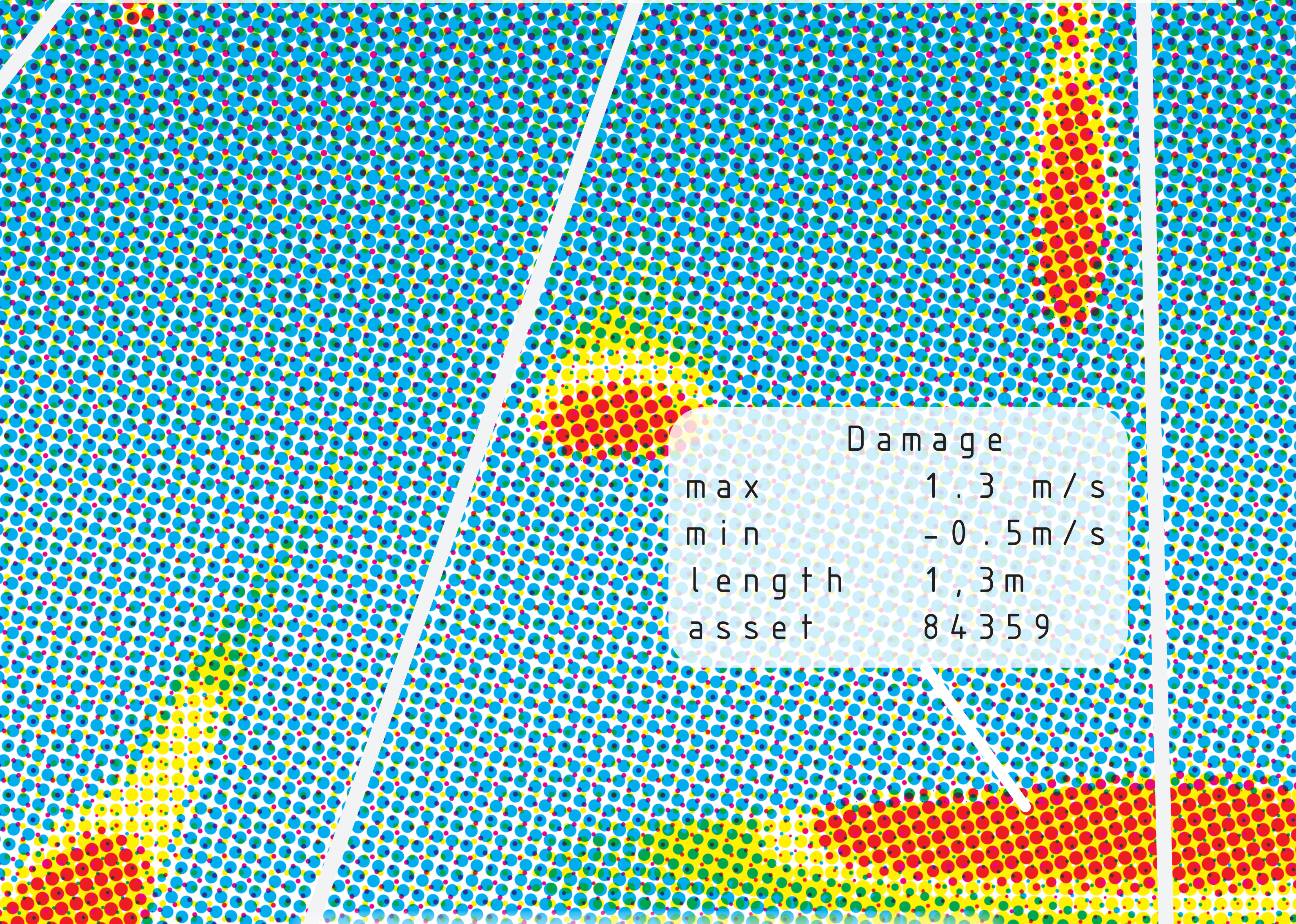
Crowdsensing as a tool for up-to-date road asset distress detection



Birgit Rosanne Ligtvoet - 3rd of July 2017
MSc Geomatics for the built Environment

Crowdsensing as a tool for up-to-date road asset distress detection

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Geomatics

by

Birgit Rosanne Ligtvoet

June 2017

The work in this thesis was made in cooperation with:

Department of OTB
Faculty of Architecture and the Built Environment
Delft University of Technology
Julianalaan 134
2628 BL Delft

VolkerInfra
Lange Dreef 19
4131 NJ Vianen

Supervisors:	Ir. E. Verbree (Edward)
	Dr.ir. B.G.H. Gorte (Ben)
	Dr.ir H.K.M. van de Ruitenbeek (Martinus)(VolkerInfra)
Co-reader:	Dr. A.A. Diakité
Delegate from the Board of Examiners:	Dr.ir. G. Bracken (Gregory)

Abstract

Contractors are moving from traditional maintenance to preventive maintenance. To apply preventive maintenance, up-to-date data is required which is unavailable. The goal is to research the applicability of crowdsensing to detect road pavement distresses, and to investigate how the positions of these detections can be used to gain information on a road pavement section.

The construction sector is currently moving from traditional maintenance contracts, where maintenance was a pay per bill business, to contracts focused on the performance and availability of assets. These types of contracts shift the risk during the maintenance phase from the government to the contractor. The contractor receives a fine if either the minimum performance or the availability is not met. This fine in combination with the risk of unforeseen maintenance costs creates a demand for more efficient maintenance methods like preventive maintenance. However, in the field of road asset management, there is a knowledge gap in how road pavement distresses develop over time. To fill this knowledge gap, up-to-date information is required. The up-to-date information cannot be provided using the current high-quality methods due to the limited availability of equipment.

In this research, raw data is gathered on the state of the road through crowdsourcing with smartphones. This data is then preprocessed by signal analysis, positioning analysis and event detection to retain only the information which describes the occurrence of a road pavement distress. Next, a database analysis consisting of clustering, asset linking and averaging is performed to gather information on the measurements of an instance of road pavement distress. The position of the instance of road pavement distress can be seen as the position of a virtual sensor. A virtual sensor groups events together from multiple smartphones. The location of the virtual sensor is linked to the location of the nearest road pavement asset, allowing asset managers, the people who plan maintenance tasks at contractors, to send their maintenance crew to the correct location.

The results from the experiments show that the detection of road pavement distress is possible by using crowdsensing. The position of the virtual sensor increases in accuracy when more data is used. However, it is possible that two different virtual sensors can be merged due to positioning errors.

It is also possible to connect the position of the virtual sensor to a real world asset location by using the Dutch national road sections and the hectometer posts. The resulting information can be used to get an indication of the status of different assets, giving asset managers insight into which areas and assets they need to focus on.

Keywords: *crowdsensing, smartphones,Asset Management, event detection, road pavement, distress, virtual sensor, clustering*

Preface

At the moment of writing this preface, it has been three-quarters of a year since starting my graduation project. It has been a wild ride with the corresponding ups and downs and of course, also lots of stress abound. Now the end is in sight and I must say it is a strange but happy feeling. I have enjoyed following the Geomatics for the Built Environment Master, and have experienced many interesting and fun moments because of it. Soon I will be leaving these wall for the last time, well, at least until I return for some random event. Now to thank some people who have helped me greatly during my graduation project.

First I would like to thank Edward and Ben, who have been very involved. I don't think I could wish for any better people to be my mentors. They took ample time to discuss every point of my research in biweekly hour-long discussion sessions where everything from my methods to physics concepts was thoroughly discussed. I always left with my head filled with ideas and lists of thing I needed to change or look into. It also caused me to stress out in confusion due to an information overload.

Next, I would like to thank Martinus and VolkerInfra for allowing me to perform my graduation research at their company. They provided me with a laid back atmosphere where I felt I could ask anything. I know I don't always make use of that, but I really appreciated it. All the people in the Future Lab were always ready for some questions, and some really fun stuff happened there during my stay.

On to some Geomatics and ex-Geomatics students. Both Bart Staats and Fanny Bot helped me during our weekly Tuesday meet-ups where we would work on our projects together. They helped me when I didn't understand something and I hoped I helped them as well. Hiske Braaksma was always available when I had some deep conceptual problem. She used her free time, and sometimes non-free time, to support me and help me with topics I didn't understand at all.

Finally, I would like my parent who supported me throughout my master and during my graduation. My mom, who always waited for me during dinner time, and my father who took the time to drive me across the same stretch op road for 3 hours straight. Thank You so much!

Contents

List of Figures

List of Tables

Glossary

asset a physical object which has potential or actual value. Also includes possible risks to the object.

Asset Management The management of physical objects to realize their values throughout a series of coordinated activities taking place throughout their lifespan.

cluster A collection of events which are used as a detection of road pavement distress at a position. A cluster is created when a minimum amount events are positioned near each other. The centre of the cluster is the position of the virtual sensor.

event A collection of measurements from a single which are used as an indicator for where road pavement distress might have occurred. An event is created if the vertical acceleration of the measurements is above a threshold.

longitudinal Along the length of an object. Does not describe a time series.

measurement Data output by the smartphone sensor at time X. The measurement consists of multiple parts like the position, the acceleration and the quaternion.

position A location is a subjective description of the position of real world object. Locations are used by people when trying to navigate instead of coordinate positions.

position The coordinates of an object gathered through GNSS. The object itself is not necessarily connected to any recognisable real world objects.

ride A single commute from start location A to destination B by 1 car.

virtual sensor A sensor which does not exist in the real world. It is connected to the center position of the detected cluster. Virtual sensors are only equipped to positions where a road pavement distress has been detected and contain information on the measurements from multiple events from different rides.

z acceleration vertical acceleration.

Abbreviations

ABS	Anti-lock Braking System
API	Application Programming Interface
ARAN	Automatic Road ANalyzer
CSV	Comma Separated Value
DBFM	Design, Build, Finance & Maintain
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DCM	Direction Cosine Matrix
FFT	Fast Fourier Transform
GNSS	Global Navigation Satellite System
GIS	Geographical Information System
GPS	Global Positioning System
HRI	Half-car Roughness Index
IMU	Inertial Measurement Unit
IoT	Internet of Things
IoV	Internet of Vehicles
IRI	International Roughness Index
LIDAR	LIght Detection And Ranging
MEMS	Microelectromechanical System
MPR	Mean Panel Rating
OGC	Open Geospatial Consortium
PBC	Performance Based Contracting
PS	Persistent scatter
PSI	Present Serviceability Index
PS-InSAR	Persistent Scatter Interferometric Synthetic Aperture Radar
RN	Ride Number
RWS	Rijkswaterstaat
SOS	Sensor Observation Service

SWE Sensor Web Enablement

TPR Taxi Passenger Refusal

UAPS Unmanned Aerial Photogrammetric System

UAV Unmanned Aerial Vehicle

WFS Web Feature Service

WPS Web Processing Service

ZOAB Zeer Open Asfalt Beton

1 Introduction

Every day thousands of people use the Dutch highway system to get to their work, to go shopping or maybe visit a theme park. These people would like to get to their destination as fast a possible without having to wait in traffic congestions. Some of the possible causes of traffic congestions are bad road pavement or maintenance works. To ensure that people experience the least amount of possible nuisance, maintenance is generally planned in the evening or night and during holidays. The maintenance planning is created by an asset manager working at a contractor, who decides what road section needs maintenance at what time. The asset manager used to plan maintenance based on whether the owner of the road contacted them. Unfortunately, an asset manager cannot plan every maintenance tasks in advance. Road pavement distresses can occur suddenly and in some cases require immediate maintenance. The risks associated with road pavement are traditionally the responsibility of the owner of the road. However, things are changing in the contracting world. The owner of the Dutch road network now wants that the contractor is more proactive in managing the state of the highway network. Let's look at how an asset manager traditionally works to create his maintenance planning in a scenario.

1.1 A current scenario in Asset Management

An asset manager named John at VolkerInfra, a contractor, is planning the new maintenance schedule for the coming months (figure 1.1). There are many maintenance jobs in the pipeline, making a complex schedule of tasks. Due to the high pressure on the maintenance schedule, John cannot afford to have sudden emergency tasks.

Unfortunately, John does not have an up to date status of the road pavement of the Dutch highway system at his disposal. Instead, he has to make due with information from daily quick visual inspections and a yearly inspection by a professional measurement vehicle. With his professional knowledge, John is able to make a maintenance planning. However, John does not have any insight into how damage develops over time because he mostly receives information when objects or assets already need maintenance (Beijer, 2016). From Johns perspective this means that damage seemingly occurs at a random time, making it difficult for him to anticipate when damage will appear. This is exacerbated by the fact that when John is checking the state of his assets, he does simple tools but for the most part he has to examine the information of each asset under his supervision.

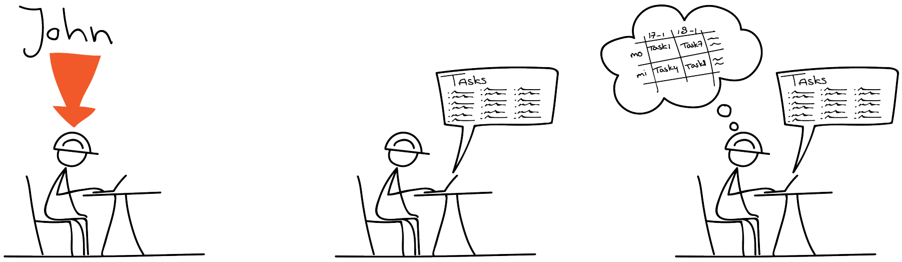


Figure 1.1: Asser manager John trying to create the maintenance planning.

1.2 Problem statement and motivation

The importance of carrying out performance-based maintenance is increasing. The building and construction sector is increasingly often using Design, Build, Finance & Maintain (DBFM) contracts and the railway sector is moving to Performance Based Contracting (PBC) for their maintenance contractors (Leijten & Koppenjan, 2010; Pohl, Schenk, & Bilt, 2013). With DBFM contracts the contract value is decided up front and the contractor has to finance both the construction and the maintenance, shifting the risks to the contractor. Efficient planning allows for repairs to be done at the least costly times of the day and week, reducing maintenance costs. In PBC the contractor is obliged to provide a certain amount of availability instead of maintenance, which is linked to financial repercussions when the obligations are not met. In both situations monitoring the performance of assets can be used to gain knowledge on when assets need maintenance, which in turn enables for more efficient planning and strategic decision-making regarding maintenance activities.

There is also interest from other organizations to know more about the state of the Dutch highway network. In 2010 the Dutch organization TNO performed a quick scan of the potential of the car as a sensor and held a workshop to gather more information on the subject (Klunder et al., 2010). The results of the quick-scan showed that there are both interested parties as well as skeptics. Both parties agreed that this type of sensor data could be used to learn about how distresses develop over time and to perform lifetime extending maintenance. Organizations in other countries are also exploring the potential of using smartphones to monitor the state of their own road network. There is, for example, the use of the Roadroid app in Sweden and the use of the StreetBump app in Boston, USA (Carrera et al., 2013; Forslöf & Jones, 2015)

1.3 A future scenario in Asset Management

What John would like is to have more up to date information on the state of the highway pavement (figure 1.2). John thinks he can use this extra information to learn more about how road distress occurs over time. The next step for John would be to use this knowledge, information and data to create a system which can predict where and when distresses are going to occur. The knowledge on how distresses occur and when they will occur can then be used to take into account these events when creating the new maintenance schedule.

However, John first needs data before he can derive information. John is unable to simply increase the frequency at which the professional measurement vehicles inspect a highway because of the high costs and the lack of hardware. Instead, he would like to use the commuters who use the highway as his inspectors through mobile crowdsensing (figure 1.2). 81% of the Dutch have a smartphone which can potentially be used to collect data through its many sensors (de Bruyckere, 2015). The data from smartphones is less accurate compared to a professional measurement vehicle. However, due to the sheer amount of up to date data which can be gathered, indications of change in the state of the road pavement might be detected.

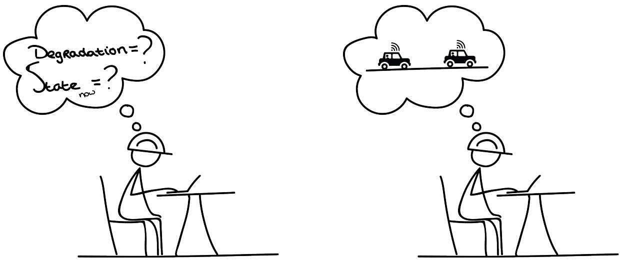


Figure 1.2: John would like to know the current state and the degradation of a road pavement asset, he thinks that this information can be gathered by using mobile crowdsensing; getting the information from commuters.

1.4 Academic relevance

The subject of using mobile crowdsensing to say something about the state of the road network has been an area of interest for researchers. In 2016 a thesis was written on the subject of using consumer-grade smartphones as a substitute for the ARAN of RWS (Lub, 2016). The results showed that the accuracy of the smartphone data is not sufficient for this type of application. Next, there is the Roadroid system which is under development by an international research team. The Roadroid system collects both accelerometer data and visual data on the status of the road (Forslöf & Jones, 2015). Furthermore, there is the Nericell system developed by Microsoft, which is used to detect potholes (Mohan, Padmanabhan, & Ramjee, 2008).

The Nericell application was designed at the start of the smartphone era, approximately 9 years ago with the focus of analyzing accelerometer data to detect if something has occurred (Mohan et al., 2008). In the Nericell the smartphone data is also reoriented to allow the comparing between datasets from differently oriented smartphones. This thesis goes further than Nericell: data between interesting sets of measurements is compared and combined to a single new measurement with a higher measurement frequency.

Compared to both the Roadroid system and Lub's thesis, this thesis is focused on just processing data which describes damaged areas. The mindset is that when no deviations can be detected, the current state of the asset is good. The state of an asset is bad when a deviation from the normal situation is detected. The information on the deviation from the normal situation is stored as measurements in an object called an event.

Another point of difference is that both Lub and the Roadroid system do not take into account the changes in orientation that are caused by for example different telephone holders in cars, which cause differences with which phenomenon is being measured on which axis. These differences can be removed, which enables the comparison between different measurement sets.

Finally, none of the projects described made an attempt to connect the data to an asset, instead, they used just the GNSS position. In this thesis, the data retains its positional information while at the same time be associated with an asset which has its own position. By associating data and information with assets, contractors can query the information based on their asset of interest ad have the option to color assets based on the linked information.

1.5 Research question

Based on asset manager John's problem of not having up to date information, and considering the related researched performed for similar scenarios, the following research question is formulated:

To what extent can the current state and the degradation of a road pavement asset be measured using mobile crowdsensing?

The research has three focus areas: the use of mobile crowdsensing methods to gather data, the detection and analysis of data which describes an event happening at a location on the road pavement asset, and the creation of the connection between the asset and the information pertaining that specific asset.

The term mobile crowdsensing is first mentioned in a paper by Mohan et al. (2008), where it is used to describe both participatory mobile crowdsensing and opportunistic mobile crowdsensing using smartphones. Crowdsensing itself is rooted in the fact that some devices are able to not just collect data, but also able to process and communicate this information or data in different ways. The mobile crowdsensing mentioned in the research questions is the opportunistic type, which requires the minimum amount of interaction of the smartphone user and is largely autonomous.

The goal of this thesis is to test the extent to which large amounts of data gathered trough mobile crowdsensing can be used to detect the state and degradation of road pavement assets quantitatively and qualitatively. This information can then be used for different applications, e.g.: add a change in the maintenance plan, make a decision to do nothing based on the data, schedule an inspection, or learn more about the features or degradation of the Dutch road network. The focus is not on delivering the most accurate profile through calculating, for example, the displacement, instead, the focus is on how to deal with the large amount of data gathered through crowdsensing by detecting events, and how information can be connected to road pavement assets.

1.5.1 Subquestions

In support of the main research question, six subquestions are defined. These subquestions have a more limited scope compared to the main research question, and cover different areas of interest.

1. Is it possible to detect road pavement distresses with consumer grade sensors?
   This sub-question is focussed on the process of pre-processing the data and analyzing multiple data streams to gather information usable for the monitor a road section asset at a single point in time. In this case specifically the current time. If it is impossible to monitor the current state of a road section, then it would be impossible to monitor the history with the same method.

2. What data and information are needed to monitor a road pavement asset?
   It is important to specify the needed data and information before data collection to prevent a mismatch between the collected data and the information which needs to be calculated from that data. A mismatch between these two factors can result in an inability to gather the wanted information, making the data useless for its intended purpose.

3. Which factors influence the accuracy of a detected event?

While there are factors which always influence the accuracy of the collected data, there are also factors which influence this specific application and the analytical methods used to gather information. During this thesis, there is the local analysis phase. The accuracy of the results from this phase, which are the events, has an impact on the subsequent database analysis or shock determination phase. There are multiple factors which can influence the accuracy of the data, some of these factors are a common occurrence. However, there are also factors which are specific to the scenario presented in this thesis.

4. What are the effects of having constant low accuracy and low precision measurements during the shock determination analysis?

In the papers of Forslöf and Jones (2015); Lub (2016); Mohan et al. (2008) there is an assumption that the quality of the information will increase if there are more measurements. However, there is also the possibility that the low accuracy and the amount of noise in a large amount of measurements leads to an unusable cloud from which no useful information can be gathered.

5. How can the collected data be structured to fit with existing standards?

Standards can be used to transfer data in a clear and structured way, with extensive documentation on the data structure. Standards also increase the interoperability and make it easier to find new uses for the data in other applications. This subquestion will be answered by performing a literature study.

6. How can the information be linked to an existing asset?

Contractors classify the different objects in their contract area as different assets. By dividing a large area into these different objects or assets, the management of this area becomes more manageable. The connection between an asset and the information enables the asset manager to get the information per asset instead of per position.

1.6 Research methodology and research approach

The methodology used in this thesis is based on iterative design processes that are based on a constant reevaluation of the created product. During this thesis, the iterative design process is executed once instead of the circular process used during the design of a product. Based on the problem statement, scenario and the related work, a research area is defined. This research area refers to the main research question, the subquestions and a short description of the design which is going to be implemented during this research. From the context of use, a research approach is created. This research approach contains a short step by step description of the actions which need to be implemented. During the implementation, the actions described in the research approach are completed through methods that are constantly adjusted and changed for better results. Next, both the implementation and the results it produces are evaluated. The research questions are answered based on the gathered knowledge and the limitations of the research is discussed. In the continuous design process, the next step would be to redefine the context of use. However, in this thesis possible topics for future work are presented instead. These topics can then be used to continue the development of a system which uses crowdsensing as a tool to detect road pavement distresses.

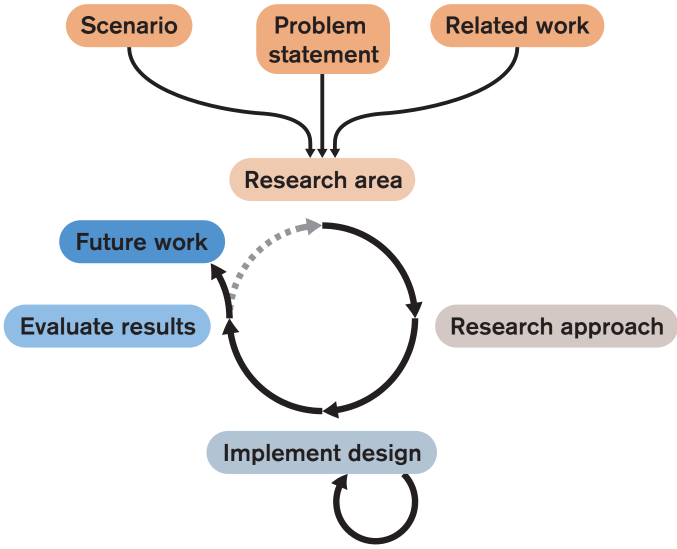


Figure 1.3: Research methodology based on iterative design cycles.

1.7 Scope

This thesis is a feasibility study rather than a design or a real world implementation of an operational system. Practical solutions to problems are be used instead of the most optimal solutions to gather results in a speedy fashion. The scope of this thesis is visualized in table 1.1

To further elaborate on the scope of the research the phases in several smaller scope related subjects are discussed.

- The locational focus of this thesis is the Dutch highway network, which consists in a large part of Zeer Open Asfalt Beton (ZOAB) types (Rijkswaterstaat, 2017c).
- There are many types of road distresses as described by the Asphalt Institute (2017) and the Research & Development division of the Highway Department (2013). However, not all of these types of distress can be detected using smartphone accelerometers as the main sensor for gathering data. Distress types which can be detected are patch failures, potholes, corrugations, delamination, and subsistence near embankments. During the research, there won't be a focus on separating the different types of distress. Instead to focus in on detecting any type of distress and its severity.
- The smartphone is oriented in an arbitrary stable position during the data capture, the owner is not using or touching the smartphone. This is to prevent noise caused by human movements.
- The creation of a custom app to gather and share the data with either a computer or a server is outside the scope of this thesis. Instead, only premade apps are used to gather data. Consequently, this can lead to constraints on the type of information which can be gathered trough these premade apps.
- For this thesis, the local analysis does not take place on the smartphone, instead it is performed on a computer. By performing the analysis on a computer, there is no need to create a

Table 1.1: A MoSCoW diagram showing the scope of the research.

Must	**Should**
Collect data containing multiple road passes through an experiment	Store temporary data in a database
Perform local analysis on a machine separate from the smartphone	Test different apps for usability before data collection
Connect the data to a road pavement asset	Use well documented algorithms for the cluster analysis
Use data from events to create additional information for a cluster or virtual sensor	Document how the data fits with existing standards
Perform database analysis	
Could	**Will not**
Use different car types during data collection	Build an app
Use different phones during data collection	Focus on getting the highest level of quality during the local analysis phase
Store final data in a database	Create a working service to exchange and receive the raw data
	Create a dashboard from which the data can be viewed
	Research an extensive business case for technology adoption
	Apply sensor standards

smartphone app.

- The analysis methods used during the local analysis are not chosen to bring out the optimal result in the field of accuracy, this is to shorten the time spent on the local analysis. On the other hand, there is also the fact that high-quality analysis methods would most likely not be applicable in a smartphone due to battery and memory management (Mohan et al., 2008).
- When transferring a large amount of data, standards are used to increase the interoperability and to provide a clear general documentation on how a system of classes is composed. During this thesis, a theoretical connection between the crowdsensing system and existing sensor standard are discussed.
- As the local analysis takes place on a computer, there will be no actual data transfer between smartphones and a temporary storage point database.

1.8 Thesis outline

This thesis consists of five chapters next to the introduction chapter. chapter 2 Theoretical background & related work contains information on the topics related to the research like Inspection methods for road quality, Smartphone sensors and Mobile Crowdsensing . It also contains information on several research papers which have common grounds with this thesis. chapter 3 Research approach is focussed on an expanded description of the different steps taken during this research. chapter 4 the Implementation contains the implementation of the different steps described in the research approach chapter, as well as an explanation on why certain decisions were made. chapter 5 Results and analysis contains the analysis of the gathered and processed data as well as an evaluation of the quality of the starting data and the end results. The final chapter, chapter 6 Conclusion contains the conclusion of the research as well as te discussion points and possible directions for future work.

2 Theoretical background & related work

Through the years many different techniques have been used to monitor the state of a road network asset. The Theoretical background identifies the current techniques used during inspections in general and in the Netherlands by RWS. Next the different types of smartphone sensors and their capabilities are discussed. Finally the origins and the meaning of the term crowdsensing is defined. The Related work section consists of three subsections describing some of the uses of mobile crowdsensing in transport networks, the role of IMU data in road network monitoring and the role of the combination of IMU data and crowdsensing together in road network monitoring.

2.1 Theoretical background

The theoretical background of this research will be covered in this section. Fist the term glsasset management will be defined as well as the relationship between it and a geo-information context. The different types of inspections used for road quality monitoring are discussed together with the methods which are used in the Netherlands (§ 2.1.2, § 2.1.3). The developments in this digital age have made smartphones into a sensing object carried by a large number of people, providing a potential to use crowdsensing as a tool (§ 2.1.4). The theory of crowdsensing is discussed in § 2.1.5

2.1.1 Asset Management

The main object of interest in Asset Management is, of course, the asset itself. In Asset Management, the asset is described as a physical object which has an economic value and which is owned by an individual or a corporation (Davis, 2013). A single asset can consist of multiple interconnected assets (ISO/TC 251 Asset management, 2014).

Asset management is focussed on the complete life cycle of an asset, from the acquirement to the commissioning, the operating phase, and lastly, the disposal. Asset management is therefore not just the maintenance of an object. The main goal of Asset Management is the realization of the value of an asset through a series of coordinated activities. The realization of the value of the asset is done by balancing the following factors:

- costs
- risks
- opportunities
- performance

(ISO/TC 251 Asset management, 2014)

2.1.2 Inspection methods for road quality

The goal of gathering data on the state of the road is to convert it to some type of profile data which describes within a pre-set accuracy the true profile of that same area. There are three major ways to collect information on the road profile; using static measurement devices, using dipsticks and using IMUs.

Static profilers are also known as rod and level measurement devices (figure 2.1) which can be used to lay out a road (Young, 1989). However, the constraints for collecting usable data for road profiling differ compared to the task of laying out a road. The absolute height of the measuring instrument is of no importance, but the distance intervals for measurements is short at 30 cm or less to perform roughness calculations (Sayers & Karamihas, 1998, page 4). The actual measurement accuracy should be equal or less than 0.5 mm.

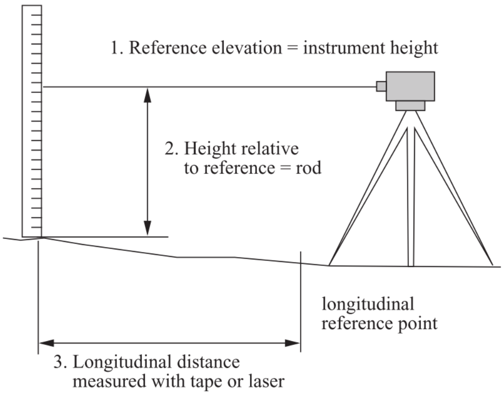


Figure 2.1: The concept of a rod and level (Sayers & Karamihas, 1998).

Dipsticks are devices which contain an onboard computer, allowing them to perform the analysis needed to return a road profile without extra analysis devices (Sayers & Karamihas, 1998). A dipstick uses the previous measurement point as a reference to collect information on the difference in height between the reference point and the new measurement point (figure 2.2). Dipsticks can collect the data needed for a road profile faster than a static profiler, and when the reference height for the first measurement point is set to the same height as static profilers, the resulting road profiles will be similar.

IMU data can be gathered by using inertial profilers, a simple schematic of an inertial profiler can be seen in figure 2.3. Inertial profilers work between the 15 km/h and the 100 km/h. The upper bounds of the profile speed are due to the decreased contact between the profiler and the road at higher speeds. Inertial profilers are capable of being deployed at the speeds needed to drive over a highway without causing major danger to other drivers. The downside to using Inertial Profilers is that different runs across the same road section will lead to a shift in location due to GNSS inaccuracies (Fujino, Kitagawa, Furukawa, & Ishii, 2005).

Next, to the different types of profilers, there are also different ways to quantify the road quality from the gathered data. the most important ways are the IRI and the Ride Number (RN). The IRI is calculated from the accumulated suspension movement from The Golden Car over a profile (Bridgelall, 2014). The Golden Car is a quarter car model consisting of pre-set values for the sprung mass m_s, unsprung mass m_u, the spring rate k_s, the tire spring rate k_u and the damper rate c_s (figure 2.4).

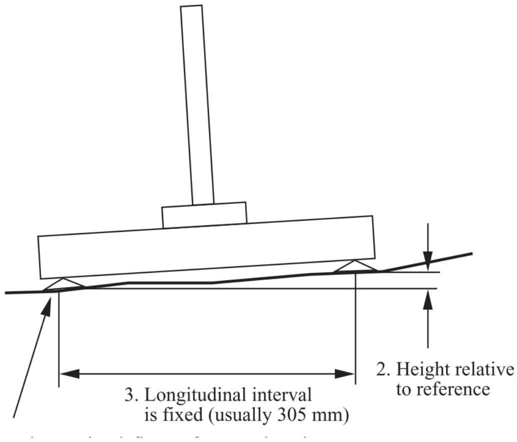


Figure 2.2: The concept of a dipstick (Sayers & Karamihas, 1998).

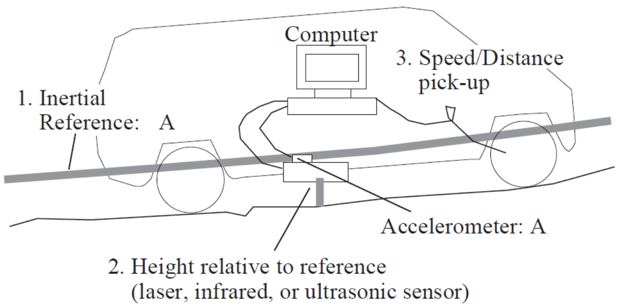


Figure 2.3: The concept of an inertial profiler (Sayers & Karamihas, 1998).

A quarter car only takes into account the suspension from one car wheel. In the case of The Golden Car, the values for the variables are chosen to exhibit behavior shown by most cars which use the highway with the exception of a higher damper rate.

The speed for an IRI calculation is set to 80 km/h due to the fact that the results of the IRI are speed dependent (Gillespie, 1981). By setting the speed for the IRI to 80 km/h, the comfort level in a car at normal driving speed on a highway can be calculated. RWS makes use of the Half-car Roughness Index (HRI) to measure comfort in the Netherlands. The IRI is closely related to three vehicle response variables: the road meter response which is used for historical continuity; the vertical passenger acceleration which is used to determine the ride quality and the tire load which is used for vehicle safety (Sayers & Karamihas, 1998).

RN is a way to quantify the quality of the road on a preset subjective scale between 0 and 5, with 0 being poor quality and 5 being a very good quality. This scale is based on the Present Serviceability Index (PSI) created by the American Association of State Highway and Transportation Officials. The RN itself was created based on research performed by the National Cooperative Highway Research Program in the 1980's. The system uses a quarter car model similar to the one used to calculate the IRI value, however, the RN uses an estimation of the Mean Panel Rating (MPR) to calculate the value

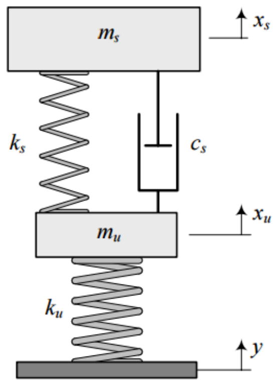


Figure 2.4: The concept of a quarter car (Jazar, 2008). For the Golden Quarter Car the values of the parameters are set to reflect the behavior typical of most highway vehicles (Sayers & Karamihas, 1998).

on a scale between 0 and 5. The MPR is the mean of the opinion of a group of people (the panel) on the state of the road; everyone in the panel rates the quality of the road between 0 and 5 from which the mean is calculated, resulting in the MPR. The relationship between the profile information and the RN is extremely non-linear, limiting its use. An example is that the RN for two consecutive road sections cannot be directly averaged (Sayers & Karamihas, 1998)

2.1.3 The current state of road network asset inspections in the Netherlands

RWS has defined different types of inspections which are performed to monitor the state of the Dutch road network assets (Rijkswaterstaat, n.d.; Schultz van Haegen, 2016). Depending on the type of road network asset, different types of inspections are performed. Generally, there are 3 types of inspections: the daily inspection, the 'toestandsinspectie' and the 'visuele inspectie'. The daily inspection pertains a drive-by visual on the traffic flow and general safety of assets. The 'toestandsinspectie' is a yearly activity to check less often used assets. The 'visuele inspectie' is the least often performed inspection, it is performed every 6 years. However, road pavement assets follow a different schedule of inspections (table 2.1). Note that the three most often performed inspections are all visual. The ARAN (figure 2.5) takes up a large part of the non-visual inspections. The profile information gathered by the ARAN is converted to HRI values instead of IRI values. The HRI takes the average suspension movement of two wheels, however RWS is researching whether they can move from HRI to IRI (Lub, 2016, Annex I, A4).

Table 2.1: An overview of the inspection types for road pavement in the Netherlands. (Klunder et al., 2010)

Subject of inspection	Method	Frequency
Roughness	Road Analyser and Recoder of Norsemeter (ROAR),Skidding wheel	(bi)yearly
Rutting	ARAN-3, Laser height profile	(bi)yearly
Longitudinal plane and transversal plane	ARAN-3, Laser height profile	(bi)yearly
Cracks	ARAN-3, video images +detection software	(bi)yearly
Ravelling	Visual	yearly
Crack formation	Visual	yearly
Local defects (holes)	Visual	daily

Figure 2.5: The ARAN of RWS in action. (Rijkswaterstaat, 1996)

2.1.4 Smartphone sensors

Smartphones are devices which have had a rapid development in the last years, increasing both their sensing and computing capabilities. This section will contain information on the different types of sensors smartphones can use to collect data.

The most common sensors are the accelerometer, gyroscope, and magnetometer (named as a compass in some cases) which are part of the IMU, GNSS receivers, microphones, cameras, and proximity sensors to detect whether a phone is near an ear (Ganti, Ye, & Lei, 2011; Kos, Umek, & Tomaic, 2016). Next, there are light sensors which are commonly used to set screen brightness, temperature sensors, and barometers (Android Open Source Project, 2017b; Ganti et al., 2011). Smartphones can also contain sensors of the same type with different set-ups or ranges, which enable them to gather data on different phenomena, examples are accelerometers which can measure acceleration and accelerometers which can measure linear acceleration, linear acceleration does not contain a gravity component.

Each sensor is associated with its own unit of measurement. For each of the sensor types this is:

Table 2.2: The different types of sensors in smartphones, what they measure and their unit of measurement (Android Open Source Project, 2017b).

Sensor Type	What is measured	Unit of measurement
3 axis Accelerometer	Acceleration on X,Y and Z	G force or meters/second^2
3 axis Gyroscope	Rotation around X,Y and Z	radians/second
3 axis Magnetometer	Magnetic field around X,Y and Z	microTesla
GNSS reciever	Latitude and Longitude ,in WGS84	degrees
Microphone	Sound	dB
Camera	Colour intensity	RGB radiance
Proximity sensor	Distance between device screen and object	centimeters
Light sensor	Light intensity	Lux
Temperature sensor	Temperature	degrees Celcius
Barometer	Air pressure	hectoPascal or millibar

2.1.5 Mobile Crowdsensing

The term mobile crowdsensing is proposed by Ganti et al. (2011) to cover mobile devices which can be used to collect and process sensor data. The main difference between devices usable for mobile crowdsensing and Internet of Things (IoT) devices is that crowdsensing has the capability to process the sensor data locally. Another aspect of mobile crowdsensing is that it is community-based and is focussed on monitoring large scale phenomena like traffic or weather. Ganti et al. (2011) identify two subtypes of mobile crowdsensing; participatory and opportunistic sensing (figure 2.6). Burke et al. (2006) defines the term participatory sensing and the architecture needed to facilitate an interactive and participatory sensor network where the sensors are always under the control of their own user. Opportunistic sensing is the ability to measure a target area within a time window of interest based on uncontrolled mobility, for which no direct user interaction is needed (Campbell, Eisenman, Lane, Miluzzo, & Peterson, 2006). In opportunistic sensing, the time at which the target area is sensed cannot be predicted.

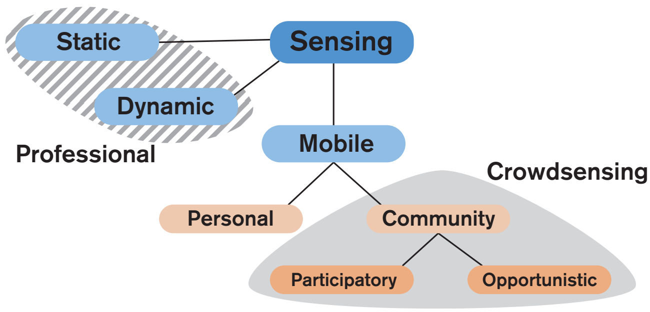


Figure 2.6: Crowdsensing is a part of mobile sensing, it is community driven and contains both participatory sensing and opportunistic sensing.

2.2 Related work

Multiple papers describing ways of monitoring both roads and other transport networks are summarized in this section t give insight into other research projects in similar areas of interest.

2.2.1 Crowdsensing in transport networks

There are many applications for crowdsensing and the user as a sensor. This section will shortly describe some of the research done in the area of automobiles, positioning and participatory mobile crowdsensing. Jin, Han, Liu, and Feng (2015) use participatory mobile crowdsensing to identify unwanted behavior of taxi drivers, namely Taxi Passenger Refusal (TPR), which is described as the act of refusing passengers after enquiring and knowing the traveling destination considering the economic benefit. On the other hand, Renganathan and Velaga (2016) use data from cars which have driven across a road to inform the cars behind it about the road condition. Qin et al. (2015) use participatory mobile crowdsensing to gather information about obstacles to improve spatial services for disabled people. An application created by a company called Sentiance monitors the behavior of road users to either lower or heighten their car insurance based on their driving style (Spruyt, 2016).

The research of Jin et al. (2015) is focussed on combining the automated gathering of Positioning data with passenger experiences. The act of a TPR can be described as the nearing of the taxi to the potential passenger, next the taxi slows down. If the taxi proceeds to drive away from the potential passenger, a TPR has occurred. TPR is a problem in large cities, however, the threshold of reporting a TPR too high for passengers and they often do not bother reporting it. Gathering proof of a TPR is also difficult for authorities. Jin et al. (2015) use the GPS data from both the taxi and the potential passenger and analyze both to detect abnormal behavior. By using this information together with Human Intelligent Tasks, the feedback from denied passengers can be connected to the evidence of the taxi driving away.

Renganathan and Velaga (2016) describe a theoretical system in which information from cars driving on a road can be shared with drivers who use the same road at a later time, informing them about the quality and condition of the road (figure 2.7). The research makes use of the Internet of Vehicles (IoV), an offshoot of the IoT where vehicles are able to constantly communicate with each other. The data is gathered by rash driving sensors, Anti-lock Braking System (ABS) deployment sensors, orientation sensors and the positioning system in the car. This data is sent to a server where it is analyzed to gather information about the quality of the road, the information is then shared with all the other cars in the area. The main focus of Renganathan and Velaga (2016) is on the identification of valid data; the data gathered from the rash driving sensor might not always indicate a bad road quality, but a bad driver.

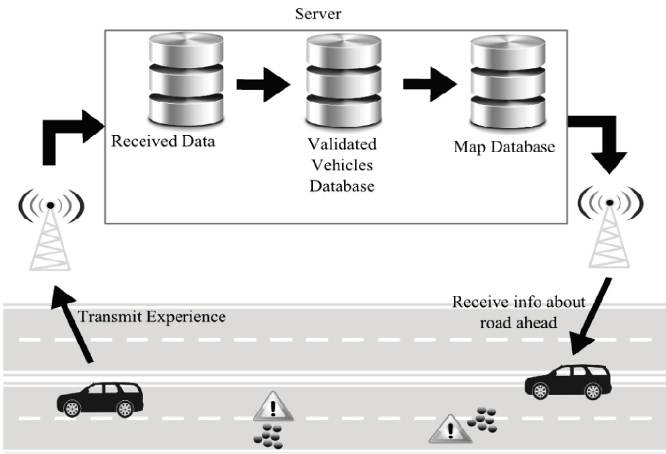


Figure 2.7: Design of a system to share passenger experience with other cars on the road. (Renganathan & Velaga, 2016)

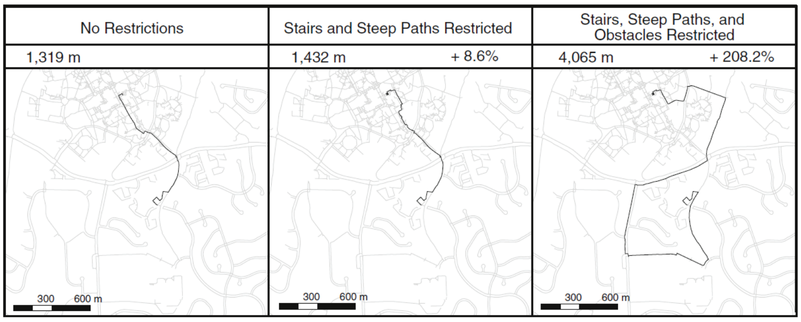


Figure 2.8: Different possible routes to a location on the George Mason University campus, Fairfax based on different movement constraints (Qin et al., 2015).

Qin et al. (2015) use participatory mobile crowdsensing to improve accessibility and routing services for disabled people. The test location used in the research is the George Mason University Campus in Fairfax, USA. One of the problems for disabled people is that they are dependent on static accessibility maps while many obstacles are transient in nature. These transient obstacles change the pedestrian routes in a day to day way, making it difficult to have an up-to-date accessibility map of the campus. By using up to date information provided by students and people on the campus, the accessibility maps can be more temporally accurate. Volunteers can mark locations which are blocked by obstacles, they can add a description as well as an estimation of duration of the obstacle. These obstacles are used to redefine the network used for routing. Different networks are created to accommodate people with different disabilities (figure 2.8). Qin et al. (2015) notes that one of the other possible applications is the identification of areas which are highly inaccessible. However, there are also downsides of relying heavily on the input of volunteers, namely the quality and reliability of the information provided by the volunteers.

Sentiance, a company specialized in machine learning, developed an application to monitor the behavior of drivers and to enrich this information with semantics e.g. where the driver has been before getting in the vehicle (Spruyt, 2016). The goal is to enable insurance companies to give out insurance policies based on behavior. The result would be that the road users who keep to the rules will get a lower monthly cost on their policy, while dangerous drivers get a higher monthly cost. The system works by detecting things like in-car phone handling, braking behavior, speed and centrifugal force. The parameters are then converted to a percentage number for different categories (Spruyt, 2016). The categories are efficiency, anticipation, phone usage and legal. The efficiency is calculated from the magnitude and duration of events like acceleration, braking and the centrifugal force in turns. The anticipation is based on how smooth the driver drives and how the driver reacts to for example to traffic lights. Phone usage describes how much time the driver spends handling his or her phone. Lastly, the legal category represents the number of speed violations. The information is presented through a dashboard, which gives an overview of a person's driving behavior (figure 2.9).

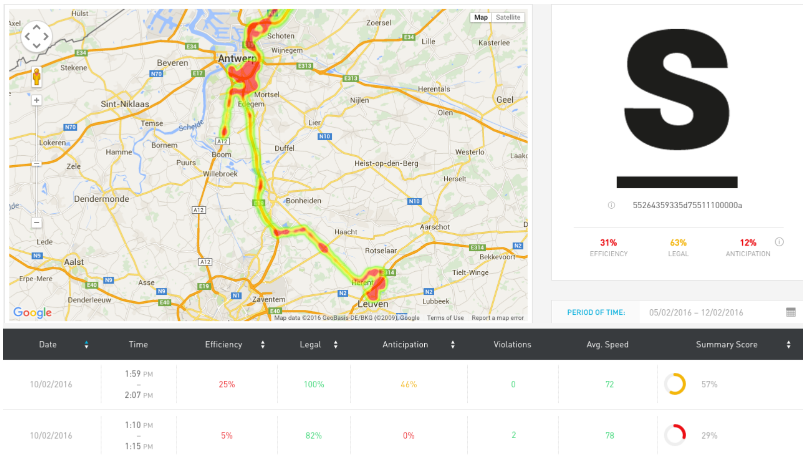


Figure 2.9: The dashboard which shows the information on the different categories used by Sentiance for a single driver (Spruyt, 2016).

2.2.2 Crowdsensing and IMU data in road network monitoring

Crowdsensed IMU data can be gathered from either smartphones or directly from the car when working together with manufacturers (Waite, Walsh, & Garcia, 2012). The usage of lower costs of devices is an advantage while the lower precision data is a disadvantage. There have been multiple researchers who have studied the applicability of these devices for several road-related applications.

Fujino et al. (2005) showed a relation between the measurements from IMUs mounted inside the car and the geometry of the road. This research confirmed that the IMU data measured from inside the car can be used as a way to gather information on the road below, even though there is a complex system of suspension between the measurement device and the area of interest.

Researchers at Microsoft developed the Nericell system based on this relationship between accelerometer data and the state of the road. The focus of the system is to use sensors in chaotic traffic situation to detect honking, breaking and bumps (Mohan et al., 2008). Data was only stored and processed after an event like breaking or honking took place. Virtual reorientation was then performed on the collected data to account for the different smartphone orientation. Multiple measurements near a single location would indicate something like a bad traffic flow, or bad road pavement.

Karamanou, Papazissi, Paradissis, and Psarianos (2009) used the relationship between IMUs inside the passenger area of the car and the geometry of the road to gather more extensive information on road centerline profiles for mapping purposes. Karamanou et al. showed that an GPS/IMU system mounted inside the passenger area can be used to calculate the vertical profile of large stretches of road. The developed system was used to analyze 600 km of road network in Greece. The results showed that the vertical accuracy was on average around 10 cm, though there is no mention of compensating for the car suspension.

Furthermore, there is the Roadroid project, which has been in development since 2002 (Forslöf & Jones, 2015). The researchers have developed an application for Android smartphones which can be used to calculate the IRI of a stretch of road and a classification of the state of the road. This information can then be viewed in an online portal (figure 2.10). The use of smartphone sensor data as a support system to high-quality, high-cost data acquisition was also defined as an area of potential for the technology. However, Forslöf and Jones (2015) also notes that the technology can be used to give early warnings of changes, provide constant monitoring possibilities and trend description. Road roughness calculated from smartphone data cannot replace high precision profiles, but it can be used to give subjective ratings of road comfort.

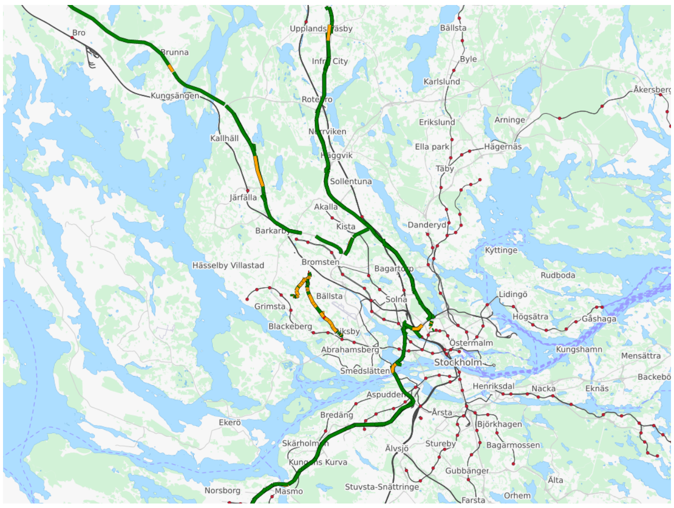


Figure 2.10: A screenshot of the working Roadroid map interface (Roadroid, 2017).

There is also the StreetBump project, which was developed for use in Boston, United States of America (Carrera et al., 2013). The city of Boston was interested in the project due to the possible cost reductions by detecting potholes early. Carrera et al. (2013) use an old saying to describe the benefit of early repairs:

> *'an 'ounce' of prevention is worth a 'pound' of cure'*

The project has maritime roots and was created to measure the turbulence caused by motorboats in the canals of Venice, Italy. Now the StreetBump application created during the project can be used to detect potholes through opportunistic crowdsensing; users install the app and have to start it when they start a journey by car, but they do not have to deliver any extra input. The project distinguishes between street bumps and normal smartphone usage through one of the powers of crowdsourcing: the chance of having multiple people using their smartphone in the car on exactly the same location is mitigated by the larger amount of people not using their smartphone. When a bump is detected, its GPS location and accelerometer information are sent to the server. The next step in the project would be to distinguish whether measurements close to each other ar similar and if they are, to increase the importance of these measurements (figure 2.11).

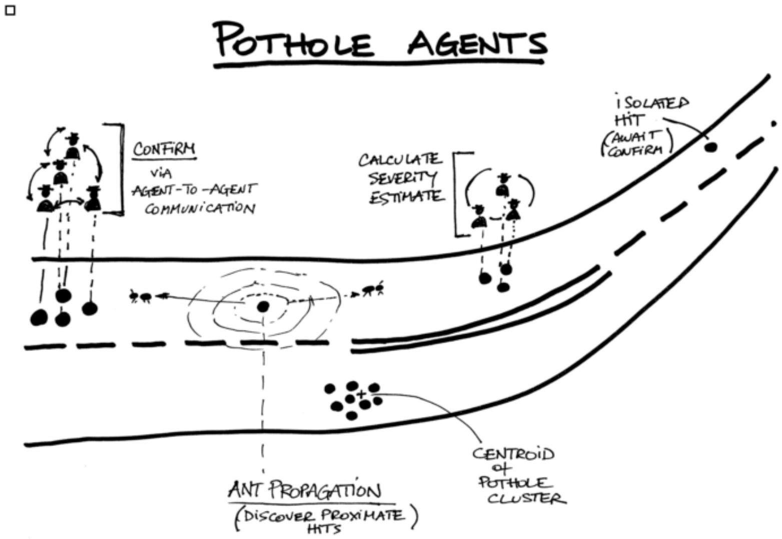


Figure 2.11: A concept drawing for the next step in the development of the StreetBump application (Carrera et al., 2013).

Finally, research at the University of Twente showed that the information gathered by smartphones is not accurate enough to replace professional measurement systems like ARAN (Lub, 2016). However, smartphone technology can be used to fill the gaps in between the high precision inspections, providing a more constant insight into the current status of road networks.

2.2.3 Other technologies in road network monitoring

There are also other innovations in the monitoring of roads. In this subsection applications used to monitor the state of a road using photogrammetry and Persistent Scatter Interferometric Synthetic Aperture Radar (PS-InSAR) are discussed. The usage of photogrammetry to monitor roads is supported by the increasing popularity of Unmanned Aerial Vehicle (UAV)'s, making the hardware needed to gather data cheaper and the technology a more viable option. The advantage of PS-InSAR is that the data is readily available through services and that the time interval between two satellites is $\pm$ 6 days depending on the system (ESA, 2017). The short interval increases the potential that the data can be used to monitor the road pavement.

Photogrammetry

Aerial photogrammetry is not commonly used to gather data about the state of the road network. This is due to the lower accuracy of a point cloud generated from photogrammetric data (Díaz-Vilariño, González-Jorge, Martínez-Sánchez, Bueno, & Arias, 2016). However as shown by Díaz-Vilariño et al., there are uses for photogrammetric imaging. The decrease in costs due to the developments in the field of Unmanned Aerial Vehicles also makes the technique interesting: a photogrammetric survey can be automated by designing a flight plan the UAV will adhere to and consumer grade cameras can be used to collect the images (Remondino, Barazzetti, Nex, Scaioni, & Sarazzi, 2011).

Díaz-Vilariño et al. used UAVs to collect data on a road section in Italy. The goal was to test the usage of an Unmanned Aerial Photogrammetric System (UAPS) for calculating road run-off. The images collected by the UAPS were converted to a DEM from which the road run-off direction was calculated by using the commonly used D8 algorithm. It was concluded that data from UAPS data with cell resolutions higher than 2 meters, will return the same result as data from a LIght Detection And Ranging (LIDAR) system 84% of the time. A downside of using photogrammetry in combination with a UAV is that there are stick rules on where a UAV is allowed to fly, as well as the limited battery time and the small weather window suitable for operations (Díaz-Vilariño et al., 2016).

2.2.4 PS-inSAR

PS-InSAR is a technology which uses an archive of satellite images to determine deformation based on changes between the individual images or an image series though time (Gehlot, Ketelaar, Verbree, & Hanssen, 2005). PS-InSAR can be used to monitor changes to surfaces, and one of its main advantages is that it can monitor large areas with a high temporal resolution compared to more conventional techniques, due to the constant data gathering from satellites. PS-InSAR makes use of Persistent scatter (PS) points which are selected based on their amplitude, phase and time series coherence, these points have a high temporal consistency and are often connected to man-made buildings (Wu & Hu, 2015). These PS points are used to cope with incoherences caused by changing landscapes through the seasons among other things. As Wu and Hu show, PS-InSAR can be used to quickly detect areas of the road network with high subsistence (figure 2.12). However as Gehlot et al. note, the deformations registered with PS-InSAR have different causes. The correct identification of the cause of subsistence has an impact on the meaning of a piece of data. PS-InSAR has not yet been used to detect road pavement distresses.

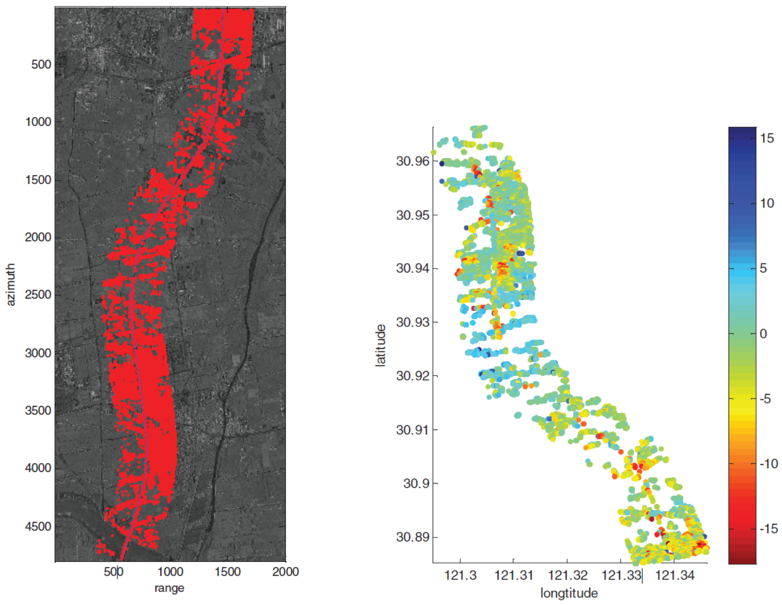


Figure 2.12: The selection of Persistent Scatter points (left) and the settlement results(right). (Wu & Hu, 2015)

3 Research approach

In this chapter, the research approach will be discussed. First, the real world context in which a system which uses crowdsensing as a tool to detect road pavement distresses is further elaborated on (§ 3.1). From this information, the steps of the research approach are defined further in the successive sections and subsections of this chapter.

3.1 Context of use

A system which gathers data through crowdsensing relies on a large group of people who are prepared to provide data. In this case, the people consist of drivers who are traveling to their destination. Each individual driver collecting data on the z acceleration their smartphone is experiencing in the car cabin (figure 3.1). From this constant stream of data, only the high peaks and low valleys are of interest as these describe possible road pavement distresses. These high and low values describe the occurrence of events. The events of all cars are grouped together based on their position. Next, the position of these groups are linked to positions like specific highway sections, these sections contain additional information like the maximum speed. The maximum speed can be used to calculate the weighted averages of all the events in a group for, for example, the maximum and minimum z acceleration. This grouped information can then be used to give information about a highway location (figure 3.1).

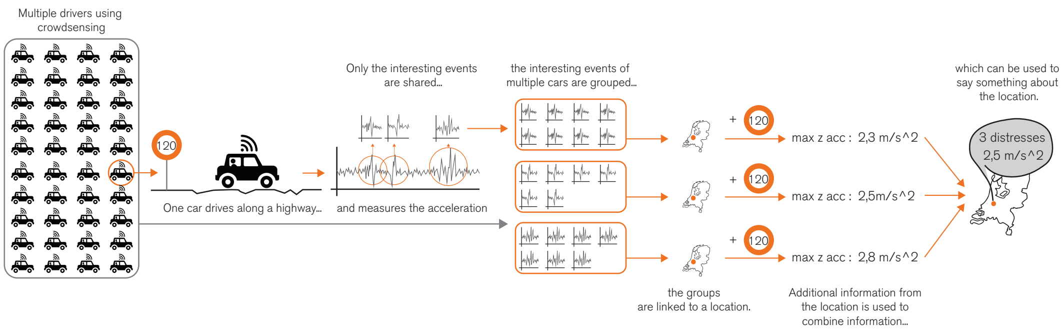


Figure 3.1: The context of use for the crowdsensing system during this research.

From the context of use described above, the research approach can be defined in phases and steps (figure 3.2,figure 3.3,figure 3.4). The phases of the research approach are based on the Geographical Information System (GIS) chain defined by Lemmens (1991) and the definition of a system which uses crowdsensing by Ganti et al. (2011).

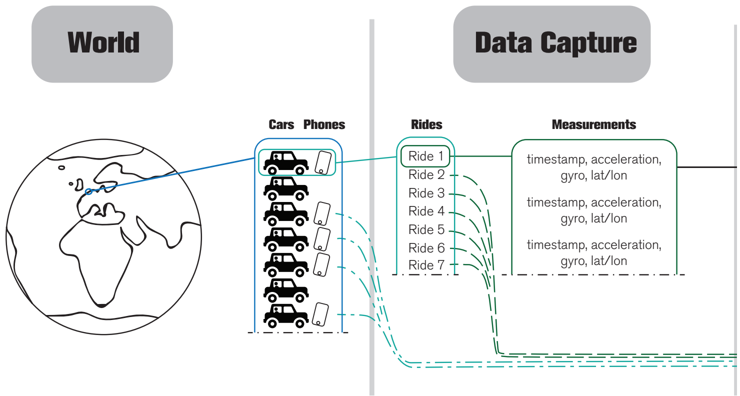


Figure 3.2: A flowchart of the research process, part 1.

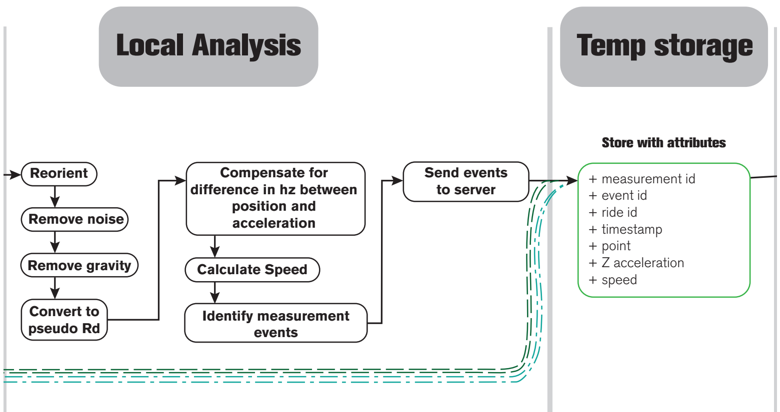


Figure 3.3: A flowchart of the research process, part 2.

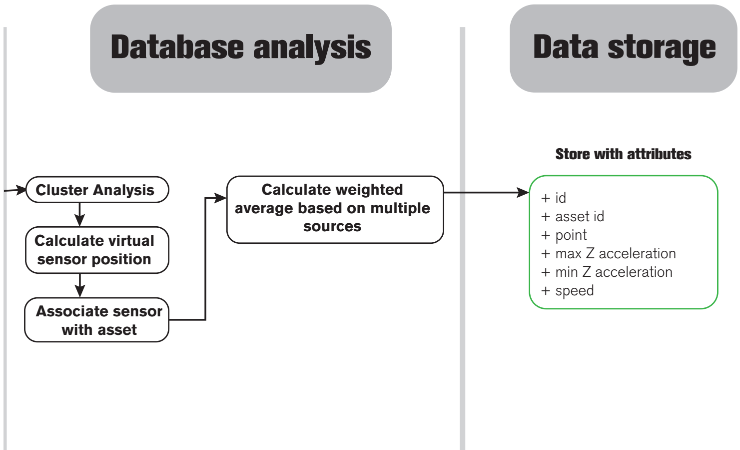


Figure 3.4: A flowchart of the research process, part 3.

3.2 Data capture

Before any data can be gathered on the state of the pavement of a road network asset, the requirements for the raw data need to be set. This is covered with the sub-question: *What data and information are needed to monitor a road pavement asset?* These raw data requirements need to be set before measuring to prevent a misalignment between the available data and the requirements from the methods which will be used to calculate the state of the road pavement.

The data gathering phase consists of choosing locations based on the existence of road pavement distresses. The data will be gathered by an app on a commercial smartphone with the basic assumption that distresses like potholes cause the car driving over it to vibrate (figure 3.5). This vibration is then measured by the smartphone. During the research, a premade app will be used which provides the raw sensor data.

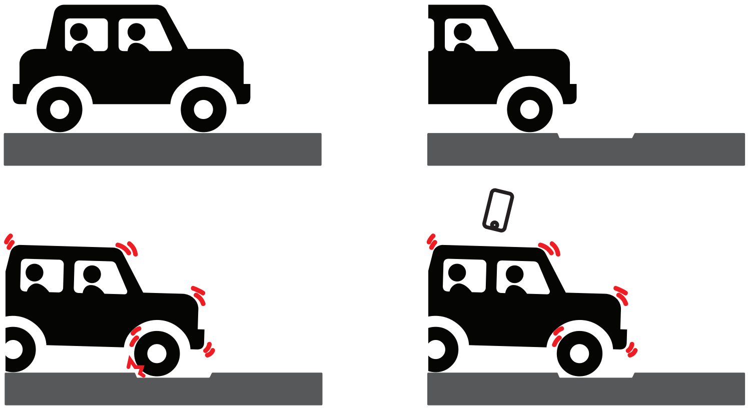


Figure 3.5: A car is driving along a highway when there is suddenly a pothole or some other form of distress. This distress makes the car vibrate. This vibration is measured with a smartphone.

3.3 Local analysis

After data has been gathered, a local analysis would take place on the smartphone (figure 3.6) (Ganti et al., 2011). The steps in this phase are performed on a local computer. In a real world implementation of a crowdsensing system, the local analysis would be performed on the smartphone.

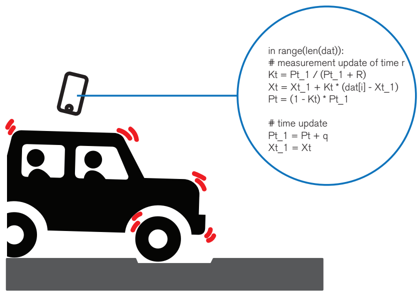


Figure 3.6: Using the processing power of the smartphone, calculations on the data can be performed directly by the device.

Using reorientation to compensate for data gathered by a smartphone in an arbitrary orientation

One of the features of using a smartphone as a measurement device is that its body frame can be moved around and rotated. This means that all individual smartphones have an arbitrary orientation. These differences in orientation cause the vertical acceleration to be measured in different or multiple body frame axes (figure 3.8, figure 3.9).

To combine the data streams from different devices, a transformation from their arbitrary body frame to a local reference frame is necessary (figure 3.7). An often used method in positioning is the use of the IMU and the gyro measurements to create a Direction Cosine Matrix (DCM) (Ibrahim & Moselhi, 2016; Renaudin & Combettes, 2014; Zheng, Han, Yue, Yuan, & Wo, 2016). This DCM can be applied to the IMU measurements to transform the accelerometer data to the same local reference frame.

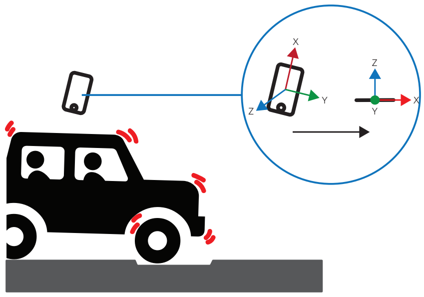


Figure 3.7: The body frame of the smartphone needs to be transformed to a local reference frame to enable the comparison between different measurements.

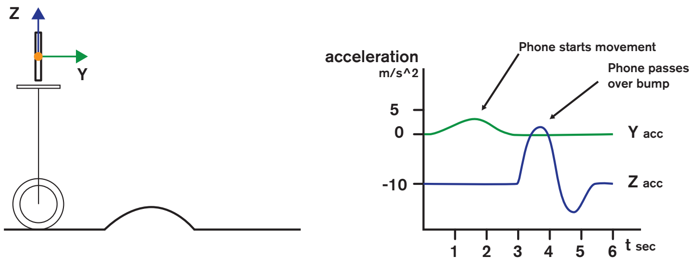


Figure 3.8: Conceptual acceleration data without noise from a smartphone oriented perpendicular to the earth's surface. Road geometry can be seen on the left, the conceptual acceleration measurements are on the right.

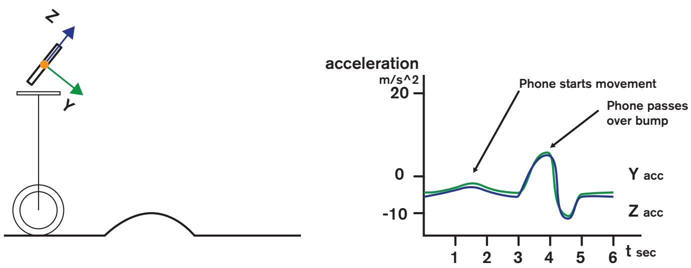


Figure 3.9: Conceptual acceleration data without noise from a smartphone tilted approximately 45 degrees. Road geometry can be seen on the left, the conceptual acceleration measurements are on the right.

Removing noise from the accelerometer data

Every sensor experiences noise, for the application used during this research the noise is caused by the small, constant vibrations in the car cabin. This noise needs to be removed to enable the recognition of events in a later step (figure 3.10). A low pass filter can be used for noise removal and is a relatively simple calculation compared to other options (Bruwer & Booysen, 2015; Cízek, Wolfgang, & Weron, 2005; Grewal, 2011).

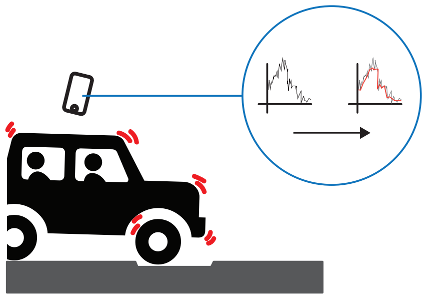


Figure 3.10: Smoothing the signal received from the smartphone sensors by removing noise.

Removing gravity from the accelerometer data

An accelerometer in an IMU constantly measures the gravity the earth is exerting on it (Sayers & Karamihas, 1998). The gravity constant needs to be removed o ensure that it does not affect the measurements in the Z-axis(the vertical acceleration) (figure 3.11). The linear acceleration (acceleration without the gravity constant) can be calculated by applying a high pass filter with a low-frequency threshold (van der Heijde & Ligtvoet, 2017).

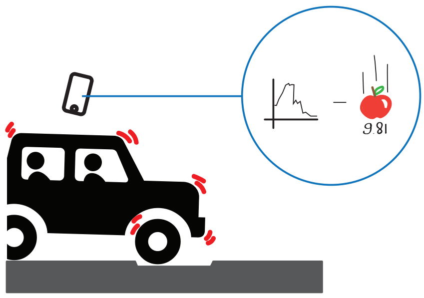


Figure 3.11: Removing the gravity constant from the accelerometer data.

Converting from WGS84 coordinate system to a local coordinate system

Most measurements gathered by a smartphone IMU can be analyzed to get information in the metrical system. Meanwhile, the position calculated by the smartphone GNSS positioning uses latitude and longitude. When combining for example coordinates and the speed to calculate a better estimate of the position, both need to have the same unit of measurement. Because the locational focus of the research is the Netherlands, the latitude and longitude of the WGS84 system are transformed to the Cartesian coordinates system of EPSG28992/Amersfoort RD-new (figure 3.12). Cartesian coordinate systems use meters as their measurement unit.

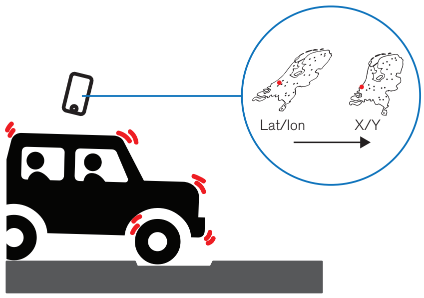


Figure 3.12: Converting between the WGS84 coordinate system and the RD-New coordinate system.

Compensating for the difference in measurement frequency between the GNSS and the accelerometer

The frequency with which a smartphone calculates its position based on the signal of the GNSS is lower than the frequency with which other sensors collect measurements. The result is that multiple measurements from a fast sensor all have the same GNSS position. The problem is that this position is not accurate for all these measurements. By combining the GNSS position with the accelerometer data, a more current position can be calculated for the times that there is no new GNSS position (figure 3.13).

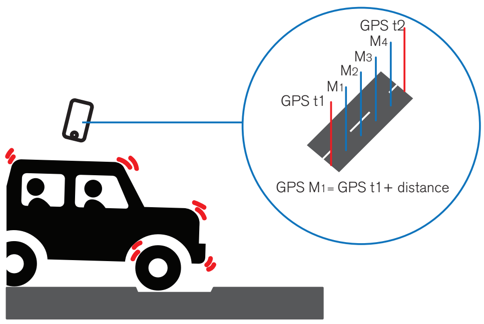


Figure 3.13: Using the last GNSS positional information to calculate a new position for the measurements at M1.

Calculating the speed for each measurement based on the GNSS position

The speed at the time of each measurement can be used to gather insight into the relevance of the vertical acceleration. As shown in figure 3.1 the speed of the road can be used to give the asset manager context to the information. In turn, the speed can be used to select which information is relevant. The experience of a car driving at 80 kilometers an hour is different compared to a car driving 120 kilometers an hour. When driving along a road where the maximum speed is 120 km/h, the experience of the second driver is arguably more indicative of the dangers of the road distress. The positional information between multiple points is used to calculate the average speed (figure 3.14). By averaging the speed across multiple points, possible speed errors due to bad positioning are given less influence.

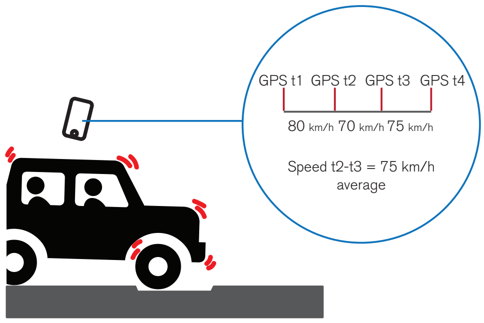


Figure 3.14: Using the last GPS positional information to calculate the average speed.

Identifying measurement events

To limit the amount of data which needs to be transferred to a temporary storage database, a distinction between interesting data and uninteresting data has to be made. The distinctions between interesting data and uninteresting data can be seen as the difference between an event happening and nothing taking place. Filters with pre-set tolerances can be used to remove data which does not describe the events of interest (figure 3.15).

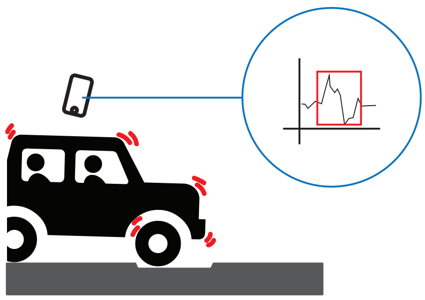


Figure 3.15: The area of interest in the measurements. Two tolerances can be set to remove all data which is not inside this area of interest.

3.4 Saving the measurement events to a database

It is beneficiary to use standards when creating class structures before transferring the data. Standards increase the interoperability and reusability due to their well-documented class structures which can always be consulted by other parties. These standards also dictate what type of information will be stored and in what way the ordered information is received by the database (figure 3.16). Sensor standards are implemented on a theoretical level during this research.

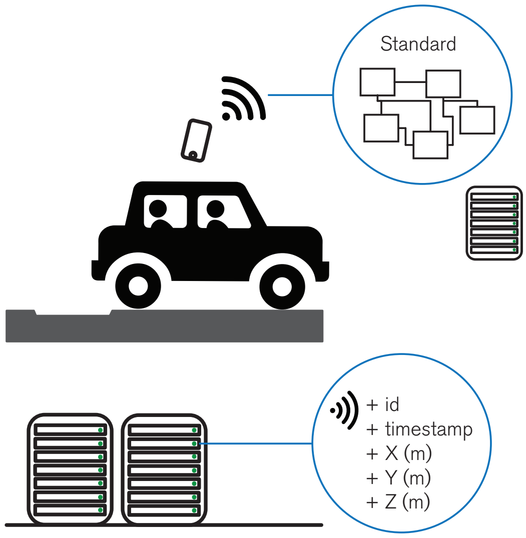


Figure 3.16: Transferring the information while adhering to the structures designed in open sensor standards.

3.5 Database analysis

The database analysis phase contains the two major actions of connecting the measurement information from multiple sources to an asset, and to convert the measurement information from multiple sources into a single measurement. Other steps are made to facilitate these two actions.

Performing a cluster analysis on the measurement events

The events detected on a highway can describe multiple instances of road pavement distress. A distinction between which events describe the same instance of road pavement distress and which do not describe the same road pavement distress is needed. To achieve this separation and grouping of events, a cluster analysis is performed (figure 3.17). There are different types of cluster analysis, an algorithm based on both the distance between events and the density of events is used.

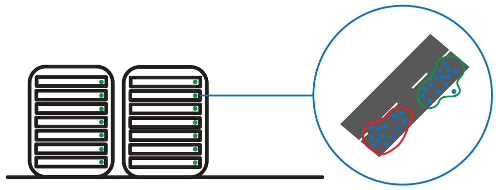


Figure 3.17: Making clusters from sensor measurements events from different cars.

Calculate virtual sensor positions

Each group created by the cluster analysis in the previous step contains multiple positions based on the event data. Now a single position needs to be calculated from all the input positioning values (figure 3.18) to create a position for a virtual sensor. The concept of the virtual sensor will be discussed in the Implementation chapter. The approach used during this research is to calculate the average position for each event. Next, the average positions from all events together are used to calculate the position of the cluster.

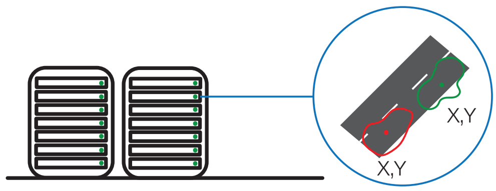


Figure 3.18: Combining the positions of the different sensor measurement events to calculate the position of the virtual sensor.

Connecting the virtual sensor to a road pavement asset

Both an asset and the information from the virtual sensor can be seen as separate classes with their own attributes. These two classes now need to be connected to allow a user to access the information related to an asset in a speedy fashion (figure 3.19). The asset geometry type and its attributes need to be established before measurement information can be associated with it. The next step is to make a connection between the virtual sensor and the road pavement asset based on both nearness and direction of the road pavement asset. The connection between the two objects needs to be stored in either the virtual sensor or the road pavement asset.

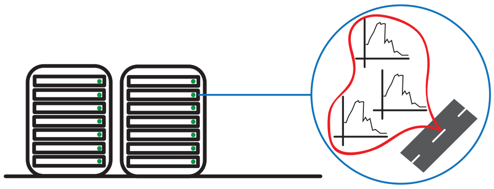


Figure 3.19: Associating several instances of sensor measurement information to an asset.

Calculate the most likely acceleration values based on multiple measurement events

The final step in the database analysis phase is to calculate the acceleration values from the events inside the virtual sensor (figure 3.20). There are multiple methods to perform the calculations, though a simple calculation might lead to a less accurate result. The measurements inside the event can be weighted based on the speed of the event compared to the maximum speed allowed along the asset.

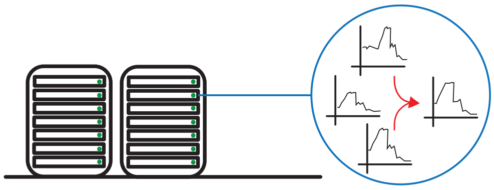


Figure 3.20: Combining the sensor measurement events in a cluster into a shock determination event.

3.6 Database storage

The final phase to be implemented is the database storage where the information is saved together with its connection to the road pavement asset. As the database analysis as a whole already takes place in a database, the resulting information from the previous step can be seen as the final stored product.

4 Implementation

In this chapter, the implementation of the research approach is discussed. The first two sections are theoretical designs describing what data and information are needed during this research (§ 4.1) and how the data and information can fit with existing standards (§ 4.2).

The other two sections (§ 4.4, § 4.6) are focussed on the technical implementation during this research. The contents of these sections give a partial answer to the subquestion:

3. Which factors influence the accuracy of a detected event?

The subquestion on the subject of how to connect information to an asset is discussed in § 4.6.3.

4.1 Defining the types of data and information

There are three main factors that influence what types of data and information are needed to monitor the road pavement: the capabilities and limits of the measurement device, the phenomenon that will be measured and what type of information needs to be provided to the asset manager. These phenomena can be used to answer the first part of the sub question:

2. What data and information are needed to monitor a road pavement asset?

4.1.1 Raw data types gathered by smartphones

First, the capabilities and limits of the measurement device, a smartphone in this scenario, will be defined. The different research papers mentioned in chapter 2 show that the most commonly used sensor data for this type of application are accelerometer data, gyroscope data and GNSS data. These types of sensors are almost always available in a smartphone, as they play an important part in supporting regular smartphone capabilities, e.g. screen rotation, positioning and navigation applications. In some cases, data from the magnetometer is also used, however, this data will not be used in this research due to the possible proximity of the smartphone to the different types of electrical devices in the car causing major interference. The sample rate for smartphones running on the iOS operating system is set to a maximum of 100 Hz, while the sample rate on Android devices is not limited in the same way. Instead, the speed depends on the type of sensor and the sensor framework (Android Open Source Project, 2017b; Apple Inc., 2016).

4.1.2 Common methods to measure the state of the road and their applicability to the scenario

The phenomenon that will be measured are road distresses like patch failures, potholes, corrugations, delamination, and subsistence near embankments (§ 1.6 Research methodology and research approach). These types of road distress have in common that there is always vertical change occurring on the pavement and that the vertical change is always transversal or near transversal to the driving direction. Vertical changes can also be described as a displacement in the continuity of the road pavement. This displacement in continuity is quantified by the IRI and qualified by the RN. However, the IRI cannot be used to describe only the displacement caused by a single event, e.g. a hole because the IRI is a calculation of the total displacement over a constant length. The possible differences between for example potholes can lead to problems. Another downside of using the IRI in the case om measuring only an event is that the outcome of the IRI describes the distress in its entirety. This attribute is beneficial in measuring and quantifying large stretches of a road network because it is a way to summarize a large amount of data for human interpretation while at the same time decreasing the amount of data which needs to be stored. However, when the goal is to monitor how the distress develops, a single number describing that same distress is too aggressive of a summary; there is a chance that different shapes of road distress can be described by the same IRI (figure 4.1). As stated by Sayers and Karamihas (1998)

> *'It is meaningless to talk about the roughness of a point. Instead, one must consider roughness as summary of deviations that occur over an interval between two points.'*

The RN is similar to the IRI in the sense that it gives a single value for a stretch of road pavement or an event, therefore the RN is also unfit for monitoring how distresses actually develops.

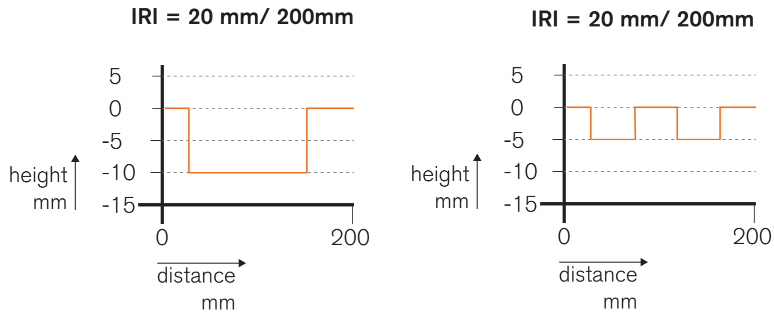


Figure 4.1: An example of similar IRI values for different shapes of distress. Note that this drawing is a simplification that does not take into account the possible frequency response of the Golden Car used to calculate a real IRI value.

It is clear that neither the IRI nor the RN can be used in the specific case of analyzing separate events with variable length and placement. Another option is to calculate the displacement. This is the base information used to in turn calculate the IRI and RN (Sayers & Karamihas, 1998). However, after consulting with an expert in the field of displacement and vibration measurements (van der Heijde & Ligtvoet, 2017), the decision was made to not attempt to convert to displacement due to the possible errors caused by double integrating or calculating on top of filtering artifacts. The

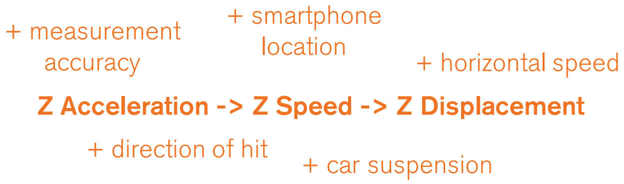


Figure 4.2: The relationship between vertical acceleration and vertical displacement together with factors which influence the measurements.

real world relationship between acceleration and vertical displacement is reliant on many factors (figure 4.2) and the relationship between the two is not linear. Figure 4.3 shows the relationship between the IRI and the speed. In turn Sayers and Karamihas (1998, page 36) state that the response of a car suspension is speed dependant. The speed dependent response together with the close relationship between IRI and vertical displacement indicate that the relationship between horizontal speed and vertical displacement is also non-linear. Consequently, making the compensation of the speed dependency a complex task.

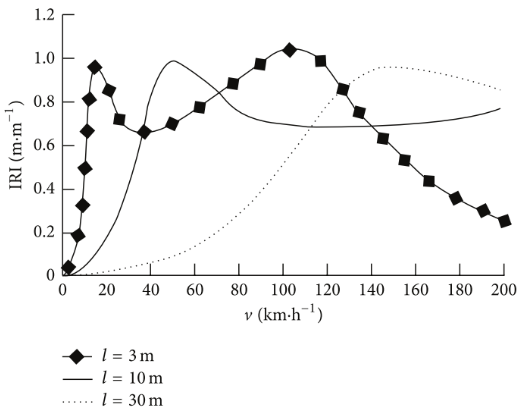


Figure 4.3: The relationship between speed and the calculated IRI. l is the wavelength of the road (Du et al., 2014).

4.1.3 A different way to measure road pavement distress

The focus of this research is to test if crowdsensing can be used to detect road pavement distresses for monitoring purposes within the future the possibility to predict at what time a road pavement distress needs to to be fixed. The accelerometer data gathered by the smartphones contains the information necessary to convert to displacement at a later stage when needed (Wilson, 1999). At the same time, it contains information on the severity of a type of distress due to sudden changes in displacement leading to larger vertical accelerations (Wilson, 1999). However, there is still the factor of speed. The horizontal speed directly influences the vertical displacement, therefore its second derivative is also affected (Sayers & Karamihas, 1998, page 34). The maximum speed connected to a road pavement asset is also of importance; road pavement assets where drivers are allowed to drive at 120 km/h should be safe for at least the same speed. In this research, a combination of vertical acceleration and the maximum speed of the road pavement asset is used as a way to monitor the

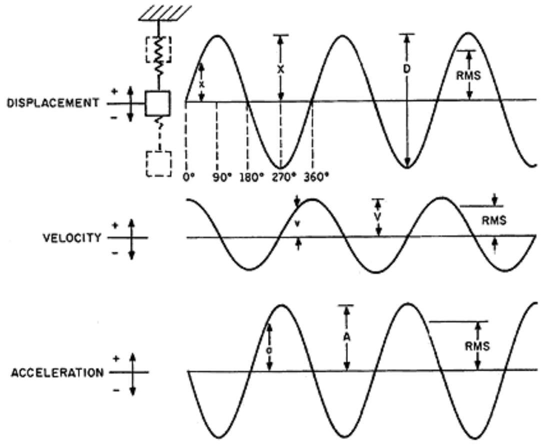


Figure 4.4: The relationship between displacement, velocity and acceleration (Wilson, 1999). The amplitude of the acceleration is dependant on the frequency of the displacement.

state of a road pavement asset. The speed of an individual car that measured an event can be used as a weight when multiple measurements are combined. To calculate the speed, both the time and positioning data is necessary. Finally, the orientation of the smartphone in quaternion is needed to compensate for possible differences in orientation between different smartphones (§ 4.4.1).

4.2 Connecting crowdsensing to sensor standards

Sensor standards promote information exchange between platforms, promote best practices and increase the usability of data for new applications. The OGC is an association that develops open standards on geospatial topics with input from professionals working in the domain. The fact that the developed standards and best practices are open source allows companies, universities and self-employed people to use them freely and provide feedback. In this section subquestion:

5. How can the collected data be structured to fit with existing standards?

will be discussed.

4.2.1 OGC sensor standards

The OGC has two working groups which have produced two different standards: the Sensor Web Enablement (SWE) and SensorThings. The difference between the two standards is the area of interest. The SensorThings API is developed specifically with the IoT and interconnectedness of devices in mind while SWE is focussed on the actual data exchange and the communication for services (Open Geospatial Consortium, 2016, 2017). In the case of this research, a high-level description is provided on how to fit the theoretical system into the SensorThings API and the possible connections to geospatial services.

4.2.2 The relationship between a sensor and road pavement distress

The relationship between sensor, smartphone, car, distress and road can be seen in figure 4.5. The sensor measures the acceleration and movement of the smartphone. As stated in § 1.6 Research methodology and research approach, the smartphone is stably positioned inside the car, meaning that the relationship between movement inside the cabin of the car and the smartphone are closely related. Distresses cause the car to move, in turn moving the smartphone and the sensor. In this system the car takes up a large role; the sensor measures the movement of the car and the position of the car is also the location of the sensor. In sensor standards, this would mean that the sensor is a dynamic sensor with a constantly changing location, which again puts a large emphasis on the car itself.

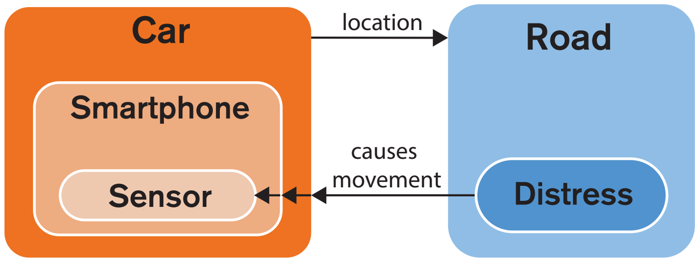


Figure 4.5: A simplified view of the relations between the sensor, smartphone, car, a road distress and the road itself. Note that the car itself takes up a large role.

The emphasis on the cars and their dynamic location is not the focus of this research, instead, the location and the severity of the distress is. To shift the focus from car to road distress, a 'virtual sensor' equipped to the center point of the distress can be used (figure 4.6). The location of this virtual sensor will be relatively stable and measures the displacement in a car cabin caused by the distress. A conversion needs to take place to deal with the difference between a car measuring multiple instances of distress, and a single instance of distress being measured by multiple cars. The position of the virtual sensor is equal to the position of a cluster created from events and is on or near the location of an instance of road pavement damage.

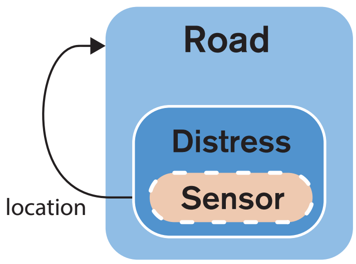


Figure 4.6: Shifting the focus from the dynamic car to the road distress by equipping a 'virtual sensor' to the location of the road distress. This 'virtual sensor' measures the displacement the distress causes on cars.

4.2.3 Implementing the virtual sensor

The first step in shifting the focus from cars to road pavement distress is to see a measurement from a smartphone in a car at location X, Y as the same as a measurement of a sensor embedded in the road pavement at location X, Y (figure 4.7). The change in location does not cause a change in the measured phenomenon: the acceleration inside the cabin of a car. The sensor embedded in the road pavement would be a ’virtual sensor’; it does not exist and the phenomenon it is measuring cannot be measured with real sensors which can be embedded in the road pavement.

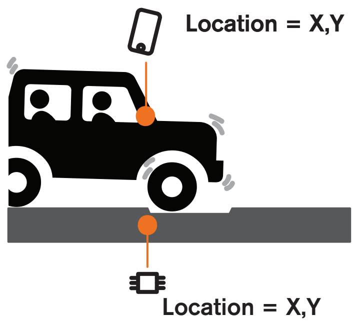


Figure 4.7: The concept that a smartphone at location X,Y is similar to a sensor in the road pavement at location X,Y.

The next step is to design how the raw data from the smartphone can be ordered in a way that it can be connected to road pavement distress. The proposal is to use the use the measurement events output from the local analysis, and apply the shift in focus during the figure 3.5 and § 3.5 steps of the Database analysis phase. The conversion to a virtual sensor makes use of activities which are already needed for both steps of the process. First there is the measurement event input in the figure 3.5 step (figure 4.8.1). Next the centre for each measurement event is calculated (figure 4.8.2). The centre of these measurement events are used to perform a cluster analysis (figure 4.8.3, § 3.5). The center of this cluster can be seen as the location of the new virtual sensor which is ’equipped’ to the distressed area, the range of the virtual sensor is a total area containing data from the measurement events (figure 4.8).

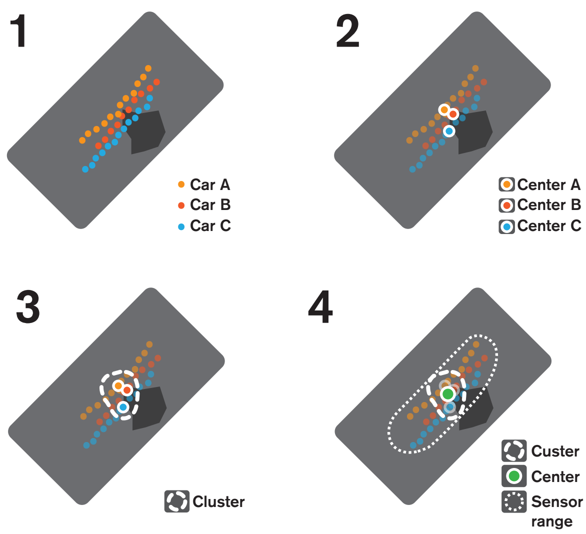


Figure 4.8: How to move to a virtual sensor connected to the road pavement distress. 1: measurement events from multiple cars, 1 event contains multiple measurements. 2: calculate the centre point for each measurement event. 3: cluster events based on their centre points. 4: calculate centre of virtual sensor.

4.2.4 Connecting the virtual sensor to the SensorThings API

The new virtual sensor and the information it provides can now be structured according to the SensorThings API, which is more flexible compared to the Sensor Observation Service (SOS) while at the same time having backward compatibility (Liang, 2016). The different classes in the data model are general classes which can be used to order data from different types of sensors (figure 4.9), these general classes can also be applied to the situation of the virtual sensor (table 4.1).

Table 4.1: Aligning the general classes of the SensorThings API and the virtual sensor.

Class	Virtual sensor scenario
Sensor	IMU
ObservedProperty	The vertical acceleration inside the cabin of the car at a location
MultiDatastream	A measurement event
Thing	Road distress
Observation	A single raw measurement from the sensor
FeatureOfinterest	Location and vertical acceleration

The advantage of using the SensorThings API is the ability to connect to and exchange data with Web Feature Services (WFS's) and Web Processing Services (WPS's). Stasch, Pross, and Jirka (2017) describe three methods on how to connect a WPS to an SOS. Because the SensorThings API is backward compatible, these methods are also applicable to this scenario. The method of having the user interact with the WPS, which in turn requests the information from an SOS or through a SensorThings application and returns the analysis results to the user, is of interest for this research. John the asset manager does not need access to the raw measurements from all the cars, instead, he needs the single measurement which describes the distress on for example a single day. By using the WPS as an intermediate, the processing can be kick started from the moment the data is requested, or it can be queued and executed at a later time.

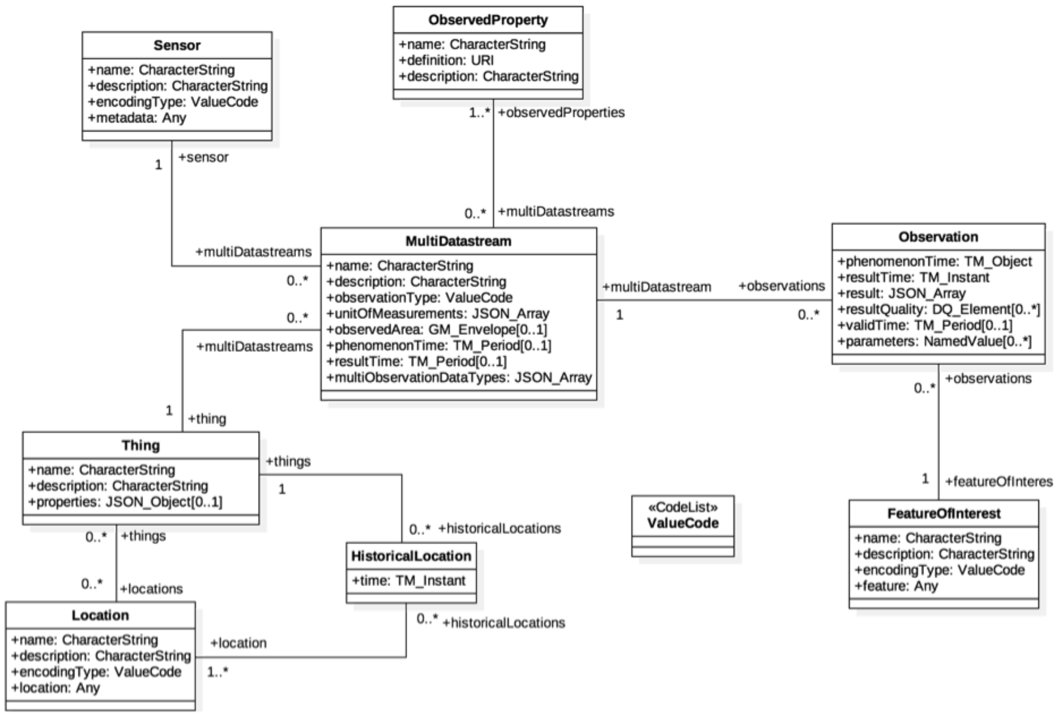


Figure 4.9: The MultiDatastream Extension Entities from the SensorThings API (Open Geospatial Consortium, 2016).

4.3 Data capture

The data was gathered by using the Sensor Data app available for iPhone models. The Sensor Data app can provide the output for the different types of sensors the smartphone is equipped with as well as some measurements which ar calculated through sensor fusion (Wavefront Lab, 2010). The outputs used during this research are acceleration, attitude quaternion, latitude/longitude, and absolute time. The measurements themselves are saved to a Comma Separated Value (CSV) file which is copied from the smartphone to the computer at the end of the data gathering phase.

The data was gathered during bus rides over the course of multiple weeks and in a car across two days. The main areas which were used for gathering measurements were the A27 between Gorinchem and Vianen, the A15 between Sliedrecht and Papendrecht and the N3 between Papendrecht and Dordrecht in the Netherlands. The bus model during data capture is the Volvo 8900, the car is a BMW 523i (figure 4.10). During the data gathering phase in the car, 4 different smartphones were used: an iPhone 5, an iPhone 5c, an iPhone 5s and an iPhone 6 (figure 4.11). The measurement frequency was set to 100 Hz for all smartphones.

Figure 4.10: The Volvo 8900 bus (left) and the BMW 523i (right) used during the data gathering phase (Volvo, 2017).

Figure 4.11: The smartphones used during the data gathering inside the car.

In total there are 40 files of data concerning the in cabin movement of the bus and 68 files concerning the in cabin movement of the car. The rides gathered by bus concern the most right lane and sporadically the emergency lane. The rides gathered by car concern the most right lane and sporadically the left lane. In total 49933596 measurements were gathered. An example of the collected data can be seen in figure 4.12.

Timestamp	Accel_X	Accel_Y	Accel_Z	Quat.X	Quat.Y	Quat.Z	Quat.W	Lat	Long
515758008.998631	0.487305	0.054138	-0.988220	0.002185	0.035758	-0.486183	0.873123	51.840896	4.713939
515758009.008631	0.565353	-0.129547	-0.980057	0.002185	0.035758	-0.486183	0.873123	51.840896	4.713939
515758009.018706	0.569519	-0.111786	-0.884033	0.002185	0.035758	-0.486183	0.873123	51.840896	4.713939
515758009.029128	0.557892	-0.121658	-0.879868	0.002185	0.035758	-0.486183	0.873123	51.840896	4.713939
515758009.038610	0.542511	-0.107513	-0.910187	0.002185	0.035758	-0.486183	0.873123	51.840896	4.713939
515758009.047712	0.542511	-0.107513	-0.910187	0.002185	0.035758	-0.486183	0.873123	51.841132	4.714005
515758009.059124	0.551941	-0.113892	-0.930771	0.002185	0.035758	-0.486183	0.873123	51.841132	4.714005
515758009.068899	0.554184	-0.124146	-0.979355	0.002185	0.035758	-0.486183	0.873123	51.841132	4.714005
515758009.078653	0.566284	-0.125107	-0.918335	0.002185	0.035758	-0.486183	0.873123	51.841132	4.714005

Figure 4.12: Raw data columns from the output csv files.

4.4 Local analysis

The local analysis phase consists mainly of data preprocessing steps. These preprocessing steps consist of signal analysis and position and speed calculations. The main objective of the local analysis phase is to perform the event detection (§ 4.4.7) and to save the data to the database § 4.5. Sections § 4.4.1 to § 4.4.3 are focussed on signal analysis and sections § 4.4.4 to § 4.4.6 are focussed on the speed and position analysis.

4.4.1 Using reorientation to compensate for data gathered by a smartphone in an arbitrary orientation

Rotations are often described in the terms Roll, Pitch and Yaw, also known as the Euler angles (Renaudin & Combettes, 2014). Euler angles can be used to convert the orientation of devices through for example a rotation matrix. However, the downside to using Euler angles is the ”gimbal lock” problem which describes the problem of having an end position which can be achieved trough multiple rotation combinations (Renaudin & Combettes, 2014). Gimbal lock is not always a problem, it depends on the constraints of the application (Sachs, 2010). For example, it is unlikely that the nose of a car is pointing upwards into the sky, or that an aircraft takes off at a 90-degree angle. However, these constraints do not apply to smartphones which can be oriented in any arbitrary direction. To cope with the arbitrary orientation, quaternions can be used instead of Euler angles. Quaternions are 4-dimensional vectors, which for simple calculations can be described as a 3-dimensional vector around which a rotation takes place (Sachs, 2010). The two main advantages of using quaternions are their insensitivity to gimbal locks, as well as the result of the rotation being dependent on the order in which it is performed.

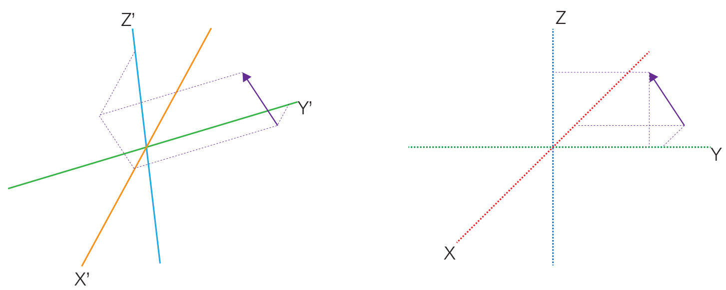


Figure 4.13: A single vector V has different notations in different reference frames. To move from reference frame X'Y'Z' to XYZ, a DCM can be used to reconstruct the same vector in the new frame.

The quaternion describes the rotation of the body frame, in this case, the smartphone compared to the gravity and the compass. A quaternion consists of 4 components: a vector along the x-axis, a vector along the y-axis, a vector along the z-axis and a rotation vector θ. The quaternion is calculated through performing a sensor fusion between the data from the sensors the smartphone is equipped with (§ 2.1.4). This quaternion can the be converted to a DCM, which describes how a single vector can be notated in a different reference frame (figure 4.13) (Farrell, 2008). A DCM is often described as the rotation matrix of a quaternion. To calculate the DCM from a quaternion the following function is used:

$$DCM(q) = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}$$

(Mathworks, 2017)

Note that the above version of converting the quaternion q to a DCM expects the rotation factor θ to be the first value q_0, while quaternions can sometimes have the rotation factor as the last value in the quaternion.

There is also a downside to using quaternions to create a DCM. The sensor fusion used for the quaternion calculation requires enough information from multiple sensors of different types. Therefore it takes some time before valid quaternions are available. Figure 4.14 shows the consequences of initialization time of the quaternion on the resulting reoriented data.

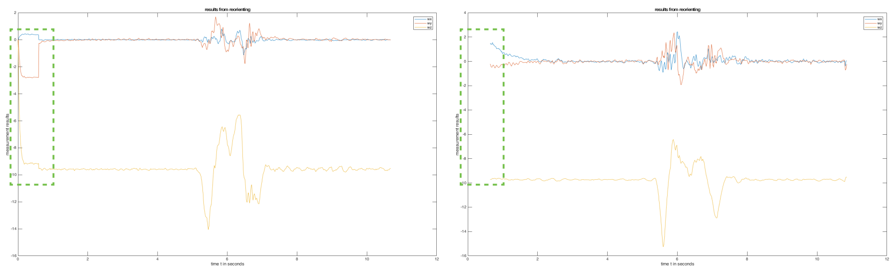


Figure 4.14: The results of using a DCM to convert the vector data on data gathered under a 45 degree angle (left) and a 90 degree angle (right).The green highlights show the errors during the time it takes to calculate a correct DCM.

4.4.2 Removing noise from the accelerometer data

There are two types of noise in the collected accelerometer data: the noise inherent in the measurement device itself, in the case of Microelectromechanical System (MEMS) accelerometer often brown noise, and the noise caused by external factors during the data gathering (Cemer, 2011). These external factors can be seen as the high-frequency vibrations in the cabin caused by the type of road pavement and the suspension. This high-frequency noise is undesirable and obscures the movement caused by actual road pavement. A can be seen in figure 4.15 the event of driving across some pavement distress takes around a second to pass. This knowledge can be used to filter the data to remove high-frequency noise. An event has the frequency of around 1 Hz and the measurement speed is 100 Hz. With a sampling rate of 100 Hz the maximum frequency which can be detected with certainty is 50 Hz, significantly higher than 1 Hz (Smith, 1997). However the higher sampling rate can be used to detect noise, and to remove it while at the same time enabling the reconstruction of the original signal.

The 1 Hz duration of the event seems to be contradictory to the size of road pavement distress at first. At 80 km/h, 22 meters is covered in 1 second, which is an unlikely size for road pavement distress. However, the 22 meters is the duration of the cabin movement due to the distress. The difference between the length of the road pavement distress and the cabin movement is caused by the suspension of the car; a spring-mass system which slowly dampens the movement caused by the pavement distress. The effect of the suspension is not just a factor in the dampening of the z acceleration, it also allows for the detection of road pavement distress due to its dampening time.

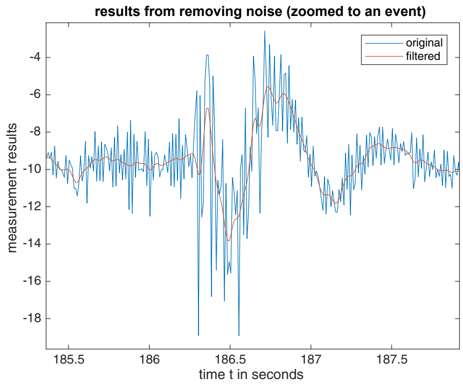


Figure 4.15: The data from a car ride zoomed in on a peak and the filtered data of that same peak.

To remove the noise from the measurements a Butterworth low pass prototype filter is used with a cut-off frequency of 10 Hz. The 10 Hz cut-off frequency can be used to remove the noise of the measurements while at the same time keeping the in cabin movement. The Butterworth filter does not have the steepest cut-off possible of all the existing prototype filters, however, its advantage is the low distortion of the data that is passed through (figure 4.16). A filter has different orders which describe the cut off response: a higher order filter has a faster frequency response time but requires more computing power. The order of the filter is set to 2 to maintain a low computing power. The resulting Fast Fourier Transform (FFT), a transformation which shows the frequency components of a signal, shows the slow cut off of the filter (figure 4.17).

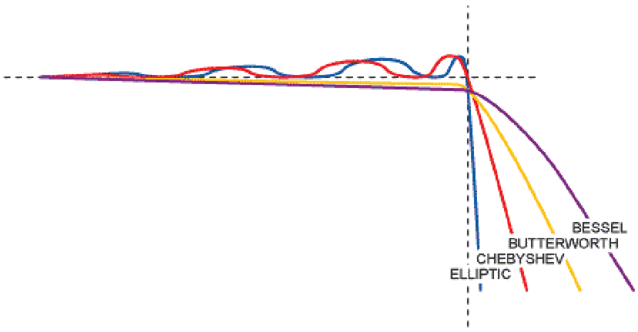


Figure 4.16: Four different types of prototype filters. Note the steepness of the filters and the smoothness of the function before and after the cut off location. Image retrieved from Maxim Integrated Products (2002)

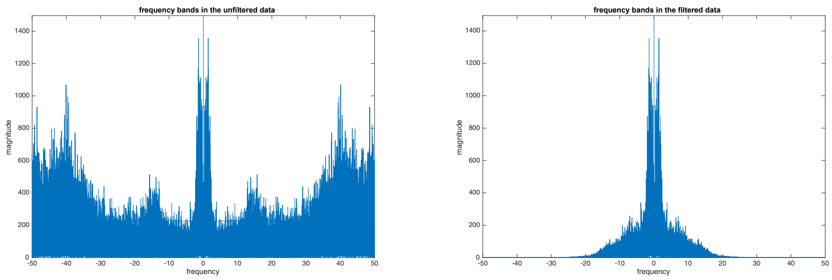


Figure 4.17: The FFT of a single car ride before (left) and after (right) filtering. Note that the Y axis is has been shortened.

4.4.3 Removing gravity from the accelerometer data

Next, there is the gravity component. After the reorientation in § 4.4.1 the gravity component is measured exclusively in the Z axis. The gravity component can be seen as a signal with the constant frequency of 0. This characteristic can be used for the removal of the gravity through a high-pass filter with a cut-off frequency of 0.1 hertz. The Butterworth high pass filter is used as the prototype filter. After testing the resulting data from different order filters, the filter order of 1 is chosen due to its performance in removing the actual gravity component completely (figure 4.19).

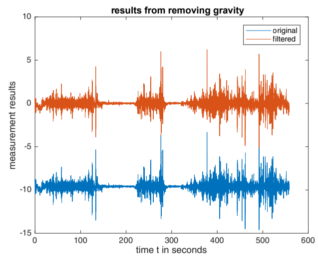


Figure 4.18: The results from filtering with a Butterworth of the 2nd order.

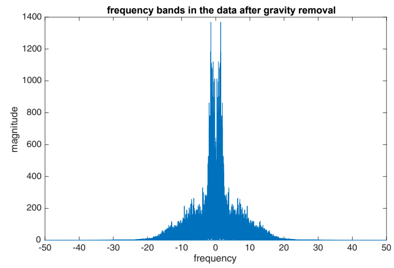


Figure 4.19: The frequency bands after removing the gravity component.

4.4.4 Converting from WGS84 coordinate system to a local coordinate system

The coordinates provided by smartphones running on Android or IOS are in the WGS84 coordinate reference system (Android Open Source Project, 2017a; Apple Inc., 2017). WGS84 uses latitude and longitude, while the data from the accelerometer is converted to meters per second. The frequency at which the position is calculated using the smartphone app is once a second, while the frequency at which the smartphone sensors can be queried is higher. This results in multiple measurements having the same position. The unique position of measurements can be approximated by finding and using the unique positions between measurements as well as using the difference in time § 4.4.5. To smoothen the process of calculating the speed, the coordinate reference system has to be converted to one which uses meters as a unit. It was stated in § 1.6 that the scope of this research in terms of location is the Netherlands, and the Amersfoort RD New system uses meters as its unit of measurement.

The OSGeo package is used to convert the coordinates from WGS 84 to Pseudo-Rd in Python. The code to perform the transformation can be found at the open documentation of Deltares (den Heijer & Santinelli, 2014). This transformation does not use the correction grid needed to make a transformation to the true Amersfoort RD New system. However, the locations used during the data gathering phase are all in areas which have errors under the 5 cm. While the inaccuracies of the Dutch RD system do not have a large effect during this research, it can cause problems when a nationwide system is created. The true description of the Amersfoort RD new projection in proj4 notation is:

$+proj = sterea + lat_0 = 52.15616055555555 + lon_0 = 5.38763888888889 + k = 0.9999079 + x_0 = 155000 + y_0 = 463000 + ellps = bessel + nadgrids = rdtrans2008.gsb + geoidgrids = naptrans2008.gtx + units = m + no_defs <>$
(Debian GIS Project, 2017)

This projection can be implemented by using the python pyproj package.

4.4.5 Compensating for the difference in measurement frequency between the GNSS and the accelerometer

As stated in the previous subsection, the speed at which a smartphone calculates its position is slower than the speed at which other sensors can gather data. The method used in this subsection consists of multiple steps: first, a change in GNSS position is detected, then the speed and vector are calculated, the combination of speed, vector, and the time difference is then used to calculate a new position. Algorithm 1 shows the operation of grouping sets of measurements which have the same position information as well as calculating the vector, the speed and the new position for each measurement. The speed is calculated by taking the position provided by the GNSS for the current group and the next group and comparing the distance between the two with the difference in time.

The grouping is performed at [1] in the algorithm, detecting the end of a group and calculating the vector, speed and new position are done at [2]. Calculating the speed for the final set of measurements at [3] poses a problem. The final set of measurements does not have a GNSS position of another group to compare with. Instead, the speed of the previous group is used.

There are several factors influencing the quality of the new position, these are the GNSS initialization time and the positioning precision. The GNSS initialization time causes a position error at the start of the data collection for a single car ride. The positioning precision is based on te quality of the smartphone clock used for the calculations. The maximum precision for the smartphone used during this research is 5 meters based on the precision indicator provided through the smartphone app. In turn, the positioning precision can influence the calculated speed between positions. With the current method, the measurements are spread evenly between the two known positions, regardless of whether or not a particular speed is within the realm of possibility (figure 4.20).

measurement_id	4b28b9f0-151b-11e7-b51c-f81654697a99
event_id	4b28b9ee-151b-11e7-9c7b-f81654697a99
datetime	1970-01-18 06:06:51.239088
Zaccel	-0.882343294902893
Speed	610.051324141755

Figure 4.20: A speed error, this speed error occurred during the initialisation phase of the GNSS device. An example of a speed error after the initialisation phase is an acceleration from 80 km/h to 120 m/h in a single second for a bus.

Algorithm 1: Derive the position of measurements based on GNSS position, vector and speed

```
Data: xcoordinates as x, ycoordinates as y, timestamps as t
Result: new X and Y coordinates for all measurements
Create list new x
Create list new y
Create list t
for i in range x do
    if i = first index then
        current x,y,t = x[i],y[i],t[i]
    else
        3if i = last index then
            append t[i] to list-t
            for j in range 1 to list-t do
                if j = 1 then
                    new x,y = current x,y + speed x,y * Δ(t[j], t[j − 1])
                    append new x,y to list new x,y
                else
                    new x,y = new x,y + speed x,y * Δ(t[j],t[j − 1])
                    append new x,y to list new x,y
        1else if x[i],y[i] = current x,y then
            append t[i] to list t
        2else
            next x,y = x[i],y[i]
            delta t = Δ(t[i], current t)
            vector x,y = next x,y − current x,y
            speed x,y = vector x,y / delta t
            for j in range 1 to list t do
                if j = 1 then
                    new x,y = current x,y + speed x,y * Δ(t[j], t[j − 1])
                    append new x,y to list new x,y
                else
                    new x,y = new x,y + speed x,y * Δ(t[j], t[j − 1])
                    append new x,y to list new x,y
            current t = t[i]
            list t = [currentt]
            append x[i],y[i] to list new x,y
            current x,y = x[i],y[i]
```

4.4.6 Calculating the speed for each measurement based on the GNSS position

As described in the previous subsection, the speed between two positions provided by the GNSS can contain errors due to inaccuracies in positioning. To decrease the effect of these errors the average speed across multiple positions is taken. The calculation of the average speed moves over all positions in a window form; for a point x the average speed is calculated across $x - 2$ up to and including $x + 2$ (figure 4.21). The code for calculating the speed has a similar structure of 'for' and 'if' statements as the code used in algorithm 1, however, the grouping clauses are different. Therefore, a separate function is used to for the calculation.

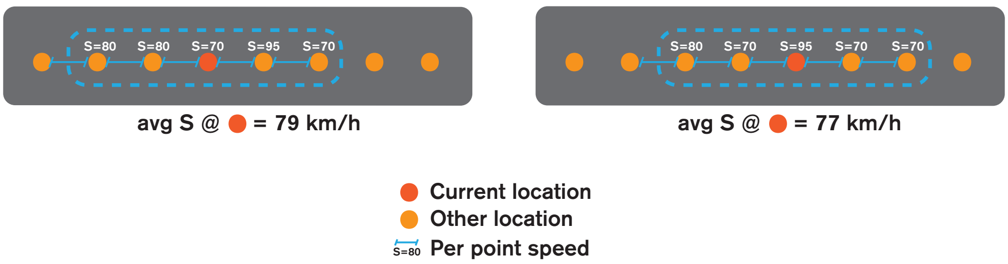


Figure 4.21: Using a window to calculate the speed at a position. The window iterates over all positions.

Using a window which uses both positions before and after the current one has a downside in that special cases are needed to deal with both the first few positions and the last few positions. The special cases are needed for the first two positions and the last two positions because of the window size of $x - 2$ up to and including $x + 2$. In these two cases, the speed is calculated by keeping the window in the start and end position. These positions can be described as the window for positions[3] and positions[-3].

4.4.7 Identifying measurement events

The results of this subsection are used to partly answer the subquestion:

2. What data and information are needed to monitor a road pavement asset?

There are many different ways to identify the events describing the occurrence of road pavement distress. An example would be to use a training set which has been classified beforehand as the input for a machine learning algorithm. This training set would then be used to make several descriptions of how an accelerometer signal behaves when driving across road pavement distress. The advantage of using this type of pattern recognition in machine learning is that events can be detected regardless of the type of car and its suspension. The downside of this type of classification is its complexity; the local analysis takes place on a smartphone which means a limited amount of resources to perform calculations (Ganti et al., 2011). The complex calculations need for the pattern recognition will require more resources, which in turn can cause problems with smartphone power management.

A more simple way to detect a signal which corresponds with road pavement distress is to use the standard deviation σ for detection. Instead of computing the relationship between the measurements, a threshold consisting of a single number is calculated. The calculation of a single number costs fewer resources compared to the pattern recognition algorithm. Applying the threshold on the raw data, in turn, costs fewer resources.

During this research, the a in $a\sigma$ is calculated by tallying the road pavement distress that can be felt in the vehicle cabin, while at the same time gathering data though the smartphone and. the value a can be approached by attempting to find the equal amount of events in the data to the events tallied. This standard deviation can then be used on all other rides. The corresponding measurement value for $a\sigma$ is calculated per ride.

By using standard deviation as a tolerance to filter out uninteresting information, only the peaks and valleys of the signal are returned. This means that the part of the measurements describing the road pavement distress is lost (figure 4.22). The peaks and valleys which are passed through are then seen as separate events. This is not a reflection of the real world situation where there are measurements with a value within the tolerance describing the area between the start and end of for example a pothole.

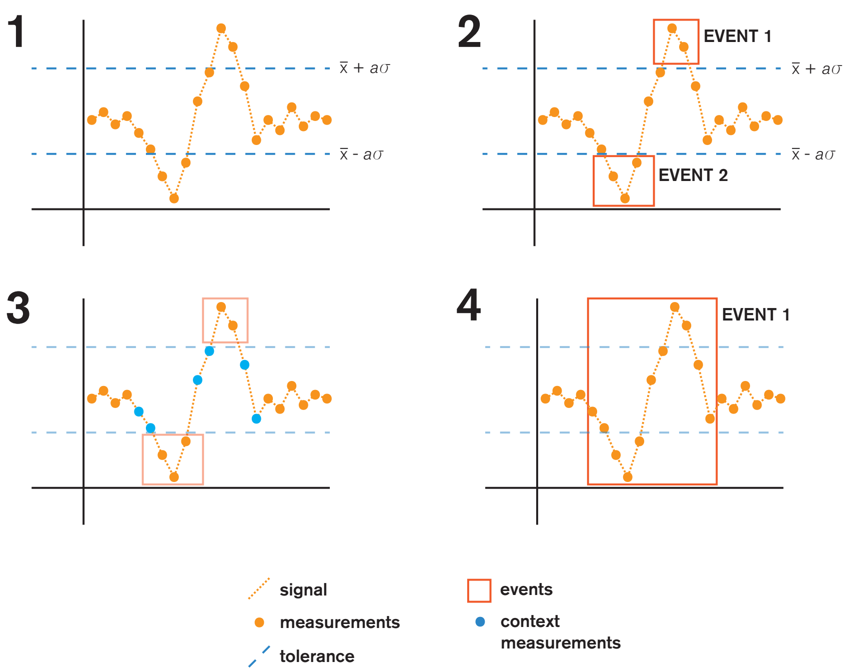


Figure 4.22: The creation of events through using tolerances and context measurements.

To overcome this problem context is needed. The context is provided by first detecting a value above or below the tolerance and then using its index value i to create a window of the size from $i-3$ up to and including $i+3$. The range of the window was chosen through experimentation (figure 4.23). This window of index values in then used to detect overlap between events. Overlapping events are combined into one and used to select the corresponding z acceleration, speed and position (figure 4.22).

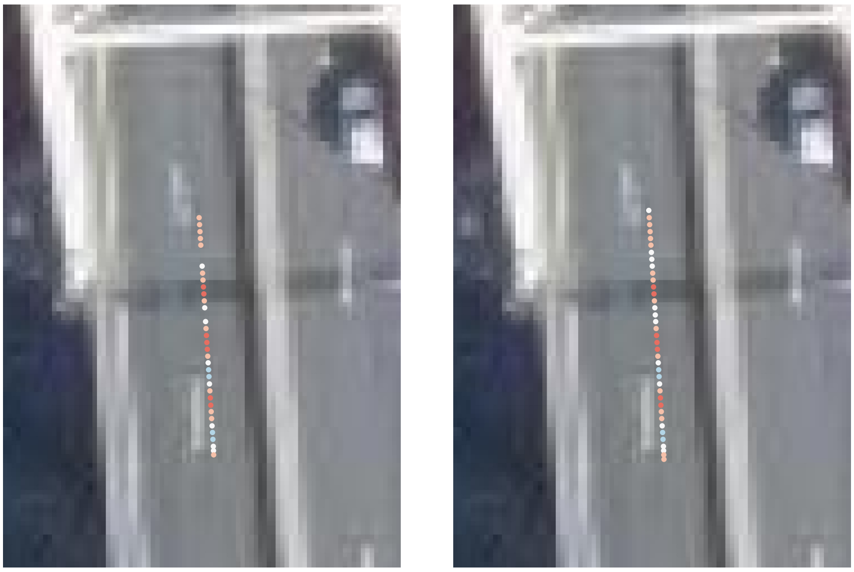


Figure 4.23: The results of a using a window of the size $i-2$ up to and including $i+2$ to stitch events together (left) and the results of using a window of the size $i-3$ up to and including $i+3$ to stitch events together (right). The location is a bridge across the Beneden Merwede along the N3.

Measurements that are not part of an event are discarded. Measurements that are part of an event ar given an event identifier, marking what event they belong to. The resulting event information saved to the measurements contain te following information: multiple a measurement id which can be used as an unique identifier, an event identifier to track which measurements are part of the same event, a ride identifier which can be used to track if two events are made during the same commute, a time stamp, the z acceleration, the speed and the position in a point geometry. The ride identifier can be used to count how many events in a cluster or instance of road pavement distress are not from the same commute from the same car.

4.5 Saving the measurement events to a database

After all the steps described in the above subsections are performed, the processed data can be stored in a local database. By saving the data to a database, the local analysis only needs to performed once on a file containing the ride data. At the same time, the data become easy to access through GIS packages like QGIS and ArcMap. By using these GIS packages, the resulting data can be checked visually for inconsistencies which can otherwise be difficult to track. An example would be the ability to check the results from calculating the new position for each measurement as described in § 4.4.5. As well as checking the results from combining events through looking at nearby measurements (figure 4.23).

The data is stored in a spatial table on a PostgreSQL database with PostGIS support. The PostGIS support gives access to built-in spatial functions which can be used in the subsequent steps of the database analysis. The spatial table is named events and contains the columns measurement_id, event_id, datetime, zaccel, speed, and point. the primary key is on the measurement_id column. The concept of the measurement_id and the event_id is that the measurement_id is unique, while the event_id indicates which event the measurement belongs to. The point column contains the encoded point geometry.

4.6 database analysis

Now that the data has been preprocessed during the local analysis phase and stored in a database, an attempt to combine the information from multiple rides can be made. The steps needed to combine the correct data together are described in this section.

4.6.1 Performing a cluster analysis on the measurement events

The first step in the database analysis is the clustering of the events. As an even already consists of multiple measurements, the center of the event needs to be calculated. By using the center of the events during clustering, the possibility of measurements in a single event being split up is solved. Calculating the position of the centroid of an event can be done by using the built-in capabilities of the PostGIS extension. The events are now ready to be clustered.

When clustering spatial data or any type of data at all, the decision on what algorithm to use has a major effect on the results from the clustering. The PostGIS plug-in supports two types of spatial clustering algorithms: the k-means algorithm and the DBSCAN algorithm. The k-means algorithm is often used for its low processing costs. However, for this algorithm, the number of classifications needs to be known beforehand. In the case of this research, each instance of road pavement distress is a new class. The algorithm is not suitable for the subject of detecting the number of road pavement distresses because this exact number needs to be known for the k-means algorithm. An algorithm usable in the scenario presented in this research needs to function without knowing the final amount of instances of road pavement distress while taking into account that a single event does not make a detection. The DBSCAN algorithm works by using two variables; one variable for the maximum distance between points to belong to a cluster and the minimum amount of points needed to make a cluster. The algorithm does not need to know the total amount of clusters, instead, the two input variables can be used to cluster based on nearness and to deal with noise i.e. an event which has an isolated position. DBSCAN is also not reliant on specific shapes of clusters and can create clusters which are positioned inside other clusters (figure 4.24). The final advantage of using the DBSCAN algorithm is that is readily available as an implemented function in PostGIS from version 2.3.0 and on. The chosen algorithm also has a major downside. It is computationally complex and does not scale well to large datasets and multidimensional datasets where other information next to the X and Y coordinates are important in the classification or clustering.

In PostGIS, the function works by having two inputs, the ”eps” and the ”minpoints”. The ”eps” is the maximum distance a point can be from a cluster to be a part of it and the ”minpoints” is the minimum amount of points needed to create a cluster. During the research, the clusters are created by using 5,4 meters based on the positioning precision of the smartphones (§ 5.1.1), and 3 as the minimum amount of points to create a cluster. The output of the DBSCAN itself is a new column which is called cluster_id with a number which for each event signifies which cluster they belong to.

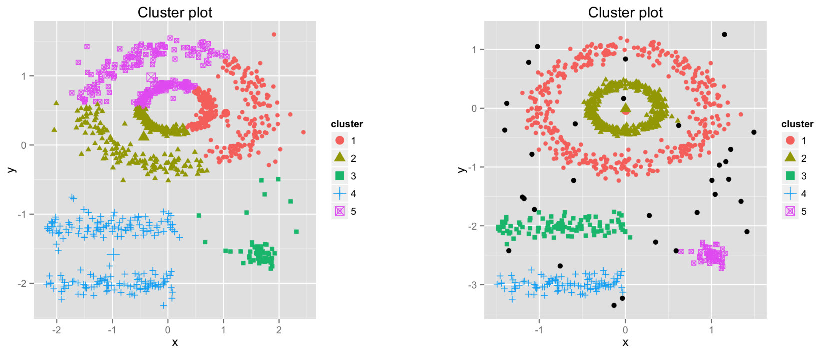


Figure 4.24: A comparison between the clustering results of a k-means algorithm (left) and a DBSCAN (right) on example data . (STDHA, 2017)

4.6.2 Calculate cluster positions

With the events now have the information on which cluster they belong to. To calculate the cluster center the centroid function from PostGIS is used again. A combined geometry is created by grouping on cluster_id value and the st_collect function. The collected geometry is then used as the input of the centroid function. The result is a single point position denoting the center of the cluster (figure 4.25). The center of the cluster is also the location of the virtual sensor.

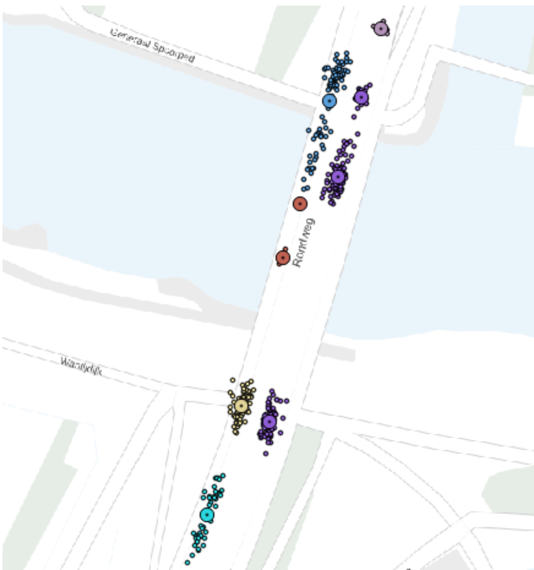


Figure 4.25: The result of the clustering with the virtual sensor positions as the centre of the cluster.

4.6.3 Connecting the virtual sensor to a road pavement asset

In this subsection the subquestion

6. How can the information be linked to existing assets?

is answered. The performance and challenges of the implementation described here are discussed in § 5.2.2.

To connect the virtual sensor to a corresponding asset, two different datasets are used; the 'maximum snelheden (wegvak)' dataset from the weggeg datastore, and the 'hectopunten' dataset from the Nationaal Wegen Bestand (Rijkswaterstaat, 2017a, 2017b). The geometries from both these datasets can be connected through the attribute named 'wvk_id', an abbreviation of the name 'wegvak id'. The 'maximum snelheden (wegvak)' contains the multiline geometry of al highways in the Netherlands while the 'hectopunten' dataset contains point geometries for all the Dutch hectometre posts on the highways as well as the provincial roads (figure 4.26). The relationship between the two datasets is that a single instance of a wegvak (road section) can have multiple hectometre posts along its length.

Figure 4.26: The two datasets on a map of Zuid-Holland with the 'maximum snelheden (wegvak)' shown in blue and the 'hectopunten' in orange.)

During this research, the shapefiles of both datasets are loaded into PostgreSQL to simplify the query process and to bypass the limitations of the public WFS services. The 'hectopunten' dataset also contains hectometre posts for provincial roads, which can be removed due to the focus on highways during this research. To remove these excess hectometre posts the existence of the 'wvk_id' in the dataset can be compared to the 'wvk_id's available in the 'maximum snelheden (wegvak)' dataset.

The connection between the virtual sensor and the asset is made by using the st_distance function from PostGIS. The distance function is used to perform a nearest neighbor analysis on the nearby assets, in this case, the analysis uses the geometries from the 'maximum snelheden (wegvak)' dataset. The nearest road section can be found by maintaining the 'wvk_id' from the road section with the lowest distance value (algorithm 2). Next, the nearest hectometre post for a virtual sensor can be found by selecting the hectometre post with the same 'wvk_id' as the nearest road section and calculating the distance from the virtual sensor to the hectometre post. The hectometre post with the lowest distance result is the closest to the sensor. The connection between hectometre post and the virtual sensor is made by saving the 'gid' attribute from the dataset to the sensor.

Algorithm 2: Connect a virtual sensor to the nearest road section and a hectometre post.

Data: A table containing cluster information with columns: cluster_id, events_in_cluster, rides_in_cluster, nr_of_rides, geom and centroid. The dataset named 'maximum snelheden (wegvak)' from the weggeg datastore and the 'hectopunten' dataset from the Nationaal Wegen Bestand datastore.

Result: A table containing virtual sensor positions linked to both a wegvak and a hectometre post.

explode geometry from 'maximum snelheden (wegvak)"

for *each geometry* **do**

  calculate direction

save to table road direction

for *each instance of a virtual sensor* **do**

  calculate direction based on event direction

  for *each geometry in road direction* **do**

    if *difference in direction between geometry and virtual sensor > 20 degrees* **then**

      discard

  calculate distance to each entry in max_snelheden

  if *distance is more than 50 metres* **then**

    discard

  calculate minimum possible distance

  if *distance of road section == minimum distance* **then**

    add th wvk_id attribute value to the cluster

  if *wvk_id of hectometre post == wvk_id of virtual sensor* **then**

    select hectometre points with the same wvk_id

    calculate distance to each hectometre point

    calculate the minimum distance

    if *distance of hectometre post == minimum distance* **then**

      add the gid attribute value to the virtual sensor

Using the road centreline as the first object to connect to, makes better use of the positional context of the virtual sensor. However, there is also another context: the driving direction. A virtual sensor can be closer to the road centreline at locations like highway intersections, causing them to be connected(figure 4.27). The driving direction can be used to connect virtual sensors to the right road section even if the position contains a large error or is mathematically closer to another road section center line.

The direction of the virtual sensor can be found by calculating the average direction between all the events which are part of the virtual sensor. To calculate the direction of the virtual sensor, lines are created between all measurements in an event assigned to the virtual sensor. The start and endpoints of these lines are used to perform two centroid calculations. From the new average centroids, a line is created, this line is used as the direction of the virtual sensor through the st_azimuth function from PostGIS. For the comparison, the direction of the road centrelines is also necessary. However, the geometry from the 'weggeg' data store consists of multiline string geometry. To calculate the correct direction for a road section the multiline strings are split up into simple line geometry through splitting by the multiline vertices. For each simple line, the direction is calculated through using the st_azimuth function. The directions of the virtual sensor and the road section are compared and if their difference in direction is lower than 20 degrees, the road section is maintained for the nearest neighbor calculation.

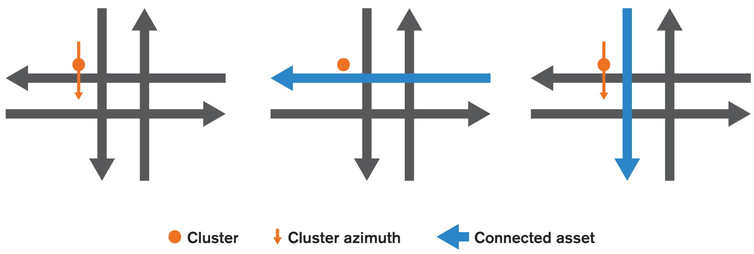


Figure 4.27: If the distance to the road centreline would be used to connect the virtual sensor to the asset, it would be connected to the westbound road (middle). However using the direction of the virtual sensor together with the miniumum distance leads to a correct connection (right).

4.6.4 Calculate the most likely acceleration values based on multiple measurement events

The virtual sensor is now connected to the nearest asset and the nearest hectometre post. The information from all the events part of the virtual can be combined to get averaged measurements of the instance of road pavement damage the virtual sensor is connected to. The chosen parameters are the minimum and maximum acceleration, the speed and the duration of the event. These parameters give an indication of the most discomfort that is experienced on average by drivers driving over the road pavement distress as well as the duration of the discomfort on average and the average speed of the drivers.

The minimum acceleration, the maximum acceleration and the duration of the events are speed dependent, combining these measurements into averages without weighing leads to a skewed result. The weighing of the measurements can be performed by using the speed at the maximum and minimum acceleration of the event and comparing this to the maximum speed allowed on that road section and using the difference to create a weight (algorithm 3). During this research the weights are set to 1 if the difference in speed is between 0 and 10 km/h, 0.8 if the difference is between the 10 and 20 km/h, 0.5 if the difference is between 20 and 40 km/h and 0.2 if the difference is above the 40 km/h. This decline is used based on the concept that the measurements gathered by a car driving closer to the maximum speed allowed give an indication of whether a road is safe enough to facilitate its maximum speed.

Algorithm 3: Calculate the minimum acceleration, maximum acceleration, speed and duration of a cluster.

Data: The resulting table from algorithm 2 and a table containing all measurements with their id's z acceleration, geometry, speed and time.
Result: the maximum and minimum acceleration, speed and length of a cluster based on the weighted events.

```
for each cluster do
    select all the events inside the cluster
    use the wvk_id to gather the maximum speed allowed
    for each event in the cluster do
        calculate the maximum positive z acceleration
        calculate the maximum negative z acceleration
        calculate the average speed during the event
        group all measurement geometries into a multipoint
        calculate the length of an event based on the multipoint geometry
        calculate the weight of an event based on the average speed during the event and the
          maximum speed allowed
    calculate the maximum positive z acceleration based on all the events in the cluster and their
      weights
    calculate the maximum negative z acceleration based on all the events in the cluster and their
      weights
    calculate the weighted speed
    calculate the weighted length
```

4.7 Database storage

The results of the database analysis steps are stored in a table named 'road_distress' which has the following attributes:

Table 4.2: The attributes for the road_distress table and their types.

Attibute	:	Type
cluster_id	:	int
wvk_id	:	double
gid	:	int
centroid	:	geometry(point)

Attibute	:	Type
min_zaccel	:	double
max_zaccel	:	double
speed	:	double
length	:	double

From the 'road_distresses' table, other calculations can be made on, for example, the amount of distresses per 'wegvak', the amount of distresses per hectometre post or the severity of these distresses per asset. At the same time the data on the distresses is retained

The table does not have a primary key, which while not a necessity during this research, is a must when implementing a real world system. A primary key can only be implemented on the 'cluster_id' column. When creating a real world system to monitor road pavement distresses trough crowdsensing, the virtual sensors need to be dated based on the timestamps of the input data, this information can then be stored in a new 'timestamp' attribute. If the database analysis is performed per day, then the 'timestamp' would be the date.

5 Results and analysis

Like in chapter 4, the structure of this chapter is based on the different steps and phases of the Research approach. During this chapter new factors of influence are discussed to answer the subquestion:

3. What factors influence the accuracy of a detected event

The other subquestion which is answered in this chapter is:

4. What are the effects of having constant low accuracy and low precision measurements during the shock determination analysis?

5.1 Local analysis

From the local analysis phase, the results and the design decisions are discussed for the compensation of the measurement speeds, the speed calculation, the event identification and the database storage steps.

5.1.1 Compensating for the difference in measurement frequency between the GNSS and the accelerometer

The method used to calculate the new positions for the measurements without their own unique positions relies heavily on the precision of the known positions provided through GNSS. The quality of this measurement is therefore also on the precision on the GNSS potions. The data gathered in buses can be used to check to GNSS positioning precision because buses most often drive in the most right lane. This near constant fact can be used to measure the precision perpendicular to the road direction(figure 5.1). Based on 5 positions and the center of 170 events along the A27 highway, the mean distance to the right lane is 3 meters on with a standard deviation of 2.4 meters. The spread of the positioning errors is not a normal distribution; 74% of the positions fall within 1 σ from the mean. This positioning precision does not take into account the different precisions different types of smartphones might have.

Another factor is the positioning dilution caused by the warm boot of the smartphones. However events with these positioning errors can be removed during three separate steps; during the calculation of the event centroid, during the clustering, and during the calculation to connect a virtual sensor to the asset. During the clustering step, the events with major errors in their position often have a centroid which is not near any other event centroids, resulting in the event being counted as noise. If there are enough events with positioning errors caused by the warm boot to create a virtual sensor, it can be dropped during the step where the sensors are connected to the asset by tuning a maximum distance for a connection. During the event centroid calculation, the events with a positioning error can be filtered out by looking at the average speed of the event.

Figure 5.1: Using the centroids of events to check the precision of the smartphone GNSS by using the distance to the middle of the most right lane as a basis.

5.1.2 Calculating the speed for each measurement based on the GNSS position

The positioning errors also have an effect on the speed as it is calculated by using the difference in time and position between the known positions. The effect of this error can be explained by using the example of a car driving at 80 km/h. The speed of the GNSS is 1 calculation per second, which means that at 80 km/h the car travels 22 meters in that same second. The average GNSS precision of 3 meters as described in § 5.1.1 translates to an error of 10 km/h. If both the first and second position have a 3 meter error, the result of the speed calculation is 100 km/h instead of 80 km/h. The speed of the car will be used in a later step to assign weights to the vertical acceleration. An error the calculation would cause a measurement getting either a too high or too low weight, influencing the resulting weighted average of for example the maximum and minimum vertical acceleration.

During the research, an implementation of a moving window is used to decrease the effects of the position error on the speed. The moving window size is set to the size of 5 and scans both in front and behind the 'current' measurement (figure 4.21). A larger size window could further decrease the effect of positioning errors on the speed but would require more processing power. Another option is to use the window to estimate the likeliness of an increase in speed is an error or a car speeding up. Large accelerations between measurements can indicate that a large positioning error has occurred. The value which is an error can then not be used during the calculation of the average speed inside the window. Another implementation used to remove speed errors is to filter out events with speed errors after they are sent to the database by filtering on the average speed per event.

5.1.3 Identifying measurement events

In § 4.4.7 the concept of using the standard deviation as a tolerance is introduced. By using standard deviation as a tolerance, cars with different types of suspensions can be used to detect road pavement distress without the stiffness of the suspension greatly hindering the detection. However, the downside of using the standard deviation is that it is a variable which changes based on the input measurements. Say for example that 0.3% of the Dutch highway network consists of road pavement distress. This can be translated to any measurements that fall outside the 3σ bounds describe an instance of road pavement distress in the Netherlands. While the 3σ variable makes for an easy translation between different cars and suspensions, the variable only counts for the whole of the Netherlands. Single roads do not necessarily adhere to the 0.3% pavement distress. In the real world, 0.9% of a stretch of road might consist of distresses while two other roads both have a 0.05% of their area consists of distresses. If a single ride in a car would drive along the road which consists of pavement distresses for 0.9% of its surface, the result would be that not all measurements describing distresses are classified. Meanwhile in the instance of driving along one of the roads with very few distresses would lead to over-classification of measurements. A possible solution to this problem would be to convert the tolerance from a standard deviation in σ to an actual measurement value over time. The gathering of the necessary information for this conversion can be seen as an initialisation phase of the smartphone app.

Another important factor in using the standard deviation as a tolerance is chosen value for it. The value a of $a\sigma$ has an effect on the number of events which are detected and sent to the database. The 'correct' value a of $a\sigma$ can be seen as the number where the relationship to the error of commission and error of omission is optimal. In the specific scenario of using the system to study how road pavement distress develops and to use the same information as an indicator to send inspectors the effects of errors of commission and omission are as follows:

- Errors of commission result in an over-classification of positions where a road pavement distress has occurred. In turn, the over classification can lead to wasting the time of road inspectors who are sent to the location for verification. Errors of commission can also cause the formulation of an incorrect relationship between the in-car acceleration and the severity of the road pavement distresses.
- Errors of omission lead to inefficient information on how road pavement distresses occur, especially if the errors of omission apply to a certain type of distress. More related to real world applications, errors of omission can also cause the contractor to be unable to meet the boundary conditions defined in the contract, in turn leading to high costs for emergency repairs and fines.

Based on both the consequences of the error of commission and the error of omission together with the case that during the research and development of a system it is more beneficial to have too much information and therefore more errors of commission. The value a of $a\sigma$ would need to include all road distresses which can be felt. The consequences are that the events which are detected are actually noise, however with real world validation by road inspectors the optimal value of a can be approximated. However, the standard deviation used to detect events is not the final way that events are qualified as describing distresses. The amount of events near a position is used to classify road pavement distress later in the process. As long as the events describing noise are not classified based on their proximity, the negative effect of introducing noise from the errors of commission is mitigated.

On the subject of calculating the value a in $a\sigma$, the data gathered during multiple bus rides is used together with a manual counting of the pavement distresses detectable inside the vehicle cabin. The locations used are the A14 and N3 highways from the access road at Schiedam towards Dordrecht up until the exit to the Albert Schweitzer hospital. During the manual counting, there were 55 instances of road pavement distresses or expansion joints. Some instances of road pavement distress were long stretches of road pavement causing continuous vibration. These long vibrations can be detected as multiple events. At $a = 3$ one location that was felt in the cabins was not detected. In this case the location contained an expansion joint on a bridge. In other instances, the distress was detected only once. This information together with the knowledge that more data can lead to more events describing pavement distresses leads to the conclusion that the value for a should be slightly lower than 3.

Three different values of $a\sigma$, $a = 3$, $a = 2.75$ and $a = 2.5$ were tested on the data gathered by bus. Based on the results the $a = 2.5$ value allowed for the detection of the missing road distress or expansion joint, however, the value also caused a decrease in clarity where an event ended and the next started figure 5.2. The absence of a clear divide between multiple events describing different instances of road pavement damage can lead to the creation of new virtual sensors or the merging of virtual sensors. The value of 2.75 shows an increase in detected events while at the same time keeping a visible divide between what are most likely different instances of road pavement distress or expansion joints.

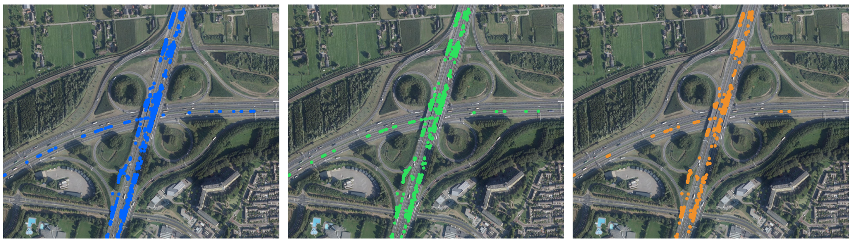


Figure 5.2: The results of event detection when using different tolerances. The blue events are the result at $a = 3$, the green events are the result at $a = 2.75$ and the orange events are the result at $a = 2.5$.

5.1.4 Saving the measurement events to a database

After the event detection, the data can be saved to the database. At this stage, an analysis can be performed to document the amount of data saved to the database and how much information is filtered out at the event detection. The events sent to the database are around 4% of the original input data with a small variance between data gathered by car and by bus (figure 5.3). The filtering of the data happens in the reorientation step and the event detection step. During the reorientation step data which does not have a quaternion is discarded, however, this amounts to around a 100 measurements at a maximum per file. Related to the event detection step the amount of data filtered out during this step is between the 0.05% and the 0.01%.

In later stages of development, there might be the possibility to filter out more data to decrease the amount of necessary server space. When using crowdsensing the filtering of data is of importance because the amount of data which can be gathered can be extremely large (Ganti et al., 2011). A slight decrease in the amount of information all smartphones are transmitting to the database can have a large positive effect on the amount of storage space needed. One potential place for more

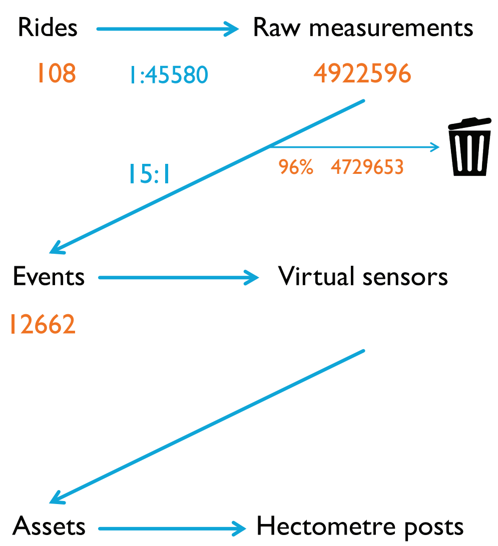


Figure 5.3: The relationship between the number of rides, raw measurements and events.

filtering is again the reorientation step. As shown in § 4.4.1 the first few quaternions from the sensor fusion are incorrect. While the amount of data removed by removing these invalid measurements is low, the quality of the information output will go up. Another option is to filter during the speed step to remove measurements which have an incorrect position due to a large GNSS error. These types of measurements can be recognized by their extremely high speeds. As shown in § 4.4.6 a window is used to minimize the effect of speed errors caused by small positioning errors. However, errors which result in a speed of more than 200 km/h are difficult to compensate. A possibility would be to remove the events which have an unrealistic speed. In the Netherlands, this would mean removing all events with a speed above for example 180 km/h. There is the possibility that the filtering would also remove events from drivers driving too fast, however, the maximum speed on the Dutch highways is 130 km/h. At 180 km/h, the driver would either be driving 40 km/h over the speed limit or the GNSS positioning error would be 12 meters with the driver driving a maximum speed. This option is implemented in the database analysis phase.

5.2 Database Analysis

In this section the results and discussion points from the cluster analysis, the connecting to assets, the calculations to get average values for a virtual sensor and the final database storage.

5.2.1 Performing a cluster analysis on the measurement events

The choice of clustering algorithm has a large effect on the result of the clustering step and the information that can be gathered from the results. The DBSCAN algorithm used in § 4.6.1 is known to decrease performance when the input data increases on the $O(n * log(n))$ scale (Kantardzic, 2011). However, during this research, the performance is adequate. The algorithm takes 436 milliseconds to calculate the clusters from the centres of 6946 events containing multiple measurements on an HP ZBook 15 with an i7-4700MQ processor and 16 GB memory.

DBSCAN is created specifically for spatial data, however, it is unable to process multidimensional data more complex than 2D at an effective rate. The limitations of the algorithm can cause problems when creating the clusters as other contexts like driving direction cannot be taken into account. An option is to connect the events to their respective asset based on both distance and driving direction and performing the clustering algorithm afterwards on a per asset base. However, there is a major downside to this method; at locations where the is an entry or an exit occurs the events are connected to the wrong asset due to the similar direction of the entry and exit lanes to the main road. This in turn results into the creation of more clusters (figure 5.4).

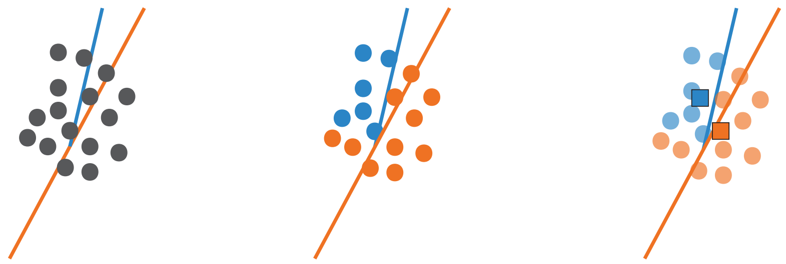


Figure 5.4: Some events have been detected at location X (left). When the events are connected to the asset before clustering, two groups are created (middle). When the cluster analysis is performed two virtual sensors are created (right).

The criteria of the DBSCAN algorithm are set to 5,4 meters for maximum distance and 3 for the minimum amount of points (minpoints). The maximum distance is based on the precision and standard deviation of the positioning system. The value for the minimum amount of points 3 is a place-holder value. This is because the optimal value to form a cluster and identify distresses relies on the amount of data that is gathered and how the effect of errors in event detection increases. A possible method of calculating the optimal value for the minpoints parameter is by using machine learning. The relationship between car rides can be calculated by using a ground truth of pavement distresses at location X and datasets of increasing size as the input of the machine learning algorithm.

Another point of interest is the accuracy of the position of the resulting virtual sensor created by the method. To test the precision, the algorithm is run on the data gathered by car on rides which are on the same side of the road along a limited area of the N3 (figure 5.5). First, the cluster calculation is run over 10 rides, then 20 rides and then 30 rides. As the car data concerns mostly the most right lane, the transversal accuracy can be calculated by the distance of the distress to the lane centreline. The lane centreline is drawn based on the Dutch aerial photography background service. Because of the different amount of rides used as an input, the number of clusters created in each situation differs slightly. At 10 rides 27 clusters are detected, at 20 rides 26 clusters are detected and at 30 rides 32 clusters are detected. The results show a clear increase in transversal precision especially with an input of 30 rides where the average distance plus the standard deviation is smaller than half the lane width (Rijkswaterstaat GPO m.m.v. Witteveen + Bos, 2015).

To calculate the longitudinal accuracy of the clusters, the exact location of the pavement distress needs to be known. Based on the results from the transversal accuracy, making an estimate of the location based on recordings, Google Streetview, and aerial photos will likely lead to large offsets compared to the actual longitudinal error. Therefore the longitudinal precision cannot be calculated reliably. However, there is a clear visual cohesion between cluster locations and real world object.

Figure 5.5: The area of the N3 used to calculate the transversal accuracy of the clusters.

Table 5.1: The transversal accuracy of the clusters.

Nr of rides	Nr of clusters	Average distance to lane centreline (m)	Standard deviation (m)
10	27	1.54	0.96
20	26	0.98	0.92
30	32	0.82	0.66

The location shown in figure 5.6 is a bridge along the N3 route. The cluster positions visually correspond with the presence of expansion joints.

The final subject of this subsection is the temporal accuracy of the clusters. More specifically how can the road pavement distress detected yesterday be connected to the road pavement distress which has been detected today in a robust way? Though this question is outside the scope of the research it is important to note that the chosen method of clustering does not take this into account. Future research can be performed to formulate a method to maintain temporal accuracy while at the same time manage the unknown number of clusters or classes

5.2.2 Connecting the measurement events to a road pavement asset

In this subsection, the design decisions behind method implemented in § 4.6.3 and the challenges of connecting to a road pavement asset are described. Some of the situations have a solution which is already implemented. The subsection continues with providing information for the subquestion:

6. How can the information be linked to existing assets?

The connection to the road section asset is created connecting to the road centrelines of the weggeg dataset. At first, tests were done by connecting the hectometre posts and the road centreline. This decision was made because performing a nearest neighbour analysis between point geometries is less complex than a nearest neighbour analysis between point geometry and line geometry. The

Figure 5.6: The position of the clusters along the N3 Merwedebrug.

relationship between the datasets containing the road centreline and the hectometre posts can be described as follows: a single road centreline can contain multiple hectometre posts, but a hectometre post can only belong to a single road centreline. This attribute can be used to connect the clusters to the road centreline through a single attribute of the hectometre posts. However, the results when connecting directly to the hectometre posts does not take into account the road geometry. In for example bends, the virtual sensor might be closest to a hectometre post from the road in the opposite direction. This phenomenon results in a connection between the virtual sensor and the incorrect hectometre post (figure 5.7). To solve this problem, the cluster is first connected to the nearest road center line, then the nearest hectometre post with the same wvk_id as the road center line is calculated. As shown in figure 5.8 the errors caused by bends in the road is solved.

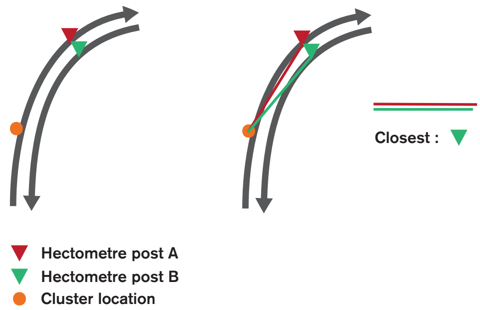


Figure 5.7: An error in connecting a cluster to a hectometre post in a bend.

Using the driving direction of a virtual sensor as a first filter to remove roads which are not of a similar direction allows for the correct connection at highway junctions and overpasses. However, a virtual sensor can still be connected to the wrong road at entrances and exits because the azimuth of these objects is very similar to the azimuth of the main road geometry. A possible solution might be to predict the lane a car is driving on. During this research, the actual lane location was outside of the research scope. Another downside to using the direction of the assets only after the cluster analysis is that events can still be connected to a cluster consisting mostly of events describing road pavement damage on the opposite road. During this research, the connecting of an event in the opposite direction happens sporadically. When more data is added, these instances can lead to clusters merging together if the algorithm is not tuned correctly (figure 5.9). Lub (2016) used a lane detection algorithm in a similar context as this research, the lane detection in combination with the event direction can deliver a possible solution to keeping clusters from merging together and connect to the wrong asset.

Figure 5.8: Connecting clusters shown as circles to the wrong hectometre post shown as triangles(left) and the result after fist connecting to the road centreline (right).

Figure 5.9: The creation of a cluster consisting of events from both road directions.

5.2.3 Calculate the most likely acceleration values based on multiple measurement events

The information per virtual sensor is a summary of the maximum vertical acceleration, the minimum acceleration, the average length and the average speed. These values are calculated by using a weighted average based on the difference in speed between the car and the maximum speed allowed on the road section. The weighting is currently unoptimized to cope with the complex relationship between horizontal speed, distress size and vertical acceleration previously mentioned in § 4.1.2. The exact relation between these phenomena can be used to create a more accurate weight distribution and can, in turn, lead to more accurate summary results.

Another factor which is of importance to a virtual sensor and its measurement values is the car suspension. The averaged values provided by the virtual sensor describe the feedback the most average Dutch car generates. The properties of this average Dutch car need to be researched to classify at what point the values of for example the maximum z acceleration go from 'good' to 'bad'. In § 2.1.2 the "Golden car" is introduced as the average car, however, the "Golden car" is not an optimal model for the scenario of using crowdsensing as it is a quarter car model based on the car properties in the nineteen eighties. The calculation of the properties of the average Dutch car is a topic which needs further research.

5.2.4 Database storage

After the analysis is performed in the database, the results are stored in the road_distress table. With the data gathered by bus and by car used for the database analysis, the table contains 526 records. The number of measurements with event_id used as input for the database analysis is 192943. The number of events used during the analysis is 12622. Therefore the proportions between measurement, events and virtual sensors is 15:1 for measurements to events and 24:1 for events to virtual sensor (figure 5.10). In turn, the virtual sensors are connected to the road section assets in a 14 sensors per asset relationship. On average the virtual sensors are connected to 6 hectometre posts along the length of the asset.

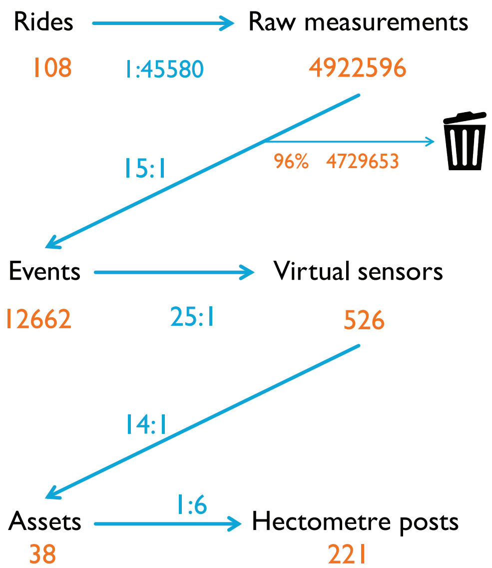


Figure 5.10: The amount of virtual sensors generated during this research and the number of assets and hectometre posts they are connected to.

The resulting clusters can now be visualized in a GIS package like QGIS. The connection made between the road pavement distresses and the road pavement assets like the road sections and hectometre posts allow for the classification of road sections and hectometre posts based on the amount of damage nearby (figure 5.13, figure 5.14). The classification per road sections is coarse in the sense that a complete road section can be multiple kilometers. When a classification on a finer level is needed, the hectometre posts give an indication on where on a road section the damage can be found. Another possibility is to classify the road sections and hectometre posts not by the amount of distress but the average severity per distress. In combination with the knowledge on how many distresses have been detected, the severity can be used to make a prediction on whether there are many small distresses or a few large distresses (figure 5.14,figure 5.16).

Figure 5.11: Visualizing the road pavement distresses and their severity.

Figure 5.12: Visualizing the distribution of road pavement distresses in a heatmap.

Figure 5.13: The road pavement distresses can be used to classify the road sections based on the amount of distresses detected per road section.

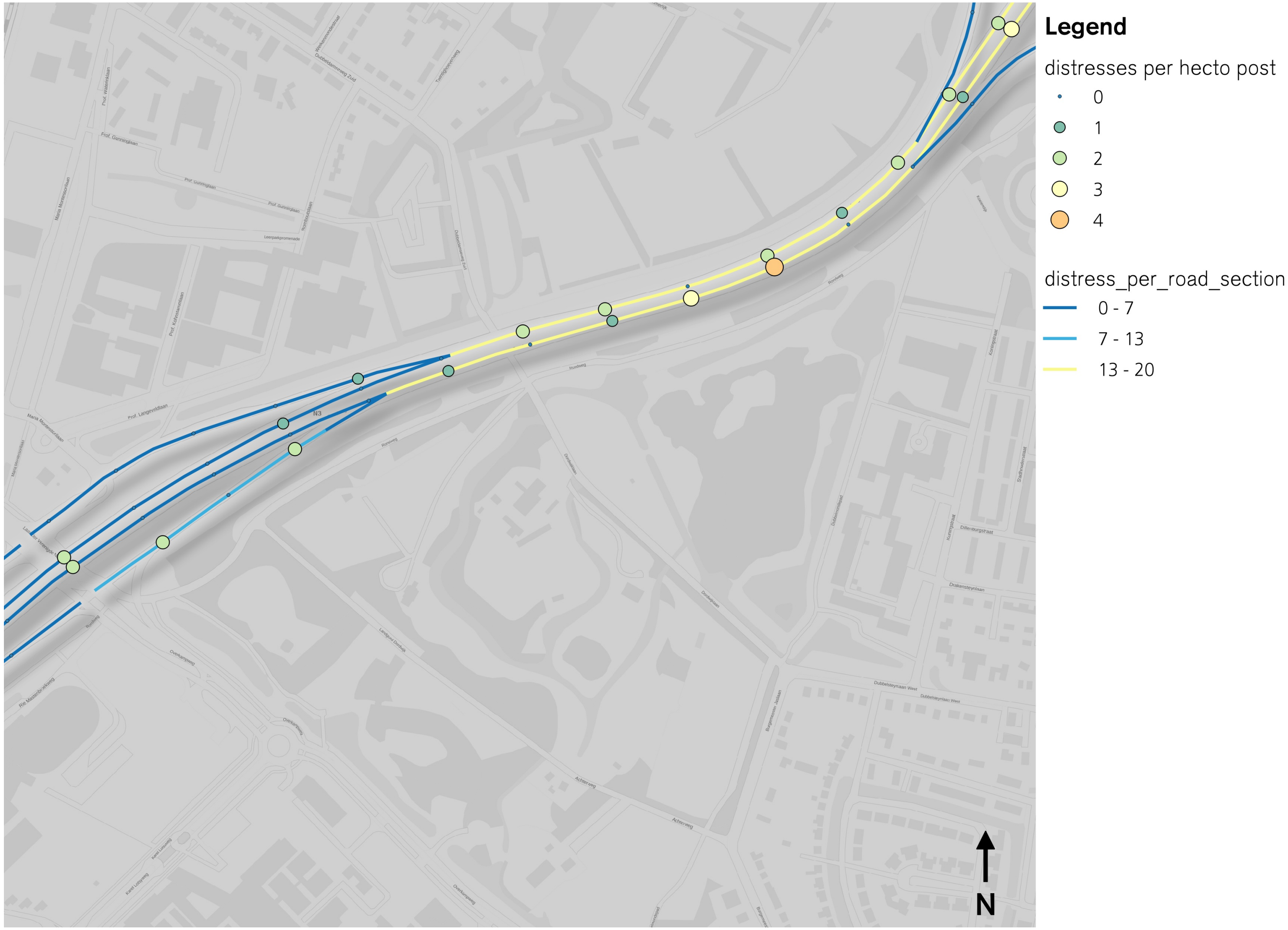


Figure 5.14: On a smaller scale the hectometre posts can be used to visualize the amount of distresses in their neighbourhoods.

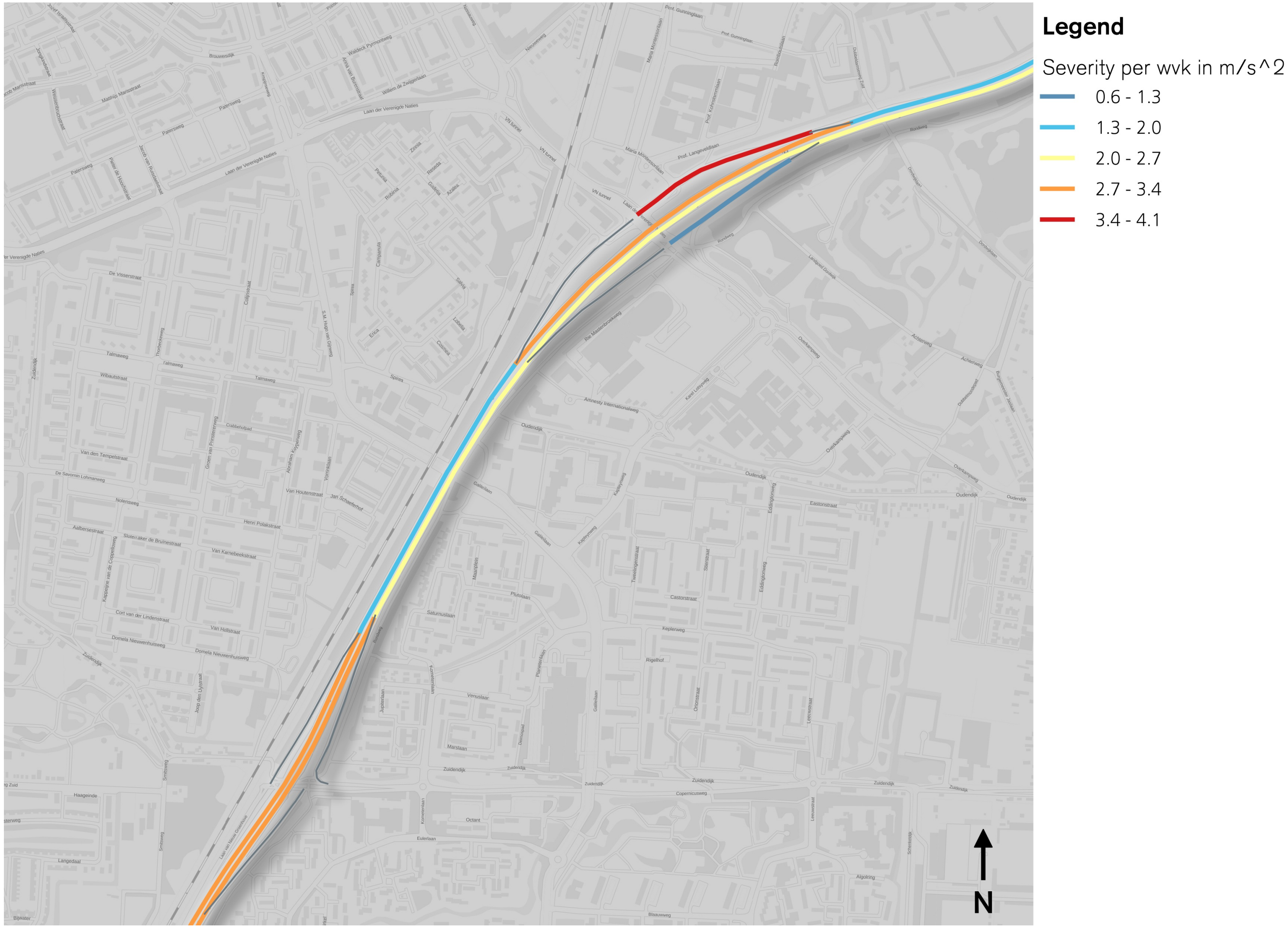


Figure 5.15: The severity of the road pavement distress can also be visualized per road section

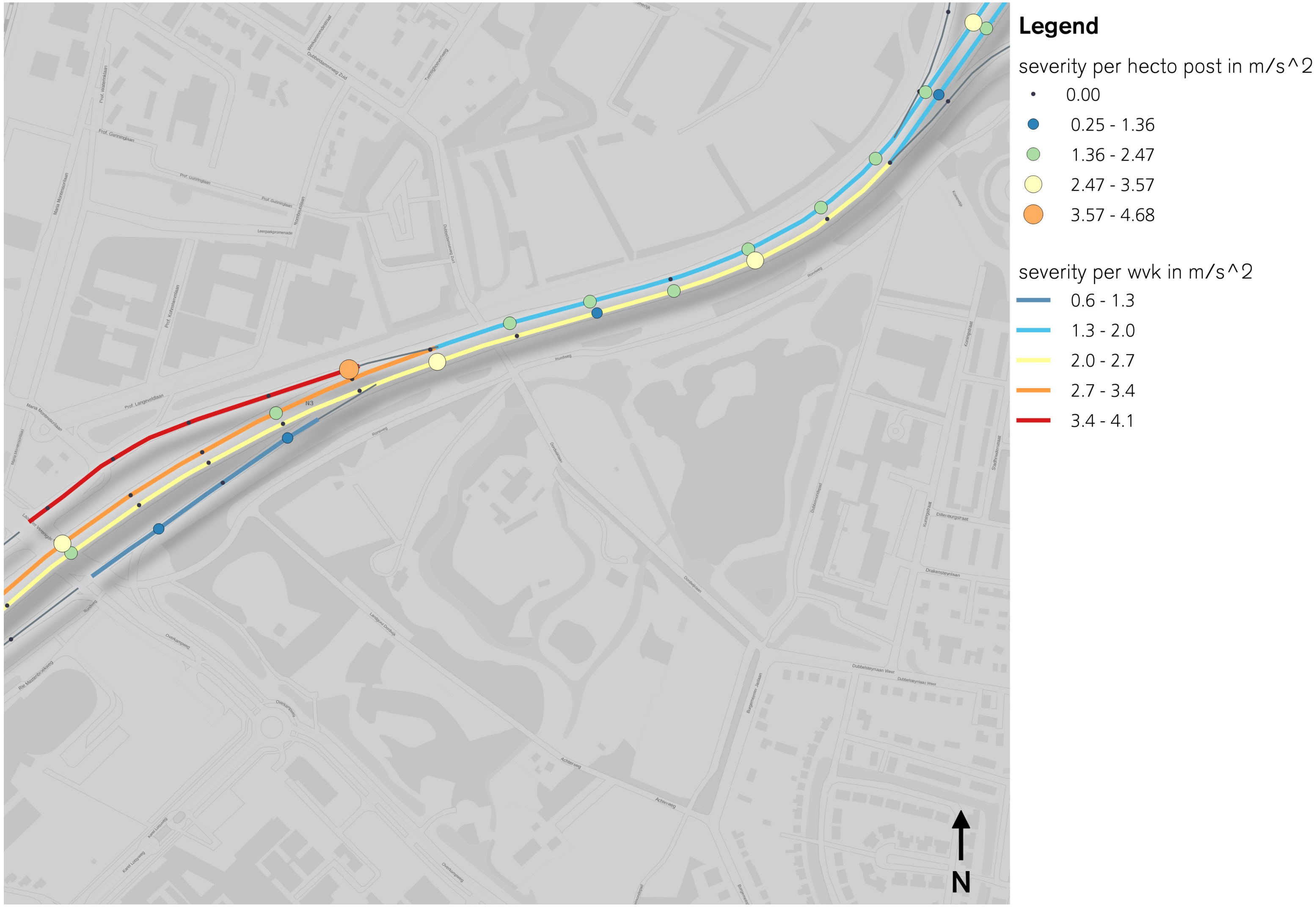


Figure 5.16: For a less general location of the severity of the pavement distresses can be visualized per hectometre post.

6 Conclusion

Based on the knowledge gathered during the research, the sub-questions introduced in § 1.5.1 are answered. The main research question, in turn, is answered based on the resulting information from the sub-questions.

Sq 2. What data and information is needed to monitor a road pavement asset?

The sub-question focusses on what data is needed from the smartphone to perform the local analysis, and what information needs to be in an event to facilitate the database analysis phase. The intensity of road pavement distress is expressed by the vertical acceleration. The vertical acceleration can be used to detect sudden changes in the road pavement. This characteristic contributes to the detection of events; road pavement distress causes peaks in the vertical acceleration while gentle undulations of the road pavement are not detected. Other types of data needed from the smartphone sensors are positioning data, timestamp and the orientation of the smartphone in quaternions. These data types contribute to the positioning of measurements, the calculation of the speed and the reorientation of the vertical acceleration data.

An event contains six types of data:

- event identifier;
- ride identifier, used to track a number of cars which have detected something;
- vertical acceleration;
- time;
- speed, used as a weight for averaging other measurements;
- positions of the measurements, used to create the virtual sensor.

Together these data types enable the events to be combined into groups which represent a virtual sensor, while at the same time retaining the individual data for every separate event.

Sq 3. Which factors influence the accuracy of a detected event?

Five main factors are found to have an influence the accuracy of an event:

- the measurement accuracy of real world phenomenon;
- the orientation of cars when hitting the road pavement distress;
- the location of the smartphone in the car;
- the speed of the car;

- the suspension properties of the car.

Of which both the speed of the car and the suspension are the most important. All the above factors are an intrinsic part of the gathered raw data and their influence is difficult to remove due to their exact effect being unknown. Consequently, these factors will also influence the accuracy of the information contained in the events.

The accuracy of the events created during this research is affected by five factors. All of these factors relate to the accuracy of the performed local analysis:

- the effectiveness of the reorientation algorithm;
- the performance of the noise filtering algorithm;
- the performance of the gravity filtering algorithm;
- the positioning precision and;
- the accuracy of the tolerance used to detect events.

These factors are aimed at the quality of the implementation. The positioning precision is the exception, though this factor can also be seen as part of the real world phenomenon factors. The effect of these factors can be decreased by optimizing the algorithms and tolerances implemented during this research.

The factors influencing the accuracy during the data gathering phase are difficult to remove from the data because their exact influence is to be further researched. This research attempts to counteract the factors influencing the measurement accuracy by using crowdsensing, the speed of the car is used for weighing measurement values. Both the way the car hits a distress and car suspension are averaged through crowdsensing even though their direct relationship to the measurement data needs to be further researched.

Sq 4. What are the effects of having constant low accuracy and low precision measurements during the shock determination analysis?

The precision of the measurements is especially important when looking at the positioning of measurements and the error size. The positioning precision affects:

- the tolerance value for the event detection;
- the maximum distance used to group events into a cluster;
- the accuracy of the clustering in the context of events of roads in opposite direction;
- the position of the virtual sensor;
- the connection between the virtual sensor and the asset or hectometre post;
- the virtual sensor values.

In the case of the accuracy of the clustering, events which were detected on roads in opposite direction would be grouped together. This is due to the positioning precision which in turn affects the position of the events and their spread.

The effect of the precision of the positioning cascades through the events to the position of the virtual sensor where it can cause failures to connect to the correct asset and hectometre post. These errors can be mitigated by taking into account the average direction of all the events of a virtual sensor.

The cascading effect of the positioning precision also affects the average values of the virtual sensor as these are calculated through using the position derived speed.

The accuracy of the measurements can be viewed as two categories:

- the accuracy of the vertical acceleration measured by the sensor;
- the transversal and longitudinal accuracy of the virtual sensor.

The accuracy of the vertical acceleration is dependant on the factors discussed in subquestion 3. As the effects of these factors on the measurement data is currently unknown, further research is needed to quantify the accuracy of the measurements.

The transversal accuracy of the virtual sensor increases when more events are used to crease a virtual sensor. The longitudinal accuracy could not be established, however, there is a clear relationship between the position of the virtual sensor and the locations of for example expansion joints.

Sq 5. How can the collected data be structured to fit with existing standards?

This research defines the virtual sensor as a sensor object which is always connected to the area where a road pavement distress was detected through crowdsensing. By using this virtual sensor the focus shifts from following the position of the car to tracking the positions and properties of road pavement distresses. It also shifts the type of sensor network from dynamic to static: instead of moving sensors, the sensors are static and turn "on" and "off" depending on whether an instance of distress has been detected.

Sq 6. How can the information be linked to existing assets?

The connection between virtual sensor and asset is made by using the official road sections of the Dutch highway network. The connection between the virtual sensor and asset is based on the direction and minimum distance. The virtual sensor direction is used as a way to remove road sections which do not have a similar direction, decreasing mismatches at intersections. Next, the distance from the road pavement distress to all remaining road sections is calculated and the nearest asset is selected based on the lowest distance. The asset and distress are linked through the road section id.

Hectometre posts can be used to get a more accurate position of the road pavement damage. The connection to the hectometre posts is based on the road section and the nearest hectometre post.

Sq 1. Is it possible to detect road pavement distresses with consumer grade sensors?

During this research data collected through crowdsensing is used to:

- create events through using vertical acceleration data;
- validate events though multiple observations;
- create virtual sensors.

The creation of virtual sensor is based on the concept that events that are grouped close together indicate the existence of a common underlying cause.

While the longitudinal accuracy of the virtual sensor created through clustering is to be further researched, there is a noticeable real world relationship between real world locations like expansion joints and the positions of the virtual sensors. There is also a relationship between the position of events and the areas where vibrations could be detected by a human inside a car or a bus. These two factors show that it is possible to detect road pavement distresses with consumer grade sensors, though there are still factors which need to be further researched. Examples are the accuracy of the position generated by clustering events and the effects of adding more data into the analysis.

To what extent can the current state and the degradation of a road pavement asset be measured using mobile crowdsensing?

This research has shown that 1), it is possible to detect road pavement distresses using smartphones, 2) multiple indications of road pavement distresses can be used as a detection, 3) a road pavement distress can be connected to a nearby asset, 4) the information from multiple indications can be combined to give more information on the instance of road pavement distress.

The points above indicate that mobile crowdsensing can be used to measure the current state of a road pavement asset. In regards to Asset Management strategies, in theory, asset manager John can also use older information to see how much the road pavement has degraded as long as there is historical information available from crowdsensing. Based on this new knowledge he can decide that the maintenance of a specific asset under his supervision need to be accelerated. John is also planning to use this new information to find trends in the degradation speed so he can perform preventive maintenance in the future.

7 Discussion & future research

In the previous chapter, the subquestions and the main research questions are answered based on the knowledge gathered during this research. In this chapter, the limitations of the research and the answers given in the previous chapter are discussed. Topics for future research are proposed to further examine some of these limitations.

7.1 Discussion

The usage of crowdsensing to collect up to date information for Asset Management is a wide research area. To give this research a clear focal point and complete the research in the set time span, four real-world factors are not taken into account. These factors are the lane position of the car when gathering more data, the possibility of smartphone usage during driving, the possibility of road pavement surfaces which are not ZOAB, how people can be convinced to gather the necessary data and the processing capabilities of the smartphones.

Other factors were touched upon, but need further research. This is applicable to some of the real world factors which influence the raw data like the car suspension, direction of hit and the placement of the smartphone inside the car. Due to crowdsensing, these three factors influence the output values of for example the maximum z acceleration of a detected road pavement distress. This value can be seen as a description of the maximum z acceleration of a car which has the average properties of the factors described above driving across the road pavement distress. To gain more insight into the meaning of the values contained in an instance of road pavement distress, these averaged factors need to be defined:

1. What are the suspension properties of the average Dutch car, and how can these properties be kept up to date?
2. What is the average location of smartphones inside the car cabin?
3. How do people react to road pavement distresses, how does the way the car hits the road pavement distress affect measurements and how can this hit direction be recognized?

The main research question also pertains the degradation of the pavement of the glsasset. It has been established that it is theoretically possible to calculate the degradation as it is possible to gather information on the current state of the asset through crowdsensing. However further research is needed to verify if this theory translates well to reality.

Finally, the scope of this research is limited by the amount of data collected. By increasing the amount of data, new problems will most likely be observed. During the research there was the instance of clusters merging together due to positioning inaccuracies, the severity of a problem such as this one will most likely increase with more data. There is also the possibility that new problems will occur with more data to analyze.

7.2 Recommendations for future work

In this section topics for future research are presented. Theses topics are divided into two categories; topics which are rooted in the implementation side and topics which are suitable for academic research.

Topics on the implementation of crowdsensing

During the data gathering phase of this research, a premade app which delivers raw data in the form of a CSV file is used. The raw data provided by the app is then moved to a computer and processed before being stored in a database. An advantage of smartphones is that computing can take place locally on the device, this is also noted as one of the strengths of crowdsensing by Ganti et al. (2011). The local analysis currently implemented on a computer needs to be moved to an app which gathers the data and pre-processes it before sending the detected events to a database. 17 A challenge in crowdsensing is how to make people adopt an app which collects data on road pavement distresses. The installing of the app is an opt-in method of collecting data. People need to be stimulated into installing the app. This stimulation can also be seen as incentives for personal gain or for social gains. An example of a combination of personal and social incentives is the Waze app where the personal gain in reporting congestions results in gaining points for a gaming element and the social gain is that users are alarmed when an incident has occurred on their route (mobile, 2017). Research needs to be done on what elements need to be present in an application for road pavement distresses to provide sufficient incentive for users to install it on their smartphones.

This research is concentrated smartphones as a data source. However, the automotive industry can also be seen as a potential source for data. Cars contain a multitude of sensors including gyros and accelerometers which are already used for position dead reckoning (Skog & Händel, 2009) and are increasingly capable of sharing information due to the rise of IoT technology (Renganathan & Velaga, 2016). By collecting the data directly from the car, users do not have to install an app. Instead, the software can be installed on the car from the factory or reseller and the driver only has to give his or her permission. Another advantage is the access to other information when the car is used to collect the data, the type of car becomes a known factor. A downside of using cars to gather data is the introduction of the car producers as stakeholders. A topic for future research is how data streams from cars can be converted to events and how an application capable fo detecting the events can be made suitable for running on the board computer of the car.

Lane detection has been outside the scope of the research, however, it can provide additional information on the location of an instance of road pavement distress. Lub (2016) has implemented a lane detection algorithm during his research. By combining lane detection with event detection there is the possibility of making a prediction on which lane the road pavement distress is situated, giving the asset manager and road inspectors more information on locations of interest. Research on how the two implementations can be combined in a robust way is needed. interest. Research on how the two implementations can be combined in a robust way is needed.

Academic research topics

As stated in the Discussion section, the use of crowdsensing depends on the averages of several factors. The exact definition of these average factors is important for comparisons between historic data and gives insight into whether the values of instances of road pavement distress can be compared.

Another research topic is the real world relation between car speed, vertical acceleration, car suspension, distress size and car weight. A test set up in a controlled environment can give insight into this relationship. The resulting information allows for the translation of measurement data between different types of cars, the prediction of the size of the pavement distress and other comparisons like data translations based on speed. The information can be used to translate classifications on for example the severity of a road pavement distress to classifications for the vertical acceleration based on their relationship.

Due to the limitations of not knowing the exact location of road pavement distresses, the accuracy of the position generated trough clustering events was not able to be calculated. Future research can be focussed on the increase of positional accuracy when using more data, in turn, this information can be used to define when a cluster contains enough events to be deemed an instance of road pavement distress.

Düzgün (2017) performed research on how asset managers like John make use of new data sources. He discovered that much of the knowledge of asset managers is intrinsic and that they rely heavily on their expert knowledge instead of the new data sources. One of the conclusions is that the new data sources provide lower level information in the form of values, which the asset managers find difficult to interpret. However, Düzgün (2017) concluded that providing higher level information can be a solution to having these asset managers making more efficient maintenance decisions. Research needs to be done on how to move from lower level information like the maximum vertical acceleration, speed and duration of road pavement distress to information like the lifetime expectancy in days and severity subjective terms.

Finally, there is the temporal consistency problem between clusters. Temporal consistency between clusters simplifies the implementation of temporal analyses such as how fast is road pavement distress X degrading and what the optimal moment is to perform maintenance on an asset. A theoretical method needs to be developed on how temporal connections between a variable number of objects with positioning errors can be made accurately.

References

Android Open Source Project. (2017a). *Location*. Retrieved 2017-04-10, from https://developer.android.com/reference/android/location/Location.html

Android Open Source Project. (2017b). *Sensors Overview*. Retrieved 2017-03-06, from https://developer.android.com/guide/topics/sensors/sensors{_}overview.html

Apple Inc. (2016). *Motion Events*. Retrieved 2017-03-15, from https://developer.apple.com/library/content/documentation/EventHandling/Conceptual/EventHandlingiPhoneOS/motion{_}event{_}basics/motion{_}event{_}basics.html

Apple Inc. (2017). *Core Location*. Retrieved 2017-04-10, from https://developer.apple.com/reference/corelocation

Asphalt Institute. (2017). *Asphalt Pavement Distress Summary*. Retrieved 2017-02-24, from http://www.asphaltinstitute.org/asphalt-pavement-distress-summary/

Beijer, B. (2016). *Progress meeting at Vialis*. Haarlem.

Bridgelall, R. (2014). Connected Vehicle Approach for Pavement Roughness Evaluation. *Journal of Infrastructure Systems*, *20*(1), 4013001. doi: 10.1061/(ASCE)IS.1943-555X.0000167

Bruwer, F., & Booysen, T. (2015). Vehicle acceleration estimation using smartphone-based sensors. In *Civil engineering* (Vol. 23, pp. 51–56). doi: 10.13140/RG.2.1.3244.4640

Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., & Reddy, S. (2006). Participatory sensing. *Center for Embedded Network Sensing*, 6. Retrieved from http://escholarship.org/uc/item/19h777qd

Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., & Peterson, R. a. (2006). People-centric urban sensing. In *Proceedings of the 2nd annual international workshop on wireless internet - wicon '06* (pp. 18–31). Retrieved from http://dl.acm.org/citation.cfm?id=1234179http://sensorlab.cs.dartmouth.edu/pubs/metrosense06.pdf doi: 10.1145/1234161.1234179

Carrera, F., Guerin, S., & Thorp, J. B. (2013). By the people, for the people: The crowdsourcing of "streetbump", an automatic pothole mapping app. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *40*(4W1), 19–23. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84924338176{&}partnerID=tZOtx3y1 doi: 10.5194/isprsarchives-XL-4-W1-19-2013

Cemer, I. (2011). *Noise Measurement*. Retrieved 2017-05-10, from http://www.sensorsmag.com/embedded/noise-measurement

Cízek, P., Wolfgang, H., & Weron, R. (2005). *Kalman Filtering and Smoothing*. Retrieved 2017-01-08, from http://fedc.wiwi.hu-berlin.de/xplore/tutorials/xlghtmlnode57.html

Davis, R. (2013). *An introduction to asset management*. Retrieved from www.eatechnology.com

Debian GIS Project. (2017). *index : pkg-grass/proj-rdnap*. Retrieved 2017-05-16, from https://anonscm.debian.org/cgit/pkg-grass/proj-rdnap.git/tree/debian/rdnap

de Bruyckere, S. (2015). *Majority of the elderly in the Netherlands has a smartphone*. Retrieved 2016-12-12, from https://www.telecompaper.com/pressrelease/majority-of

-the-elderly-in-the-netherlands-has-a-smartphone--1088067
den Heijer, K., & Santinelli, G. (2014). *Python convert coordinates.* Retrieved 2017-04-20, from https://publicwiki.deltares.nl/display/OET/Python+convert+coordinates
Díaz-Vilariño, L., González-Jorge, H., Martínez-Sánchez, J., Bueno, M., & Arias, P. (2016). Determining the limits of unmanned aerial photogrammetry for the evaluation of road runoff. *Measurement: Journal of the International Measurement Confederation*, *85*, 132–141. doi: 10.1016/j.measurement.2016.02.030
Du, Y., Liu, C., Wu, D., & Jiang, S. (2014). Measurement of international roughness index by using Z -axis accelerometers and GPS. *Mathematical Problems in Engineering*, *2014*. doi: 10.1155/2014/928980
Düzgün, B. C. (2017). *New Data Sources in Road Infrastructure Management* (Unpublished doctoral dissertation). Delft University of Technology.
ESA. (2017). *Revisit and Coverage.* Retrieved 2017-06-22, from https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario
Farrell, J. (2008). Computation of the Quaternion from a Rotation Matrix. Retrieved from http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Computation+of+the+Quaternion+from+a+Rotation+Matrix{#}0
Forslöf, L., & Jones, H. (2015). Roadroid: Continuous Road Condition Monitoring with Smart Phones. *Journal of Civil Engineering and Architecture*, *9*(4), 485–496. Retrieved from http://www.davidpublisher.org/index.php/Home/Article/index?id=7647.html doi: 10.17265/1934-7359/2015.04.012
Fujino, Y., Kitagawa, K., Furukawa, T., & Ishii, H. (2005). Development of Vehicle Intelligent Monitoring System (VIMS) Yozo. In M. Tomizuka (Ed.), *Smart structures and materials 2005: Sensors and smart structures technologies for civil, mechanical, and aerospace systems* (Vol. 5765, pp. 148–157). Retrieved from http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=862371 doi: 10.1117/12.601727
Ganti, R., Ye, F., & Lei, H. (2011). Mobile crowdsensing: Current state and future challenges. *IEEE Communications Magazine*, *49*(11), 32–39. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-81355138524{&}partnerID=40{&}md5=4c0b9f4b26ffd71cd8915cbea38816a6 doi: 10.1109/MCOM.2011.6069707
Gehlot, S., Ketelaar, V. B. H., Verbree, E., & Hanssen, R. F. (2005). Conceptual Framework for {PS-InSAR} Deformation Interpretation Assisted by Geo-Information Technology. *ISPRS Workshop, High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany, 17–20 May 2005*, 5. Retrieved from http://doris.tudelft.nl/Literature/gehlot05.pdf
Gillespie, T. D. (1981). *TECHNICAL CONSIDERATIONS IN THE WORLDWIDE STANDARDIZATION OF ROAD ROUGHNESS MEASUREMENTS* (Tech. Rep.). Michigan: Highway Safety Research Institute The University of Michigan. Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/366
Grewal, M. S. (2011). Kalman Filtering. In *International encyclopedia of statistical science* (pp. 705–708). Berlin: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/referenceworkentry/10.1007{%}2F978-3-642-04898-2{_}321 doi: 10.1007/978-3-642-04898-2_321
Ibrahim, M., & Moselhi, O. (2016). Inertial measurement unit based indoor localization for construction applications. *Automation in Construction*(71), 13–20. doi: 10.1016/j.autcon.2016.05.006
ISO/TC 251 Asset management. (2014). *ISO 55000:2014 Asset management — Overview, principles and terminology.* Geneva: International Organization for Standardization. Retrieved from https://www.iso.org/standard/55088.html
Jazar, R. N. (2008). Quarter Car. In *Vehicle dynamics: Theory and application* (pp. 931–975).

Boston, MA: Springer US. Retrieved from http://dx.doi.org/10.1007/978-0-387-74244-1{_}15 doi: 10.1007/978-0-387-74244-1_15

Jin, L., Han, M., Liu, G., & Feng, L. (2015). Detecting Cruising Flagged Taxis' Passenger-Refusal Behaviors Using Traffic Data and Crowdsourcing. *Proceedings - 2014 IEEE International Conference on Ubiquitous Intelligence and Computing, 2014 IEEE International Conference on Autonomic and Trusted Computing, 2014 IEEE International Conference on Scalable Computing and Communications and Associated Sy*, 18–25. doi: 10.1109/UIC-ATC-ScalCom.2014.110

Kantardzic, M. (2011). Cluster Analysis. In *Data mining: Concepts, models, methods, and algorithms,* (Second Edi ed., pp. 249–279). Hoboken, NJ: John Wiley & Sons. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781118029145.ch9/summary doi: 10.1002/9781118029145

Karamanou, A., Papazissi, K., Paradissis, D., & Psarianos, B. (2009). Precise Estimation of Road Horizontal and Vertical Geometric Features Using Mobile Mapping Techniques. *Boletim de Ciencias Geodesicas, 15*(5), 762–775. Retrieved from http://apps.webofknowledge.com/full{_}record.do?product=UA{&}search{_}mode=GeneralSearch{&}qid=1{&}SID=Q1m51rtTqWGQYQrIIlx{&}page=1{&}doc=1

Klunder, G., van Aalst, W., Willemsen, D., van Vliet, D., Leegwater, G., & Ruijs, P. (2010). *Auto als Sensor – Quick-scan naar detectiemogelijkheden van wegdekschade Inhoudsopgave* (Tech. Rep.). Delft: TNO.

Kos, A., Umek, A., & Tomaic, S. (2016). Comparison of Smartphone Sensors Performance Using Participatory Sensing and Cloud Application. *Proceedings - 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things, IIKI 2015*, 181–184. doi: 10.1109/IIKI.2015.45

Leijten, M., & Koppenjan, J. F. M. (2010). Asset management for the dutch railway infrastructure. In *3rd international conference on next generation infrastructure systems for eco-cities, infra 2010 - conference proceedings* (pp. 1–5). Shenzhen. doi: 10.1109/INFRA.2010.5679222

Lemmens, M. J. P. M. (1991). GIS: The Data Problem. In *European conference on geographical information systems* (pp. 626–636). Delft: Faculty of Geodesy.

Liang, S. (2016). *Comparison of SensorThings API and Sensor Observation Service.* Retrieved 2017-03-29, from http://www.sensorup.com/blog/tutorials/comparison-of-sensorthings-api-and-sensor-observation-service/

Lub, R. (2016). *Smartphones als meetinstrument voor langsonvlakheid* (Master Thesis). University of Twente.

Mathworks. (2017). *Quaternions to Direction Cosine Matrix.* Retrieved 2017-04-07, from https://nl.mathworks.com/help/aeroblks/quaternionstodirectioncosinematrix.html

Maxim Integrated Products. (2002). *Filter Basics: Anti-Aliasing.* Retrieved 2017-04-19, from https://www.maximintegrated.com/en/app-notes/index.mvp/id/928

mobile, W. (2017). *Get the best route, every day, with real–time help from other drivers.* Retrieved 2017-05-19, from https://www.waze.com/

Mohan, P., Padmanabhan, V. N., & Ramjee, R. (2008). Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. *Proceedings of the 6th ACM conference on Embedded network sensor systems - SenSys '08*, 323. Retrieved from http://research.microsoft.com/pubs/78568/Nericell-Sensys2008.pdf{%}5Cnhttp://research.microsoft.com/en-us/people/padmanab/nericell-sensys2008.pdf{%}5Cnhttp://portal.acm.org/citation.cfm?doid=1460412.1460444 doi: 10.1145/1460412.1460444

Open Geospatial Consortium. (2016). *OGC SensorThings API Part 1: Sensing.* Open Geospatial Consortium. Retrieved from http://docs.opengeospatial.org/is/15-078r6/15-078r6.html{#}30

Open Geospatial Consortium. (2017). *SWE Common Data Model Encoding Standard.* Retrieved 2017-03-28, from http://www.opengeospatial.org/standards/swecommon

Pohl, I., Schenk, S., & Bilt, P. V. D. (2013). *PPP in Rail Infrastructure – The Dutch Case* (Tech. Rep.). Rotterdam: Rebel Group. Retrieved from www.rebelgroup.be/file/244

Qin, H., Rice, R. M., Fuhrmann, S., Rice, M. T., Curtin, K. M., & Ong, E. (2015). Geocrowdsourcing and accessibility for dynamic environments. *GeoJournal, 81*(Golledge 2001), 699–716. Retrieved from http://link.springer.com/article/10.1007/s10708-015-9659-x{%}5Cnhttp://link.springer.com/10.1007/s10708-015-9659-x doi: 10.1007/s10708-015-9659-x

Remondino, F., Barazzetti, L., Nex, F., Scaioni, M., & Sarazzi, D. (2011). Uav Photogrammetry for Mapping and 3D Modeling – Current Status and Future Perspectives –. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVIII-1*(C22), 25–31. doi: 10.5194/isprsarchives-XXXVIII-1-C22-25-2011

Renaudin, V., & Combettes, C. (2014). Magnetic, acceleration fields and gyroscope quaternion (MAGYQ)-based attitude estimation with smartphone sensors for indoor pedestrian navigation. *Sensors (Switzerland), 14*, 22864–22890. doi: 10.3390/s141222864

Renganathan, J., & Velaga, A. S. (2016). Converting Obsolete Data into Progressive Data (COP) to assist Lonely Drivers. In *International conference on advances in electrical, electronics, information, communication and bio-informatics* (pp. 1–5).

Research & Development division of the Highway Department. (2013). *Guidance Notes Catalogue of Road Defects* (Tech. Rep. No. January). Hong Kong: Highway Department. Retrieved from https://www.hyd.gov.hk/en/publications{_}and{_}publicity/publications/technical{_}document/guidance{_}notes/pdf/GN015B.pdf

Rijkswaterstaat. (n.d.). *De inspectie van een brug.* Retrieved 2016-11-28, from https://staticresources.rijkswaterstaat.nl/binaries/ArtikelInspectievaneenbrug{_}tcm21-65315.pdf

Rijkswaterstaat. (1996). *De Aran in actie op RW4 deze wordt gebruikt voor langs dwars en vlakheidsmeting van het wegennet ID139600.* Retrieved 2017-01-05, from https://beeldbank.rws.nl/MediaObject/Details/139600

Rijkswaterstaat. (2017a). *Index of /geodata/dmc/nwb-wegen/geogegevens/shapefile/Nederland_totaal.* Retrieved 2017-05-03, from https://www.rijkswaterstaat.nl/apps/geoservices/geodata/dmc/nwb-wegen/geogegevens/shapefile/Nederland{_}totaal/

Rijkswaterstaat. (2017b). *Index of /geodata/dmc/weggeg/geogegevens/shapefile/weggeg_wegvakniveau.* Retrieved 2017-05-03, from Indexof/geodata/dmc/weggeg/geogegevens/shapefile/weggeg{_}wegvakniveau

Rijkswaterstaat. (2017c). *Zoab.* Retrieved 2017-03-08, from https://www.rijkswaterstaat.nl/wegen/wegbeheer/aanleg-wegen/zoab.aspx

Rijkswaterstaat GPO m.m.v. Witteveen + Bos. (2015). *Richtlijn Ontwerp Autosnelwegen 2014* (Tech. Rep.). Rijkswaterstaat. Retrieved from https://staticresources.rijkswaterstaat.nl/binaries/RichtlijnOntwerpAutosnelwegen2014{_}tcm21-76103.pdf

Roadroid. (2017). *Roadroid Map.* Author. Retrieved from http://www.roadroid.com/Map

Sachs, D. (2010). *Sensor Fusion on Android Devices: A Revolution in Motion Processing.* GoogleTechTalks. Retrieved from https://www.youtube.com/watch?v=C7JQ7Rpwn2k

Sayers, M. W., & Karamihas, S. M. (1998). *The little book of profiling - Basic Information about Measuring and Interpreting Road* (No. September). UMTRI - Transport Research Institute of the University of Michigan. Retrieved from http://www.umtri.umich.edu/content/LittleBook98R.pdf

Schultz van Haegen, M. (2016). *Beantwoording aanvullende vragen van het lid Bruins (ChristenUnie) over de sluiting van de Merwedebrug A27 voor vrachtverkeer.* The Hague:

Ministerie van Infrastructuur en Milieu.

Skog, I., & Händel, P. (2009). In-car positioning and navigation technologiesa survey. *IEEE Transactions on Intelligent Transportation Systems, 10*(1), 4–21. doi: 10.1109/TITS.2008.2011712

Smith, S. W. (1997). ADC and DAC. In *The scientist and engineer's guide to digital signal processing* (1st ed., pp. 35–66). Retrieved from http://www.dspguide.com/ch3.htm

Spruyt, V. (2016). *Driving behavior modeling using smart phone sensor data.* Retrieved 2017-03-14, from http://www.sentiance.com/2016/02/11/driving-behavior-modeling-using-smart-phone-sensor-data/

Stasch, C., Pross, B., & Jirka, S. (2017). Processing in the Sensor Web – Coupling SOS and WPS. In *Ogc tc meeting, delft 21.03.2017* (p. 17). Delft: 52 North.

STDHA. (2017). *DBSCAN: density-based clustering for discovering clusters in large datasets with noise - Unsupervised Machine Learning.* Retrieved 2017-04-24, from http://www.sthda.com/english/wiki/dbscan-density-based-clustering-for-discovering-clusters-in-large-datasets-with-noise-unsupervised-machine-learning

van der Heijde, M., & Ligtvoet, B. R. (2017). *Phone conversation with Marcel van der Heijde on the 1st of March.* Zwijndrecht.

Volvo. (2017). *Volvo 8900 Specificaties.* Retrieved 2017-05-09, from http://www.volvobuses.be/nl-be/our-offering/buses/volvo-8900/specifications.html

Waite, A., Walsh, R., & Garcia, D. (2012). *Enabling a Secure Environment for V2V and V2I Transactions - Public Workshop* (No. June). McLean: U.S. Department of Transportation.

Wavefront Lab. (2010). *Sensor Data.* Retrieved 2017-05-09, from http://www.wavefrontlabs.com/Wavefront{_}Labs/Sensor{_}Data.html

Wilson, J. (1999). *A Practical Approach to Vibration Detection and Measurement Part 1: Physical Principles and Detection Techniques.* Sensors Online. Retrieved from http://www.sensorsmag.com/components/a-practical-approach-to-vibration-detection-and-measurement-part-1-physical-principles

Wu, J., & Hu, F. (2015). Ground subsidence monitoring of high speed roads by using PS-InSAR method. In *Proceedings of the 2015 ieee 5th asia-pacific conference on synthetic aperture radar, apsar 2015* (pp. 853–858). Singapore. doi: 10.1109/APSAR.2015.7306337

Young, G. (1989). Geodetic leveling. In *Geophysics* (pp. 470–474). Boston, MA: Springer US. Retrieved from http://dx.doi.org/10.1007/0-387-30752-4{_}56 doi: 10.1007/0-387-30752-4_56

Zheng, Z., Han, S., Yue, J., Yuan, L., & Wo, Y. (2016). Compensation for Stochastic Error of Gyros in a Dual-axis Rotational Inertial Navigation System. *The Journal of Navigation*(69), 169–182. doi: 10.1017/S037346331500051X

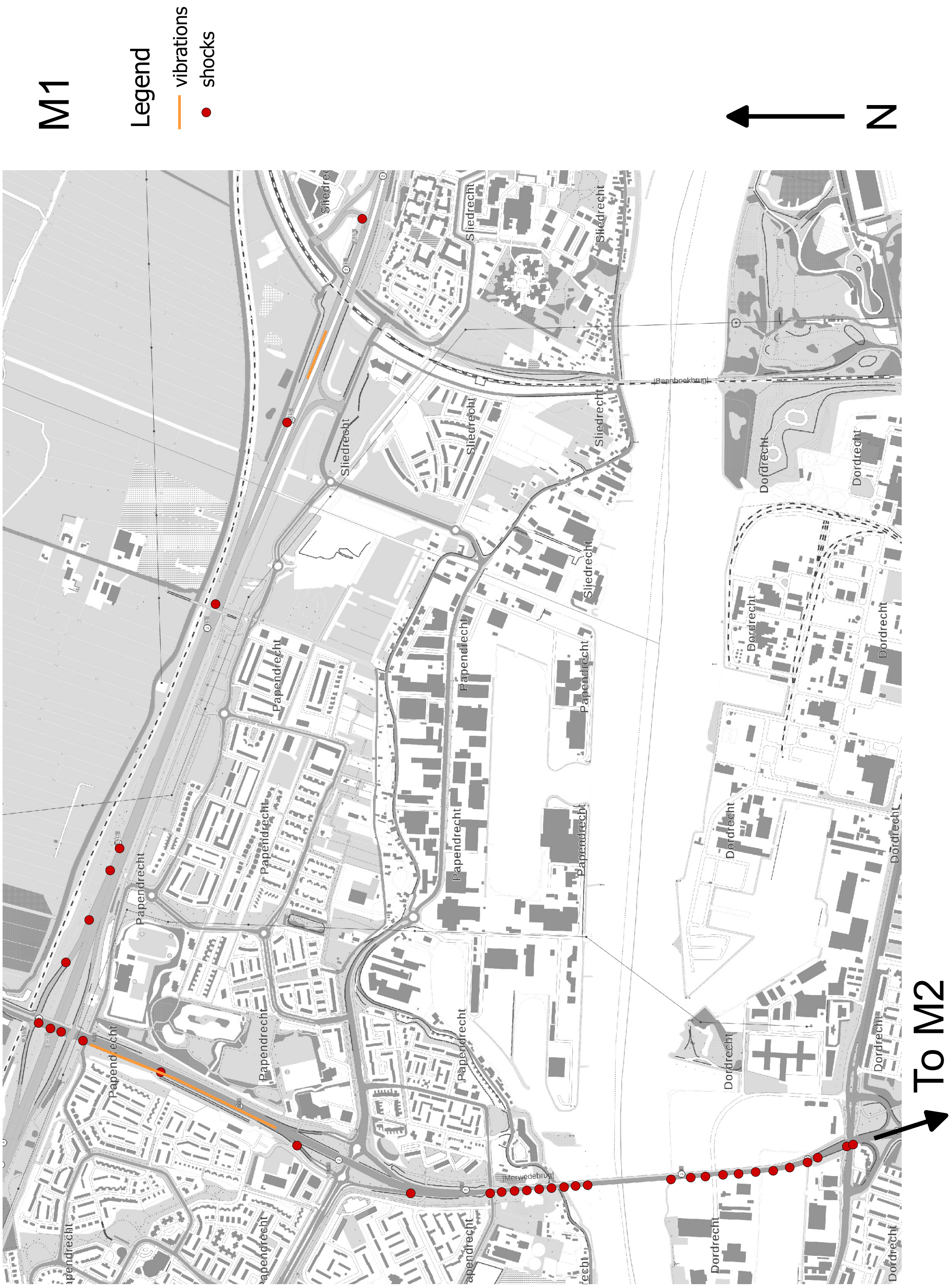
Appendices

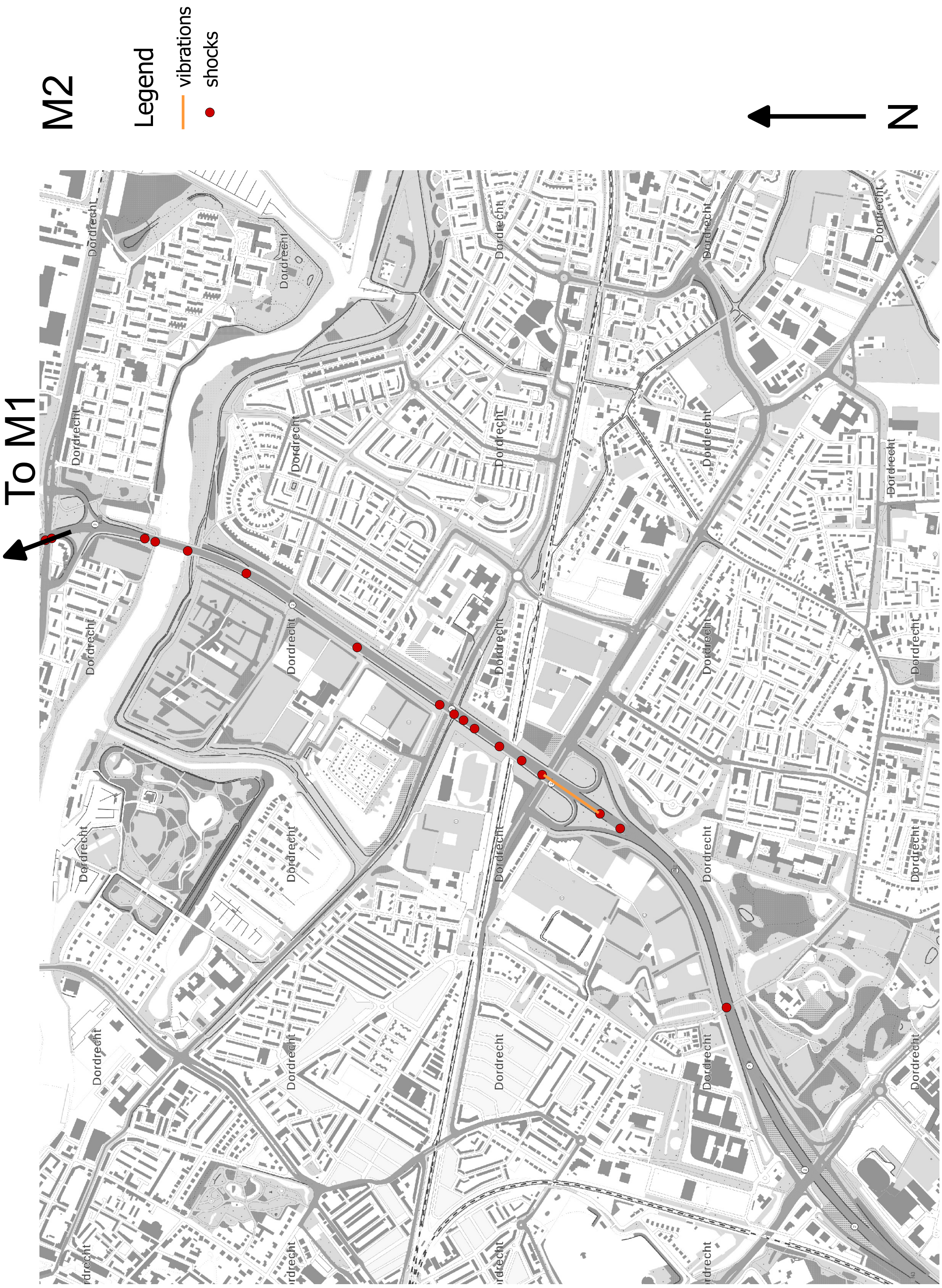
A Phone conversation with Marcel van der Heijde

This appendix contains a bullet point summary of a phone conversation with Marcel van der Heijde of Techno Fysica B.V, a company that specializes in applying high accuracy sensors to collect data on vibrations and strains in for example parts of boat engines. The conversation took place on the 1st of March 2017.

- A major problem with randomly trying to measure something is that you cannot know for sure that what you are measuring contains the information you need. There is the possibility that you are measuring noise or a completely different phenomenon. The other possibility is that what you are trying to measure has the same frequency as the noise you are also measuring, making it difficult to separate the two factors. Therefore you need to have knowledge of what you can expect to see in the data you are collecting as well as knowledge on the phenomenon you are trying to measure.
- When using prototype filters you need to keep in mind the following points:
  - The frequency characteristics of a Butterworth filter are 3 Db point.
  - To test the frequency response of the filter, numerical test data can be used to emulate a pulse, a hard change, and a sinusoid.
  - By using a filter, a phase shift takes place on the output data.
- There is also the possibility to filter in the frequency spectrum through Fast Fourier Transform (FFT). If this is done, then the x-axis is changed from 2πF to m/s. When using an FFT there are some points to take into account:
  - The time series needs to be near constant, i.e. a repeating signal.
  - An FFT needs to have a start and end value of 0 in the measurement values. If this is not true for the input data, then these first and last values are changed to fit this condition for performing an FFT.
- Using an FFT on a signal which is neither constant nor has a fitting start and end is not recommended. The data is changed to fit the requirements of an FFT and any errors which are made during the filtering in the frequency domain are unrecognizable in the data when moving back to the time series domain.

B Tally results for A15 and N3


M1
Legend
vibrations
shocks
N
To M2
Papendrecht
Sliedrecht
Dordrecht
Merwedebrug


M2
Legend
vibrations
shocks
N
To M1
Dordrecht

C Reflection

This appendix consists of a reflection on the research process and planning during the graduation and the resulting product. Topics on the relevance of the research in the field of geomatics and the real world are discussed based on the 3 topics provided in the Master Geomatics graduation manual of 2016-2017.

Birgit Ligtvoet 4096614

Process

The subject for this graduation research was proposed by myself to my mentors. I enjoyed the subject, however using a self-chosen subject does bring some challenges with it. One major hurdle was the focus of the research. The terms crowdsensing and roads only give a general direction which does not necessarily have anything to do with Geomatics specifically. This also caused problems during the P2 presentation where the actual need for the research remained unclear together with how the subject would fit with the field of Geomatics.

The chosen focus area of the research encompasses many different subjects next to subjects which can be ascribed to Geomatics. The major example is the signal analysis performed during the research. The broad scope of the research oftentimes caused confusion on what parts were specific to the field of Geomatics.

Planning

Gantt charts were made multiple times throughout the graduation. However, they were not followed strictly, instead, they were used more as a guideline. In certain cases, tasks were exchanged to maintain a high motivation when faced with a task which had slowed down due to small setbacks. The schemas did help with keeping an inventory of uncompleted tasks. Another personal advantage was that the tasks in the charts were all allotted te amount of time needed in a worst case scenario. While this caused some distress for all parties, mentors included, it also created a clear focus to get all the work done.

Product

I am happy with the result of the research, a personal goal at the start of the research was to get to information which can be viewed in a GIS package, which has been achieved. Having information which can be viewed in a GIS package enabled me to perform a visual analysis on the results of the analysis steps. I would have liked to have more time to test the precision and accuracy of some of the analyzed information, though that was not necessarily the focus of the research. Another subject which I would have liked to explore further was the effects of large amounts.

The relationship between the methodological approach of the Master Geomatics and the method chosen in this research

The methodological approach of the Master Geomatics of the built environment is described as containing a part data processing, a part analysis, and a part visualization. These three points are applicable to this research though the focus has been more on the data gathering and the analysis. The visualization is mainly used to verify the results in for example the positions of the virtual sensors. However, the resulting information from this research has been prepared to be loaded into GIS packages.

Another relationship between methodologies is the usage of the GIS chain proposed by Lemmens. The GIS chain consists of 7 parts: the world, the data capture, the pre-processing, the analysis the presentation, the quality control, and the resulting management steps. During this research data was captured to monitor a real world phenomenon, this data was then preprocessed during the local analysis phase, the analysis step took place during the database analysis, quality control was performed both the pre-processing and analysis steps and the results were visualized for extra insight into the resulting data.

The relationship between the research and the field of Geomatics

The usage of crowdsensing to detect distresses brings certain challenges with it. While different research papers on using crowdsensing have been published, they lack specific insights which are discussed during the Master Geomatics for the Built Environment. Some examples are the idea that too much data is not always desirable, the usage of spatial data standards for the exchange of information, and the inclusion of extra information on the existing semantics of road assets which can be used to influence the decision-making process for the maintenance planning in a direct or indirect fashion.

The relationship between the project and the wider social context

The relationship between this research and the wider social context is also the main reason this research is of interest to contractors. The building sector is switching to a new type of contract where maintenance is integrated into the brief of the tenderer. The goal of this research is to provide information which can be used to gain new insights into the current state of a road asset. This information can eventual lead to better maintenance plans, saving money for the contractor, decreasing the hindrance to road users, increasing the safety of road users, decreasing the amount of damage caused by bad roads and possibly decreasing congestion.