CTIPS Center for Transformative Infrastructure Preservation and Sustainability

USDOT Region 8 University Transportation Center

Leveraging Rideshare Electric Vehicle Data for Road Condition Monitoring: A Crowdsourcing and Machine Learning Approach

CTIPS-021

Approved 7/16/2024

University

Colorado State University

Principal Investigators

Gaofeng Jia, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering
Colorado State University
Phone: (970) 491-6580
Email: gaofeng.jia@colostate.edu
ORCID: 0000-0001-9419-8481

Research Needs

One key part of pavement management is to assess the road condition and identify pavement distresses such as cracks and potholes. These road distresses, if not identified and repaired timely, could compromise road safety, cause expensive damage claims, and also lead to more expensive later repairs (e.g., in the form of either reactive maintenance or emergency maintenance). Accurate, robust and fast detection of cracks and potholes is critical to enabling timely and cost-effective pavement maintenance and management, thus helping preserve good state of the transportation system and improve road safety.

To assess pavement condition, pavement condition data need to be collected first and then processed for distress detection and assessment. However, traditional pavement data collection still relies on manual or specialized vehicles equipped with expensive sensors (such as 3D laser scanning (e.g., LiDAR), ground penetrating radar (GPR)) and requires personnel driving along each road in the road networks of interest. Therefore, traditional road inspection methods are often costly, labor-intensive, and sporadic with limited coverage, leading to delayed maintenance and compromised safety. Recently, 2D image-based sensing (e.g., using cameras from phone and in private vehicles) has been investigated as a low-cost option for rapid road condition data collection and has been found useful for detecting cracks and potholes. Different machine learning algorithms have also been developed for image-based road damage detection [1-5, 8]. These investigations have shown the potential and promise of using images/videos collected by cameras and machine learning for automated pavement condition assessment. *Recent advancements in machine learning (ML) and the proliferation of electric vehicles (EVs) equipped with various sensors offer a promising avenue for revolutionizing road condition assessment practices.*

Despite the growing prevalence of electric vehicles (EVs) equipped with advanced sensors, their potential for contributing to infrastructure maintenance, specifically road condition assessment, remains largely untapped. Instead of installing specialized sensors, there is the possibility of using the sensors (e.g., cameras and GPS) for pavement data collection and using the collected images and/or videos for pavement condition assessment. This would help reduce the cost of pavement data collection. Continuous advancements in vehicle technologies and sensor capabilities may further enhance the feasibility and accuracy of using EVs for infrastructure monitoring and assessment in the future.

On the other hand, crowdsourcing (e.g., private EVs and rideshare EVs) has the benefit of extensive coverage of roads, frequent and up-to-date data collection, as well as reduced cost compared to traditional way of data collection. With more and more EVs in rideshare services (e.g., Uber, Lyft), there is an opportunity to leverage rideshare EV data for road condition monitoring.

While promising, the feasibility of using images/videos captured by cameras on EVs for road condition monitoring has yet been investigated. There are also knowledge gaps related to data collection protocol, processing, use for road condition assessment, and integration with existing pavement condition assessment systems or platforms. For example, although the cameras on EVs capture images and videos that include the road surface, their primary design is oriented towards navigation and safety. This means the focus is not specifically on detailed road surface analysis (like identifying small cracks or potholes), and the viewing angles might not always be optimal for this purpose. The data will need to be processed, analyzed, and integrated using suitable algorithms to assess the road conditions from the captured footage.

This project will thoroughly investigate the viability of using rideshare crowdsourcing EV data for road condition assessment and will develop machine learning algorithms that uses data from EVs to automatically identify road damages such as cracks and potholes. The project has the potential to offer a more efficient, cost-effective, and real-time approach to road and pavement condition monitoring over large road networks and provide critical information for timely maintenance.

Research Objectives

The goal of this project is to develop an effective solution for road condition assessment using rideshare crowdsourcing EV data, addressing both the technological and practical challenges involved. The following major objectives are designed to meet this goal.

1. Establish a standardized data collection and processing framework that facilitates the efficient aggregation, anonymization, and preparation of high-quality rideshare EV data, ensuring its relevance and quality for road condition analysis.
2. Design, implement, and validate advanced machine learning algorithms for accurate and robust identification and assessment of road damages, such as cracks and potholes, from the processed rideshare EV data, with an emphasis on integration of information from multiple cameras, scalability and adaptability to diverse road and environmental conditions.
3. Evaluate the feasibility and overall effectiveness of using crowdsourced rideshare EV data for real-time road condition assessment.
4. Develop integrated automated tools to facilitate the implementation and adoption of the proposed approach for road condition assessment.

Research Methods

To address the aforementioned research needs, this project will establish a framework for collecting and processing rideshare crowdsourcing EV data and develop machine learning algorithms that uses data from EVs to automatically assess road conditions and identify road damages such as cracks and potholes. Figure 1 shows an overview of the proposed research.

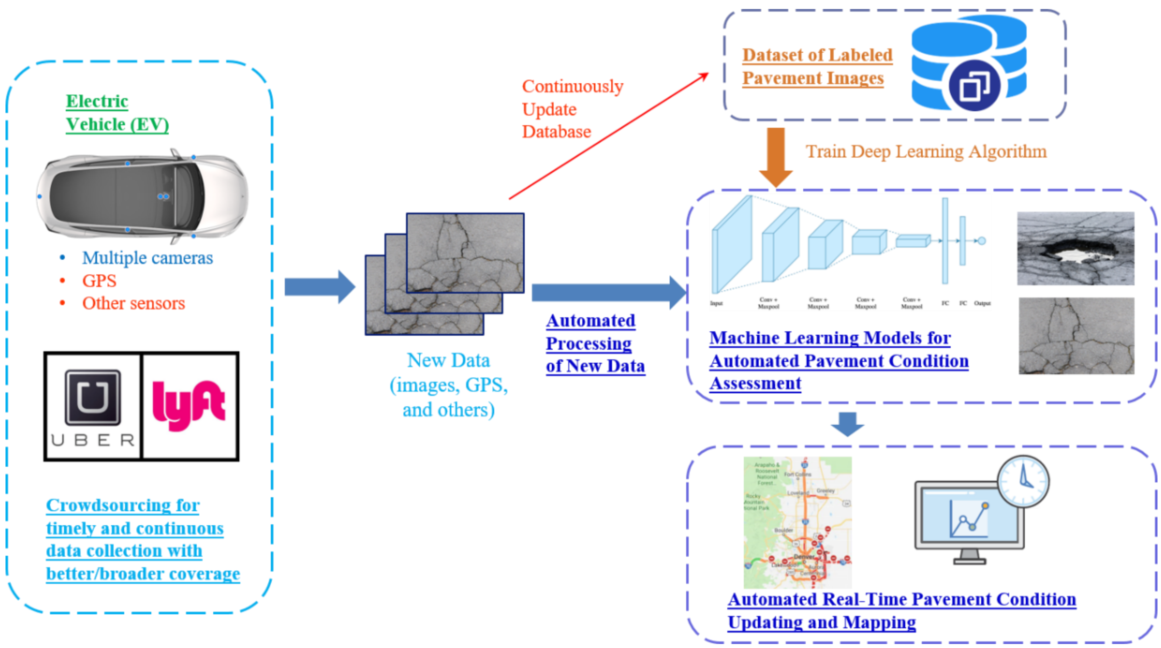

Figure 1. Overview of the proposed research

The first phase involves developing a standardized data collection framework to capture, anonymize, and preprocess data from rideshare EVs. This includes GPS coordinates, images, and sensor data relevant to road conditions. The framework will ensure data privacy through anonymization techniques and compliance with data protection regulations. EV rideshare drivers will be recruited to participate in the data collection. Data preprocessing will be conducted to first determine which cameras and images could be used for road condition assessment, because images from some of the cameras may not be useful for road condition assessment. Then the images will be processed (e.g., remove irrelevant information and standardize formats) using existing machine learning algorithms. Features relevant to road condition assessment will be identified and extracted from the preprocessed data to prepare the dataset for subsequent labeling and analysis. Then the pavement images will be labeled using tools such as LabelMe establish the training data. Other opensource data will also be used to augment training data to improve the training of the machine learning algorithms.

Second, deep learning algorithms capable of classifying pavement conditions and identifying pavement distress (e.g., cracks, potholes) will be trained using the training data. More specifically, based on the labeled images, deep learning algorithms (e.g., deep convolution neural network, which has been shown to be highly effective in extracting and identifying features from images) will be trained. Data augmentation (including rotation, blurring, and contrast enhancement) will first be applied to increase the size of the training sets. Different deep learning algorithms will be investigated (e.g., Mask R-CNN [6], YOLO [7]). Model training will involve splitting the dataset into training, validation, and test sets. Performance metrics such as accuracy, precision, recall, and F1 score will be used to evaluate the model's reliability and generalizability. We will analyze the effectiveness of the collected data and the trained ML model in classifying and identifying road conditions and damages, focusing on accuracy, reliability, and scalability.

Thirdly, automated software tools will be developed with the trained deep learning algorithms as the engine. Newly collected geo-tagged images/videos by cameras from crowdsourcing EVs can be directly fed into the software to efficiently and automatically assess pavement conditions. The results will be further fed to the visualization and decision-making tools for automated real-time pavement condition updating and mapping. The map and information database will be automatically updated once new data from crowdsourcing rideshare EVs are collected and processed through the automated deep learning algorithm. Also, the robustness of the deep learning algorithms will be iteratively and continuously improved with newly collected data. The tool will be tested using several rideshare EVs to demonstrate the coverage improvement offered by crowdsourcing.

Relevance to Strategic Goals

This study will use sensors in EVs and crowdsourcing for pavement condition assessment, and develop AI enabled automated tools for accurate and robust pavement distress detection, mapping, and updating. The tools are expected to help enhance the capabilities of state DOTs and highway maintenance team in timely and cost-effective maintenance of highways and strengthen existing pavement asset management systems and practices by reducing costs for road condition monitoring. Hence the project directly contributes to USDOT strategic goal of Economic Strength and Global Competitiveness by helping “strengthen asset management systems and practices to reduce the costs of managing assets throughout their lifecycle.”

Educational Benefits

A graduate research assistant will be hired to conduct the research described in this proposal. The student will be involved in setting up the data collection protocol, processing and labeling the images, and applying deep learning techniques for pavement distress detection, as well as developing automated tools to map out the condition of road network. The student will gain valuable experience in applying deep learning techniques for image processing and road damage detection. The procedure and example from this project will be used in a graduate level course that will be developed by the PI on machine learning in engineering applications.

Outputs through Technology Transfer

The developed automated tools will be presented to CDOT maintenance teams to promote the adoption of the developed tools. The research findings will also be disseminated through technical publications in conferences and journals as well as presentations in conferences and seminars (including virtual delivery via live webinars and in-person delivery).

Expected Outcomes and Impacts

The expected outcomes include:

1. A set of procedures to collect and process image data from EVs to support pavement condition assessment.
2. A unique and valuable database of geotagged and labeled images from EVs for training pavement distress detection algorithms.
3. Performance and feasibility analysis of using crowdsourcing rideshare EV data and machine learning to improve the timeliness and coverage of pavement condition data collection and pavement condition assessment.
4. Automated tools for pavement distress detection, mapping, and updating for use by state DOTs and highway maintenance team.

A full report documenting procedures for data collection using sensors in EVs, data processing, training and application of deep learning algorithms, and development of the tools will be provided, and one or more journal papers will be published.

Work Plan

The work plan of this project includes the following six major tasks:

<u>Task 1: Recruitment of Rideshare EV Drivers</u>

EV rideshare drivers will be recruited to participate the data collection, and compensations will be provided to incentivize participation. Initially, we will focus on specific EVs (such as Tesla) to ensure consistency in the data collected by different drivers. The expected completion date for this task is 3 months from the project start date.

<u>Task 2: Development of a Standardized Data Collection Framework</u>

We will establish guidelines on what data to collect from the various sensors on the EV (e.g., images, video, GPS, accelerometer data) and how to collect/save and upload the data to designated cloud storage, focusing on consistency and privacy. Data from different rideshare drivers and EVs will be anonymized to protect privacy. The expected completion date for this task is 6 months from the project start date.

<u>Task 3: Data Preprocessing and Analysis</u>

We will develop computer algorithms to automate the process of cleaning the raw data, removing irrelevant information and standardizing formats for analysis. Features relevant to road condition assessment will be identified and extracted from the preprocessed data. Initial analyses will be performed to understand the quality, coverage, and potential of the collected data for assessing road conditions. Since cameras on EVs are not specifically designed for detailed road surface analysis, the viewing angles might not always be optimal, therefore, the images captured will need to be processed and analyzed with suitable algorithms so that they can be used for subsequent analysis. The expected completion date for this task is 12 months from the project start date.

Task 4: Train Machine Learning Model for Road Condition Assessment

The collected data will be labeled to create a training dataset, incorporating data of diverse road and environment conditions. Based on the training data, deep learning algorithms based on convolutional neural network (CNN) will be developed and trained to classify different road conditions, and image segmentation ML algorithms will be used to identify and locate the damages (e.g., cracks, potholes). Different algorithms will be used and compared to select the one with best performance. For this task, PI Jia's prior experience in developing ML algorithms for image fusion and image-based road damage identification will be leveraged. The expected completion date for this task is 18 months from the project start date.

Task 5: Feasibility Study and Analysis

We will analyze the effectiveness of the collected data and the trained ML model in classifying and identifying road conditions and damages, focusing on accuracy, reliability, and practicality. We will also document any challenges faced during data collection and model training, including data quality, variability, and model/data limitations. Based on the findings, recommendations will be provided for scaling up the approach, improving data collection and analysis, and potential areas for further investigation. The expected completion date for this task is 24 months from the project start date.

Task 6: Develop and Test Automated Tools

This task will package the data processing algorithms, the trained deep learning algorithms, and road condition maps into an automated tool for road condition assessment and mapping using crowdsourcing EV data. Newly collected geo-tagged images/videos by cameras from crowdsourcing EVs can be directly fed into the software to efficiently and automatically assess pavement conditions. The tool will be tested using several rideshare EVs to demonstrate the coverage improvement offered by crowdsourcing. The developed automated tools will be presented to CDOT maintenance teams to promote the adoption of the developed tools. The expected completion date for this task is 24 months from the project start date.

Project Cost

Total Project Costs:	$90,000
CTIPS Funds Requested:	$45,000
Matching Funds:	$45,000
Source of Matching Funds:	Colorado State University

References

[1] Majidifard, Hamed, Peng Jin, Yaw Adu-Gyamfi, and William G. Buttlar. "Pavement image datasets: A new benchmark dataset to classify and densify pavement distresses." *Transportation Research Record* 2674, no. 2 (2020): 328-339.

[2] Maeda, Hiroya, Yoshihide Sekimoto, Toshikazu Seto, Takehiro Kashiyama, and Hiroshi Omata. "Road damage detection and classification using deep neural networks with smartphone images." *Computer-Aided Civil and Infrastructure Engineering* 33, no. 12 (2018): 1127-1141.

[3] Arya, Deeksha, Hiroya Maeda, Sanjay Kumar Ghosh, Durga Toshniwal, Alexander Mraz, Takehiro Kashiyama, and Yoshihide Sekimoto. "Deep learning-based road damage detection and classification for multiple countries." *Automation in Construction* 132 (2021): 103935.

[4] Arjapure, Surekha, and D. R. Kalbande. "Deep learning model for pothole detection and area computation." In *2021 international conference on communication information and computing technology (ICCICT)*, pp. 1-6. IEEE, 2021.

[5] Ahmed, Khaled R. "Smart pothole detection using deep learning based on dilated convolution." *Sensors* 21, no. 24 (2021): 8406.

[6] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." In *Proceedings of the IEEE international conference on computer vision*, pp. 2961-2969. 2017.

[7] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7464-7475. 2023.

[8] Chen, Wei-Hsiang. "Visible & thermal imaging and deep learning based approach for automated robust detection of potholes to prioritize highway maintenance." MS Thesis, Colorado State University, 2023.