Marcin Badurowicz[1], Jerzy Montusiewicz[2], Tomasz Cieplak[3]

Community-Driven Road Quality Assessment for Users and Territorial Government Organizations

1. Introduction

While the problem of finding and warning users about possible road hazards, including speed breakers, different road artefacts like potholes or points of surface degradation has been already discussed, and there are similar community-driven systems for monitoring police patrols or places of current road accidents, a more open question is if and how such data may be used by local government units.

The authors' previous works on road artefacts detection[4] with commonly used devices such as smartphones[5], concentrated on the detection of single road artefacts during one time drive over a road fragment. Similar aspects have been discussed in different works by several researchers in the last ten years, in different countries[6], and often a possibility for the creation of a community data acquisition system is also proposed[7]. While the community is already working

1 Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Institute of Computer Science.

2 Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Institute of Computer Science.

3 Lublin University of Technology, Faculty of Management.

4 M. Badurowicz, J. Montusiewicz, *Identifying Road Artefacts with Mobile Devices*, in: *Information and Software Technologies*, G. Dregvaite, R. Damasevicius (Eds.), "ICIST 2015", Springer-Verlag 2015, CCIS 538, pp. 504–514.

5 M. Badurowicz, J. Montusiewicz, T. Cieplak, *The Cloud Computing Stream Analysis System for Road Artefacts Detection*, "Computer Networks 23 rd International Conference, CN 2016", P. Gaj, A. Kwiecień (Eds.), Springer-Verlag 2016, CCIS 608, pp. 360–369.

6 J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, H. Balakrishnan, *The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring*, in: MobiSys 2008: Proceeding of the 6th International Conference on Mobile Systems, Applications, and Services, ACM, New York 2008, pp. 29–39.

7 P. Aksamit, M. Szmechta, *Distributed, Mobile, Social System for Road Surface Defects Detection*, in: Computational Intelligence and Intelligent Informatics (ISCIII), 2011 5th International Symposium, pp. 37–40.

on similar solutions themselves[8], the amount of such data is very limited and restricted to the most commonly used road fragments.

The authors would like to introduce such community-driven system reporting about road surface quality which may be useful not only for common users of the roads but also for local government units in future road maintenance plans – especially concentrating on smaller roads not included in central governmental planning. The problem of getting measurable data about road quality will be solved in the proposed solution by the usage of multiple devices providing the system with data (community-driven), while the road quality will be calculated from this data in the system (cloud computing based) in real time (on the fly).

There are already existing solutions on the road surface quality estimation using professional, specialized machinery. These methods are, however, very expensive and cannot be performed continuously, while the proposed solution will be able to distribute measurements, achieving less accuracy, but at a much lower cost for the benefits both for end-users (drivers) as well as road managers.

2. The System Architecture

The proposed system consists of a smartphone application, which is recording data about acceleration when mounted in a stable way in a car. Acceleration sensors are available in most currently available smartphones, as well as location sensors (using the Global Positioning System – GPS) and data connection.

Every user which has a proposed application installed is a data source for the system. The application, when running, records a set of data which is being directly transferred to the central processing system. Now, in the prototype solution, the authors are using MQTT protocol, while the prototype application is available for Windows Phone, Windows 10 and Android platforms. The proposed architecture is further presented in Figure 1:

[8] Scyscraper City.com, http://www.skyscrapercity.com/showthread.php?t=1310317 (31.10.2016).

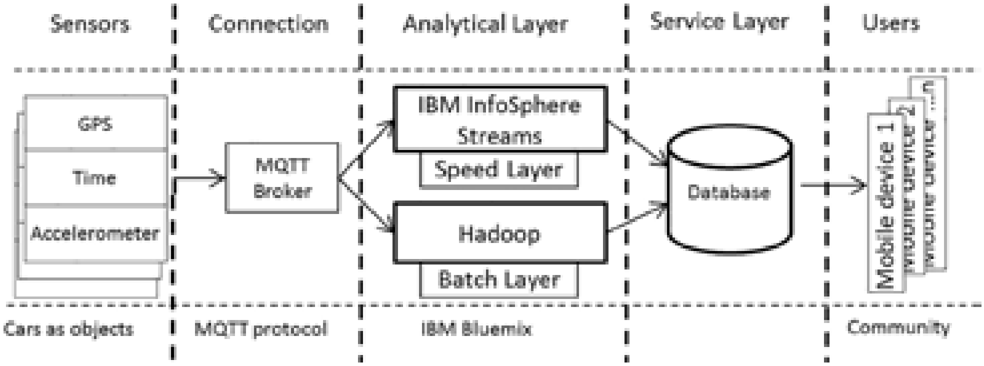

Figure 1. General System Architecture
Source: the authors' own work.

Every device sending data to the system is uniquely identified, but the solution does not require the user to perform any kind of registering in, allowing for anonymous access. Data is also generalized through the processing engine, extracting only information relevant to the problem.

The massive amount of data – as they are recorded multiple times per second, as well as a potentially massive number of devices, requires the usage of Big Data techniques. In the case of the system presented here, the Lambda architecture was used, enabling different analysis techniques for the same data streams.

The Lambda architecture is a new Big Data architecture as defined by Nathan Marz. It is a robust framework for ingesting streams of real-time data as well as providing efficient analytics of historical data. The assumption made for the system is that immutable data flows in one direction: into the system. The aim of the architecture is to ingest, process and query the data. Laying on the structure of the architecture, which consists of three layers, the processed data can be analyzed in real-time as well as by predefined ad-hoc queries. The Lambda architecture provides a general-purpose approach to implementing an arbitrary function on an arbitrary dataset and having the function return its results with low latency[9].

The system architecture, as presented in Figures 1 and 2, consists of three layers: batch, speed and service. The batch layer is a "data lake" for all data that flows into the system. That layer stores the master copy of the dataset. On the other hand, the batch layer precomputes the batch view per queries defined

[9] N. Marz, J. Warren, *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*, Manning Publications, 2015.

in the system as well as processing queries defined ad-hoc. The two main functions of the batch layer i.e. storing an immutable dataset growing with time, and computing arbitrary functions on the dataset are accomplished by means of batch-processing systems like e.g. Hadoop. The computing process with an arbitrary function can be achieved with the use of the MapReduce algorithms.

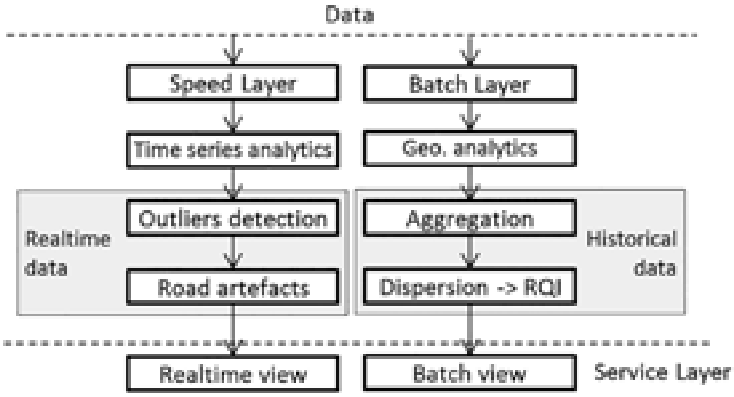

Figure 2. Data Flow in the Proposed System

Source: the authors' own work.

As opposed to the batch layer, the speed layer's objective is not to store data but to process it in real time. This data has just arrived in the system and has not yet been processed by the batch layer. It provides this data in the form of real time views. Subsequently, the data is transferred to the batch layer for further computation. The tasks of the speed layer are accomplished by means of stream computation systems, e.g. IBM InfoSphere Streams, Azure Stream Analytics or Apache Storm.

For example, the speed view can utilize the message publication/subscription mechanism, and this is proposed in the system the authors are working on, and the batch layer can employ different kinds of database systems[10].

Situated at the end, the service layer is responsible for serving the computed data to the users. In the proposed solution, where a set of users is providing the system with their data, to achieve a level of "fairness", the service layer sends alerts about possible imminent road artefacts or road quality deterioration directly back to the users' devices in the form of push messages.

10 B. Twardowski, D. Ryzko, *Multi-Agent Architecture for Real-Time Big Data Processing*, in: WI–IAT 2014, no. 3, pp. 333–337.

A more important aspect in the case of communication with territorial government units is that the service layer could be also prepared for providing data about current road quality conditions in a selected territory. In the current experiments, several generalized views of data are proposed, which may be useful for local governmental units or NGOs, and will be described later.

3. Location-Based Aggregation

As mentioned earlier, data is sent along with its time and geographical location, allowing us to perform historical data aggregation into groups over each geographical location saved in the system. This method was chosen to cope with the huge issue of getting multiple time-series based data: users may be driving in different times, with different vehicles over every possible road. When using location data as an aggregation source, acceleration data, which is a base of the system, always refers to geographical points, no matter what the time or speed or direction was of the car which was driving over that point.

The issue when using the location-based aggregation is obvious when the GPS system accuracy is taken into consideration: the location data are discrete values with variable accuracy. That means the system needs not only to aggregate data rows with the same longitude and latitude values, but also the points being significantly close (less than the current GPS receiver accuracy) being aggregated along. That also means it is not possible to calculate road quality from our acceleration data for the whole road, but rather for a series of points, which are close enough for the driver to treat them as the road as a whole: during the experiments, the GPS accuracy value was ranging from 10 meters down to 3 meters, depending on the device used and the parameters of the car. There is a possibility of achieving higher accuracy values using external GPS devices[11], however, the authors believe that even the low 10 meter value is high enough to estimate road quality properly for end users.

After aggregation, the first generalized value being calculated is the Dispersion (D): an absolute difference between maximum and minimum Z-axis acceleration values for the current location and measurement.

[11] M. Afrin, M. Redowan, A. Razzaque, *Time Detection of Speed Breakers and Warning System for on-road Drivers*, in: 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE).

The absolute value of dispersion is directly proportional to the road quality. Obviously, the index of road quality cannot be determined after one or a few measurements. To obtain an actual idea about road quality, the index should be calculated over many measurements. The graphical representation of dispersion value is presented in Figure 3.

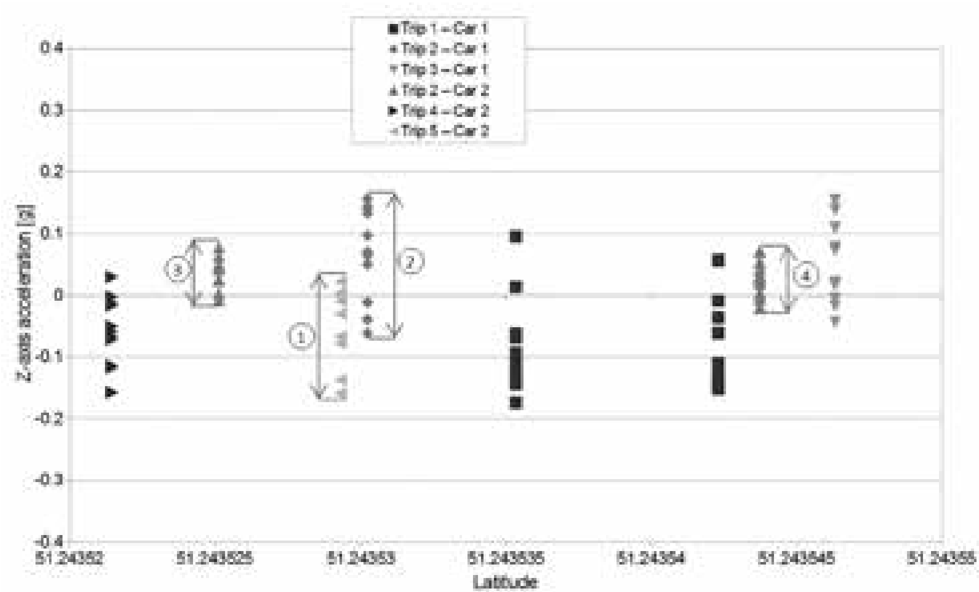

Figure 3. Dispersion of Measurement Points
Source: the authors' own work.

For every measurement, every location point and every acquisition device, dispersion was calculated separately, and was then aggregated again by calculating the average value of dispersion for the current location by the acquisition device. The separation between acquisition devices was introduced to cope with one more issue which was spotted during the planning phase of the experiment – the acceleration values for different cars will be different.

The issue was confirmed when comparing the results from different cars for the same data point. This is because of different types of suspension systems, different lengths of cars and wheel diameters.

This causes the need for the correction function, including the variables mentioned, which must be used to properly scale the dispersion value between multiple cars. Such a function was not introduced for the experimental phase, but will be an interesting point for future work.

The dispersion value is one of the metrics of the road quality: when corrected by the car factor, dispersion values will be aggregated over the current location from multiple data points. This Corrected Average Dispersion will be one of the metrics the authors would like to include in the final road assessment scale for end users, dubbed the Road Quality Index (RQI). The calculation of the RQI is the authors' next goal in the research performed.

The experiments to confirm the proposed quality indicator were performed during 4 trips over selected roads using two different cars, with different wheel diameters, a different distance between wheels and different wear, mileage and year of manufacture. The acquisition devices, Nokia Lumia 820 and 920 smartphones, were stably mounted in the car in the holders.

Those results were then analyzed, and previously presented dispersion and car correlation was found. The data was also presented in a two-dimensional graphical form, as presented in Figure 4:

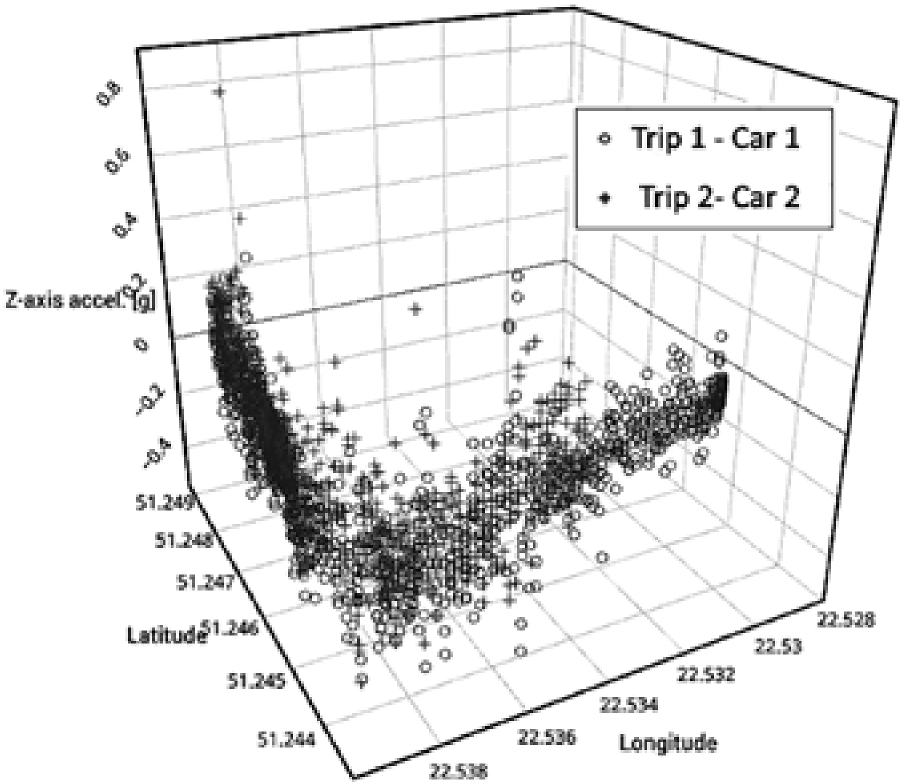

Figure 4. Z-Axis Acceleration vs the Road Profile

Source: the authors' own work.

The two-dimensional form is identical to the real road profile, as the road pro-file which was analyzed was a turn. The pure Z-axis acceleration for every trip and every car was aggregated also in a one-dimensional form, including only the latitude for easier analysis.

4. Road Quality Indicators Monitoring over Time

While the Corrected Average Dispersion (CAD) and further road indicators, like the proposed RQI (Road Quality Index) will be useful for both end-users and local governmental units in a fast assessment of road state, another aspect of the system is also very important, especially for territorial government units – monitoring of the road quality change over time.

The CAD, as an average value, will be recalculated every time any user is driving over already known point in the system. The new value of the CAD for the current location is saved in a specified database along with a timestamp of its change, allowing us to check out the road quality indicator value at the current point in time. Figure 5, below, shows the simulated approach on how the CAD changes over time could be presented to the users. In the simulated results, the value of dispersion is rising – which is a direct response to the road quality deterioration in this physical location.

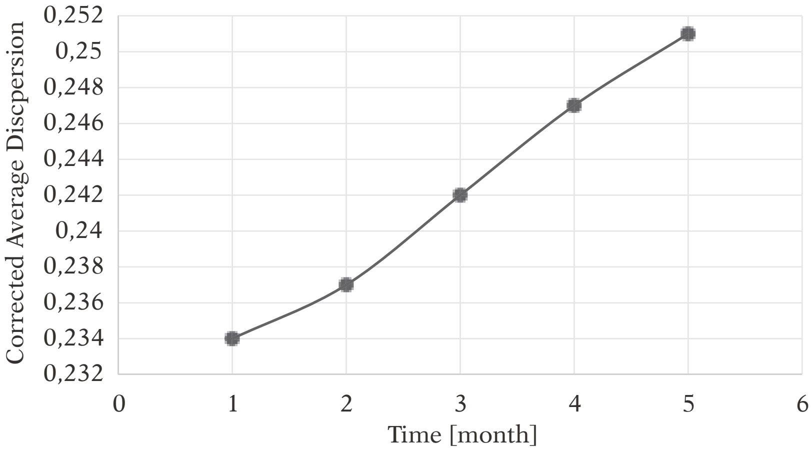

Figure 5. Changes of Dispersion for a Specific Location as Presented over Time (Simulated)

Source: the authors' own work.

This one of the proposed solutions may be especially interesting for territorial government units – the system will be able to calculate changes of road quality over time. This allows detecting the progressing deterioration of roads. This category of data, when correlated with another sets of data, e.g. a measured number of heavy devices, may be used in the decision-naking system about how the traffic (especially heavy vehicles) should be handled to minimize road maintenance costs. The authors believe such data will be interesting to territorial government units, especially at the beginning of the "smart cities" era.

5. Summary and Future Research

The presented solutions have been planned and some of the experiments were already performed, proving that the authors' approach is correct in the boundaries defined. Several new aspects were introduced, which should be considered: how to calculate the correction function to apply for results standardization, which will be valid research points in the future.

The system proposed, while the authors are concentrating on end-users, i.e. drivers at the moment, is also a valid and important concept for the governmental units – allowing road quality estimation in real time, measuring changes over time and extracting conclusions to be introduced. While the system is not as detailed and accurate as the specialized machines for road surface quality estimation, its centralization and low cost of implementation may be useful.

Bibliography

Afrin M., Redowan M., Razzaque A., *Time Detection of Speed Breakers and Warning System for On-Road Drivers*, in: 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE).

Aksamit P., Szmechta M., *Distributed, Mobile, Social System for Road Surface Defects Detection*, in: Computational Intelligence and Intelligent Informatics (ISCIII), 2011 5th International Symposium.

Badurowicz M., Montusiewicz J., *Identifying Road Artefacts with Mobile Devices*, in: Information and Software Technologies, G. Dregvaite, R. Damasevicius (Eds.), "ICIST 2015", Springer-Verlag 2015, CCIS 538.

Badurowicz M., Montusiewicz J., Cieplak T., *The Cloud Computing Stream Analysis System for Road Artefacts Detection*, in: "Computer Networks 23 rd International Conference, CN 2016", P. Gaj, A. Kwiecień (Eds.), Springer-Verlag 2016, CCIS 608.

Eriksson J., Girod L., Hull B., Newton R., Madden S., Balakrishnan H., *The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring*, in: MobiSys 2008: Proceeding of the 6th International Conference on Mobile Systems, Applications and Services, ACM, New York 2008.

Marz N., Warren J., *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*, Manning Publications, 2015.

Twardowski B., Ryzko D., *Multi-Agent Architecture for Real-Time Big Data Processing*, in: WI–IAT 2014, no. 3.

Web sources

Scyscraper City.com, http://www.skyscrapercity.com/showthread.php?t=1310317 (31.10.2016).

* * *

Społecznościowy system oceny jakości drogi z punku widzenia użytkowników oraz jednostek samorządowych

Streszczenie

Artykuł prezentuje koncepcję wykorzystania społecznościowego systemu informatycznego do gromadzenia danych dotyczących stanu nawierzchni drogi. Dane są zbierane za pomocą telefonów typu smartphone wyposażonych w odpowiednie czujniki, a następnie opracowywane za pomocą przetwarzania strumieniowego w chmurze obliczeniowej, tak by możliwe było wyznaczenie współczynnika oceny drogi w danym punkcie fizycznym. Informacje tego typu są aktualizowane wraz z każdym kolejnym przejazdem dowolnego użytkownika przez dany punkt, co pozwala na wyznaczenie tego, w jaki sposób konkretna lokalizacja pogarsza się lub polepsza w czasie z punktu widzenia użytkownika drogi. Te informacje mogą być wykorzystywane przez jednostki samorządowe do planowania niezbędnych modernizacji lub zlecania dokładnych badań stanu nawierzchni drogi.

Słowa kluczowe: przetwarzanie strumieniowe, chmura obliczeniowa, jakość drogi, czujniki, urządzenia mobilne