken from the sensors due to the experimental focus on any correlation between acceleration and gyroscope data with the existence of road anomalies, and then GPS data was coordinated with the breakout data in order to create possible locations for any anomalies.

Once the systems were determined to be in working condition after assembly, the systems were both placed in a 2005 Dodge Grand Caravan. Any anomalies would show up twice in one encounter as one system in the rear of the vehicle would hit an anomaly immediately after the front system did. Once the systems were put into place, they were connected to laptops that received data through the serial monitor, which could then be moved to the proper files after collection.

2.2 Data Collection

With the system functional and set up for use, tests for road anomalies during normal driving could began. The location that was picked for this experiment was the town of Toms River, New Jersey, where the vehicle drove for fifteen trials at a constant speed of 16 m/s. Each set of three trials were considered as its own cluster of data for processing and each trial had its own csv file.

2.3 Data Processing

Once the data were physically collected and able to be processed, all data were placed into Microsoft Excel files based on their cluster. Data from each cluster were then fed into a separate application that removed any outliers that could not be anomalies, such as points where the speed or coordinates were zero, and placed back into Excel as the complete dataset for the cluster. Using this data, five number summaries were found for all three axes of the accelerometer and gyroscope and any extreme outliers were marked as potential anomalies. Any data points marked as potential anomalies were then printed to another csv file so the coordinates could be mapped and verified by following the original route and comparing each point to the actual condition of the road. Once the data were confirmed, Student's t-tests were done on each of the axes of the accelerometer and gyroscope in each cluster to confirm the validity of the results and a percentage of anomalies confirmed compared to potential anomalies were calculated.

2.4 Filtering Methodology

When data were collected and placed in the proper cluster, the validity of the data were solely reliant on the integrity of the system and the testing environment. In order to further validate the data and ensure that any outliers unrelated to the purpose of the experiment were removed, several Java classes were implemented in each cluster in two steps. The first class used was a NoiseHandler which took in the raw data and printed it again if two identifiers were false, if data points where the speed and date were equal to zero or data points where the speed was less than 1 m/s and the difference between the coordinates above or below the data point were zero. The first filter, speed and date being equal to zero, was chosen because the only time both values would ever equal zero is in the event of the GPS not having a fix and thus not being able to accurately record data. The second filter, speed being below 1 m/s and the difference between surrounding data points' coordinates being 0.0°, was chosen because these conditions only occurred when the car was stopped which would not provide any useful data about the road and would skew the data with excessively low numbers.

With the data being stripped, Excel formulas were used in order to find the five number summaries and IQRs (interquartile ranges) for the acceleration and angular velocity on the x, y, and z axes. Upper and lower bounds were determined for each category by adding or subtracting 3 times the IQR from the respective quartile instead of 1.5 times the IQR such that only the worst outliers would be present. From these bounds, boolean values were generated for each category for each point based on whether or not the value is within the proper bounds and a seventh boolean value was generated that would be equal to true if any of the booleans based on the acceleration and angular velocity data were false for that data point. The new data was then fed into another Java class, ParsedDataIQR, which took all of the data in a cluster and checked how many outliers had matching coordinates and of those outliers how many instances of the same outlier existed. If it was determined that an outlier existed at a location four or more times, the coordinates were printed to another csv file so the coordinates could be fed into a mapping program and visual verification could begin.

3. Trial Runs

The system relies on placement that allows changes from either side of the car to be felt by the gyroscope and accelerometer. As long as the system is placed correctly and the GPS has a fix, the data should reflect any potential road anomalies that the vehicle encounters. This section describes the placement of the system relative to the vehicle as well as the collection of the data throughout the trials.

3.1 System Placement

In order to ensure that the results could be replicated in another experiment, the two systems were placed in a 2005 Dodge Grand Caravan. Both systems were horizontally centered and then placed as close to the tire rods as possible, with onesystem being slightly forward from the back tire rod and the other being placed slightly backward from the front tire rod and directly in front of the console. In an effort to make the results from each device appear as identical as possible, the systems were placed facing the front of the vehicle. After a brief test run to ensure that the devices' responded similarly to movement, the vehicle began driving on the route specified in Figure 2 starting from the corner of South Main St. and East Water St.

Figure 2: A map of the route in Toms River, NJ. It goes from Water St. to Hooper Ave., then to NJ Rt. 37, then to Washington St., then to Main St., then back to Water St.

3.2 Experimental Conditions

The route that was used for the trials was 7.220 km (4.4863 mi) long and was traveled at a rate of 16 m/s, with some minor fluctuations due to traffic and road conditions. Fortunately, most of the trials were done between 2:30 UTC and 5:30 UTC on December 9th, 2017 which meant that the roads were almost completely empty throughout most of the experiment and the vehicle did not have to change its speed or move over for other vehicles often. The trials for most of cluster four and five occurred between 13:15 UTC and 14:25 UTC on the same day but were only slightly more impacted by traffic than the previous trials.

Due to the similarity in road and traffic conditions, 14 of the 15 trials happened with similar results. One trial that acted as an exception in the data collection had a minor incident on Hooper Avenue approaching NJ Route 37 where the vehicle had to make several sudden stops due to increased traffic. This incident belonged to a trial in cluster four, which was impacted by the variation in data collected and thus negatively affected the accuracy of the road anomaly locations collected from the data.

3.3 Data Analysis

The data in Table 4 are the average results from each cluster as well as the five number summaries, IQRs, and upper and lower bounds.

While the minimums and maximums were not used in determining the possible outliers, the upper and lower bounds that were found were compared to each data point that was recorded. For example, if at a specific moment in trial 3 the acceleration on the y-axis registered as $-6.03 m/s^2$ then it would be considered outside of the normal range and flagged as an outlier and potential road anomaly. The data in the tables were taken after the processes to remove any abnormalities in order to prevent the data being skewed by data with no GPS fix or stops. As a result of these filters being used, the vehicle could stop for an extended period of time in future tests without the data being affected as there should no longer be any data points where the speed is that low.

In an effort to prevent turns affecting the outcome filtering, any data from the z axis was eliminated. If the data had been kept in however, cluster four would have been affected by the anomaly much more than it already was due to its minimum and maximum being over ±100 dps. Besides that one exception, the data throughout the trials appeared similar and were then tested to via Student's t-tests (shown in Table 2), in which 29/30 tests reject the null hypothesis ($P < 0.05$). The one case that was not significant to the data was the acceleration of the y-axis in trial 4, which is likely a result of the series of stops impacting the significance of the data.

Taking these conclusions into consideration, a list of coordinates of potential anomalies was created using the ParsedDataIQR class, which were then visually checked to determine how accurate both the filtering method and the system was. As shown in Table 3, the system was shown to be 66.80% accurate with cluster four being the most accurate at 83.33% and cluster three being the least accurate at 57.14%.

Unlike previous analyses of the data, cluster four is not adversely affected by the anomaly, but this could

μ	0.5136	-0.2312	9.7856	1.7546	-0.1167	-2.1784
x̄						
t score	47.3286	40.0890	2.7112	-4.5413	-2.6340	-17.0543
p value	0	0	0.006710423	5.62E-06	0.00844566	9.10E-65
p value < 0.05	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
x̄						
t score	39.9069	34.4654	4.9536	-6.7540	-3.6150	-11.0016
p value	0	1.24E-252	7.35E-07	1.48E-11	0.000301162	4.55E-28
p value < 0.05	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
x̄						
t score	124.7010	52.4729	-32.7926	-17.0908	-1.9791	-13.3654
p value	0	0	1.80E-229	5.31E-65	0.047822	1.47E-40
p value < 0.05	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
x̄						
t score	-32.7369	1.6352	-12.5778	-9.0734	3.7762	-10.2245
p value	1.51E-228	0.102017	3.93E-36	1.27E-19	0.00016	1.79E-24
p value < 0.05	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
x̄						
t score	-258.4320	-180.0580	38.8440	55.3862	6.1505	55.1492
p value	0	0	0	0	7.86E-10	0
p value < 0.05	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Table 2: The results of the Student's t-test on each cluster. Each cluster is separated by a white space, all acceleration values are measured in $m/2^2$, and all angular velocity values are measured in dps.

also be insignificant due to the results of the t-test for this cluster.

4. Discussion

This section is intended to connect the results to the original purpose of the experiment as well as assessing the weaknesses of the system and any potential improvements that could be made in future tests.

4.1 Goal Analysis

The goal for this experiment remains to be finding a cost-effective means of precisely and accurately finding road anomalies. As far as cost goes, the price of two systems was higher than expected in the end, but was definitely still cheaper than the example shown in Table 1. Considering that the materials bought were actually for two systems, the cost could be brought down to $135.15, if enough materials were bought for only one system. Halving the resources at work could reduce the accuracy further than it is however, the accuracy was high

	Anomalies Confirmed	Anomalies Predicted	Percent Accuracy
Subset 1	38	48	79.17%
Subset 2	35	55	63.64%
Subset 3	4	7	57.14%
Subset 4	10	12	83.33%
Subset 5	90	147	61.22%
Total	177	269	65.80%

Table 3: The accuracy resulting from each subset (cluster) as well as the average accuracy.

enough to detect the most damaging anomalies along the route. A precise outcome was also reached, with only cluster four having outliers insignificant enough that they potentially skewed the results. Overall, the data shows a successful outcome in all three aspects of the experiment's analysis with small amounts of error that could be fixed by modifying the experimental procedure in future attempts.

4.2 Future Work

Given enough vehicles and road, an instance of this experiment could theoretically run forever. The design of this system allows so that the Arduino could collect data from all over a country and the filtering algorithms would eventually have enough data from smooth roads and road anomalies that only the worst road anomalies would show up in a dataset. By proxy, this also means that running the experiment over a longer period of time should increase the accuracy of the system as it learns what constitutes a road anomaly. This should also resolve most of the issues involving precision. A filter could also be created that checks for positive spike for angular velocity and a negative spike for acceleration on the axis that would be affected by a sudden stop, signifying both the sudden stop and forward lean that would accompany it, and possibly resolving the issue that skewed the results of cluster four. While it would make the experimental design slightly more complex, the cost of the experiment could be decreased by replacing the sensors with individual sensors that did not come pre-assembled like the GPS shield did. Another approach to the accuracy issue could be to create a fixed container for the system to ensure it is always facing the correct direction without any adjustment. Furthermore, the experiment could be

Mean values										
lat	long	alt	course	spd	Accelx	Accely	Accelz	Gyrox	Gyroy	Gyroz
39.9559079	-74.18376183	9.005772521	157.5611921	12.53506702	0.748142627	0.110044479	9.799480424	1.627308275	-0.16653942	-2.606622513
Minimum					-7.5	-7.01	2.55	-64.35	-149.54	-28.5
Maximum					8.16	6.13	19.6	128.41	37.58	22.14
Quartile 3					1.16	0.97	10.04	2.7	0.63	-0.86
Quartile 1					0.39	-0.71	9.52	0.37	-0.95	-3.91
IQR					0.77	1.68	0.52	2.33	1.58	3.05
Upper Bound					3.47	6.01	11.6	9.69	5.37	8.29
Lower Bound					-1.92	-5.75	7.96	-6.62	-5.69	-13.06

Mean values										
lat	long	alt	course	spd	Accelx	Accely	Accelz	Gyrox	Gyroy	Gyroz
39.9558	-74.1835	9.345008	160.6736	12.86574	0.691792497	0.049317863	9.807567511	1.647721526	-0.166777148	-2.474056403
Minimum					-2.84	-4.51	3.05	-16.74	-32.93	-29.56
Maximum					7.91	4.53	16.73	22.97	22.92	22.39
Quartile 3					0.95	0.92	10.06	2.76	0.6	-0.81
Quartile 1					0.45	-0.88	9.53	0.41	-0.92	-3.88
IQR					0.5	1.8	0.53	2.35	1.52	3.07
Upper Bound					2.45	6.32	11.65	9.81	5.16	8.4
Lower Bound					-1.05	-6.28	7.94	-6.64	-5.48	-13.09

Mean values										
lat	long	alt	course	spd	Accelx	Accely	Accelz	Gyrox	Gyroy	Gyroz
39.95377	-74.1839	14.87741	160.8119	12.92537	1.63127	0.161381	9.618703	1.480239	-0.14321	-2.54672
Minimum					-2.72	-7.77	-3.65	-31.6	-44.75	-32.19
Maximum					7.16	6.09	14.58	42.91	15.68	24.11
Quartile 3					2.57	0.86	9.99	2.69	0.63	-0.79
Quartile 1					0.79	-0.64	9.24	0.35	-0.92	-3.88
IQR					1.78	1.5	0.75	2.34	1.55	3.09
Upper Bound					7.91	5.36	12.24	9.71	5.28	8.48
Lower Bound					-4.55	-5.14	6.99	-6.67	-5.57	-13.15

Mean values										
lat	long	alt	course	spd	Accelx	Accely	Accelz	Gyrox	Gyroy	Gyroz
39.95564	-74.1833	12.93234	170.9223	13.73186	0.080873471	-0.21635	9.717748	1.599332	-0.05199	-2.46792
Minimum					-5.85	-5.13	4.37	-27.78	-32.45	-124.75
Maximum					6.27	5.43	15.84	27.37	31.32	136.55
Quartile 3					1.02	0.67	10.08	2.68	0.8	-0.67
Quartile 1					-1.21	-1.23	9.36	0.51	-0.93	-4.16
IQR					2.23	1.9	0.72	2.17	1.73	3.49
Upper Bound					7.71	6.37	12.24	9.19	5.99	9.8
Lower Bound					-7.9	-6.93	7.2	-6	-6.12	-14.63

Mean values										
lat	long	alt	course	spd	Accelx	Accely	Accelz	Gyrox	Gyroy	Gyroz
39.95572	-74.1847	18.55612	148.7506	14.04687	-0.46655	-1.14886	9.955379	2.335932	-0.05797	-0.95608
Minimum					-3.92	-4.78	1.87	-9.14	-11.18	-30.03
Maximum					3.06	3.2	16.19	18.88	12.8	23.2
Quartile 3					-0.26	-0.95	10.2	2.91	0.52	-0.55
Quartile 1					-0.68	-1.47	9.7	1.66	-0.67	-1.05
IQR					0.42	0.52	0.5	1.25	1.19	0.5
Upper Bound					1	0.61	11.7	6.66	4.09	0.95
Lower Bound					-1.94	-3.03	8.2	-2.09	-4.24	-2.55

Table 4: The average values for all recorded values except time and date, as well as five number summaries, IQRs, and bounds for each axis of the accelerometer (m/s^2) and gyroscope (dps).

repeated in a hilly area and a test could be done to see how different filters impact the accuracy of the system in comparison to the accuracy of the system on an average roadway.

4.3 Conclusion

In this paper, Arduinos were applied as a cost-effective method of measuring large jolts on a roadway that could signify the existence of an anomaly. Through this, a method of detection was found that was easy to use or modify if needed. This system allowed road anomalies to be accurately detected on a local road 65.80% of the time, which was then verified to be based on accurate data with one minor exception in the group known as cluster four using Student's t-tests. The determination was then made that these lower percentages could likely improve over time, as training the system further would allow it to become less sensitive to false positives and more sensitive to highly damaging road anomalies. Overall, the system achieved the goal of being cost-effective, accurate, and precise with room and adequate means to improve on all counts.

5. References

[1] "Build software better, together," *GitHub*. [Online]. Available: http://www.github.com/.

[2] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol," *Proceeding of the 6th international conference on Mobile systems, applications, and services - MobiSys '08*, 2008.

[3] J. Kim, N. Gucunski, and K. Dinh, "Similarities and differences in bare concrete deck deterioration curves from multi NDE technology surveys," *Health Monitoring of Structural and Biological Systems 2016*, Jan. 2016.

[4] "Roads," *ASCE's 2017 Infrastructure Report Card*. [Online]. Available: https://www.infrastructurereportcard.org/cat-item/roads/