A PAVEMENT RATING SYSTEM BASED ON MACHINE LEARNING

Charalambos Kyriakou and Symeon E. Christodoulou
University of Cyprus, Nicosia, Cyprus

ABSTRACT

The evaluation of roadways utilizing complex contemporary datasets is currently conducted periodically because of the collection methods' high cost. The study presents a data-driven framework on the use of a vehicle, a smartphone, an on-board diagnostic (OBD) device and machine learning for the rating of pavement surfaces. The proposed system architecture has been field-tested for the detection of pavement anomalies and the classification of five rating categories. Further, the proposed system may provide daily information on roadway pavement surface conditions, which can be used by engineers for automating the planning of pavement maintenance operations and improving public safety.

INTRODUCTION

The frequent monitoring of roadway pavements and their reliable evaluation has gained high significance during recent years, for many reasons. Firstly, the pavement surface condition is one of the key metrics for providing a safe and high-quality driving experience (Federal Highway Administration, 2010). Secondly, roadway anomalies such as speed bumps (if they are not constructed under regulations), potholes and patches are the most annoying obstacles faced by vehicles, drivers and passengers. Thirdly, roadway surface anomalies can damage vehicles and, most often, be the reason for car accidents. In fact, the most road accidents are caused by the poor condition of roads (World Health Organization, 2015). Fourthly, bad pavements are a big problem for vehicles and drivers because the deterioration of the roadway network leads to more expensive maintenance, not only for the roadway itself but also for vehicles.

To date, the methodology used for collecting pavement distress data utilises surveys which address an evaluation or a detailed measurement of distress (Walker et al., 2002). The collection of the pavement data can be either manual, from a moving vehicle or by 'walking' the roadway, or automated, by the use of vehicles fitted with specialized cameras and sensors (AASHTO, 1990). Another approach is to use existing technologies such as light detection and ranging (LIDAR) or other commercial products (Roadscanners, 2020). Currently, in the USA are used specialized high-end vehicles capable of detecting and mapping roadways conditions. The overall cost and maintenance of the above equipment are very high which reduces the system's feasibility. As a consequence, pavement agencies evaluate roadway quality approximately only once per year, mainly because the existing practices of assessment pavement network are high-priced (high-end vehicle cost: approximately one million dollars) (Fugro Roadware, 2015).

The principal goal of this research is to devise a low-cost vibration-based data acquisition and pavement anomaly detection method and a reliable rating system for categorising pavement surface anomalies.

Currently, the widespread usage of smartphone technology has gained noteworthy consideration within the infrastructure, transportation and vehicle industries for many reasons. Given the escalating popularity of smartphones, their high processing power and their ability to transfer data over wireless or General Packet Radio Service (GPRS) networks, smartphone-based technologies and applications have emerged as an efficient and low-cost alternative to traditional approaches. This is possible because standard-model smartphones come with a variety of built-in sensors such as accelerometer, gyroscope and Global Positioning System (GPS) sensors. For the above reasons, this research utilizes smartphone applications (e.g. the DashCommand™ app) coupled with On-Board Diagnostic (OBD) Bluetooth devices (e.g. OBD II ELM 327) and an architecture which enables the real-time communication of smartphone and vehicle system (Controller Area Network, CAN, bus) sensor data, complimented with machine-learning algorithms for subsequently processing the sensed data and deducing knowledge from the pavement sensing process.

With regard to the methodology employed in this research, due to the reduced precision of smartphone devices compared to expensive hardware for road profiling, the proposed system is based on the use of participatory sensing, signal processing techniques and machine learning for pattern recognition, knowledge extraction and reinforcement. This information when analyzed can be timely and periodically disseminated to travellers and involved agencies in the form of pavement-rating maps of the roadway network.

In addition to this short introduction, a state of the art section presents an outline of past and ongoing work related to the development of a rating system for roadway pavements by use of smartphone accelerometers and gyroscopes. A section on the proposed methodology presents the data collection system and the analysis method, while the results and discussion section includes the processes and tools used to rate roadways based on the sensed data. The paper concludes with the key findings and with an outline of future work.

LITERATURE REVIEW

In developed countries, pavement agencies utilize specialized platforms for the assessment of roadway pavements, which are hosted on expensive pavement evaluation vehicles. These platforms typically use machine-vision and laser technologies, and they can detect all surface distresses types, providing a reliable standard rating system for pavement surface. The costs related to existing technologies are comprised of several cost components (e.g. software, data collection, database set-up, consulting services, training and personnel costs) (AASHTO, 1990) and the typical combined unit cost of pavement imaging and analysis ranges from $15 to $52 per kilometre (McGhee, 2004). Because of this high unit cost, the evaluation of road networks by current automated methods is typically conducted no more than once a year. In such an evaluation process, very important are the stages related to data collection, the classification and the spatial mapping of pavement conditions. For that reason, a plethora of researches have already published works on low-cost pavement surface condition assessment using smartphones and machine learning techniques. Related work is discussed below.

Seraj et. al. (2014) proposed a system that detects road surface anomalies utilizing smartphones equipped with accelerometers and gyroscopes (sampling rate of 93Hz). They applied wavelet transform analysis for signal processing of inertial sensor signals and Support Vector Machine (SVM) for anomaly detection and classification. The obtained results showed a consistent accuracy of 90% on detecting severe anomalies regardless of vehicle type and road location.

Allouch et al. (2017) utilized an accelerometer, a gyroscope and a GPS sensor for collecting roadway data and for plotting the road location trace in Google map. Three algorithms were tested: decision tree C4.5, SVM and Naïve Bayes. The decision tree C4.5 algorithm showed consistent accuracy of 98.6%.

Souza et al. (2018) proposed a low-cost system to monitor and evaluate pavement surface conditions in real-time utilizing smartphone sensors, signal processing and machine learning algorithms. Accelerometer sensors were employed to measure the vehicle vibration while driving and these data were used to evaluate roadway condition. Their system achieved a classification performance of 90% in a five-class problem considering the following road qualities: Good, Average, Fair, Poor and obstacles.

Alam et al. (2020) aimed to develop a system that detects three road events, speed-breakers, potholes and broken road patches, over smooth and rough roads. The system's first phase used robust auto-orientation and auto-tune thresholding algorithms, the second phase utilized decision tree based classifier to reduce false-negative and false-positives and the third phase applied a k-medoids clustering to geo-localized detected events over a map service.

Kyriakou et al. (2018,2019) proposed a low-cost pavement monitoring system to obtain up-to-date information about the most common road surface anomalies with the use of a typical smartphone and vehicle. Robust regression analysis, Artificial Neural Networks (ANN) and bagged trees were used to analyze smartphone sensors data. The proposed system was field-tested (accuracy levels higher than 90%) and it is currently expanded to include a bigger number of pavement surface anomalies.

As shown in Table 1, the aforementioned researchers used several machine-learning techniques to evaluate pavement surface conditions.

Table 1: Comparison of related machine-learning research in pavement condition monitoring.

Reference	Detected/ Categories	Detection/ Classification Method	Reported Accuracy
Seraj et al (2014)	Span, Mild, Severe	Wavelet Analysis, SVM	90%
Allouch et al. (2017)	Smooth, Potholed	C4.5 Decision Tree Classifier, SVM, Naïve Bayes	98.6%
Souza et al. (2018)	Good, Average, Fair, Poor, Obstacles	Signal Processing Techniques, SVM	90%
Alam et al. (2020)	Speed breakers, potholes over a smooth road and rough road	Thresholding Algorithms, Decision Trees, K-medoids clustering	90%
Kyriakou et al. (2018, 2019)	Cracks, Rutting ravelling, patching and potholes	Robust Regression, ANN, Bagged Tress	98%

METHODOLOGICAL SETUP

System Overview and data collection methodology

The paper focuses on the detection of speed bumps, patches, potholes and the development of a five-class rating pavement surface condition system utilizing nine different sections of roadways (of 10 Km total distance; approximately 20 data points per GPS location; 21059 total data points, of which 220 data points referred to speed bumps, 60 data points to patches and 100 data points to potholes). The system architecture, shown schematically in Figure 1, utilizes data from a smartphone and server-based unsupervised learning algorithms to analyze the acquired data. The processed information on the location of the detected speed bumps, potholes and patches and on the deduced rating category of pavement surface are then disseminated to pavement agencies and road users.

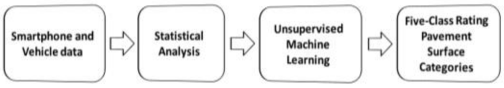

Figure 1: System components

Experimental design

The experimental setup for the data acquisition is as shown in Table 2. The data collection was performed on several urban roads using a car, a smartphone fitted with the DashCommand^{TM} application and an OBD II Bluetooth reader (ELM 327). Vehicle and smartphone sensor data were transmitted through the smartphone application to a data server for processing (and storing) via a digital cellular or Bluetooth connection. The smartphone was mounted on the car's windshield and for visually verifying the existence of the pavement condition, the smartphone had also its video camera active for recording the routes travelled. The collected dataset, containing several motions and vibration datafields, is collected at intervals of 0.1 seconds and relates to both uni-dimensional (e.g. speed) and two-dimensional indicators (e.g. the vehicle's roll and pitch values).

Table 2: Experimental setup summary

	Parameters considered
Location	Lakatamia, Cyprus
Distance	10 Km
Detection	Speed bump, Patches, Potholes
Vehicle	Nissan Qashqai
Smartphone	Samsung Galaxy S8
Mount point	Dashboard

System Design and Methodology

Mathematically, the proposed methodology is based on rigid-body dynamics and the ability to express any three-dimensional rotation as a combination of yaw, pitch and roll rotations. A variety of regression analyses was performed to investigate the statistical significance of each datafield, with robust regression analysis eventually selected (Kyriakou et al. 2019) because it analyses highly contaminated data by detecting outliers from both dependent and independent variables (R^2=1). As reported in Kyriakou et al. (2019), the forward and lateral acceleration, and the vehicle roll and pitch variables were the most significant statistical variables (with p-values $\leq$0.05) from the initial dataset (of about 190 variables). The reduced dataset of the aforementioned four variables was then fed into several unsupervised machine-learning algorithms for clustering purposes. A simplified flowchart of the data collection and analysis process used is depicted in Figure 2.

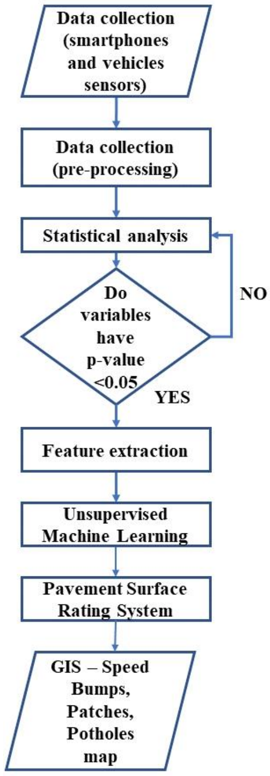

Figure 2: System flowchart

RESULTS AND DISCUSSION

To prepare the data for the clustering phase, during which the sensed roadway pavement features are clustered into five pavement categories based on the four most statistically significant factors according to the robust-regression analysis, the sensed vibration data was aggregated by GPS location and the dataset's dimensionality reduced.

Feature Extraction

Feature extraction is the process of dimensionality reduction by which an initial set of raw data (in our case, 21059 raw data x 4 variables) is reduced to more manageable groups for processing (in our case, 1031 data x 4 variables for K-means clustering, or 1031 data x 1 variable for K-medoids clustering). The study presented herein examined several time-domain features as data aggregators (e.g. the mean, standard deviation and variance of the original signals at each GPS location). For example, a standard-deviation analysis was performed for each geographical point (thus, the original 20 datapoints per geographical point were converted into a single point – 21059 / 20 total data points ~ 1031 total data points) in which analysis each GPS point is characterized by the standard deviation of the observed signal values at the point.

Clustering

Upon completion of the feature extraction phase, the data is fed into an unsupervised machine-learning algorithm for clustering analysis. Clustering analysis is one of the main analytical methods in data mining and it is of wide use and great importance, and clustering is, in essence, a classification process in which data are grouped into groups of similar features. A cluster is thus a collection of data which are similar to each other, but they are dissimilar to the data belonging to other clusters. The prevalent clustering methods can be classified into five categories (partitioning, hierarchical, density-based, grid-based and model-based methods), with the partitioning methods being the most well-known and commonly used clustering methods. Partitional clustering techniques create one-level partitioning of the data points, and most widely partitioning methods are the k-means and the k-medoids method, and their variations. Both techniques are based on a single centre point which presents a cluster. Nowadays, cluster analysis tools based on k-means, k-medoids and several other methods have been built into several statistical analysis software (such as SPSS and NCSS).

K-means clustering

At first, the smartphone-based datasets were fed into the k-means clustering algorithm for unsupervised machine learning. K-means is the most popular partition-based clustering algorithm and in the k-means method, a centroid is used which is the mean or median of a group of points. It should be noted that the cluster's centroid rarely corresponds to an actual data point. First, upon definition of the number of clusters to use (k), the k-means algorithm randomly selects k points as the initial centroids. Secondly, the algorithm assigns all data points to the closer centroid. Thirdly, it recomputes the centroid of each cluster based on the assignment of points to the clusters. Finally, it repeats the previous steps until the computed centroids in two successive iterations do not change.

The objective (E) of the k-means clustering is to minimize the Euclidean sums of squared deviations of objects from the cluster mean.

$$\mathrm{E} = \sum_{i=1}^{k} \sum_{xeC_i} |x - m_i|^2 \qquad (1)$$

E = the sum of the square error of all objects in the data set
c = cluster
x = the point in space representing the given object
m = the mean of the cluster

In the case-study dataset, system-best results were obtained when clustering the standard deviation of the forward & lateral acceleration and the vehicle pitch & roll variables into five classes (i.e. k=5), with k=1 the class corresponding to the 'very good' pavement condition and k=5 the class corresponding to the 'very bad' pavement condition. As ground truth for verifying the method's detection accuracy of speed bumps, patches and potholes, the smartphone's recorded video was used. The resulting clusters correctly classify speed bumps, patches and potholes in category five (worst category). By contrast, the k-means method misclassifies the other four categories because the algorithm is sensitive to outliers, since an object with an extremely large value may substantially distort the distribution of the data. Furthermore, the method requires several passes on the entire dataset, which can make the whole process expensive and time-consuming. For the above reasons, the smartphone-based datasets were subsequently fed into the k-medoid clustering algorithm for unsupervised machine learning, to increase the accuracy and applicability of the method.

K-medoids clustering

In the k-medoid clustering method, a medoid is used, which is the most representative (central) point of a group of points. The method attempts to determine k partitions from n objects, and each cluster is represented by one of the objects in the cluster. In contrast to the k-means algorithm, instead of taking the mean value of the object in a cluster as a reference point, the method uses the medoid which is the most centrally located object in a cluster.

The k-medoid clustering technique is simple and is based on the search for k medoids among the objects of the dataset. These medoids represent the structure of the data. The target is to find k objects which minimize the sum of the dissimilarities of the objects to their closest medoid. At first, the algorithm selects k initial medoids. These medoids are the candidate medoids and are intended to be the most central points of their clusters. Then the distance of each non-selected points from the closet candidate medoid is calculated, and this distance is summed over all points. The distance represents the cost of the current selection, and all possible swaps of a non-selected point for a selected one are considered. The cost of each selection is calculated and the configuration with the lowest cost is selected.

$$\mathrm{E} = \sum_{i=1}^{k} \sum_{xeC_i} |x - o_i| \qquad (2)$$

E = the sum of the distances for all objects in the data set
x = the point in space representing the given object
o = is the medoid of the cluster

If this is a new configuration then the effect of replacing one of the selected medoids with one of the non-selected medoids is considered.

In the case-study dataset, system-best results were obtained when clustering the variance of vehicle pitch variable into five classes (Table 3). The algorithm correctly classifies speed bumps, patches and potholes in category four and five. As with the k-means case, the ground-truthing to verify the accuracy of the k-medoids method was accomplished by visually verifying the existence of speed bumps, patches and potholes from the smartphone's recorded video. In contrast to the k-means algorithm, the k-medoids method classifies the other three categories correctly. To make an example, in two different scenarios (e.g., no visible distress and potholes) the system returns respectively the “excellent (1)” and “failed (5)” ratings. The reason is that k-medoid is more robust because it minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances. Further, the k-medoids is faster than the k-means algorithm, for the k-medoids requires the computation of the distance between every pair of objects only once and uses this distance at every stage of iteration.

Table 3: K-medoids Rating System

Surface Rating	Visible Distress/Road Features	Data
1 Excellent	None	282
2 Good	Longitudinal/ Transverse Cracks	596
3 Fair	Severe ravelling/ Block cracking	107
4 Poor	Patches / Occasionally potholes	35
5 Failed	Potholes / Speed bumps	10

Spatial Decision Support System (DSS) and pavement condition assessment mapping

The outcome of the unsupervised machine learning algorithms and the resulting pavement surface condition assessment scores can be spatially mapped, pointing out the areas of concern in a roadway network (Figure 3). The generated maps can then be used by the pavement management agencies to investigate the condition of pavement surfaces and to plan the construction maintenance operations.

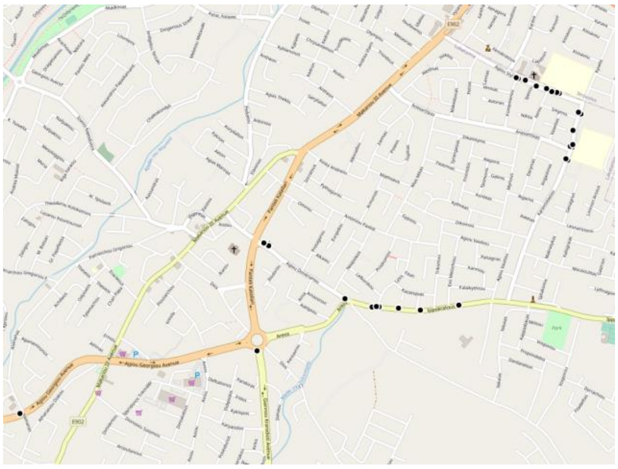

Figure 3: Spatial condition-assessment mapping (classes 4&5)

LIMITATIONS

As perceived limitations of the method are: (1) a less objective system evaluation rather than standardized evaluation system used (in contrast to the existing IRI rating system), and (2) the need to predefine the number of clusters (i.e. the value of k) used in the analysis.

CONCLUSIONS AND FUTURE WORK

The paper described a methodology based on low-cost devices (such as a smartphone) and unsupervised machine learning methods by which roadway pavement condition assessment can be achieved. The research motivation was to create a system that automatically assesses the quality of the pavements based on a triaxial accelerometer, a gyroscope and a clustering method, map the existence of roadway anomalies (using GPS) and save all sensed datapoints (and their features) for subsequent reinforcement learning through participatory sensing of multiple probe vehicles. The methodology used has the advantages of being low-cost and highly scalable, as the number of smartphone users increases day by day. Further, the smartphone-based approach is very worthwhile because it removes the need to deploy special sensors and expensive specialized vehicles. Finally, the methodology does not require any expertise in pavement condition assessment and allows for speedier and continuous evaluations of the roadway networks.

The proposed methodology and developed python-based decision support system platform are currently field-tested with multiple probe vehicles, more kilometres of roadway pavements and multiple data dates, for increasing the accuracy of the method and finetuning the data aggregation and clustering processes.

ACKNOWLEDGMENTS

This work is partially funded by the European Regional Development Fund and the Republic of Cyprus through the Cyprus Research & Technology Foundation (‘Restart 2016-2020’) (Grant No. INTEGRATED/0918/0056).

REFERENCES

AASHTO (1990) Guidelines for Pavement Management Systems. American Association of State Highway and Transportation Officials, Washington, D.C.

Alam, M.Y., Nandi, A., Kumar, A., Saha, S., Saha, M., Nandi, S. & Chakraborty, S. (2020) Crowdsourcing from the True crowd: Device, vehicle, road-surface and driving independent road profiling from smartphone sensors. Pervasive and Mobile Computing, vol. 61, pp. 101103.

Allouch, A., Koubâa, A., Abbes, T. & Ammar, A. (2017) Roadsense: A smartphone application to estimate road conditions using accelerometer and gyroscope. IEEE Sensors Journal, vol. 17, no. 13, pp. 4231-4238.

Federal Highway Administration (2010) Conditions and Performance – Policy. U.S Department of Transportation, Washington D.C, pp502.

Fugro Roadware (2015) ARAN 9000- Integrated Systems Retrieved from http://www.roadware.com/products/survey_equipment/aran_9000/.

Kyriakou, C., Christodoulou, S.E. & Dimitriou, L. (2019) Smartphone-based pothole detection utilizing artificial neural networks. ASCE - Journal of Infrastructure Systems, 25 (3), 04019019.

Kyriakou, C., Christodoulou, S.E. & Dimitriou, L. (2018) Detecting and Classifying Roadway Pavement Cracks, Rutting, Ravelling, Patching and Potholes Utilizing Smartphones. TRB 96th Annual Meeting Compendium of Papers, Washington, DC United States, paper #18-02674.

McGhee, K.H. (2004) Automated pavement distress collection techniques. Transportation Research Board.

Roadscanners (2020) Beyond the surface. Retrieved from https://www.roadscanners.com/.

Seraj, F., van der Zwaag, Berend Jan, Dilo, A., Luarasi, T. & Havinga, P. (2014) RoADS: A road pavement monitoring system for anomaly detection using smartphones. International Workshop on Machine Learning for Urban Sensor Data, Nancy, France.

Souza, V.M., Giusti, R. & Batista, A.J. (2018) Asfault: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning. Pervasive and Mobile Computing, vol. 51, pp. 121-137.

Walker, D., Entine, L. & Kummer, S. (2002) Pavement Surface Evaluation and Rating: PASER manual. University of Wisconsin, Transportation Information Center, Madison WI.

World Health Organization (2015) Global status report on road safety. World Health Organization.