Real-time Quantitative Visual Inspection using Extended Reality

Zaid Abbas Al-Sabbag University of Waterloo, ON, Canada
Jason Paul Connelly University of Waterloo, ON, Canada
Chul Min Yeum University of Waterloo, ON, Canada
Sriram Narasimhan University of Waterloo, ON, Canada
Email: {zaalsabb, jpconnelly, cmyeum, sriram.narasimhan}@uwaterloo.ca

Abstract

In this study, we propose a technique for quantitative visual inspection that can quantify structural damage using extended reality (XR). The XR headset can display and overlay graphical information on the physical space and process the data from the built-in camera and depth sensor. Also, the device permits accessing and analyzing image and video stream in real-time and utilizing 3D meshes of the environment and camera pose information. By leveraging these features for the XR headset, we build a workflow and graphic interface to capture the images, segment damage regions, and evaluate the physical size of damage. A deep learning-based interactive segmentation algorithm called f-BRS was deployed to precisely segment damage regions through the XR headset. A ray-casting algorithm is implemented to obtain 3D locations corresponding to the pixel locations of the damage region on the image. The size of the damage region is computed from the 3D locations of its boundary. The performance of the proposed method is demonstrated through a field experiment at an in-service bridge where spalling damage is present at its abutment. The experiment shows that the proposed method provides sub-centimeter accuracy for the size estimation.

1 Introduction

The goal of vision-based structure inspection methods, in general, is to utilize images to conduct remote visual inspections of large-scale structures that are difficult to access (e.g., rough terrains, spanning water bodies) and have large surface areas to be inspected. These methods aim to automate key steps typical in visual inspections, namely, collection, identification, localization, quantification, and documentation of damage regions to reduce time for inspections and monetary costs.

Recent advances in computer vision technologies: new vision sensors, sensing platforms, and high-performance computing, have transformed how structures are inspected. As sensors become smaller, lower cost, and more powerful, a larger volume of high-quality visual data can be captured from structures with high spatial and temporal granularity. Powerful computer vision methods and machine learning algorithms enable automatic extraction of visual features and semantic information to detect visual changes of structures, which may be an indicator of damage in structures.

Over the past several years, different vision-based visual inspection techniques have been actively developed with the intent of applying them to various civil structures. There are two key steps in vision-based visual inspection: data collection and feature extraction. Visual data collection can be challenging for large scale civil structures, so remote and automated data collection systems such as drones, mobile ground robots, and surveillance cameras [1, 2] have been proposed to allow access to inspection regions that are difficult to reach. Advances in computer vision techniques such as deep neural networks (D-NN) [3–5] and 3D scene reconstruction [6] have allowed for the extraction and localization of damage features based on images.

However, despite technological opportunities, their adoption in the field has been quite limited. One of the main obstacles is the lack of real-time interaction between the inspector and the technology during the process of inspection. Existing vision-based inspection methods are heavily asynchronous, which means data processing takes place hours or days after field-inspection and data collection. As a result, key outcomes are not known at the time of inspection and hence cannot inform either the inspector to allow for immediate investigation or intervention.

A potential solution to address the current limitation is to exploit an extended reality technology (XR) for real-time processing and visualization of inspection data. XR is a term referring to a technology combining reality and virtual spaces for immersive data visualization and enabling human-machine interaction. Microsoft HoloLens 1 & 2 headset (HL2) are examples of XR technology enabled via a wearable headset. For the domain applications of vision-based inspection, a camera and depth sensor equipped in an XR device facilitates the real-time detection, analysis of visual damage, and visualization of results in real-time. Also, the depth sensor permits scanning and reconstructing the 3D geometry of the scene so that such 3D maps can be used for quantitative damage analysis.

In this study, we propose a vision-based quantitative damage measurement method through an XR wearable device, where the physical size of the damage is estimated by processing sensor data collected from the XR device (HL2 in this study). A deep learning-based interactive segmentation algorithm called f-BRS was deployed to precisely segment visual damage regions from the images and engaging the human in this process to improve the quality of the result. To measure the size of damage, the segmented damage region is geometrically evaluated using 3D scene geometry and camera pose information obtained from the XR device. Also, to support computationally or memory intensive algorithms, we build a pipeline for offloading heavy computations to a remote server. To demonstrate the capability of the proposed method and feasibility of its application in the field, an experiment study is conducted using HL2 at an actual bridge where spalling damage is present at its abutment.

2 System Overview

The process pipeline for the proposed system to quantitatively evaluate the size of damage area by leveraging the XR headset and DNN-based interactive segmentation algorithm is shown in Fig. 1. The user starts by selecting segmentation seed points inside and outside the target damage region through a hand gesture in the XR device. The user simply clicks those locations with their fingers and the XR headset automatically anchors those points, denoted X_p, to the spatial mesh environment using a ray-casting algorithm, denoted *rayCast*. Once the point selection is completed, the user then takes an image (I) that includes the view of the entire target damage region and X_p. The 2D pixel coordinates (x_p) corresponding to X_p are obtained by multiplying X_p by the projection matrix (P) of I. Then, I and x_p are sent to the computational server. An interactive segmentation algorithm uses x_p as seed points to segment damage region using the DNN-based interactive segmentation algorithm (*Segment* in Fig. 1), and obtain the pixel coordinates of its boundary (x_s) using a contour extraction algorithm (*findContours* in Fig. 1). The pixel information of the extracted damage boundary, x_s, will be sent back to the XR headset. x_s is then back-projected from the camera center (C) of I using the pseudo-inverse of P, denoted P^+, to the spatial mesh using *rayCast* to obtain the 3D world coordinates of the damage boundary edges (X_s). The headset then displays a holographic overlay of the damage region and anchors the graphics to the spatial mesh so that the user can determine if the target damage region is properly detected for size computation. Then, the user decides on the quality of the segmentation output through the graphic overlay, illustrated as the decision symbol in Fig. 1. If the segmentation is not satisfactory, the user needs to refine the segmentation by adding more seed points and repeat the process for segmentation. If the segmentation is satisfactory, the physical area (A_s) of the damage region is calculated by computing the area of the polygon joined by X_s. Here, since X_s is not perfectly placed on a single plane due to errors in the spatial mesh, X_s is projected on a best-fitting plane before computing the polygon area. Finally, the headset generates a text overlay of the physical size of the target damage region and anchors it to the center of the detected damage area.

3 Interactive Segmentation

Most state-of-the-art vision-based methods for detecting and localizing damage rely on automated image segmentation algorithms

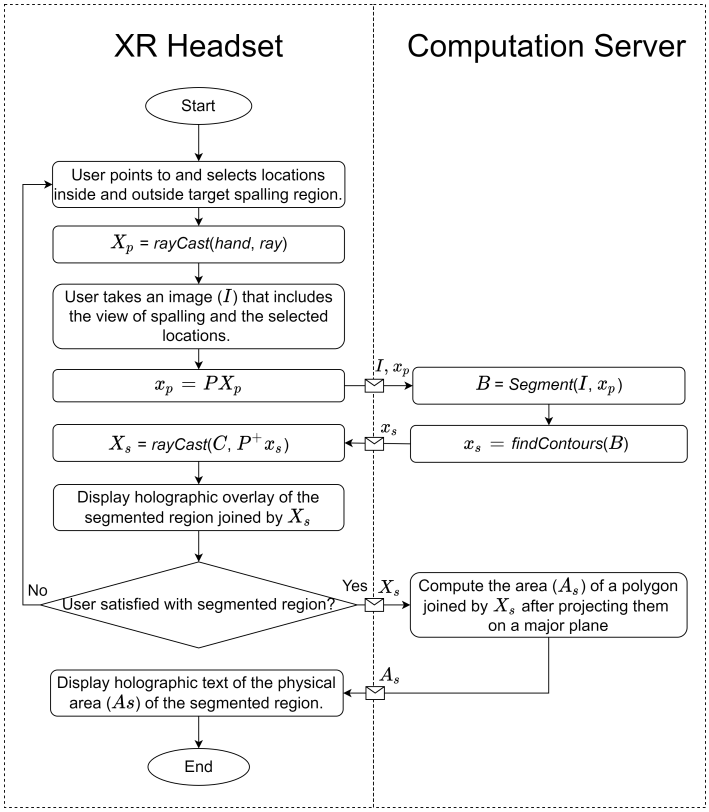
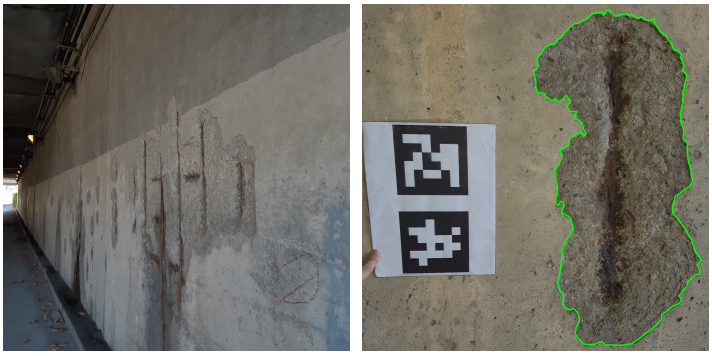


Fig. 1: Real-time damage size estimation pipeline using an XR headset and computation server.



(a) *(b)*

Fig. 2: Experimental validation using actual spalling damage on an in-service bridge (a) Overview of a bridge abutment wall for testing and (b) Ground-truth measurement of a target spalling area using a marker and manual segmentation.

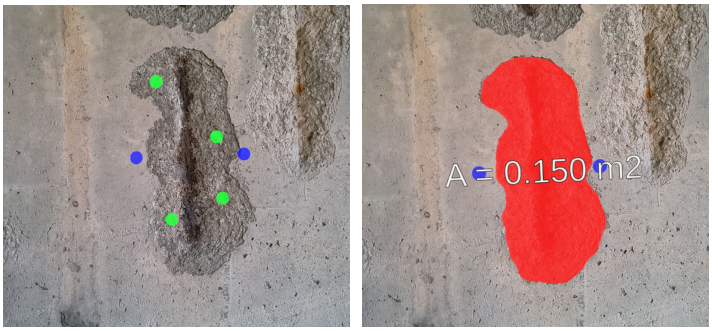
Fig. 3: Overview of test setup: A Microsoft HoloLens 2 headset wirelessly connected to a laptop for computation.

that either use DNN-based semantic segmentation [3–5] or hand-crafted features [7–9]. Although the performance of such algorithms have improved over the years, they often fail to capture the clear boundary of target damage under real-world lighting, texture variations or clutter conditions. Such incorrect segmentation will result in erroneous visual inspection results as an under- or over-estimation of damage size. Thus, in this study, rather than applying automated algorithms, we implemented an interactive segmentation algorithm, called feature back-propagating refinement scheme (f-BRS) [10]. The XR headset can fully support data visualization and interactive operations so we can deploy f-BRS for reliable real-time segmentation of damaged regions.

The user provides seed points inside and outside the damage region, which are then used as initial conditions to find a region in the image that is distinct from the background and includes the positive points and excludes the negative points. f-BRS builds upon the back-propagating refinement scheme (BRS) [11], which improves segmentation accuracy by optimizing network inputs to minimize squared error at seed point locations after each click. However, f-BRS only optimizes scale and bias variables in intermediate layers of the network and achieves comparable results to BRS with much lower time per click. We show that f-BRS can segment spalling damage (target damage type used for experiment demonstration) using a ResNet-34 network trained on the SBD benchmark dataset [12]. The pre-trained model still achieves good segmentation results (approximately more than 0.78 intersect-over-union (IoU) when more than 6 positive and negative points are used as the seed) even though this model was not trained on images with spalling damage.

4 Experimental Validation

The proposed system was tested on spalling damage present on an abutment of an in-service bridge in Fig. 2a. There are several spalling regions, but we randomly selected one target defect location for evaluation purposes. The ground-truth area of the target spalling is simply measured by capturing an image with a square marker reference present. Then, the boundary of the spalling on the image is manually segmented, which is a green line in Fig. 2b. The surface areas in pixels is computed using the Shoelace formula and converted to a physical area using the scale obtained from the marker with a known dimension. The ground-truth size of the target spalling is 0.143 m^2. The experiment was conducted using the HL2 device which was wirelessly connected to a personal laptop using a Wi-Fi hotspot. The laptop served as the computation server in the proposed system (see Fig. 3). The time required from start to finish in Fig. 1 is around 30 sec, while little over 20 sec was consumed for the image segmentation task. This is primarily the result of the laptop not being equipped with a power GPU compute capability. It has been reported that when a GTX 1080 Ti GPU is used, the processing time of the f-BRS segmentation algorithm can be up to 0.32 sec per seed point, or less than 2 sec if 6 seed points are selected and used [10]. The remainder of the time was spent on interactive seed point selection. The *rayCast*, *findContour*, and area computation processes are executed in almost real-time. In the actual testing, four and two seed points from inside and outside spalling respectively are selected, respectively. Then, the image is captured to include a full view of the spalling and selected seed points. Fig. 4a shows the image with the selected seed points, which is sent to the laptop for segmentation. The placement and the number of points can be adjusted through trial and error to obtain the best segmentation result, if the segmentation result is not satisfactory, as mentioned in Fig. 1. Finally, the area of the segmented region is automatically computed, and the holographic overlay of the region and area value are displayed as shown in Fig. 4b. For this test, the estimated spalling area is 0.15 m^2 and the difference from the ground-truth measurement is less than 0.007 m^2, which is less than 4% error.



(a) *(b)*

Fig. 4: Outcome of spalling area estimation: (a) after selecting seed points, and (b) final visualization of the spalling segmentation and its area. Note that the graphics are holograms that are anchored in the spatial mesh, so the graphics are overlaid on the same physical locations regardless of changing HL2 locations or viewpoints.

5 Conclusion

In this study, we propose a vision-based quantitative damage measurement method that leverages XR technology to measure the size of structural damage in real world units by processing data from the built-in camera and depth sensor. The proposed system was deployed on a XR wearable device, Microsoft HoloLens 2. Using this headset, a field experiment was conducted at an in-service bridge for spalling damage size estimation. The results of the experiment show that the proposed system can segment damage areas accurately and achieve less than 4% error compared to ground-truth marker-based damage area measurements.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2020-03979].

References

[1] V. Hoskere, J.-W. Park, H. Yoon, and B. F. Spencer Jr, "Vision-based modal survey of civil infrastructure using unmanned aerial vehicles," *Journal of Structural Engineering*, vol. 145, no. 7, p. 04019062, 2019.

[2] N. Charron, E. McLaughlin, S. Phillips, K. Goorts, S. Narasimhan, and S. L. Waslander, "Automated bridge inspection using mobile ground robotics," *Journal of Structural Engineering*, vol. 145, no. 11, p. 04019137, 2019.

[3] S. Li, X. Zhao, and G. Zhou, "Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 7, pp. 616–634, 2019.

[4] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 731–747, 2018.

[5] Y.-z. Lin, Z.-h. Nie, and H.-w. Ma, "Structural damage detection with automatic feature-extraction through deep learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 12, pp. 1025–1046, 2017.

[6] B. F. Spencer Jr, V. Hoskere, and Y. Narazaki, "Advances in computer vision-based civil infrastructure inspection and monitoring," *Engineering*, vol. 5, no. 2, pp. 199–222, 2019.

[7] S. German, I. Brilakis, and R. DesRoches, "Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments," *Advanced Engineering Informatics*, vol. 26, no. 4, pp. 846–858, 2012.

[8] T. Dawood, Z. Zhu, and T. Zayed, "Machine vision-based model for spalling detection and quantification in subway networks," *Automation in Construction*, vol. 81, pp. 149–160, 2017.

[9] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 196–210, 2015.

[10] K. Sofiiuk, I. Petrov, O. Barinova, and A. Konushin, "f-brs: Rethinking backpropagating refinement for interactive segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8623–8632.

[11] W.-D. Jang and C.-S. Kim, "Interactive image segmentation via backpropagating refinement scheme," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5297–5306.

[12] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik, "Semantic contours from inverse detectors," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 991–998.