International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 2 | ISSN : 2456-3307

Detecting and Alerting Damaged Roads Using Smart Street System

S. Shathish Kumar[1], Dr. A. Jayachandran[2]

[1]PG Scholar, Department of M.Sc(Software Engineering), PSN College of Engineering & Technology, Tirunelveli,Tamilnadu,India
[2] Professor & Head, Department of M.Sc(Software Engineering), PSN College of Engineering & Technology, Tirunelveli,Tamilnadu,India

ABSTRACT

Develop an infrastructure-free approach for anomaly detection and identification based on data collected through a smartphone application (SMART STREET). The approach is capable of effectively finding the damaged roads and effectively classifying roadway obstacles and knowing its type using machine learning algorithms, and accelerometer in smartphone ,as well as prioritizing actionable ones in need of immediate attention based on a proposed "anomaly index." We explore some algorithms that combine clustering with classification and introduce appropriate regularization in order to concentrate on a sparse set of most relevant features, which has the effect of reducing over fitting.I introduce, combines novel metrics of obstacle irregularity computed based on the data captured and alerting system by the smartphone application (Smart Street). It Results by capturing the location of damaged road and transferring it to the corporation by an alert message .The data collector in corporation will receive the alert message and instruct the corporation to take necessary action for repairing the road.
Keywords: Connected Autonomous Vehicles, Support Vector Machine, Mean Squared Error, Sparse Support Vector Machine, Receiver Operating Characteristics, Area under the ROC Curve.

I. INTRODUCTION

As of 2014, 54% of the earth's population resides in urban environments, a percentage that is expected to reach 66% by 2050. This increase would amount to about 2.5B people added to urban populations . At the same time, there are now 28 mega-cities (with _10M people) worldwide, accounting for 22% of the world's urban population and projections are for more than 41 mega-cities by 2030. It stands to reason that managing urban areas has become one of the most critical challenges our society faces today. The emerging prototype for a Smart City is one of an urban environment with a new generation of innovative services for transportation, energy distribution, health care, environmental monitoring, business, commerce, emergency response, and social activities. The term "Smart City" is used to capture this overall vision as well as the intellectual content that supports it. From a technological point of view, at the heart of a Smart City is a cyber-physical infrastructure with physical elements (e.g., roads, vehicles, power lines) which are continuously monitored through various sensors to observe, for instance, air/water quality, traffic conditions, occupancy of parking spaces, the structural health of bridges, roads, buildings, as well as the location and status of city resources including transportation vehicles, police cars, police officers, and municipal workers. The data collected need to be securely communicated (mostly wirelessly) to information processing and control points. These data may be shared and the control points can cooperate to generate good (ideally, optimal) decisions regarding the safe operation of these physical elements (e.g., vehicles guided through the city).

It is important to emphasize that what ultimately makes the city "smart" is not simply the availability of data but the process of "closing the loop" consisting of sensing, communicating, decision making, and actuating. It is a highlevel illustration of this process, which must take place while taking into account important issues of privacy, security, safety, and proper energy management necessitated by the wireless nature of most data collection and actuation mechanisms involved. Finally, equally important as the development of a cyber-physical infrastructure is the necessity for Smart Cities to engage – but not coerce – their citizens. Unlike other organizations (e.g., corporations or military units) which can often assume compliance of their human constituents, cities must resonate with their population's goals, means, desires, and freedom of choice. Thus, there is a crucial trade-off between technological efficiency and user engagement and an associated challenge of integrating a crucial social aspect into the Cyber-Physical environment that constitutes a Smart City. Although a cyber-physical infrastructure is instrumental in realizing the Smart City vision described above, such infrastructure comes at a significant cost. Embedding sensors in an urban environment (e.g., induction loops in roadways to measure traffic flows or sensors monitoring the state of power lines underground) does not only entail an installation expense, but also significant maintenance costs. For wirelessly networked sensors, for instance, battery life is limited, so that a battery replacement plan must be in place or additional intelligence must be present in the sensors to manage their energy usage. As another example, to monitor the structural health of roads, one approach is to build specialized vehicles heavily equipped with a variety of sophisticated sensing devices and design patrol paths in urban environments through which such vehicles can perform this function. Clearly, the cost of building and maintaining such vehicles is significant, not to mention their operation cost. An exciting feature of Smart Cities, however, is the potentialto exploit the ubiquitous availability of wireless devices and new technologies embedded in vehicles in order to meet several Smart City goals in an infrastructure-free manner. The majority of urban dwellers nowadays carry a smartphone, a device that contains three important functionalities: (i) the ability to locate itself through GPS, (ii) an accelerometer which can provide several forms of movement information, and (iii) a wireless Internet connection which enables it to communicate with other devices or with servers in an already existing network infrastructure. Finally, the shear volume of these devices provides the opportunity to process such "big data" in ways that can bypass inaccuracies or errors. Looking into the not-so-distant future, the connected vehicle initiative will be transforming vehicles into mobile nodes in a network which does not require a Smart City to build or maintain it, but simply to take advantage of the vast amount of data from the vehicles which will allow them to be self-driven. Recommendation The Machine learning, Data Mining methods are described, as well as several applications of each method to cyber intrusion detection problems. The complexity of different machine learning and data mining algorithms is discussed, and the paper provides a set of comparison criteria for machine learning and data mining methods and a set of Problem to solve Cyber security is the set of technologies and processes designed to protect computers, networks, programs, and data from attack, unauthorized access, change, or destruction. Cyber security systems are composed of network security systems and computer security systems. Each of these has, at a minimum, a firewall, antivirus software, and an intrusion detection system .Intrusion detection system s help discover, determine, and identify unauthorized use, duplication, alteration, and destruction of information systems. The security breaches include external intrusions attacks from outside the organization and internal intrusions.

There are three main types of cyber analytics in support of intrusion detection systems: misuse-based, anomaly-based, and hybrid. Misuse-based techniques are designed to detect known attacks by using signatures of those attacks. They are effective for detecting known type of attacks without generating an overwhelming number of false alarms. They require frequent manual updates of the database with rules and signatures. Misuse-based techniques cannot detect novel attacks. Anomaly-based techniques model the normal network and system behaviour, and identify anomalies as deviations from normal behaviour. They are appealing because of their ability to detect zero-day attacks. Another advantage is that the profiles of normal activity are customized for every system, application, or network, thereby making it difficult for attackers to know which activities they can carry out undetected. Additionally, the data on which anomaly-based techniques alert can be used to define the signatures for misuse detectors. The main disadvantage of anomaly-based techniques is the potential for high false alarm rates because previously unseen system behaviours may be categorized as anomalies.

This paper focuses primarily on cyber intrusion detection as it applies to wired networks. With a wired network, an adversary must pass through several layers of defence at firewalls and operating systems, or gain physical access to the network. However, a wireless network can be targeted at any node, so it is naturally more vulnerable to malicious attacks than a wired network. The Machine learning and data mining methods covered in this paper are fully applicable to the intrusion and misuse detection problems in both wired and wireless networks. The reader who desires a perspective focused only on wireless network protection is referred to papers such as Zhang et al. , which focuses more on dynamic changing network topology, routing algorithms, decentralized management, etc.

II. METHODS AND MATERIAL

1. RELATED WORK

The authors SongnianLi, Suzana Dragicevic, et al. in [6] made review on various geospatial theory and methods used to handle geospatial big data. Given some special attributes, authors considered that customary data taking controlling methodologies and techniques are lacking and the following domains were recognized as in requirement for further advancement and examination in the control. This incorporates the advancements in calculations to manage real-time analytics and to support ongoing flooding data, as well as improving new spatial indexing techniques. The improvement of theoretical and methodological ways to deal with transfer of big data from illustrative and parallel research and applications to ones that investigates easygoing and illustrative connections. Boosting is an ensemble supervised learning method that constructs a classifier as a linear combination of simpler weak classifiers [. In this work, we will use decision stumps as the component classifiers used by Adaboost. A decision stump makes a prediction based on the value of a single input feature. AdaBoost maintains a distribution of weights for the training sample points. During each iteration, a weak classifier is trained by focusing on the data points that have been misclassified by the previous weak classifier, and the weights get updated based on the misclassification error. When these iterations terminate, AdaBoost combines the decisions of these weak classifiers using an optimally weighted majority vote. The number of iterations is selected through cross-validation.

Random Forests: Bagging is a technique for reducing variance of an estimated predictor by averaging many noisy but approximately unbiased models. A random forest is an ensemble of de-correlated trees [13]. Each decision tree is grown on a training set constructed by sampling (with replacement) a random subset of the original data. On each split, amongthe full set of the original variables only a subset of fixed size is considered and the best split using these is selected to split the node. Each tree is fully grown until a minimum size is reached, i.e., there is no pruning. While the predictions of a single tree are highly sensitive to noise in its training set, the average of many trees is not, as long as the trees are not correlated. Bootstrap sampling achieves de-correlating the trees by constructing them using different training sets. To make a prediction at a new point, Random Forests take the majority vote among the outputs of the grown trees in the ensemble. Random Forests run very efficiently on large datasets, do not have the risk of overfitting as Adaboost does and can handle data sets with unbalanced classes. The number of trees in the ensemble is selected through cross-validation.

Because bumps naturally fit into different categories with distinct signatures, we introduce a new hierarchical approach that first clusters the bumps into a pre-determined number of clusters L and then trains a different classifier for each cluster. For clustering, we use the widely used method of k- means++ that is based on a heuristic to find centroid seeds for k-means clustering. For clustering, we employ Sparse Support Vector Machines (SSVMs), that base the classification decision on a subset of features only. Due to the limited size of the data set we have in our disposal, the use of a sparsity inducing classifier is critical; a classifier that uses all features (like SVM) would not be able to learn all its parameters from a small training set. We will use the notation C-SSVM to refer to this clustering and classification method. We conduct various experiments to select the optimal number of clusters L. For clustering, a correlation-based distance metric yielded the best results. Specifically, for any two bumps i; j with feature vectors f (i); f (j) we use 1 cov(f (i); f (j))=(_(f (i))_(f (j)) as their distance metric, where _(f (i)) is the sample variance of the feature vector f (i).

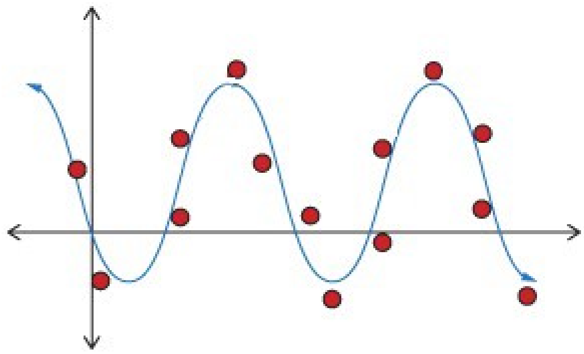
6) A Decision Support System for Prioritizing Anomalies: All the methods we have outlined in this subsection, classify a test bump by comparing a decision function of its features g(f (i)) to a threshold _. The bump is classified as actionable if g(f (i)). The distance g(f (i)) can in fact be used to prioritize among actionable bumps; the larger this distance the more confident we are about the bump being actionable. From the methods we considered, logistic regression provides explicitly the likelihood of a bump being actionable, which can then be used to order actionable bumps. Another way to make more robust predictions is to informally combine several methods, seeking for instance consensus among various classifiers in order to declare a bump as actionable.

Unsupervised Anomaly Detection Methods

In the second approach, which focuses on anomaly detection methods, we define a "normal" bump signature in two different ways: (i) a normal signal with a sinusoidal pattern, or (ii) a normal signal with an expected range of amplitude. We then propose metrics to measure how different a test bump is from a normal pattern.

Sinusoidal Fitting and a Mean Squared Error Metric:

The key idea is that the _-filtered signature defined in exhibits a pattern very similar to a sinusoidal function for nonactionable bumps, whereas actionable bumps do not exhibit this behavior. To explore this apparent separation, we fit a sine (or cosine) function to the _-filtered signature of a bump (see the red curves) and calculate a Mean Squared Error (MSE) as a goodness-of-fit metric. Specifically, the lower the MSE is, the more actionable we expect a bump to be. Android was built from the ground-up to enable developers to create compelling mobile applications that take full advantage of all a handset has to offer. It was built to be truly open. For example, an application can call upon any of the phone's core functionality such as making calls, sending text messages, or using the camera, allowing developers to create richer and more cohesive experiences for users.



Classification and Anomaly Detection System Comparison

The decision support and the anomaly detection systems can be seen as complementary but distinct approaches to the same problem. The first distinction is goal-oriented: the decision support system focuses on differentiating actionable bumps from non-actionable ones, while the anomaly detection system concentrates on identifying the most urgent actionable bumps. Another key distinction relates to how these systems can be used. The first (decision support) system is based on machine learning/classification methods which are supervised, that is, they require a "labeled" training set to learn the various classifier parameters and thresholds. The anomaly detection system, on the other hand, is unsupervised; it simply ranks bumps based on the anomaly index we introduced. It provides no guarantees but suggests that higher ranked bumps are more likely to be actionable. The consistency of the results of the two methods can be assessed by comparing whether the bumps at the top of the ranked anomaly detection list are also classified as actionable based on the decision support system. Moreover, as we discussed in Sec. III-A6, the supervised classification methods can also provide a metric of confidence the classifier has in a positive decision and this can be used to evaluate the top ranked bumps from the anomaly detection system.

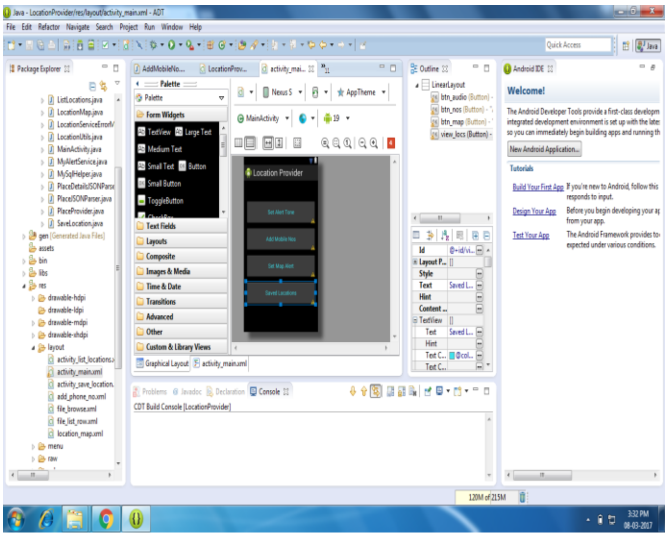
In [13] Yuehu Liu, Bin Chen et al. have proposed another technique for overseeing gigantic remote sensing image data by utilizing HBase and MapReduce framework. At first they have divided the actual image into various tiny pieces, and store the blocks in HBase, which is dispersed in a gathering of hubs. They have used MapReduce programming model on handling the stored blocks, which can be simultaneously executed in a group of hubs. The hubs in Hadoop cluster have no prerequisites for high performance and accuracy so that they can be exceptionally inexpensive. Besides, as a result of the high versatility of Hadoop, it is anything but difficult to add new hubs to the cluster, which was normally exceptionally troublesome in general ways. Finally they notice that the speeds of data commerce and processing increase because the cluster of HBase grows. The outcomes demonstrate that HBase is extremely reasonable for large image information stockpiling and handling.

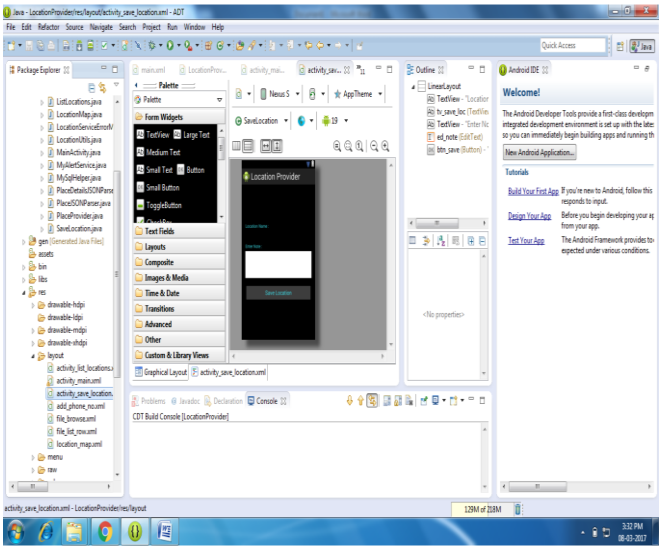
The authors Chaowei Yang, Michael Goodchild et al. in [14] have projected a replacement paralleling storage and access methodology for big scale NetCDF scientific information that is enforced dependent on Hadoop. The recovery technique is actualized dependent onMapReduce. The Argo data is utilized to exhibit the proposed technique. The execution is looked at under a disseminated domain taking into account PCs by utilizing distinctive data scale and diverse task numbers. The examinations result demonstrates that the parallel strategy can be utilized to store and retrieve the vast scale NetCDF productively.

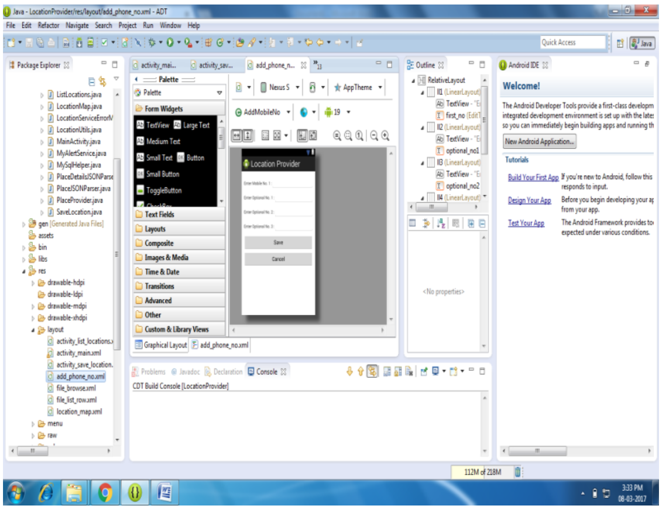
Big data has turned into a noteworthy center of worldwide interest that is progressively pulling in the acknowledgment of the educated community, industry, government and other association. The incremental development in volume and changing

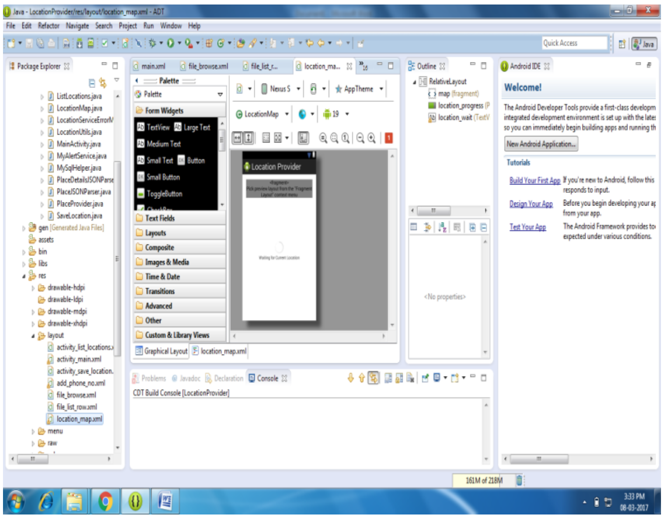
III. EXPERIMENTAL RESULTS

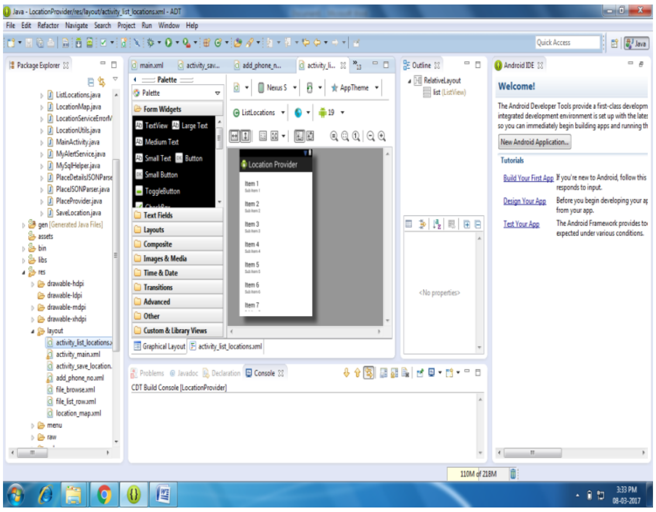
The implementation results can be shown as figure below











A. Web Tool Platform

Conventional data processing sometimes consider data from one domain. In this big data era, everyone has to make wide selection of datasets from totally different sources in several domains. Each of these datasets comprise of various strategies such as alternate representation, measurements, scale, dissemination, and consistency. Removing the force of information from numerous diverse (however conceivably associated) data sets is an extraordinary arrangement in big data research, which incorporates basically isolating big data from customary data mining undertakings. Which itself prompts propelled procedures that can combdata fusion and conventional data fusion contemplated in the database group.

B. Modelling Platform

The term crowd sourcing means to data acquirement by vast and various gatherings of individuals, who much of the time are not prepared measurer and who don't have exceptional PC learning, utilizing web innovation. In this way, these information are exchanged to and put away in a typical computer architecture e.g. a focal or a combined database, or in a distributed computing environment. The ensuing undertaking of programmed data incorporation and handling are vital to produce additional data. Envisioning and checking the nature of data. There are wide assortments of methods accessible and adjusted to imagine, dissect, control and composite big data to make this sort of data volume reasonable. Some of these procedures are data fusion,

An assortment of data mining techniques can be applied to find associations and regularities in data, extract knowledge in the forms of rules and predict the value of the dependent variables. Common data mining techniques which are used in almost all the sectors are listed as: Naive Bayes, Decision Tree, Artificial neural network (ANN), Bagging algorithm, K- nearest neighborhood (KNN), Support vector machine (SVM) etc. Data mining is an important step of knowledge discovery in databases (KDD) which is an iterative process of data cleaning, integration of data, data selection, pattern recognition and data mining knowledge recognition. KDD and data mining are also used interchangeably. Data mining encompasses association, classification, clustering, statistical analysis and prediction. Data mining has been widely used in areas of communication, credit assessment, stock market prediction, marketing, banking, education, health and medicine, hazard forecasting, knowledge acquisition, scientific discovery, fraud detection, etc but data mining holds significant presence in every field of medical for the diagnosis of several diseases such as diabetes, skin cancer, lung cancer, breast cancer, heart disease, kidney failure, kidney stone, liver disorder, hepatitis etc. Data mining applications include analysis of data for better policy making in health, prevention of various errors in hospitals, detection of fraudulent insurance claims early detection and prevention of various diseases, value for more money, saving costs and saving more lives by reducing death rates.With a wired network, an adversary must pass through several layers of defence at firewalls and operating systems, or gain physical access to the network. However, a wireless network can be targeted at any node, so it is naturally more vulnerable to malicious attacks than a wired network. The Machine learning and data mining methods covered in this paper are fully applicable to the intrusion and misuse detection problems in both wired and wireless networks. The reader who desires a perspective focused only on wireless network protection is referred to papers such as Zhang et al. , which focuses more on dynamic changing. The ensuing undertaking of programmed data incorporation and handling are vital to produce additional data. envisioning and checking the nature of data. There are wide assortments of methods accessible and adjusted to imagine, dissect, control and composite big data to make this sort of data volume reasonable.

IV. CONCLUSION

The goal of this paper is to demonstrate how the ubiquitousavailability of wireless devices can enable the development of effective infrastructure-free approaches for solving problems in Smart Cities. In particular, we have concentrated on the problem of detecting and classifying roadway obstacles (bumps) so as to differentiate between actionable bumps which correspond to obstacles that require immediate attention, and non-actionable bumps (e.g., cobblestone streets, speed bumps) for which no immediate action is needed. This classification enables City officials to efficiently and effectively prioritize repairs. We developed two complementary methods to that end. The first method uses classification algorithms. The second method introduces an anomaly index which captures the degree of regularity of a bump, and uses this index to differentiate between more “normal” bumps (non-actionable) from the “anomalous” (actionable) bumps. As a next step of this work, it is important to be able to differentiate between different types of obstacles; for example, to distinguish a pothole from a poorly repaired sunk casting. The vision is that the accelerometer and GPS data collected by the app can be used in additional applications. An example is detecting wet or icy road conditions or obstacles causing vehicles to experience abrupt motions in a horizontal/lateral, rather than vertical direction. All these results, combined with the ones by our decision support system, could potentially be integrated to create a global “road smoothness” or “road quality” metric, available to all citizens through appropriate web sites, or specialized apps, or even integrated into map/navigation applications (Google maps, Waze, Apple maps, etc.), that can then be used to select the best route.

V. REFERENCES

[1]. "World"s population increasingly urban with more than halfliving in urban areas," Report on World Urbanization Prospects, United Nations Department of Economic and Social Affairs, July 2014, http://www.un.org/en/development/desa/news/population/worldurbanization- prospects-2014.html.

[2]. K. Dresner and P. Stone, "Multiagent traffic management: a reservation based intersection control mechanism," in Proceedings of the Third International Joint Conference on Autonomous Agents and MultiagentsSystems, 2004, pp. 530–537.

[3]. "A multiagent approach to autonomous intersection management," Journal of artificial intelligence research, pp. 591–656, 2008.

[4]. A. Fortelle, "Analysis of reservation algorithms for cooperative planning at intersections," 13th International IEEE Conference on Intelligent Transportation Systems, pp. 445–449, Sep. 2010.

[5]. S. Huang, A. Sadek, and Y. Zhao, "Assessing the mobility and environmental benefits of reservation-based intelligent intersections using an integrated simulator," IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 3, pp. 1201,1214, 2012.

[6]. K. Zhang, A. D. L. Fortelle, D. Zhang, and X. Wu, "Analysis and modeled design of one state-driven autonomous passing-through algorithm for driverless vehicles at intersections," 2013 IEEE 16th International Conference on Computational Science and Engineering, pp. 751–757, 2013.

[7]. Y. J. Zhang, A. A. Malikopoulos, and C. G. Cassandras, "Optimal control and coordination of connected and automated vehicles at urban traffic intersections," in American Control Conference, 2016, submitted.

[8]. Y. Geng and C. Cassandras, "New smart parking system based on resource allocation and reservations," Intelligent Transportation Systems,IEEE Transactions on, vol. 14, no. 3, pp. 1129–1139, 2013.

[9]. Y. Geng and C. G. Cassandras, "Multi-intersection traffic light control with blocking," J. of Discrete Event Dynamic Systems, vol. 25, no. 1, pp. 7–30, 2015.

[10]. http://boston.cbslocal.com/2014/04/09/massachusetts-to-set-aside-40-million-to-fix-potholes