Virtual Ground Truth in Vehicular Sensing Experiments: How to Mark it Accurately

Girts Strazdins*†, Artis Mednis*†, Reinholds Zviedris*†, Georgijs Kanonirs*, Leo Selavo*†
*Digital Signal Processing Laboratory
Institute of Electronics and Computer Science
14 Dzerbenes Str., Riga, LV 1006, Latvia
†Faculty of Computing
University of Latvia
19 Raina Blvd., Riga, LV 1586, Latvia
Email: {firstname.lastname}@edi.lv

***Abstract*—Road surface quality monitoring is an important requirement for efficient, safe and comfortable transportation. However, the data collection is made difficult by the scope of the data source. Therefore, participatory sensing is a promising approach for road damage assessment. We are developing a vehicular participatory sensing application using Android smart-phones for pothole detection. This paper describes lessons learned from our field tests, which have exposed the deficiencies in terms of collected data quality. Nevertheless, the tests provide invaluable experience for planing future field tests and improvements to the test execution procedure for vehicular sensing researchers. Based on empirical and analytical results, we conclude, that semi-automated ground-truth reference point recording by a human observer in a moving vehicle while doing the actual data collection is imprecise as a consequence of multiple technical and human factors. We also discuss the motivation, why careful pothole position marking and categorization by walking along the test track is capable of providing highly accurate ground-truth.**

***Keywords*-real-world experiment experience; data quality; vehicular sensing; participatory sensing; Android OS.**

I. Introduction

Road surface damage (potholes, bumps, gaps, etc.) are a serious issues causing distractions for safe and comfortable transportation. Both drivers and road maintainers are interested to fix the problems as soon, as possible. To fix them, they first have to be identified. Centralized road inspection is difficult due to the scope of road infrastructure. Several pothole reporting systems already exist, including web sites, such as [1] [2]. However, they rely entirely on manual reports of individuals. The requirement for manual human interaction implies the low report rate, which, we believe, can in turn be increased dramatically by automated pothole detection and reporting systems.

We are developing a pothole detection system based on participatory sensing using Android smartphones in driving vehicles. Accelerometers and Global Positioning System (GPS) are used to detect and geotag potholes encountered during the ride. The system is envisioned to be added as a service to navigation systems, such as Waze [3], which people use on daily basis. That would enable large-scale pothole detection, reports to responsible authorities as well as publicly available pothole visualization map. Systems architecture and general principles are described in the Section II.

To evaluate our approach, a set of field tests (described in Section III) with multiple Android-based smartphones in a driving vehicle were performed and acceleration sensor data for more than five hours of driving were collected. The analysis of the data revealed both positive and negative results. On one hand, the initial ground-truth marking methodology was deficient. On the other hand invaluable knowledge about vehicular sensing experiment planning, execution and accurate ground-truth marking, and practical experience with multiple, distinct, Android devices was gained. We believe, that our concluding recommendations (described in Section IV) will improve the quality and reduce the time investment for experimental evaluation process of vehicular sensing application researchers. We state the proposed recommendations as our main contribution in this paper.

II. System Architecture

The system (see Figure 1) consists of Android smartphones located in vehicles, acting as mobile agents in urban environment; and central server, which aggregates the data and provides web interface. 3-axis accelerometers are required for pothole detection, GPS for event geotagging and communication (either WiFi or Cellular) for data exchange with the server. The requirements are not too restrictive, as the components are available on most of Android phones.

The phone should be either fully charged or connected to a car charger, as the data collection process may require significant amount of energy. The most consuming components are the touchscreen, which can be turned off if no user interaction is required, and GPS, which can be optimized by combination of other localization methods [4].

The sensor sampling and event detection (in the particular case - potholes, but other types of events can be supported in general) is performed on the phone. Only reported events

 ISBN: 978-1-61208-144-1

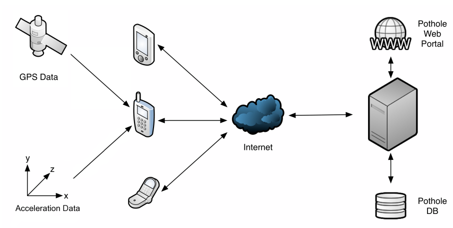


Figure 1. Our system architecture. Mobile phones collect sensor data, extract events which are exchanged with the central server, where data is aggregated and web interface is provided.

are sent to the server, without raw sensor data. The server aggregates reported data and sends back other participant collected events upon request. Hence the data transmission channel utilization is very low, and the system has high scalability, which is an important factor for platforms intended for public and large-scale usage.

The system is intended for use in urban environments, at speeds up to 70km/h ($\approx$ 37mph).

The event detection involves accelerometer data processing. We have adapted multiple algorithms from our own [5] and other researcher prior work experience. Briefly, our algorithms include vertical-axis acceleration and its standard deviation thresholding by amplitude. More in depth algorithm analysis is out of the scope of this paper.

III. Field Tests

To evaluate the approach, data collected in real-world urban environment is required. Therefore a set of field tests were performed. Data was collected by multiple Android devices in a driving vehicle. Multiple drives were performed. Additionally, a laptop with external GPS device and a microphone was used as a reference, detecting potholes by audio data processing, using RoadMic methodology, described in our previous work [6].

Online event detection and audio notification was implemented for debugging reasons. Additionally, all the raw data was recorded for offline processing and analysis, to tune and assess the detection algorithms and collected data quality.

The test track (Figure 2) was selected due to three significant features it possesses: short enough to repeat multiple laps; diverse road segments, both very smooth and very rough; located in a realistic urban environment.

A. Test setup

The experimental setup and collected data characteristics is shown in Table I. Overall, the track is 4.4km long (2.73 miles). Five different Android smartphones were used: Samsung i5700, HTC Desire, Samsung Galaxy S, HTC Desire Z, and HTC HD2. Five different test drives were performed (25th and 28th of January, 10th and 28th of February, and 24th of March, 2011), using three different vehicles: BMW 323 Touring, Mitsubishi Space Wagon, and

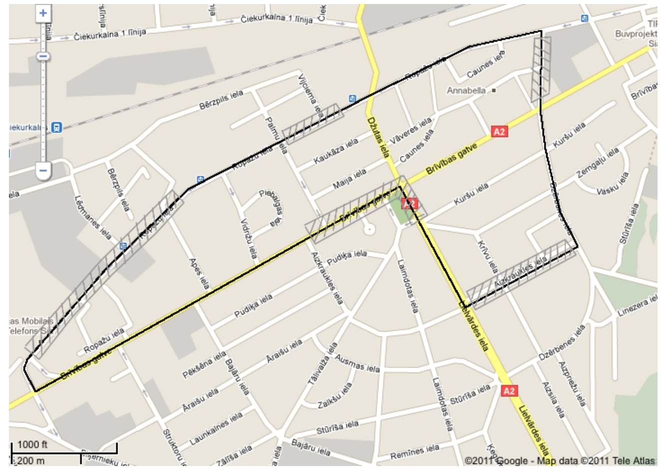


Figure 2. Used experimental test track, 4.4km long, single and multi lane streets in urban environment. Regions marked have most of the potholes.

Mazda 323F. Besides accelerometer sensor, microphone data was recorded for future use. The first and second drives contained 3 test laps each, the third - one lap. The fourth and fifth drives contained 10 laps each.

During the first two drives no ground-truth was marked and the corresponding collected data is therefore only usable to get an impression of Android smartphone peculiarity, distinct device, vehicle and environmental factor impact on sensor data quality.

The goal of the third drive was an insight into ground-truth marking. A PC laptop with an external GPS and our custom built ground-truth marking application was used in the same driving vehicle, where Android phones were collecting accelerometer data. A human operator pressed spacebar each time he experienced a street damage initiated shake, and the software recorded local system time (with millisecond accuracy) accordingly. Offline pothole detection was performed, using multiple detection algorithms on the data collected in the third drive. The detected events were compared against ground-truth points. The results were unsatisfactory - only about 65% of the detected events were in ground-truth point vicinity. Unfortunately, it was unable to distinguish whether the source of the error is ground-truth inaccuracy, detection algorithm inappropriateness, or data set size. And the situation on the road had changed since the first drives, the ground-truth was not usable as a reference for the older data.

Therefore the fourth drive was performed, which was planned to overcome the drawbacks of the first drives. Ten laps of data collection and ground-truth recording with four different Android devices were performed, a total of 44 kilometers, more than two hours of data. Unfortunately inconsistencies in the used semi-automated ground-truth marking in the driving vehicle were discovered. It led to a conclusion, that manual ground-truth pothole marking is a

 ISBN: 978-1-61208-144-1

Table I
TEST DRIVE CHARACTERISTICS

Date	Vehicle	Phones	Laps	Minutes	Kilometers	Ground truth
2011-01-25	BMW 323 Touring	Samsung i5700, HTC Desire, Samsung Galaxy S	3	32	13.2	No
2011-01-28	Mitsubishi Space Wagon	Samsung i5700, Samsung Galaxy S	3	35	13.2	No
2011-02-10	Mazda 323F	Samsung i5700, HTC Desire	1	11	4.4	Yes, incomplete
2011-02-28	Mazda 323F	Samsung i5700, HTC Desire, Samsung Galaxy S, HTC Desire Z	10	131	44.0	Yes, incomplete
2011-03-24	BMW 323 Touring	Samsung i5700, HTC Desire, Samsung Galaxy S, HTC HD2	10	119	44.0	Yes, complete
Total	3	5	27	328	118.8	3

better method for accurate reference point selection. A more in-depth analysis is described in Section III-B.

The fifth drive was performed in conjunction with manual ground-truth marking. A total of 108 potholes were marked on the 4.4km long test track (Figure 4c). The points were recorded on pedestrian sidewalks and GPS signal was interrupted by high buildings. Therefore offline position correction was performed, by calculating simple perpendiculars to the mean trajectory of all the 10 laps of the fifth test drive, as shown in Figure 3. Preliminary analysis shows that collected data and marked ground-truth in the fifth drive are sufficiently accurate to perform further pothole detection algorithm evaluation.

The next two subsections describe the negative and positive inferences of the collected data.

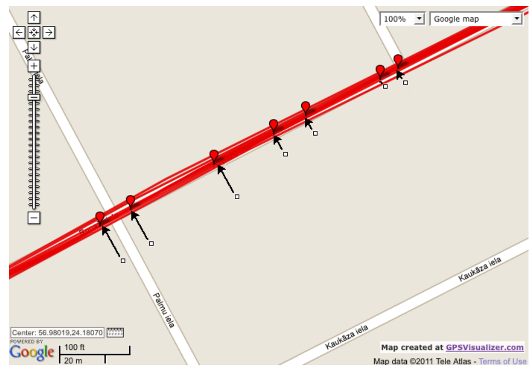


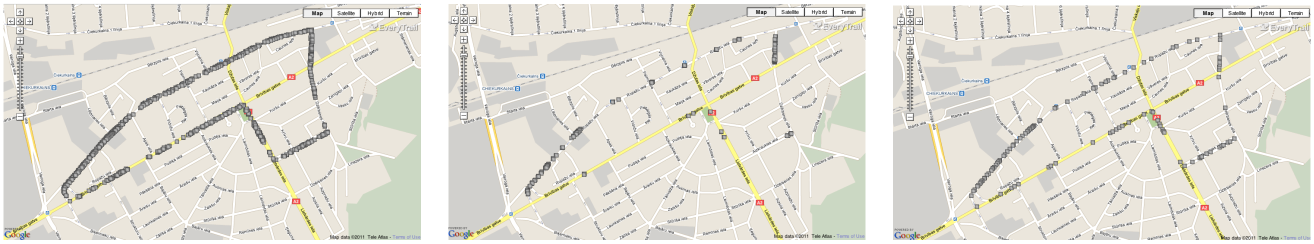
Figure 3. Manually marked ground-truth position improvement, by calculating perpendiculars to the test drive trajectory

B. The Dark Side

The offline data processing revealed several deficiencies of the fourth test drive. Semi-automated ground-truth marking mechanism seemed to be attractive due to two reasons. First of all, in such a way only the encountered potholes would be recorded. Second, it seemed a natural approach, compared to additional two hour walk with manual point-of-interest marking on the GPS device (*Walking GPS* approach [7]). As it turned out, the capabilities of semi-automated ground-truth point marking were overrated. It was imperfect both in terms of accuracy and time-efficiency.

Overall 1326 locations were marked during the 131 minute drive, shown in Figure 4a. Such a set of points cannot be used as ground-truth directly, therefore it was refined: only the points which had at least c other points (from other, distinct, laps) in their vicinity (no further than d meters) remained. A total of 273 ground-truth locations (21% of initial 1326 locations) remained after the refinement procedure with parameters $c = 6$ and $d = 5$ (Figure 4b). Although visually the refined locations correspond to expectations, more in depth analysis showed, that their usability is doubtful due to the following reasons:

1) There is no classification possibilities. All the potholes are considered equal, regardless of their actual size and significance. Multiple pothole type support in ground-truth marking software would be practically useless, as the human operator would not be able to categorize the potholes in a short period of time.
2) There is always a distance between the real pothole and the location of the button press, which is, unfortunately, undetermined and cannot therefore be eliminated by simple shift operations.
3) During a two hour drive the human ground-truth marking operator is loosing attention. Hence the last laps have less accurate ground-truth locations.
4) Lack of precisely defined methodology (when to press the button) leads to inconsistent results - a particular pothole is recorded in some laps, ignored in the rest.
5) GPS inaccuracy has a greater impact compared to Walking GPS approach, where it can be mitigated by standing a while at the same location and allowing GPS signal to stabilize.
6) The amount of shake the human perceives depends not only on the size and type of road irregularity, but also on vehicle speed and technical condition. Hence insignificant potholes might get recorded and significant ones ignored.
7) During the fourth test drive audio signal on a laptop was recorded and pothole detection by previously evaluated and reliable RoadMic [6] methodology was performed. And the results contained contradiction: while increasing algorithm threshold, detection accu-



(a) 1326 ground-truth points recorded semi-automatically during the 131 minute long 4th drive

(b) 273 refined ground-truth points

(c) 108 manually marked ground-truth points

Figure 4. Ground-truth, acquired semi-automatically in the fourth drive (a), its refined version (b) and manually marked ground-truth of the fifth drive (c)

racy had to increase and reach 100%, but it decreased. That caused a strong suspicion about the ground-truth accuracy.

8) Needless to say, the refinement procedure design, implementation and tuning took more time and effort than would Walking GPS approach require.

The arguments listed above led to an unpleasant conclusion - the collected ground-truth is inaccurate and cannot be used as a reference for further algorithm analysis.

Another problem was related to incorrect data storage. It was assumed, that acceleration values will not exceed $32m/s^2$ ($\approx 3.27g$). They were stored in 16-bit variables, supporting values -32.768..32.767m/s^2. However, Samsung i5700 reported values greater than $32m/s^2$ (which, we believe, is a bug of this specific device). It led to a number of overflows, which required additional time and effort.

Besides systematic problems unexpected ones were also encountered. Two of four Android devices experienced a reboot during the fourth ride. The data collection was restarted as soon as possible, but significant amount of data was lost: 47 and 16 minutes. The reboot reason is unclear, but it is probably related to incorrect OS handling of intense audio data recording, requiring large data buffers and fast flushing to SD card. In the fifth drive, audio recording was disabled. But it turned out to be a bad design decision, as it involved last-minute code modifications and consecutive bugs which resulted in partial data loss for two devices: 47 and 20 minutes. Additionally, one of the phones ran out of battery in the fourth drive, 19 minutes of data were lost.

Preliminary audio data analysis showed, that RoadMic approach cannot be transferred to Android phones directly due to dynamic range compression used, but further analysis must be performed to evaluate potential of pothole detection from audio signal using Android phones.

C. The Bright Side

Although the collected data did not satisfy all of our research needs, it was valuable for a number of reasons.

First of all, an insight into Android OS impact on sensor data collection was acquired. On one hand, Android handles a lot of processes which had to be done manually on an embedded system. It converts raw values to SI system, handles chip management, provides simple programming interface. On other hand, it limits the freedom available on customized embedded devices. Accelerometer sensor data is reported *as fast as possible*, without any guarantees of minimum or maximum latency. The achieved sampling rate was relatively low and device dependent.

The most unexpected difference between Android devices was accelerometer sensor output. Although the four devices have two common vendors (Samsung and HTC), every device had different sensor sampling rate, ranging from 26Hz up to 98Hz, see Table II. The sensor sensitivity and noisiness was also different. For comparison, we calculated standard deviation for vertical axis acceleration for the same 500 second time period from the fifth drive (fourth drive for HTC Desire Z, as it was not present on the fifth drive) for all the devices. Samsung i5700 has significantly higher deviation compared to other devices.

GPS accuracy was better than we expected, having mean Android-reported error under 7 meters with more than 7 satellites visible on average (Table II).

The diversity of the test drive environment (snow and potholes on the road varied a lot during the two month period) and setup (different vehicles) provide data for vehicle and environment variety impact assessment. But it will only be usable, when the correctness of pothole detection algorithms will be proved.

And, last but not least, the rather negative experience transforms into conclusions on how to perform the experiments in higher quality and with higher success rate.

IV. Experience and Recommendations

This section describes recommendations based on the performed vehicular sensing experiments on the Android platform.

1) Do not mark ground-truth positions in a driving vehicle: detailed motivation is described above, in Section III-B. The recommended solution is to walk along the test track with a GPS device, stay a couple of seconds at each position of interest allowing GPS signal to stabilize, record the

Table II
SENSOR DATA DIFFERENCES BETWEEN DISTINCT ANDROID SMARTPHONES. AVERAGED OVER 10 MINUTE DRIVE.

Device	Accelerometer sampling rate, Hz	Accelerometer Z-axis StdDev, g	GPS accuracy, m	GPS locations missing, %	GPS visible satellites	GPS avg. visible satellites
Samsung i5700	26	0.3076	-	4	4-9	6.99
HTC Desire	52	0.1215	4.21	0	4 - 11	9.68
Samsung Galaxy S	98	0.1171	6.35	0	5 - 8	6.35
HTC Desire Z	73	0.1536	3.41	0	7 - 11	9.63
HTC HD2	47	0.1242	1.78	7	4 - 10	7.73

position, and add an event category to it. Offline correction may be required.

2) Ground-truth is temporary accurate: even in a few days situation on the road can change. Therefore ground-truth must be recorded as close to the actual data, as possible.

3) Android devices are different: a complete experimental evaluation must be performed on more than one device to get valid results.

4) There is a non-deterministic delay between actual sensor sampling and data reception in the software: this fact must be taken into account in situations where the subject is moving at significant speeds.

5) During long test drives miscellaneous errors can occur preventing the data collection: It is advisable to monitor the devices continuously.

6) Although Android is an open platform, it has remarkable hardware access restrictions compared to customized embedded platforms (sensor motes): it is reasonable to invest the saved application creation time to develop more sophisticated data processing algorithms.

7) Device power supply is an underrated problem: for each device a decision has to be made - should it require a power supply during the ride or not? And a supply must be provided if necessary.

8) Insignificant data recording may distract the essential data collection: collect just the required information. Otherwise, software bugs and hardware limitations could degrade the sensing process.

9) Different value scales cause problems with data interpretation: it is advisable to convert all the data to a unified scale. Human-readable values are preferred over raw values.

10) It is hard to get the overall understanding of large data set and location data: visualization techniques are recommended to simplify pattern perception.

11) Unsynchronized time for devices will lead to inefficient post-processing: it is advisable to synchronize time of all the devices before the experimental tests. Note, that ”Use network-provided values” under date&time settings on Android devices does not mean synchronization with NTP servers!

V. DISCUSSION

One of the main problems with ground-truth is the lack of objectiveness. If data processing and event detection system is considered as a formal logic, ground-truth serves as a set of axioms used to prove further algorithm accuracy. It is, however, hard to argue on the correctness of the ground-truth itself. Three basic requirements for ground-truth as an axiomatic system for vehicular sensing experiments:

1) *Usability* - Ground-truth must represent real environmental and road conditions
2) *Consistency* - Algorithm analysis and comparison to ground-truth should contain no contradiction
3) *Completeness* - Ground-truth should be usable as etalon for analysis of any algorithm that detects the particular event type

According to Goedel’s Incompleteness theorem, consistency and completeness cannot be guaranteed simultaneously in general case.

For Walking-GPS approach, usability follows from its definition: we mark positions of encountered real potholes on city streets. For semi-automated approach it is arguable: we mark positions, where it *feels similar to pothole*. Evaluation of usability is, however, subjective - locations of particular events are recorded by humans and represent their viewpoint.

Consistency in this context requires clear response to ”what kind of event is located at x?” for each x. If a location exists for which it is not clear, whether there is a pothole or not, it is a contradiction. Semi-automated approach suffers from such inconsistencies, as described above. Walking-GPS, and any other methods contain a certain degree of position deviance due to GPS inaccuracy. But for each recorded ground-truth point it is clear - there is a real pothole in close vicinity.

Completeness includes multiple event category support, as some of the detection algorithms might be intended for event segregation or intensity detection. Manual ground-truth marking satisfies this requirement.

An alternative approach would be to use a well-known, previously proved algorithm as a ground-truth, to which all the new algorithms are compared. However, in reality, perfect algorithms with 100% accuracy are rarity. If there is a well-known algorithm with 90% accuracy as a reference and a new algorithm matches its results with 90% accuracy, how accurate really is the new algorithm, $90 * 90 = 81\%$? Or is it 100% accurate and the 10% are lost due to reference algorithm’s imperfection?

 ISBN: 978-1-61208-144-1

Accordingly, we argue, that it is hardly possible to prove that a particular ground-truth is objective and accurate, as it contains significant amount of subjectiveness. Still, manual ground-truth marking provides a reasonable degree of trust.

VI. Related Work

Vehicular sensing systems have been proposed by researchers previously, including BusNet [8], Pothole Patrol [9], Nericell [10], RoadMic [6]. Besides vehicular applications, other types of mobile participatory sensing dealing with environmental monitoring do exist, including BikeNet [11]. However, in this paper we do not concentrate on data analysis algorithms or communication and content delivery protocols, rather on effective and efficient data collection process and recommendations for improved real-world experiments and deployments.

This paper shares common structural and ideological patterns with prior research papers concentrating on real-world wireless sensor network experiences and deployment recommendations in different application areas, including precision agriculture [12], environmental [13] and wild animal monitoring [14], [15]. To the best of our knowledge, this is the first paper describing field test recommendations especially for vehicular sensing applications using mobile phones, having specific requirements and characteristics.

VII. Conclusion and Future Work

In this paper, we described our experience from urban vehicular sensing experiments with Android smartphones detecting potholes by analyzing accelerometer data. Although the collected data is deficient, we draw multiple highly valuable conclusions. First of all, we admit, that semi-automated ground-truth location marking by a human operator pressing a button during the test drive is subject to multiple errors due to both technical limitations and human factors. Manual pothole marking and categorization while walking along the test track is recognized as the right method for ground-truth recording, but offline position correction is recommended. We also detect differences between distinct Android devices, most significant of which is the difference in accelerometer sampling rate and deviation.

We believe, that our experience will help to improve efficiency and reduce time and effort fur further experiments using Android platform for vehicular sensing researchers. Our future work includes further evaluation of our pothole detection algorithm accuracy.

VIII. Acknowledgement

This work has been supported by European Social Fund, grants Nr. 2009/0219/ 1DP/1.1.1.2.0/APIA/VIAA/020 ”R&D Center for Smart Sensors and Networked Embedded Systems” and Nr. 2009/0138/ 1DP/1.1.2.1.2/09/IPIA/ VIAA/004 ”Support for Doctoral Studies at the University of Latvia”. Special thanks to Rinalds Ruskuls and Vadims Kurmis who provided phones for testing.

References

[1] http://potholes.co.uk, Last access: 27.05.2011.

[2] http://bedrukarte.lv, Last access: 27.05.2011.

[3] http://world.waze.com/, Last access: 27.05.2011.

[4] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, “VTrack: Accurate, Energy-Aware Road Traffic Delay Estimation Using Mobile Phones,” in *Proc. SenSys’09*, 2009, pp. 85–98.

[5] G. Strazdins, A. Mednis, G. Kanonirs, R. Zviedris, and L. Selavo, “Towards Vehicular Sensor Networks with Android Smartphones for Road Surface Monitoring,” in *Electronic Proceedings of CONET’11*, 2011.

[6] A. Mednis, G. Strazdins, M. Liepins, A. Gordjusins, and L. Selavo, “RoadMic: Road Surface Monitoring Using Vehicular Sensor Networks with Microphones,” in *Proc. of Networked Digital Technologies, NDT’10*, 2010, pp. 417–429.

[7] R. Stoleru, T. He, and J. A. Stankovic, “Walking GPS: A Practical Solution for Localization in Manually Deployed Wireless Sensor Networks,” *Proc. IEEE LCN’04*, pp. 480–489, 2004.

[8] K. De Zoysa, C. Keppitiyagama, G. Seneviratne, and W. Shihan, “A public transport system based sensor network for road surface condition monitoring,” in *Proc. NSDR’07*, 2007, pp. 9–14.

[9] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, “The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring,” in *Proc. MobiSys’08*, 2008, pp. 29–39.

[10] P. Mohan, V. N. Padmanabhan, and R. Ramjee, “Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones,” in *Proc. SenSys’08*, 2008, pp. 357–358.

[11] S. Eisenman, E. Miluzzo, N. Lane, R. Peterson, G. Ahn, and A. Campbell, “BikeNet: A mobile sensing system for cyclist experience mapping,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, pp. 6:1–6:39, December 2009.

[12] K. Langendoen, A. Baggio, and O. Visser, “Murphy loves potatoes: Experiences from a pilot sensor network deployment in precision agriculture,” in *IPDPS’06*, 2006, pp. 1–8.

[13] G. Barrenetxea, F. Ingelrest, G. Schaefer, and M. Vetterli, “The Hitchhiker’s Guide to Successful Wireless Sensor Network Deployments,” in *Proc. SenSys’08*, 2008, pp. 43–56.

[14] R. Zviedris, A. Elsts, G. Strazdins, A. Mednis, and L. Selavo, “Lynxnet: Wild animal monitoring using sensor networks,” in *Proc. REALWSN’10*, 2010, pp. 170–173.

[15] P. Zhang, C. M. Sadler, S. A. Lyon, and M. Martonosi, “Hardware design experiences in ZebraNet,” in *Proc. SenSys’04*, 2004, pp. 227–238.

 ISBN: 978-1-61208-144-1