DEFECTRA: ROAD DEFECT SEVERITY ASSESSMENT AND CLASSIFICATION

A Special Problem Proposal
Presented to
the Faculty of the Division of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Visayas
Miag-ao, Iloilo

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science by

BELEBER, Benz Vrianne
CATALAN, Perserose
SENCIL, Kristian Lyle

Francis DIMZON
Adviser
Jumar CADONDON
Co-Adviser

December 9, 2024

Contents

List of Tables

Chapter 1

Introduction

1.1 Overview

According to the National Road Length by Classification, Surface Type, and Condition of the Department of Public Works and Highways (DPWH), as of October 2022 approximately 98.97

In an interview with the Road Board of DPWH Region 6 it was stated that road condition assessments are mostly done manually with heavy reliance on engineering judgment. In addition, manual assessment of roads is also time consuming which leaves maintenance operations to wait for lengthy assessments (J. Chua, Personal Interview. 16 September 2024). In a study conducted by Ramos, Dacanay, and Bronuela-Ambrocio (2022), it was found that the Philippines' current method of manual pavement surveying is considered as a gap since it takes an average of 2-3 months to cover a 250 km road as opposed to a 1 day duration in the Australian Road Research Board for the same road length. Ramos et al. (2022) recommended that to significantly improve efficiency of surveying methods and data gathering processes, automated survey tools are to be employed. It was also added that use of such automated surveying tools can also guarantee the safety of road surveyors (Ramos et al., 2022).

If the process of assessment on the severity of road defects can be automated then the whole process of assessing the quality of roads can be hastened up which can also enable maintenance operations to commence as soon as possible if necessary. If not automated, the delay of assessments will continue and roads that are supposedly needing maintenance may not be properly maintained which can affect the general public that is utilizing public roads daily.

1.2 Problem Statement

Roads support almost every aspect of daily life, from providing a way to transport goods and services to allowing people to stay connected with their communities. However, road defects such as cracks and potholes damage roads over time, and they can increase accident risks and affect the overall transportation. The current way of inspecting the roads for maintenance is often slow as it is done manually, which makes it harder to detect and fix defects early. The delay in addressing these problems can lead to even worse road conditions (J. Chua, Personal Interview. 16 September 2024). There are several research studies into automated road defect classification that have advanced in recent years but most of them focus on identifying the types of defects rather than assessing their severity or characteristics like depth. Without reliable data on the depth of the defect, road maintenance authorities may underestimate the severity of certain defects. To address these challenges, advancements are needed across various areas. An effective solution should not only detect and classify road defects but also measure their severity to better prioritize repairs. Failing to address this problem will require more extensive repairs for damaged roads, which raises the cost and strains the budget. Additionally, road maintenance would still be slow and cause disruptions in daily activities. Using an automated system that accurately detects, classifies, and assess the severity of road defects by incorporating depth are necessary to efficiently monitor road quality.

1.3 Research Objectives

1.3.1 General Objective

This special problem aims to develop an automated system that will accurately detect and assess the severity of potholes on road surfaces by using image analysis, depth measurement technologies, and a combination of machine learning and computer vision techniques. The system will focus on measuring the depth of potholes to assess their severity, enabling faster and more accurate road maintenance decisions.

1.3.2 Specific Objectives

Specifically, this special problem aims:

1. To collect high-quality images of road surfaces that capture potholes including their depth in various lighting and weather conditions.

2. To develop and train a machine learning model to detect and assess the severity of potholes from images.

3. To measure the accuracy of the system by comparing the depth measurements against ground truth data collected from actual road inspections

4. To develop a prototype system that can detect and measure road potholes from image input, analyze their depth, and assess their severity.

1.4 Scope and Limitations of the Research

This system will focus solely on detecting and assessing the severity of potholes through image analysis and depth measurement technologies. The scope includes the collection of pothole images using cameras and depth-sensing tools under various lighting and weather conditions, ensuring the data captures real-world variations. High-quality and diverse image datasets will be crucial for training the model to accurately assess pothole severity based on depth.

Depth-sensing tools, such as stereo cameras, will be used to record the depth of potholes specifically. The system will not address other road defects like cracks or other surface deformations; therefore, it will detect and analyze only potholes. Additionally, only accessible potholes will be measured, meaning those that are filled with water or obscured by debris may not be accurately assessed.

The machine learning model developed will focus exclusively on detecting potholes and assessing their severity through depth measurement. The accuracy of the model's depth measurements will be evaluated by comparing them against data collected from actual field inspections. However, this comparison will be limited to selected sample sites, as collecting field data over a large area can be time-consuming and resource-intensive.

Environmental factors such as lighting, road surface texture, and weather conditions may impact the model's performance. The accuracy and reliability of the model will depend on the quality and variety of the training dataset. Its ability to generalize to unseen pothole images will need to be carefully validated.

1.5 Significance of the Research

This special problem aims to be significant to the following:

Computer Science Community. This system can contribute to advancements in computer vision and machine learning by using both visual and depth data to assess the severity of road defects. It introduces a more comprehensive approach compared to the usual image-only or manual inspection methods. This combination can be applied to other fields that need both visual and depth analysis like medical imaging.

Concerned Government Agencies. This system offers a valuable tool for road safety and maintenance. Not only can this detect and classify anomalies, it can also assess the defect's severity which allows them to prioritize repairs, optimal project expenditures, and better overall road safety and quality.

Field Engineers. In the scorching heat, field engineers are no longer required to be on foot unless it requires its engineering judgement when surveying a road segment. It can hasten the overall assessment process.

Future Researchers. The special problem can serve as a baseline and guide of researchers with the aim to pursue special problems similar or related to this.

Chapter 2

Review of Related Literature

2.1 Related Literature

This section of the chapter presents related literature that is considered essential for the development of this special problem.

2.1.1 Deep Learning

Kelleher (2019) states that deep learning is inclined on making large-scale neural networks geared towards creating data-driven decisions. Furthermore, it was also argued that deep learning is oriented towards large-scale, complex data.

2.1.2 YOLOv5

According to Solawetz (2024), YOLOv5 is a model from a family of computer vision models used for object detection. YOLOv5 is reported to perform comparably to state-of-the-art techniques. It is designed to extract features from raw input images, used primarily in training object detection models alongside various data augmentation techniques.

2.1.3 Image and Video Processing

Kumar (2024) defines image processing as a process of turning an image into its digital form and extracting data from it through certain functions and operations. Usual processes are considered to treat images as 2D signals wherein different processing methods utilize these signals. Like image processing, Riches Resources (2020) defines video processing as being able to extract information and data from video footage through signal processing methods. However, in video processing due to the diversity of video formats, compression and decompression methods are often expected to be performed on videos before processing methods to either increase or decrease bitrate.

2.1.4 Stereo Vision

MathWorks (n.d.) defines stereo vision as a process of utilizing multiple 2D perspectives in order to extract information in 3D. In addition, most uses of stereo vision involve estimating an objects distance from an observer or camera. The 3D information is stated to be extracted with stereo pairs or pair of images through estimation of relative depth of points in a scene which are then represented through a stereo map that is made through the matching of the pair's corresponding points.

2.2 Related Studies

This section of the chapter presents related studies conducted by other researchers wherein the methodology and technologies used may serve as basis in the development of this special problem.

2.2.1 Automated Detection and Classification of Road Anomalies in VANET Using Deep Learning

In the study of Bibi et al. (2021) it was noted that identification of active road defects are critical in maintaining smooth and safe flow of traffic. Detection and subsequent repair of such defects in roads are crucial in keeping vehicles using such roads away from mechanical failures. The study also emphasized the growth in use of autonomous vehicles in research data gathering which is what the researchers utilized in data gathering procedures. With the presence of autonomous vehicles, this allowed the researchers to use a combination of sensors and deep neural networks in deploying artificial intelligence. The study aimed to allow autonomous vehicles to avoid critical road defects that can possibly lead to dangerous situations. Researchers used Resnet-18 and VGG-11 in automatic detection and classification of road defects. Researchers concluded that the trained model was able to perform better than other techniques for road defect detection (Bibi et al., 2021). The study is able to provide the effectiveness of using deep learning models in training artificial intelligence for road defect detection and classification. However, the study lacks findings regarding the severity of detected defects which is crucial in automating manual procedures of road surveying in the Philippines.

2.2.2 Smartphones as Sensors for Road Surface Monitoring

In their study, Chapman, Li, and Sattar (2018) noted the rise of sensing capabilities of smartphones which they utilized in monitoring road surface to detect and identify anomalies. The researchers considered different approaches in detecting road surface anomalies using smartphone sensors. One of which are threshold-based approaches which was determined to be quite difficult due to several factors that are affecting the process of determining the interval length of a window function in spectral analysis (Chapman et al., 2018). The researchers also utilized a machine learning approach adapted from another study. It was stated that k-means was used in classifying sensor data and in training the SVM algorithm. Due to the requirement of training a supervised algorithm using a labeled sample data was required before classifying data from sensors, the approach was considered to be impractical for real-time situations (Chapman et al., 2018). In addition, Chapman et al. (2018) also noted various challenges when utilizing smartphones as sensors for data gathering such as sensors being dependent on the device's placement and orientation, smoothness of captured data, and the speed of the vehicle it is being mounted on. Lastly, it was also concluded that the accuracy and performance of using smartphone sensors is challenging to compare due to the limited data sets and reported algorithms.

2.2.3 Road Anomaly Detection through Deep Learning Approaches

The study of Guo, Luo, and Lu (2020) aimed to utilize deep learning models in classifying road anomalies. The researchers used three deep learning approaches namely Convolutional Neural Network, Deep Feedforward Network, and Recurrent

Neural Network from data collected through the sensors in the vehicle's suspension system. In comparing the performance of the three deep learning approaches, the researchers fixed some hyperparameters. Results revealed that the RNN model was the most stable among the three and in the case of the CNN and DFN models, the researchers suggested the use of wheel speed signals to ensure accuracy. And lastly, the researchers concluded that the RNN model was best due to high prediction performance with small set parameters (Guo et al., 2020).

2.2.4 Road Surface Quality Monitoring Using Machine Learning Algorithms

The study of Bansal et al. (2021) aimed to utilize machine learning algorithms in classifying road defects as well as predict their locations. Another implication of the study was to provide useful information to commuters and maintenance data for authorities regarding road conditions. The researchers gathered data using various methods such as smartphone GPS, gyroscopes, and accelerometers. Bansal et al. (2021) also argued that early existing road monitoring models are unable to predict locations of road defects and are dependent on fixed roads and static vehicle speed. Neural and deep neural networks were utilized in the classification of anomalies which was concluded by the researchers to yield accurate results and are applicable on a larger scale of data (Bansal et al., 2021). The study of Bansal et al. (2021) can be considered as an effective method in gathering data about road conditions. However, it was stated in the study that relevant authorities will be provided with maintenance operation and there is no presence of any severity assessment in the study. This may cause confusion due to a lack of assessment on what is the road condition that will require extensive maintenance or repair.

2.2.5 Assessing Severity of Road Cracks Using Deep Learning-Based Segmentation and Detection

In the study of Ha, Kim, and Kim (2022), it was argued that the detection, classification, and severity assessment of road cracks should be automated due to the bottleneck it causes during the entire process of surveying. For the study, the researchers utilized SqueezNet, U-Net, and MobileNet-SSD models for crack classification and severity assessment. Furthermore, the researchers also employed separate U-nets for linear and area cracking cases. For crack detection, the researchers followed the process of pre-processing, detection, classification. During preprocessing images were smoothed out using image processing techniques. The researchers also utilized YOLOv5 object detection models for classification of pavement cracking wherein the YOLOv51 model recorded the highest accuracy. The researchers however stated images used for the study are only 2D images which may have allowed higher accuracy rates. Furthermore, the researchers suggest incorporating depth information in the models to further enhance results.

2.2.6 Roadway pavement anomaly classification utilizing smartphones and artificial intelligence

The study of Christodoulou,Dimitrio, and Kyriakou (2016) presented what is considered as a low-cost technology which was the use of Artificial Neural Networks in training a model for road anomaly detection from data gathered by smartphone sensors. The researchers were able to collect case study data using two-dimensional indicators of the smartphone's roll and pitch values. In the study's discussion, the data collected displayed some complexity due to acceleration and vehicle speed which lead to detected anomalies being not as conclusive as planned. The researchers also added that the plots are unable to show parameters that could verify the data's correctness and accuracy. Despite the setbacks, the researchers still fed the data into the Artificial Neural Network that was expected to produce two outputs which were "no defect" and "defect." The method still yielded above 90% accuracy but due to the limited number of possible outcomes in the data processing the researchers still needed to test the methodology with larger data sets and roads with higher volumes of anomalies.

2.2.7 Stereo Vision Based Pothole Detection System for Improved Ride Quality

In the study of Ramaiah and Kundu (2021) it was stated that stereo vision has been earning attention due to its reliable obstacle detection and recognition. Furthermore, the study also discussed that such technology would be useful in improving ride quality in automated vehicles by integrating it in a predictive suspension control system. The proposed study was to develop a novel stereo vision based pothole detection system which also calculates the depth accurately. However, the study focused on improving ride quality by using the 3D information from detected potholes in controlling the damping coefficient of the suspension system. Overall, the pothole detection system was able to achieve 84% accuracy and is able to detect potholes that are deeper than 5 cm. The researchers concluded that such system can be utilized in commercial applications. However, it is also worth noting that despite the system being able to detect potholes and measure its depth, the overall severity of the pothole and road condition was not addressed.

2.3 Chapter Summary

The reviewed literature involved various techniques and approaches in road anomaly detection and classification. These approaches are discussed and summarized below along with their limitations and research gaps.

Study	Technology/ Techniques Used	Key Findings	Limitations
Automated Detection and Classification of Road Anomalies in VANET Using Deep Learning	Resnet-18 and VGG-11	Trained model is able to provide the effectiveness of using deep learning models in training artificial intelligence for road defect detection and classification.	Lacks findings regarding the severity of detected defects.
Smartphones as sensors for Road surface monitoring	Machine Learning, Smartphones	Approach was considered impractical for real-life applications.	Sensors are dependent on device's placement and orientation, smoothness of data, and speed of vehicle it is mounted on. Accuracy of results is difficult to compare.
Road Anomaly Detection through Deep Learning Approaches	Convolutional Neural Network, Deep Feedforward Network, and Recurrent Neural Network	Identified that RNN was the best deep learning approach due to high prediction performance.	Data collection is considered too difficult and complicated to execute due to sensors being mounted on an integral part of the vehicle.
Assessing Severity of Road Cracks Using Deep Learning-Based Segmentation and Detection	SqueezNet, U-Net, YOLOv5, and MobileNet-SSD models	YOLOv51 model recorded the highest accuracy.	Only 2D images are used for the study which may have allowed higher accuracy rates, and the study also lacked depth information.
Stereo Vision Based Pothole Detection System for Improved Ride Quality	Pair of stereo images captured by a stereo camera	System was able to achieve 84% accuracy and is able to detect potholes that are deeper than 5 cm.	Overall severity of the pothole and road condition was not addressed.

Table 2.1: Comparison of Related Studies on Road Anomaly Detection using Deep Learning Techniques and Stereo Vision

Chapter 3

Methodology

This chapter outlines the systematic approach that will be taken to address the problem of classifying and assessing road defects using artificial intelligence. The methodology will be divided into key phases: data collection, algorithm selection, design, testing and experimentation, and challenges and limitations. Each phase will play a crucial role in accurately classifying and assessing road defects. Each phase will be essential for accurately classifying and assessing road defects.

3.1 Research Activities

3.1.1 Data Collection

The researchers conducted initial inquiries to understand the problem domain and existing road maintenance practices. This phase included consulting the engineers under the Road Maintenance Department of the government agency Department of Public Works and Highways (DPWH). An interview with Engr. Jane Chua provided a comprehensive overview of the DPWH's road maintenance manual, which was crucial in aligning this project with existing standards. This collaboration with DPWH provided insights into road pothole classification standards, ensuring that the collected data will align with industry standards. The researchers will also manually annotate the pilot dataset based on these standards, ensuring local relevance.

3.1.2 Algorithm Selection

Potential solutions, algorithms, and system architectures were discussed by the researchers and the special problem adviser in this phase. These sessions, conducted in class and virtually via Zoom, helped narrow down the overview of the system, leading to the selection of the main architecture YOLOv5 for pothole detection and Epipolar Spatio-Temporal Networks (ESTN) for depth estimation.

Pothole Detection

YOLOv5 was selected due to its high accuracy and ability to process images in real-time, making it suitable for detecting road defects in dynamic environments. Its architecture is optimized for speed and performance, which is crucial for large-scale deployment in road inspections.

Severity Assessment

The Multi-view Depth Estimation using Epipolar Spatio-Temporal Networks was selected due to the high cost and limited accessibility of LiDAR technology. By applying epipolar geometry and temporal consistency across sequential frames, this approach provides an accurate depth estimation from standard video footage (Long et al., 2021).

3.1.3 Design, Testing, and Experimentation

This section outlines both the design and testing of the system, as well as the experimentation process to validate the selected methodologies.

Model Design

The system was designed to operate with two core components: YOLOv5 for pothole detection and ESTN for severity assessment. The model architecture was chosen based on the real-time processing capabilities and the need for accurate depth estimation from standard video footage. The design ensures that the system can detect defects and provide severity assessments in a seamless workflow.

Data Set

The YOLOv5 model was trained using two datasets from Universe Roboflow. One of the data sets was posted by a user named Eric Tam. It was also stated that the images from the dataset are sourced from a Crowdensing-based Road Damage Detection Challenge from 2022 in Japan. The challenge involves contestants being required to submit road damage datasets, shortlist their data set, and use the data set for road damage detection and classification models. The use of this data set in training models for road damage detection and classification ensures that the data is viable for training the YOLOv5 model. The dataset contains various road defects in Japan. Another data set used in training the YOLOv5 model was also uploaded in Universe Roboflow by a user named Atikur Rahman Chitholian which was stated to be part of his undergraduate thesis. The dataset is comprised of 665 images with potholes being labeled. It was also stated that the data set can be utilized in automatically detecting and categorizing potholes found in the streets of cities. Data preprocessing techniques were applied to both datasets to improve model accuracy and generalization. These included resizing images to a uniform size, applying augmentation techniques (flipping, rotation, and color adjustment) to increase dataset variability, and normalizing pixel values to ensure consistency across images.

Performance Metrics

The performance of the YOLOv5 model will be evaluated using mean Average Precision (mAP). mAP is a widely used metric in object detection tasks and is particularly useful for assessing models that need to detect and classify multiple object categories. In this case, mAP will provide a comprehensive evaluation of the model's ability to detect and classify potholes, offering an aggregated score across the relevant detection thresholds. This ensures a balanced assessment of both detection accuracy and classification performance, which is essential for accurately identifying potholes across varying conditions. The effectiveness of mAP for this task is well-established in object detection literature (Everingham et al., 2015; Lin et al., 2014).

For the accuracy of depth estimation using the Epipolar Spatio-Temporal Networks (ESTN), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) will be used. RMSE is chosen for its ability to penalize larger errors more heavily, making it suitable for assessing depth estimation performance where larger deviations from the ground truth are more significant (Zhang et al., 2018). MAE is also employed to provide a straightforward measure of average error magnitude, offering a complementary evaluation of depth estimation without emphasizing larger errors as much (Zhang et al., 2020).

Testing and Validation

The testing process will begin with a detailed testing plan that includes both simulated and real-world testing scenarios. Initially, the model will be tested in controlled environments to ensure it can detect and assess road defects accurately. Following this, real-world testing will be conducted using the StereoPi kit on local roads, specifically at the University of the Philippines Visayas Miagao Campus. The system's performance will be validated by comparing its predictions with ground-truth data collected from manual inspections.

Documentation

Throughout the research activities, thorough documentation will be maintained. This documentation will capture all methods, results, challenges, and adjustments made during the experimentation phases. It ensures the reproducibility of the work and provides transparency for future research endeavors.

3.1.4 Challenges and Limitations

Availability of Local Datasets

The lack of locally labeled datasets for road defects has posed a challenge in training accurate models. The majority of available datasets are sourced from international locations, which may not fully represent the road conditions found in the study area. To address the lack of locally labeled datasets, the researchers will create a pilot dataset from local roads within the University of the Philippines Visayas Miagao Campus. This dataset will be manually annotated according to DPWH's classification standards, ensuring local relevance.

Data Quality and Variability

Variations in the quality and resolution of the data collected from different sources may impact the performance of the trained models. In particular, images captured under varying weather conditions or lighting may affect the accuracy of pothole detection. To address this, the researchers plan to use the StereoPi kit to capture images under optimal weather and lighting conditions, such as mid-morning or early afternoon on clear days, ensuring consistent image quality for stereo vision analysis. The kit's stereo cameras will be calibrated for uniform resolution and focus. Data augmentation techniques will also be applied to simulate varying conditions, and pre-processing steps like noise reduction and contrast enhancement will be used to improve the quality of the captured data. This approach aims to minimize the impact of environmental factors on the accuracy of road pothole detection and depth estimation.

3.2 Calendar of Activities

Table 1 shows a Gantt chart of the activities. Each bullet represents approximately one week's worth of activity.

Table 3.1: Timetable of Activities for 2024

Activities (2024)	Aug	Sept	Oct	Nov	Dec
Pre-proposal Preparation	●●●●				
Literature Review	●●●	●			
Data Collection	●●	●●			
Algorithm Selection		●●			
System Design		●	●●	●●	
Preliminary Testing				●●	●
Documentation and SP Writing	●●●●	●●●●	●●●●	●●●●	●●

Table 3.2: Timetable of Activities for 2025

Activities (2025)	Jan	Feb	Mar	Apr	May	Jun
Data Collection	●●●●					
System Design	●●●	●●	●●			
Model testing	●●●	●●●●	●●●●			
Results Analysis			●●	●●●●		
Conclusion Formulation				●●	●●●	
Documentation and SP Writing	●●●●	●●●●	●●●●	●●●●	●●●●	●●

References

Fedkiw, R., Stam, J., & Jensen, H. W. (2001). Visual simulation of smoke. In E. Fiume (Ed.), *Proceedings of siggraph 2001* (pp. 15–22). ACM Press / ACM SIGGRAPH.

Jobson, D. J., Rahman, Z., & Woodell, G. A. (1995). Retinex image processing: Improved fidelity to direct visual observation. In *Proceedings of the is&t fourth color imaging conference: Color science, systems, and applications* (Vol. 4, pp. 124–125).

Kartch, D. (2000). *Efficient rendering and compression for full-parallax computer-generated holographic stereograms* (Unpublished doctoral dissertation). Cornell University.

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., ... Fulk, D. (2000). The digital michelangelo project. In K. Akeley (Ed.), *Proceedings of siggraph 2000* (pp. 131–144). New York: ACM Press / ACM SIGGRAPH.

Park, S. W., Linsen, L., Kreylos, O., Owens, J. D., & Hamann, B. (2006, March/April). Discrete sibson interpolation. *IEEE Transactions on Visualization and Computer Graphics*, *12*(2), 243–253.

Parke, F. I., & Waters, K. (1996). *Computer facial animation.* A. K. Peters.

Pellacini, F., Vidimče, K., Lefohn, A., Mohr, A., Leone, M., & Warren, J. (2005, August). Lpics: a hybrid hardware-accelerated relighting engine for computer cinematography. *ACM Transactions on Graphics*, *24*(3), 464–470.

Sako, Y., & Fujimura, K. (2000). Shape similarity by homotropic deformation. *The Visual Computer*, *16*(1), 47–61.

Yee, Y. L. H. (2000). *Spatiotemporal sensistivity and visual attention for efficient rendering of dynamic environments* (Unpublished master's thesis). Cornell University.