

# DETECTING PAVEMENT PATCHES UTILIZING SMARTPHONES TECHNOLOGY AND VEHICLES

Charalambos Kyriakou<sup>1</sup> and Symeon E. Christodoulou<sup>2</sup>

**Abstract:** Presented herein is a study on the utilization of low-cost technology for detection of roadway pavement anomalies (patches and potholes), by use of sensors on smartphones and of automobiles' on-board diagnostic (OBD-II) devices for the collection and analysis of vibration-related data while vehicles are in movement. The mobile data collection kit consists of a triaxial accelerometer, a gyroscope and a global positioning sensor. The smartphone-based data collection is complimented with robust regression analysis and a bagged-trees classification model for the classification of pavement anomalies. The proposed system is readily available, low-cost and adequately accurate, and can be utilized in crowd-sourced applications for pavement monitoring. Further, the proposed methodology has been field-tested, exhibiting detection accuracy levels higher than 90% for pavement patches, and it is currently expanded to include larger datasets and a bigger number of pavement defect types.

**Keywords:** Pavement anomalies, Detection and classification, Smartphones technology, Robust Regression, Bagged Trees.

## 1 INTRODUCTION

One of the most important indicators for road quality is the pavement surface condition, which is determined by the anomalies in the pavement surface that have an effect on the ride quality of a vehicle. Nowadays, pavement management agencies evaluate pavement surface quality almost once per year, because current practices are expensive and laborious. Distributed mobile sensing utilizing smartphones could provide a viable solution to pavement assessment at a much lower cost, as continuous monitoring of pavement performance will significantly reduce risks and provide an adequate volume of timely data to enable accurate maintenance forecasting.

As a result of the rapid and intense development and usage of smartphones in current years, smartphone technology has gained noteworthy consideration within the transportation, infrastructure, and automotive industries. Modern smartphones come with a range of built-in sensors, such as accelerometer, gyroscope and GPS sensors. Further, OBD Bluetooth devices (ELM 327 Bluetooth Car Diagnostic Scanner) coupled with smartphone applications make possible the real-time monitoring of, among others, the GPS latitude and longitude, forward and lateral acceleration, vehicle roll and vehicle pitch.

The vision for pavement anomaly detection by use of smartphone sensors, OBD devices and vehicles data is set in parallel with the idea that such technology can be applied in geographic information systems (GIS)-based pavement management systems (PMS). A

---

<sup>1</sup> Doctoral Candidate, Department of Civil and Environmental Engineering, School of Engineering, University of Cyprus, Cyprus, [kyriakou.charalambos@ucy.ac.cy](mailto:kyriakou.charalambos@ucy.ac.cy)

<sup>2</sup> Associate Professor, Department of Civil and Environmental Engineering, School of Engineering, University of Cyprus, Cyprus, [schristo@ucy.ac.cy](mailto:schristo@ucy.ac.cy)

sufficient number of vehicles collecting this crowd-sourced data can then be used to generate a georeferenced event at points where vehicles encounter abnormal behaviour on a pavement network. Even though multiple vehicles might present conflicting data regarding pavement surface quality, the joint knowledge fed by sensors provides a precise model of the pavement surface condition in relation to how an average driver experiences the roadway network condition.

In addition to this short introduction, a literature review section presents a framework of past and ongoing work related to the detection of pavement anomalies by use of smartphones. A section on methodology setup presents the build-up data collection system and methods, while the results and discussion section includes the processes and tools used to classify the data. The paper concludes with key findings and with an outline on future research directions.

## 2 LITERATURE REVIEW

De Zoysa et al. (2007) proposed a public transport system called “BusNet” to monitor pavement surface condition by adding acceleration sensors on top of public buses. The acceleration sensors identify potholes through changes in the vertical acceleration and determine the car speed modification using the horizontal acceleration. Erikson et al. (2008) used seven taxis running in the Boston area and developed a mobile sensor system called “Pothole Patrol”. Each taxi needed a computer running the Linux operating system, a Wi-Fi card for transmitting collected data, an external GPS and a 3-axis accelerometer. Strazdins et al. (2011) proposed a method requiring an Android smartphone with GPS, 3-axis accelerometer and a communication channel (cellular or Wi-Fi). The system consists of two application components, one for the Android device and one for a data server. Vittorio et al. (2014) proposed a system based on a simple application for smartphones that uses a GPS receiver and a three-axis accelerometer to collect acceleration data stemming from a vehicle’s motion on anomalous roadways. The high-energy events (anomalies) are identified by monitoring and measuring the vertical acceleration impulse. Seraj et al. (2014) proposed a system that detects road anomalies using mobile equipped with inertial accelerometers and gyroscopes sensors. They applied a method to remove the effects of speed, slopes and drifts from sensor signals. For future work, they aim to apply this method for road anomalies detection in participatory sensing, using clustering by geo-coordinates. Alessandrini et al. (2014) described a system which included a combination of a custom mobile application and a georeferenced database system. The roughness values computed and stored into a back-end geographic information system enable visualization of road conditions. This proposed approach introduced an integrated system for monitoring applications in a scalable, crowdsourcing collaborative sensing environment. Mohamed et al. (2015) proposed a road condition monitoring system that detected such as speed bumps. They suggested the gyroscope around gravity rotation as the primary indicator for road anomalies, in addition to the accelerometer sensor, in order to avoid false-positive indications; especially when there is a sudden stop or sudden change in motion acceleration. Jang et al. (2016) proposed an innovative method to collect up-to-date data about potholes and sunken manholes by using a mobile data collection kit mounted on vehicles. At a back-end server, a street defect algorithm which based on a supervised machine learning technique enhances the performance of the proposed monitoring system by integrating data collected from multiple sensor-equipped vehicles. The above systems, despite hardware differences in terms of GPS accuracy and accelerometer sampling rate

and noise, show that pothole detection by use of smartphones sensors on moving vehicles is possible.

Bridgelall (2015) developed theoretical precision bounds for a ride index called the road impact factor and demonstrated its relationship with vehicle suspension parameter variances. Lately, the Jaguar Land Rover automaker make known that is testing a new connected vehicle technology which permits a vehicle to spot hazardous potholes in the roadway and then distribute this data in real time with other Jaquar Land Rover vehicles (Nick O'Donnell 2015). Kyriakou et al. (2016, 2017) explored the use of smartphone sensor technology and of data collected while vehicles were in movement, for the detection and classification of pavement surfaces anomalies. Their proposed system architecture, which was complimented with artificial neural network techniques and bagged trees for classifying detected roadway anomalies, was trained, validated and tested against three types of common roadway anomalies exhibiting above 90% accuracy rate.

### 3 METHODOLOGICAL SETUP

The paper focusses on the detection of pavement patches (Figure 2), albeit being part of a larger effort on the detection of vibration-inducing pavement anomalies, as documented in Kyriakou et al. (2016). The data used in this study is collected from six different roads (of 5Km total distance and 8 different pavement patch types). Data of pavement patches is collected in-situ by use of a car equipped with a smartphone (mounted on the car's windshield) and with an OBD-II reader attached to the car. The data, collected at intervals of 0.1 seconds, included a total of 31 uni-dimensional (e.g. X, Y, Z accelerations, speed, etc.) and two-dimensional indicators (e.g. the smartphone's roll and pitch values). The smartphone was also fitted with the DashCommand application for recording sensor readings of taken data. Vehicle system data is transmitted through the OBD-II reader to the smartphone device and then transferred for either processing or storing, via a digital cellular connection or other means. At the back-end server, a defect detection algorithm based on random forests (bagged trees classification model) enhances the performance of the proposed monitoring system, by integrating data collected from multiple sensors and deducing knowledge from these participatory sensors.

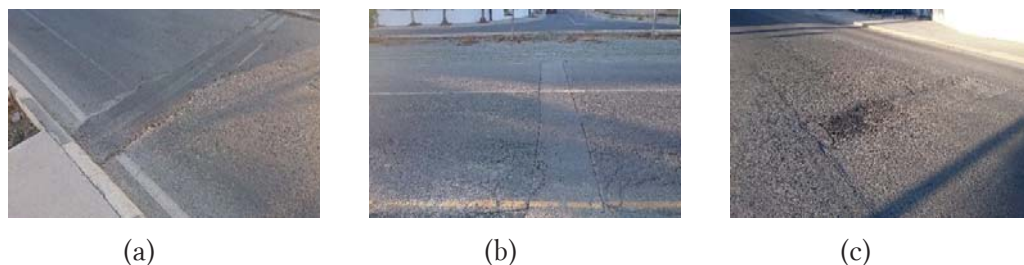


Figure 2: Pavement patches examined for detection.

Mathematically, the proposed method is based on rigid-body dynamics. Any three-dimensional rotation can be described as a sequence of yaw, pitch and roll rotations (Figure 3). Pitch is defined as counter clockwise rotation of  $\theta_{pitch}$  about the y-axis. Roll is defined as counter clockwise rotation of  $\theta_{roll}$  about the x-axis. Yaw is defined as counter clockwise rotation of  $\theta_{yaw}$  about the z-axis. The rotation matrices for pitch, roll and yaw give equations 1,2 and 3 respectively.

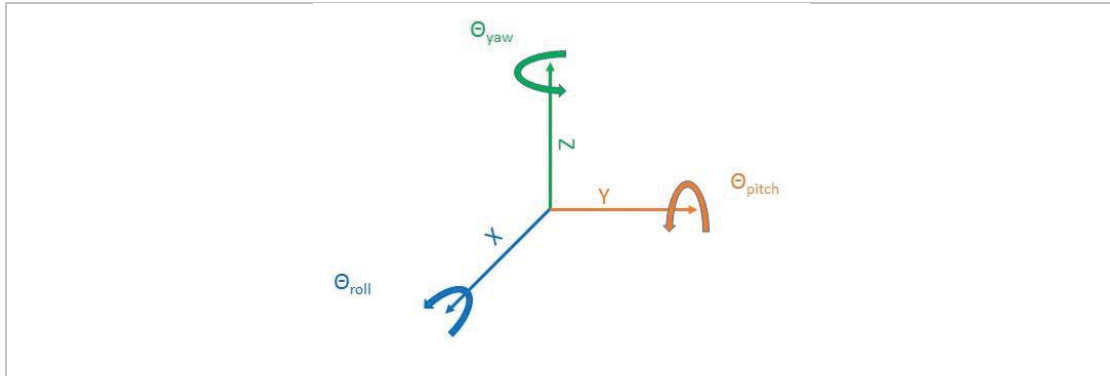


Figure 3: Yaw, pitch and roll rotations.

$$R_y(\theta_{pitch}) = \begin{bmatrix} \cos(\theta_{pitch}) & 0 & \sin(\theta_{pitch}) \\ 0 & 1 & 0 \\ -\sin(\theta_{pitch}) & 0 & \cos(\theta_{pitch}) \end{bmatrix} \quad (1)$$

$$R_x(\theta_{roll}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_{roll}) & -\sin(\theta_{roll}) \\ 0 & \sin(\theta_{roll}) & \cos(\theta_{roll}) \end{bmatrix} \quad (2)$$

$$R_z(\theta_{yaw}) = \begin{bmatrix} \cos(\theta_{yaw}) & -\sin(\theta_{yaw}) & 0 \\ \sin(\theta_{yaw}) & \cos(\theta_{yaw}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

The smartphone's roll and pitch values (Figure 4a) can be associated with the host car's roll and pitch values (Figure 4b). In essence, the roll metric refers to a car's acceleration variation between its left and right front wheels, while the pitch metric refers to a car's acceleration variation between its front and rear wheels. Concurrently, the roll and pitch values define the way in which the host car is off balance (sideways and front/back).

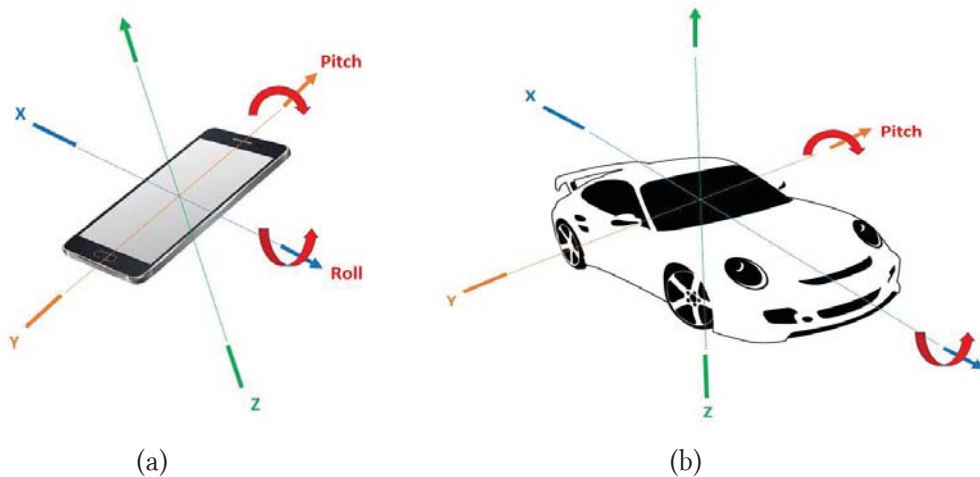


Figure 4: Smartphone's (a) roll and pitch directions and (b) relation to car's wheels' differential.

Robust regression analysis is used to detect the most significant variables and the underlying regression equation. Tukey’s Biweight robust influence function was used. It is an M-Estimator that uses iteratively reweighted least squares. This M-estimator completely downweights observations with large outliers until their weight is set to zero. It provides protection against heavy-tailed error distributions. The regression analysis outputs are shown below, in Figure 5 and Table 1, listing the most important statistical variables (with p-values ≤0.05).

Table 1: Most important variables according to robust regression.

Variable	Variable Name	Variable Description
VAR_1	Aux.Accel.Forward (Gs)	Forward Acceleration
VAR_2	Aux.Accel.Lateral (Gs)	Lateral Acceleration
VAR_3	Aux.Rotation.Pitch (Â°)	Vehicle Pitch
VAR_4	Aux.Rotation.Roll (Â°)	Vehicle Roll

Robust regression determines the coefficients  $\beta_i$  of each factor  $x_i$ , and the residual error  $\epsilon_j$  for each observation  $j$ , in the equation

$$y_j = \beta_0 + \beta_1x_{1j} + \beta_2x_{2j} + \dots + \beta_px_{pj} + \epsilon_j, \tag{4}$$

which is required to relate the dependent variable,  $y_i$ , to the independent variables  $x_i$ . The independent variables are the thirty-one parameters talk about above, and the dependent variable relates to the classification status of the pavement (‘1’ for no defect, ‘2’ for pavement patches). The subscript ‘j’ denotes the observation (row) number. Robust regression was preferred instead of other regression analysis because it examines highly contaminated data by detecting outliers from both dependent and independent variables.

**Run Summary Report**

Item	Value	Rows	Value
Dependent Variable	C7	Number Processed	1583
Number Ind. Variables	10	Number Used in Estimation	1583
Weight Variable	None	Number Filtered Out	0
Robust Method	Tukey's Biweight	Number with X's Missing	0
Tuning Constant	4.685	Number with Weight Missing	0
MAD Scale Factor	0.6745	Number with Y Missing	0
		Sum of Robust Weights	1231.013
<b>Run Information</b>	<b>Value</b>		
Iterations	30		
Max % Change in any Coef	0.067		
R <sup>2</sup> after Robust Weighting	0.9993		
S using MAD	0.00		
S using MSE	0.00		
Completion Status	Normal Completion		

**Estimated Equation**

C7 =  
 1.04207970696665 - 0.540982699787313 \* AUX\_ACCEL\_FORWARD\_Gs + 0.0103553795865275 \*  
 AUX\_ROTATION\_PITCH\_Â - 0.00443716027684323 \* AUX\_ROTATION\_ROLL\_Â + 0.493609495820244 \*  
 AUX\_ACCEL\_FORWARD\_Gs\*AUX\_ACCEL\_FORWARD\_Gs - 0.0115693903894446 \*  
 AUX\_ACCEL\_FORWARD\_Gs\*AUX\_ROTATION\_PITCH\_Â - 0.0523867849404161 \*  
 AUX\_ACCEL\_FORWARD\_Gs\*AUX\_ROTATION\_ROLL\_Â - 0.130901743860689 \*  
 AUX\_ACCEL\_LATERAL\_Gs\*AUX\_ROTATION\_ROLL\_Â - 3.46584911775038E-05 \*  
 AUX\_ROTATION\_PITCH\_Â \*AUX\_ROTATION\_PITCH\_Â + 0.00122836523185265 \*  
 AUX\_ROTATION\_PITCH\_Â \*AUX\_ROTATION\_ROLL\_Â - 0.00157108064468633 \*  
 AUX\_ROTATION\_ROLL\_Â \*AUX\_ROTATION\_ROLL\_Â

Figure 5: Robust regression summary



## 4 RESULTS AND DISCUSSION

The collected raw data exposes the complexity of the detection problem, as variables such as the vehicle pitch expected to highlight roadway anomalies are not as conclusive as originally thought. For that aim, the datasets are then fed into classification models. With the purpose of classifying and validating the data, an evaluation comparison was performed between various supervised learning algorithms for multiclass problems, to test their performance, including decision trees using various classifiers, discriminant analysis, support vector machines, logistic regression, nearest neighbours, and ensemble classification.

Supervised machine learning was performed by feeding a known set of predictors data (known pavement condition data) and known responses to the data (known pavement condition category) into classification models. The classification model predictors are the parameters listed in Table 1, while the responses are the values of '1' for no defect and '2' for pavement patches. The classification model architecture (shown in Figure 7) was implemented in MATLAB.

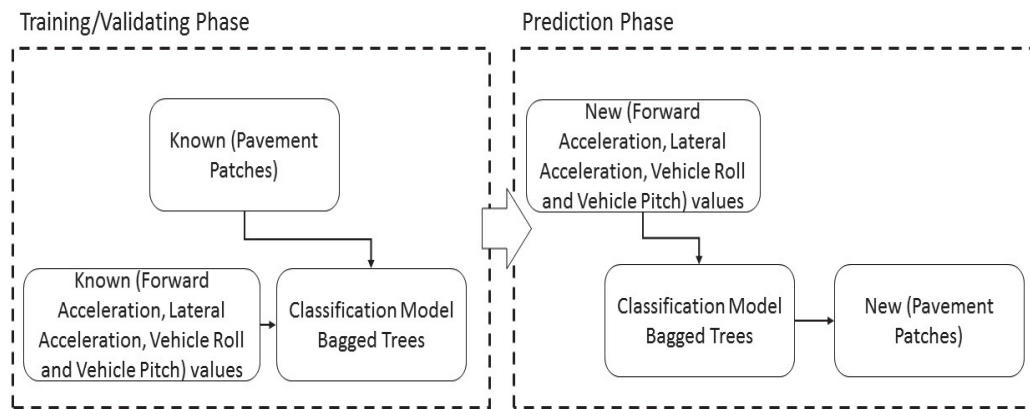


Figure 7: Classification models' architecture

Upon evaluating the various models against their performance on a given dataset and comparing the resulting cross-validation errors of each method, the best model for the studied problem and dataset was chosen to be the bagged trees. The dataset in study contained approximately 2000 data, of which 25 percent was held out for cross validation purposes. Bagged trees procedure is simple: (1) takes a sample from the dataset, (2) fits the tree to this data set, (3) repeat steps 1 and 2 many times (typically 50-1000), and (4) makes predictions for new data using each of the fitted models and average the predictions. Bagged trees use Breiman's 'random forest' algorithm. 'Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large.' (Breiman 2001).

As shown in Figure 8, the bagged trees classify pavement patches (target classes '2') while also distinguishing the 'no defect' condition (target class '1'). For each pavement surface anomaly, data is used for training, for validating and for predicting the bagged trees, with an outline of the obtained classification results presented in (Table 3).

Table 2: Bagged Trees training, validation and prediction statistics for the various roadway anomaly cases examined

Classifier Type	Accuracy	(%)
Bagged Trees	Train	98
Bagged Trees	Validation	100
Bagged Trees	Prediction	100
Classifier Type	Accuracy	(%)



Figure 8: Bagged Trees confusion matrices.

## 5 CONCLUSIONS AND FUTURE WORK

Pavement agencies can enrich the condition and operation of their roadway networks by putting into practice a PMS that uses smartphones and OBD devices. The paper puts forward a study on the use of smartphones and vehicular data for the detection of pavement patches by applying robust regression analysis and bagged trees classification models. The applied system is instantly available, low-cost and adequately accurate, and can be utilized in crowd-sourced applications. The presented methodology confirms the detection of pavement patches, exhibiting high detection accuracy levels. In future research, GPS data refinement methods will be implemented to improve the accuracy of GPS positioning data, which might affect the results of the classification analysis. Further, ongoing work investigates the effects of vehicle type and of car suspension condition on the accuracy of detection, and how the accuracy can be improved using crowdsourced data. Ongoing work includes the development of a spatial decision support system (DSS) for pavement surface condition-assessment mapping. The outcome of the bagged trees analysis and the subsequent pavement surface condition assessment score will be mapped spatially, pointing out the areas of concern and in need of rehabilitation actions.

## 6 REFERENCES

- Alessandroni, G., Klopfenstein, L., Delpriori, S., Dromedari, M., Luchetti, G., Paolini, B., Seraghiti, A., Lattanzi, E., Freschi, V., and Carini, A. (2014). SmartRoadSense: Collaborative Road Surface Condition Monitoring. In: *UBICOMM, The Eighth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. 24-28 August. Rome: Italy.
- Breiman, L. (2001). *Random Forests. Machine Learning*, 45(1), pp. 5-32.
- Bridgelall, R. (2015). Precision Bounds of Pavement Deterioration Forecasts from Connected Vehicles. *J. Infrastruct. Syst*, 21(1),1-7.
- De Zoysa, K., Keppitiyagama C., Seneviratne, G. P., and Shihan, W. (2007). A Public Transport System Based Sensor Network for Road Surface Condition Monitoring. In: *SIGCOMM, Proceedings of the 2007 Workshop on Networked Systems for Developing Regions*, Kyoto, Japan, 27-31 August 2007. New York: USA.
- Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S. and Balakrishnan, H. (2008). The pothole patrol: using a mobile sensor network for road surface monitoring. In: *MOBISYS, The 6th International Conference on Mobile Systems, Applications, and Services*, Breckenridge, CO, USA, 17-20 June, 2008, New York: USA.
- Jang, J., Yang, Y., Smyth, A., Cavalcanti, D., and Kumar, R. (2016). Framework of Data Acquisition and Integration for the Detection of Pavement Distress via Multiple Vehicles. *J.Comput.Civ.Eng.*, 04016052.
- Kyriakou, C., Christodoulou, S. E., and Dimitriou, L. (2016). Road Anomaly Detection and Classification Using Smartphones and Artificial Neural Networks. In: *TRB, The Transportation Research Board 95th Annual Meeting*, Washington D.C, U.S.A. 10-14 January, 2016. Washington D.C: U.S.A.
- Kyriakou, C., Christodoulou, S. E., and Dimitriou, L. (2017). Detecting and Classifying Roadway Pavement Anomalies Utilizing Smartphones, On-Board Diagnostic Devices and Classification Models. In: *TRB, The Transportation Research Board 96th Annual Meeting*, Washington D.C, U.S.A. 8-12 January, 2017. Washington D.C: U.S.A.
- Mohamed, A., Fouad, M., Elhariri, E., El-Bendary, N., Zawbaa, H. M., Tahoun, M., and Hassanien, A. E. (2015). RoadMonitor: An intelligent road surface condition monitoring system. *Intelligent Systems' 2014*, Springer, (323), pp. 377-387.
- O'Donnell, N. (2015). *Jaguar Land Rover Announces Technology Research Project to Detect, Predict and Share Data on Potholes*, [online] Available at: <, [http://newsroom.jaguarlandrover.com/en-in/jlr-corp/news/2015/06/jlr\\_pothole\\_alert\\_research\\_100615/](http://newsroom.jaguarlandrover.com/en-in/jlr-corp/news/2015/06/jlr_pothole_alert_research_100615/)> (Accessed 15 July 2015).
- Seraj, F., Van Der Zwaag, Berend J., Dilo, A., Luarasi, T., and Havinga, P. (2014). RoADS: A Road Pavement Monitoring System for Anomaly Detection using Smart Phones. In: *International Workshop on Modelling Social Media*, Springer, 2014, pp. 128-146.
- Strazdins, G., Mednis, A., Kanonirs, G., Zviedris R., and Selavo, L. (2011). Towards Vehicular Sensor Networks with Android Smartphones for Road Surface Monitoring. In: *CONET, The 2nd International Workshop on Networks of Cooperating Objects*, 11-12 April 2011.Chicago, USA.
- Vittorio, A., Rosolino, V., Teresa, I., Vittoria, C. M., Vincenzo, P. G., and Francesco, D. M. (2014). Automated Sensing System for Monitoring of Road Surface Quality by Mobile Devices. *Procedia - Social and Behavioral Sciences*, 111 242-251.