Smart City Road Maintenance: A LiDAR and AI-Driven Approach for Detecting and Mapping Road Defects

Giovanni Nardini*, Roberto Nucera*, Alessandro Ulleri*, Stefano Cordiner†,
Eugenio Martinelli‡, Arianna Mencattini‡, Iulian Gabriel Coltea*
*Key to business s.r.l., Department of †Industrial and ‡Electronic Engineering, University of Rome Tor Vergata
Rome, Italy
*e-mail: info@key2.it

***Abstract*—This work is focused on the development of an integrated system designed to detect, map, and analyze road surface defects, contributing to Smart City infrastructure maintenance. The system is installed on vehicles and leverages a multi-sensor approach, combining Light Detection and Range (LiDAR) point clouds, visual information from Red-Green-Blue (RGB) cameras, inertial data and Global Navigation Satellite Systems (GNSS) coordinates. Road defects such as potholes and alligator cracks are detected in RGB images by a custom deep learning model based on instance segmentation. The scene understanding is committed to a second Artificial Intelligence (AI) model based on semantic segmentation in order to perceive objects locations and the overall structure of the road. Afterward, all results are processed together and translated into the 3D domain of LiDAR data. This can be done through a proper camera calibration procedure and LiDAR-Camera data alignment with the estimation of intrinsic and extrinsic parameters. Then, AI segmentation results are projected to 3D point clouds in order to isolate the detected items from the rest of the point cloud and obtain three-dimensional models of each of them, enabling measurements like the affected surface extension, depth and volumes. GNSS and inertial data are fused together to obtain the correct orientation and location of the system, enabling geographic positioning of all detected items on the map. Results are displayed on a map-based portal, enabling easy access to near real-time defect data. This approach advances road monitoring by automating the mapping and analysis of surface conditions, enhancing urban infrastructure management. In addition, the strengths of this approach are the possibility of deploying the pipeline in edge devices enabling real-time computation, the use of pre-existing training datasets based on RGB images alone, and good accuracy on the geographical localization and estimation of defect measurements.**

Keywords-cities; road maintenance; LiDAR; AI; computer vision.

I. INTRODUCTION

As urban areas grow the need to monitor road conditions efficiently becomes crucial for keeping infrastructure intact and promoting road safety. The conventional methods of inspecting roads are laborious, time consuming and frequently fall short of providing the accuracy required for repairs. However, recent progress in sensor technology, artificial intelligence and data integration present fresh opportunities for monitoring road conditions. Over the past few years, many approaches have been explored. Sometimes using inertial data [1], pure machine learning and computer vision methods [2][3], sometimes exploiting more sophisticated deep learning models [4], and other times combining vision and depth sensing together with spatial AI [5][6]. The technologies that have been tested for depth estimation are based on stereoscopy, Red-Green-Blue-Depth (RGB-D) cameras and LiDAR. However, each has its own disadvantages: stereoscopy generally does not work with feature-poor surfaces, RGB-D cameras based on Time of Flight (ToF) technology, while achieving good accuracy, drop their performance in outdoor environments and are limited to a range of few meters, while LiDAR provides the most long-range and accurate measurements but at the expense of lower point density and the need for an additional imaging system to obtain the scene picture. Furthermore, approaches using RGB-D images as input for AI detection models, while achieving good performance due to depth information, are strongly affected by the context, sensor position and framing of the training data, and therefore require the acquisition of huge amounts of images from every possible angle and distance, in order to replicate all possible setups. Our approach, on the other hand, bases AI inference solely on RGB images and transports the detection information to the LiDAR domain, via camera-LiDAR registration, as shown in Figure 1.

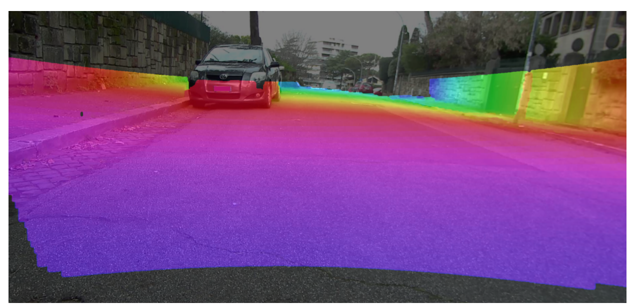

Figure 1. Camera-LiDAR Registration.

This allows the use of pre-existing datasets without having to create a custom dataset and re-labelling all images. This work suggests a setup (see Figure 2) that utilizes LiDAR technology along with RGB imaging, inertial and GNSS data within a framework based on Robot Operating System (ROS), as shown in Figure 3, in order to identify and pinpoint road surface issues efficiently.

From an economic standpoint, the system's adaptability to city vehicles, including public transport, could potentially transform routine operations into continuous, cost-effective road monitoring. Combining this distributed sensing with on-the-ground human supervision, such as cleaning personnel, creates a hybrid model that optimizes resource use and enhances data accuracy, leading to efficient urban road maintenance.

In Section 2, the methods employed are detailed, including the hardware components, the software architecture, the design and training of the AI models. In Section 3, the results of the system's validation are presented, focusing on the performance metrics of the AI models and the accuracy of defect measurements and positioning. In Section 4, the paper concludes with a discussion of the system's contributions and potential future developments.

II. Methods

The proposed system integrates the following hardware components: an Hybrid Solid-State LiDAR with 128-channels of resolution, a global shutter camera sensor with 4k resolution at 30 fps, a navigation system with 9-axis accelerometer INS and dual antenna GNSS and a Nvidia Jetson AGX module where the software runs. The LiDAR, camera and navigation system are mounted on the vehicle's roof, while the Jetson unit is installed inside the cabin and connected to the vehicle's power supply.

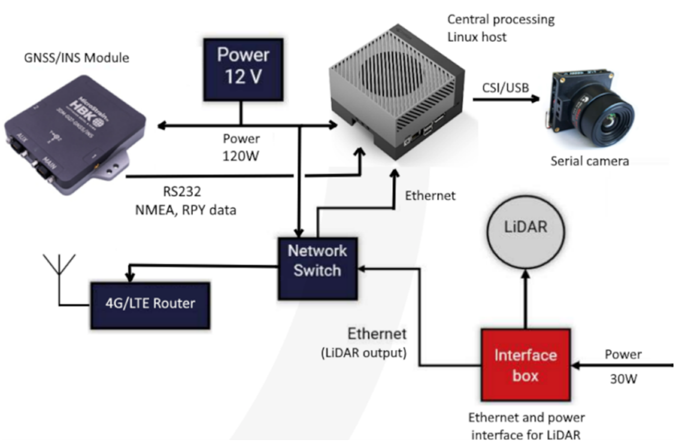

Figure 2. Hardware setup.

The software architecture is based on ROS and is made up of the following nodes: driver nodes to collect data from each sensor and publish to topics, data processing nodes to apply AI model inference on images and get results, projection nodes to map defects from RGB domain into 3D domain and make measurements and navigation nodes to estimate precise latitudes and longitudes of each defects. All collected results are then submitted to the visualization platform.

There are two types of custom trained AI models: an instance segmentation model based on the You Only Look Once (YOLOV8-Seg) small architecture [7] and a semantic segmentation model based on the SegFormerB1 architecture [8]. The former was trained on the RDD22 dataset [9]: since it is an object detection dataset, it was necessary to re-label the annotations with the addition of pothole and alligator crack segmentation mask. In order to speed up the process, the Segment Anything Model (SAM) [10] was adopted, enabling a quick annotation of the images from the bounding boxes using a

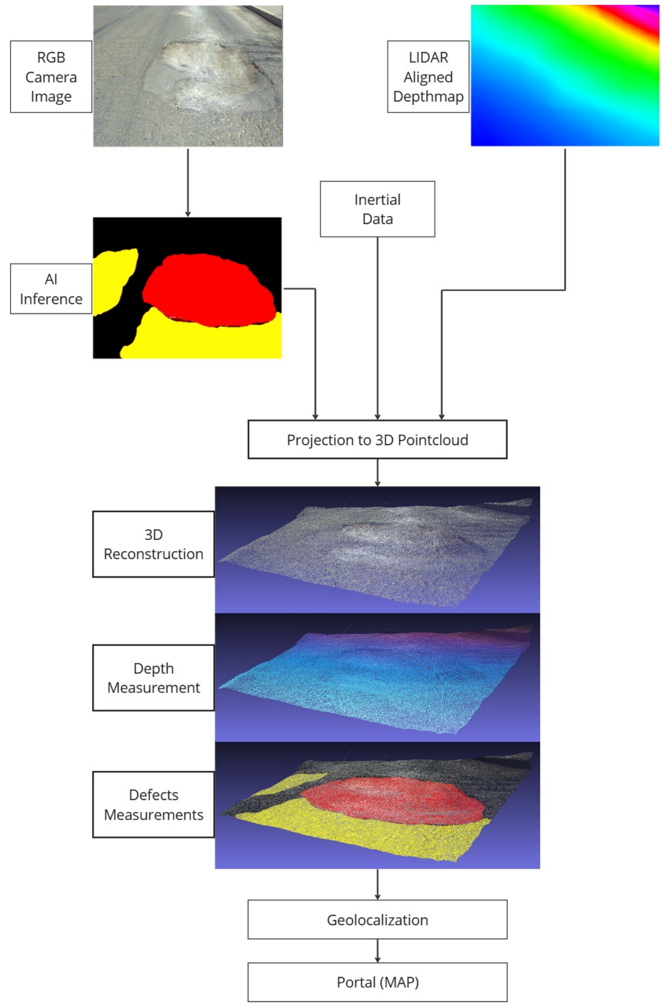

Figure 3. Software architecture components in ROS framework.

dedicated tool. The latter was trained on the Cityscapes dataset [11], which provides over 5000 densely annotated images with 30 segmentation classes. The purpose of the two models is to identify potholes and alligator cracks in road images captured by the camera, along with their segmentation masks, and to verify their placement within the "road" class of the semantic segmentation model, in order to limit false positives. Qualitative results for both models are shown in Figure 4.

III. Results & Discussion

The system was rigorously tested on various urban road sections, demonstrating strong performance in identifying potholes and alligator cracks. The validation of the AI models on dedicated test sets yielded key computer vision metrics that underline their effectiveness. For the YOLO model, a mean Average Precision (mAP) of 0.56 at thresholds of Intersection Over Union (IoU) ranging from 0.5 to 0.95 reflects its robustness in detecting and segmenting defects across different scales and conditions. The mAP is calculated with

 ISBN: 978-1-68558-251-7

the following equation:

$$\mathrm{mAP} = \frac{1}{N_c} \sum_{c=1}^{N_c} \frac{1}{N_{\mathrm{IoU}}} \sum_{i=1}^{N_{\mathrm{IoU}}} \mathrm{AP}_c^{(i)} \tag{1}$$

where N_c is the total number of classes, N_{IoU} is the number of IoU thresholds and $\mathrm{AP}_c^{(i)}$ is the average precision for class c at IoU threshold i.

Additionally, an F1-score of 0.57 indicates a balanced performance in terms of precision (reducing false positives) and recall (capturing true positives). The formula of F1-score is the following:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{2}$$

where P is the Precision and R is the Recall value. For the SegFormerB1 model used in road segmentation, a mean Intersection over Union (meanIoU) score of 0.43 demonstrates its capacity to accurately delineate the "road" class, while an exceptional F1-score of 0.98 highlights its precision and reliability in avoiding misclassifications. The following formulas provide the way for calculating meanIoU for semantic segmentation:

$$\mathrm{IoU}_c = \frac{|A_c \cap B_c|}{|A_c \cup B_c|} \tag{3}$$

$$\mathrm{meanIoU} = \frac{1}{N_c} \sum_{c=1}^{N_c} \mathrm{IoU}_c \tag{4}$$

where A_c is the predicted segmentation for class c, B_c is the ground truth segmentation for class c and N_c is the number of classes. These metrics are considered strong, given the complexity of urban environments and variability in road textures.

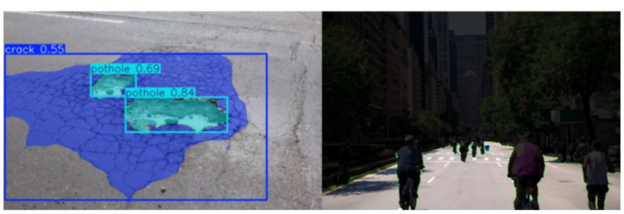

Figure 4. YOLOV8s-Seg results for pothole and crack segmentation (left), SegFormerB1 road segmentation results (right).

Moreover, the integration of segmentation results with LiDAR data allowed for accurate 3D reconstruction and spatial measurements, achieving an error margin of less than 10% for defect dimensions (surface area and depth). The system's navigation module further enhanced functionality, delivering geolocation with Global Positioning System (GPS) accuracy suitable for effective road management applications at the city scale. Finally, deployment on the Nvidia Jetson AGX 64 GB Edge device and model optimization using TensorRT enabled real-time processing. The YOLO model achieved a remarkable throughput of 312 Frames-Per-Second (FPS), while the SegFormer model delivered 18 FPS, ensuring a processing rate exceeding 10 Hz—well-aligned with the LiDAR's sampling rate. This ensures that the system can operate seamlessly in real-time, offering both efficiency and scalability.

IV. Conclusion

Our work presents an advanced, integrated system for detecting and mapping road surface defects, marking a significant step forward in Smart City infrastructure maintenance. By leveraging a multi-sensor approach, including LiDAR, RGB cameras, inertial data, and GNSS, the system achieves precise localization and accurate measurements of defects like potholes and alligator cracks. The innovative application of RGB-based AI models combined with LiDAR domain projection enables the use of existing datasets, minimizing the need for extensive retraining. Deployment on an edge device ensures real-time processing, while the ROS-based framework facilitates seamless data integration and visualization. The achieved accuracy in defect detection, spatial measurement, and geolocation demonstrates the system's potential for scalable implementation in urban road management. Future developments could further enhance adaptability to diverse environments, driving even greater efficiency in urban infrastructure maintenance.

Acknowledgement

This project has been co-financed by the European Union through the PR FESR 2021–2027 RSI program of Regione Lazio, managed by LazioInnova. The authors would like to thank the European Union and Regione Lazio for their support in enabling this research. Additionally, we extend our gratitude to all partners and collaborators who contributed to the successful implementation and validation of the proposed system.

References

[1] S. Girisan, T. V. Sreelakshmi, N. V. Swetha, V. Suresh, and K. M. Vipin, “Pothole Detection Based on Accelerometer Method“, International Journal of Research in Engineering, Science and Management, vol. 3, issue 5, May 2020.

[2] A. Ahmadi, S. Khalesi, and M. R. Bagheri, “Automatic road crack detection and classification using image processing techniques, machine learning and integrated models in urban areas: A novel image binarization technique“, Journal of Industrial and Systems Engineering, vol. 11, pp. 85-97, 14th International Industrial Engineering Conference, 2018.

[3] K. L. Qi, “Fast and Accurate Road Crack Detection Based on Adaptive Cost-Sensitive Loss Function“, Institute of Electrical and Electronics Engineers (IEEE), vol. 53, pp. 1051–1062, IEEE Transactions on Cybernetics, February 2023.

[4] Y. Li, C. Yin, Y. Lei, J. Zhang, and Y. Yan, “RDD-YOLO: Road Damage Detection Algorithm Based on Improved You Only Look Once Version 8“, Applied Sciences, vol. 14, April 2024.

[5] E. Tran et al., “SHREC 2022: Pothole and crack detection in the road pavement using images and RGB-D data“, SHREC2022 3D Shape Retrieval Challenge, 2022.

[6] S. A. Talha, “A LiDAR-camera fusion approach for automated detection and assessment of potholes using an autonomous vehicle platform“, Innovative Infrastructure Solutions, vol. 8, September 2023.

[7] J. Zeng, H. Ouyang, M. Liu, L. U. Leng, and X. Fu, “Multi-scale YOLACT for instance segmentation“, Journal of King Saud University - Computer and Information Sciences, vol. 34, October 2022.

[8] E. Xie et al., “SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers“, Neural Information Processing Systems (NeurIPS), 2021.

 ISBN: 978-1-68558-251-7

[9] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, “RDD2022: A multi-national image dataset for automatic Road Damage Detection“, Crowdsensing-based Road Damage Detection Challenge (CRDDC), 2022.
[10] A. Kirillov et al., “Segment Anything“, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015-4026, 2023.
[11] M. Cordts et al., “The Cityscapes Dataset for Semantic Urban Scene Understanding“, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3213-3223, 2016.