

DAV - DATA ANALYTICS AND VISUALIZATION SYSTEM FOR ROADS

Bhavana Gannu*, K.S. Rajan

Lab for Spatial Informatics, International Institute of Information Technology Hyderabad -
gannu.bhavana@research.iiit.ac.in; rajan@iiit.ac.in

Commission IV, WG IV/3,IV/4

KEY WORDS: Analytical, Crowd sourced, Data amalgamation, Segment, Stretch, Track, Visualization, WebGIS

ABSTRACT:

This paper proposes a system for monitoring of condition and surface of roads in developing countries like India. This system will be used by government agencies to monitor municipal activities like road laying and planning. The system utilizes a database created by geo-citizens or government workers as an input. The heavy machinery in existing systems is not an optimized solution to this problem. Some existing systems use GPS and accelerometer data for determining such artifacts. So, it is evident that there is a need for a system that generates robust, frequent and accountable geo-tagged data. We propose a new collaborative model for such a purpose by fusion of data from multiple sensors hosted on smart-phones of several active geo-citizens. The system focuses mainly on volunteered geographic information, in which users can use their respective smart-phones to collect the data required and upload it for further analysis. The server side of the system infuses this data into a PostGIS database and displays the road condition on a near real-time basis over a WebGIS. The strength of a good visualization in imparting insight to decision-makers is widely recognized. We advance the paper by assessing procured road data and displaying it in an easy to understand format. In addition to visualization, the WebGIS component also provides for timeline analysis of changes in road conditions, which may help in the improved management of road infrastructure.

1. INTRODUCTION

A crucial aspect of urban infrastructure management in cities is the continuous monitoring and maintenance of roads. Specific departments employed by governing authorities track the road deterioration and ensure timely repair and maintenance of roads. Traditional methods in this domain use high power laser or radar sensors and thereby go on to become intricate and demanding. Regardless of the data collection method, the process of gathering, assimilating and post-processing of data is an expensive and time-consuming task with low coverage and often in need of assistance from experts. Additionally, frequent collection of data is essential to generate efficient and accountable results. High-end systems are inefficient in covering large areas at regular intervals of time which leads to complexity in activity planning.

To address these challenges in an ever-expanding urban environment, we propose an innovative collaborative system for road condition monitoring leveraging the existing technologies by data fusion approach. Derived from spatial citizenship, we envision an ecosystem operating through a mobile application on smart-phones of geo-citizens for large data collection and executing centralized processing. Prime advantages of such a platform are the multiplicity of data points and crowd-sourcing model.

Data collection from numerous sources ensures statistical advantage and expandable coverage for the monitoring system. Crowd-sourcing model provides reliable coverage and eliminates the need for deploying specialized road monitoring tools. This system is also dynamic to meet the increasing demand and prioritizes the areas of high usage. On the whole, the Data Analytics and Visualization (DAV) system can disrupt

the existing traditional methods and serve as an alternate innovative solution.

This paper primarily presents the architecture and system design of the two main subsystems of the DAV system viz. visualization and data fusion.

2. RELATED WORK

There is a wide range of existing systems that leverage different methods for road monitoring. Smartphones today are powerful devices that are capable of executing intensive tasks. We can make use of their multiple sensors for road monitoring. Complex LASER and RADAR based systems may be profitable in some cases but smartphone-based systems can be easily scaled up to run vigorous applications. Existing smartphone-based systems Nericell (Mohan et al., 2008) and Wolverine (Bhoraskar et al., 2012) detect road artifacts based on a change in accelerometer readings along the direction of gravity (Z direction). Roadroid (Forslöf, Jones, 2015) is another innovative approach to this problem that uses Root Mean Square (RMS) vibration analysis and quarter-car simulation techniques to determine the road quality.

The data fusion approach as mentioned in Maargha (Rajamohan et al., 2015) and (Gannu, Rajan, 2018) was a proof of concept developed to address the classification of roads by multi-sensor fusion uniquely and simply. Most of the systems mentioned above use a single source analysis to arrive at the result. To achieve an accountable source of data and to design a sustainable system, data source from several participants is necessary. In addition to this, such a system can also be used to perform temporal analysis of the road structure.

Monitoring based on crowd-sourced data has been a part

*Corresponding author

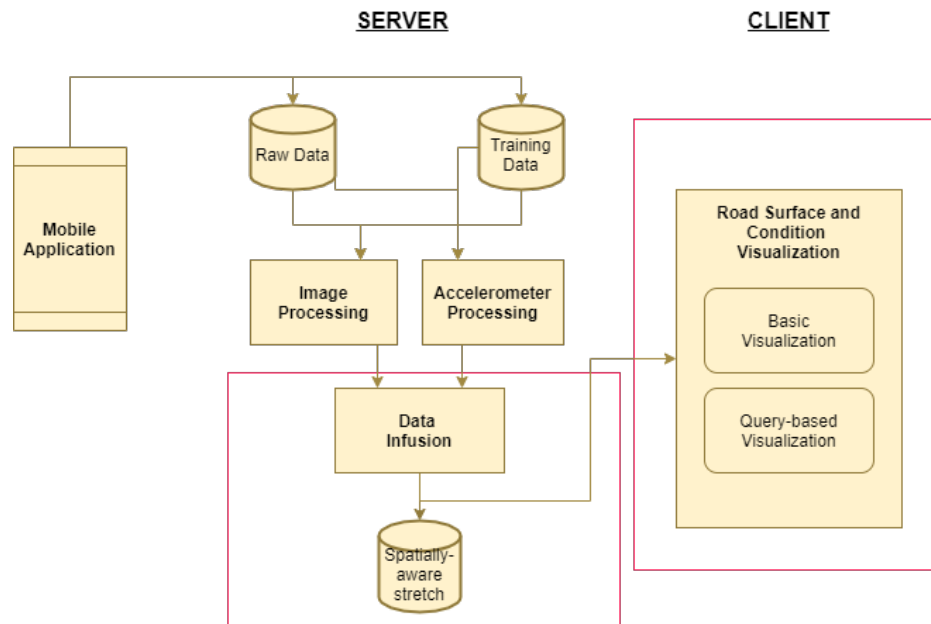


Figure 1. System Architecture

of many projects over the recent past. Air-quality monitoring based on crowd-sourced data has been gaining popularity. AirCloud (Cheng et al., 2014) uses a heterogeneous set of data sources as inputs. This data is stored and analyzed by the air quality analytics engine in the cloud, to provide accurate device calibration and fine-granularity estimation based on GPS location. Crowd-based data is also used in systems and methods for providing passive crowd-sourced alternate route recommendations (Curtis et al., 2013).

The combination of spatial data from the variety of sources on the web, being either legislative, commercially or voluntarily driven, is a major requirement for the establishment of a fully integrated geospatial web. Therefore, spatial data fusion techniques need to be linked to current web-developments, in particular on Spatial Data Infrastructures and the Semantic Web, to allow for standardized and effective use of combined spatial data for information retrieval. (Schmitt, Zhu, 2016).

3. SYSTEM DESIGN

The system architecture (Figure 1) is a distributed consumer-server software structure. The mobile application is used to acquire and train. The data collected is then uploaded to a server that runs an ML-based classifier on this data. The classified data is then fused and stored as different segments. This is the input to the visualization dashboard.

3.1 Mobile Application

The mobile application is built on the popularly used Android platform and is compatible with smartphones integrated with an accelerometer and a camera with at least 5 MP resolution. The application primarily has two modes; Data collection mode and Training mode. The data collection mode is used for the collection of necessary multi-sensor data that includes road images, accelerometer readings in all three axes, GPS coordinates, and GPS speed. The user shall securely mount the smartphone on the car's dashboard such that the target road is within the camera's view. A help screen assists the user to

use the app in data collection mode conveniently. Each user can collect data from a single lane of a road, we assume that the multiplicity of users will allow for complete coverage of the road. This also gives an advantage that even if a single segment is unsatisfactory, but other segments are classified as satisfactory, the overall majority rank of the segment can be ranked as satisfactory.

The second mode called Training mode is introduced to optimize the algorithm for data detection and storage. It works for both accelerometer data acquisition and image training. This is necessary to address the challenges of multi-sensor data fusion, especially in a crowd-sourcing environment.

3.2 Processing Server

In crowdsourcing systems, the fusion of information has new challenges to convey semantic meaning. They are 1) Source diverseness, where data obtained is from different smartphones of various specifications 2) storage diverseness, where the semantics of sensor data is originating from different forms like GPS sensors, accelerometer, and images. 3) multi-modal data, where the environment of the collection of data plays an important role. In summary, the fusion of multi-sensor data is feasible in a crowdsourcing based environment but leads to many challenges as listed above.

To address these challenges, this paper designs an efficient python based system to fuse the data as well as presents the results of the analysis in a web-based GIS framework, which constitutes the following two sections.

The classification of the data is performed using the algorithm reported in (Rajamohan et al., 2015). To summarize the process, tar roads vs. mud/concrete roads based on the intensity distribution of the scene and the mud roads are differentiated from concrete roads depending on the colorfulness of the image. Accelerometer readings across x,y,z are classified using k-NN algorithm. The road condition classification is based on the International Roughness Index (IRI) classes, viz., Good, Satis-

userId	segmentId	timestamp	roadCondition	surfaceType
15c38ad0	d14ff75c-5cd6-43e3-a52b-d91ab99921cb	1531655799	good	tar
11b89e1a	d14ff75c-5cd6-43e3-a52b-d91ab99921cb	1534330136	unsatisfactory	tar
5bcba494	d14ff75c-5cd6-43e3-a52b-d91ab99921cb	1535167767	good	tar
0633f3be	c1eac101-4a5c-41b5-a665-dda9add6574f	1539334166	poor	mud
5bcba494	c1eac101-4a5c-41b5-a665-dda9add6574f	1539340586	poor	mud

Table 1. Sample of the final processed segment output table, as stored in PostGIS

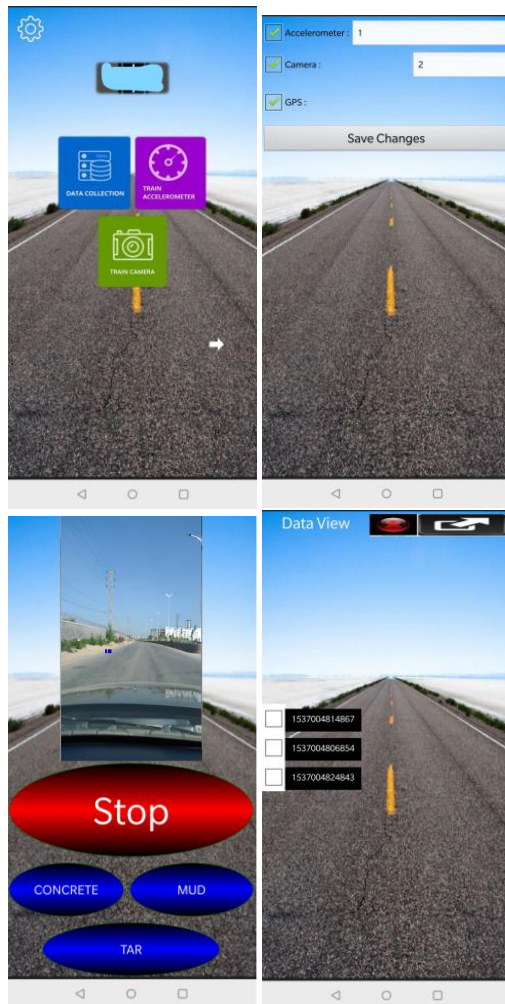


Figure 2. Mobile Application

factory, Unsatisfactory, and Poor. The final road surface types are Bitumen (tar), Concrete, and Mud.

If we observe the data in Table 1, multiple users can contribute to the single segment. This problem is resolved using majority voting.

3.3 Visualization Dashboard

In this paper, we present a WebGIS dashboard comprising a back-end python server that analyses and aggregates information from multiple inputs and showcases it on a OpenStreet-Map layer. For systems such as traffic monitoring systems, only tracking, and positional information is used. The proposed system not only using the tracking information but also the image information to give the optimized result. Hence, our system

processes positional information and attaches on the fly analysis of the data collected from different sensors. This is achieved by collating multiple users' data efficiently.

4. DATA FUSION MODEL

Data fusion involves achieving three objectives which are demarcated as a. *compactness* of data that deals with how any particular data is uniquely and concisely represented, b. *extensiveness* of the data that measures the number of attributes associated and c. *factuality* of the data that shows how true the data is to the real world. The proposed system intends to conform to a novel model that adheres to the aforementioned principles.

4.1 Challenge in data fusion

As the current system deals with a diverse stream of data, there arises a problem of standardizing this data to conclude some meaningful information. We segregate the problems into two categories a. *accuracy* which means that using low-cost devices we should get the accuracy close to high-end systems and b. *result conflict* which arises due to the multiplicity of the data, in which case there are multiple conclusions made for the same attribute of the data.

4.2 Terminology

To solve the above issues we propose the following model.

4.2.1 Segment: A segment is the smallest tangible unit that can be classified into one of 12 categories as listed in Figure 3.

	Mud	Concrete	Bitumen Tar
Good			
Satisfactory			
Unsatisfactory			
Poor			

Figure 3. Legend for different classes

4.2.2 Track: A track is comprised of multiple classified segments. Each track is contributed by an individual user. For example, Figure 4 refers to a track contributed by a single user for the Gachibowli road.

4.2.3 Stretch: A stretch also is termed as a spatially-aware stretch, is comprised of multiple segments of data contributed by multiple users. We use a maximum voting algorithm to determine the classes assigned to the different segments in a stretch.



Figure 4. Gachibowli road

4.2.4 Road: In our system, a road consists of multiple sequential spatially-aware stretches.

4.3 Amalgamation Algorithm

Instead of high-accuracy single-track data, getting multiple tracks and amalgamating them into a single stretch in the crowd-sourced model needs to be done so that the information from "N" sources helps improve the data consistency and hence it's accuracy.

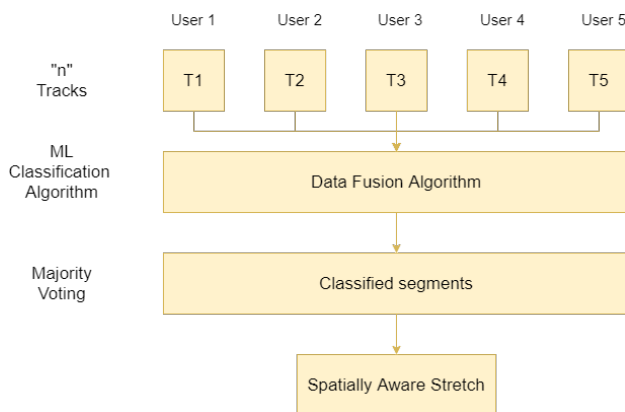


Figure 5. Data fusion flowchart

Algorithm 1 User data processing

Require: Data collected by a user through Android App
Ensure: Data is correctly stored in a Database

STEPS:

- User records road data on their smartphone
- User uploads track data to the database
- Any training data recorded by the user is uploaded to the training database
- Each track is first map-matched to remove any GPS error and correct the track information
- Corrected track is classified into different classes using k-nn algorithm
- Classified Data is divided into segments
- Segments are stored in a table

4.3.1 Database Design The Figure 6 are the tables we have designed to efficiently store the data and extract information.

Algorithm 2 Amalgamation of segments

Require: SegmentIDs of segments belonging to different users
Ensure: Each segment has an associated class

STEPS:

- For each stretch, get list of associated classified segments
- By principle of majority voting, assign the final class to each segment
- Arrive at the final spatially-aware stretch by combining all the final segments and storing in a separate table

A majority ranking model based on the number of collaborators is used to rank each of the segments. The visualization is the most recently recorded road condition and surface in a color-coded way based on the IRI scale.

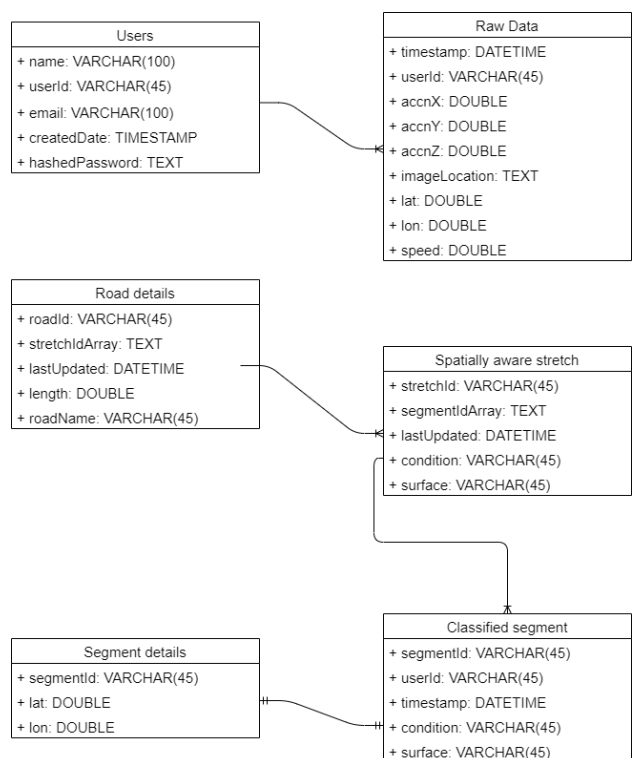


Figure 6. Database Details

5. ANALYTICAL TOOL

5.1 Case Study

The purpose of this case study was to thoroughly analyze the system which could reveal factors or information otherwise ignored or unknown.

The proof of concept of the system has been deployed and data was collected on a stretch of 15 km in Gachibowli and Tarnaka areas of Hyderabad, India on a nearly monthly basis for 5 unique users for 1.5 years for validating the concepts and verifying the proposed models. The analysis in the next section is the outcome of this study area.

Road Name	%Good	%Satisfactory	%Unsatisfactory	%Poor	Date
Old Mumbai Highway	93	0	0	7	10-12-2018
Old Mumbai Highway	95	0	0	5	21-09-2019
Osmania University Road	85	9	0	6	11-1-2018
Osmania University Road	94	0	1	5	03-01-2019
IIIT Main Road	98	0	1	1	10-12-2018
IIIT Main Road	97	0	0	3	21-09-2019
Gachibowli Road	78	13	0	9	10-12-2018
Gachibowli Road	67	17	0	16	21-09-2019

Table 2. Current state of roads in Hyderabad city

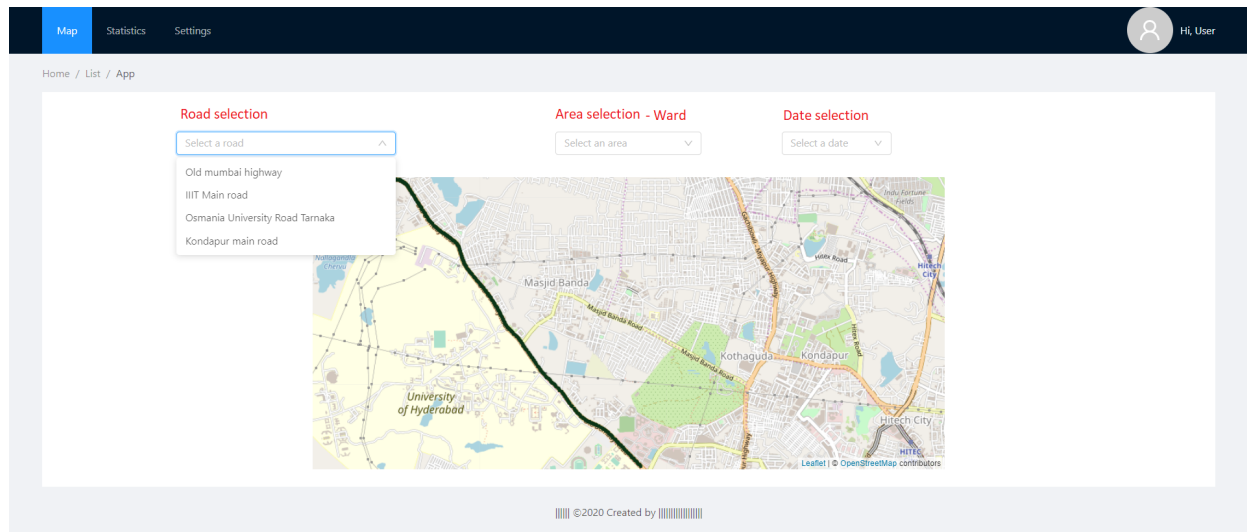


Figure 7. Snapshot of the visualization tool

5.2 Query Engine

An assumption has been made that even-though commuters travel through different lanes on a multi-lane road, we do not distinguish the lanes to provide the outcome. Let us take an example of the Old Mumbai highway in Hyderabad, which is a 3-lane road on both sides. In our experiments we have observed that commuters prefer to take the lane that is free from potholes. Hence, a majority vote of all user data collected for the stretch of road is taken for each segment and final verdict is allocated.

BASIC VISUALIZATION

The basic visualization shows the most current classification according to the 12 classes.

QUERY-BASED VISUALIZATION

The system supports two types of queries as given below: **Spatial Queries:** Analysis of the road based on municipality ward can be opted.

Temporal Queries: Queries based on different time periods can also be performed for different roads.

Using this tool we are able to answer the following queries as per data from Table 2 which represents data collected by

volunteers for different parts of the city:

Spatial Queries

Road length: 10 km

Most stable road in any city: Old Mumbai highway, IIIT Main road

Analysis of current road : Old Mumbai Highway - This road is 95% good.

Temporal Queries

Road length: 10 km

Analysis of status change of a road over a period of time
Status change of Old Mumbai Highway over the past 1 year - negligible

Current state of roads in any city In Hyderabad, over 80% of the roads have good condition.

We can see from the above data that we have built a spatially-aware knowledge base from the existing system.

6. CONCLUSION

A low cost sensor based visualization framework is a solid match for city districts since it removes the requirement for specific high-end checking infrastructure. This system can be utilized to produce geo-labeled insights on road infrastructure. For

instance, faster deterioration of some roads over others can be easily determined. The framework presently gathers and shows the information which gives different organizations a prepared reference about surface and condition qualities. This framework proposes an intuitive citizen supported model to screen road surface and condition. Crowd-sourced model for such applications can be actualized by 1) Geo-citizens; 2) Office staff of districts or city enterprises; 3) Campaign or strategic activities which are onetime occasions; 4) Cabs and commercial vehicle operations units. The present system is a convenient methodology that can be utilized in any of these methods. Most of the existing systems only provide information regarding the condition of the road, but this system also covers the surface information. By making such a system available, we show a true example of utilizing GIS in governance.

REFERENCES

- Bhoraskar, R., Vankadhara, N., Raman, B., Kulkarni, P., 2012. Wolverine: Traffic and road condition estimation using smartphone sensors. *2012 fourth international conference on communication systems and networks (COMSNETS 2012)*, IEEE, 1–6.
- Cheng, Y., Li, X., Li, Z., Jiang, S., Li, Y., Jia, J., Jiang, X., 2014. Aircloud: a cloud-based air-quality monitoring system for everyone. *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, 251–265.
- Curtis, S., Halber, E. P., Evans, G. M., Petersen, S. L., 2013. Passive crowd-sourced map updates and alternate route recommendations. US Patent 8,620,532.
- Forslöf, L., Jones, H., 2015. Roadroid: Continuous road condition monitoring with smart phones. *Journal of Civil Engineering and Architecture*, 9(4), 485–496.
- Gannu, B., Rajan, K., 2018. Maarg: Geo-processing and information fusion approach for road condition and surface monitoring. <http://toc.proceedings.com/47496webtoc.pdf>.
- 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*, 323–336.
- Rajamohan, D., Gannu, B., Rajan, K. S., 2015. MAARGHA: a prototype system for road condition and surface type estimation by fusing multi-sensor data. *ISPRS International Journal of Geo-Information*, 4(3), 1225–1245.
- Schmitt, M., Zhu, X. X., 2016. Data fusion and remote sensing: An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, 4(4), 6–23.