BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 69(3), 2021, Article number: e137123
DOI: 10.24425/bpasts.2021.137123

The measurements of surface defect area with an RGB-D camera for a BIM-backed bridge inspection

Bartosz WÓJCIK* and Mateusz ŻARSKI

Department of Mechanics and Bridges, Faculty of Civil Engendering, Silesian University of Technology,
ul. Akademicka 5, 44-100 Gliwice, Poland

Abstract. Bridge inspections are a vital part of bridge maintenance and the main information source for Bridge Management Systems is used in decision-making regarding repairs. Without a doubt, both can benefit from the implementation of the Building Information Modelling philosophy. To fully harness the BIM potential in this area, we have to develop tools that will provide inspection accurate information easily and fast. In this paper, we present an example of how such a tool can utilise tablets coupled with the latest generation RGB-D cameras for data acquisition; how these data can be processed to extract the defect surface area and create a 3D representation, and finally embed this information into the BIM model. Additionally, the study of depth sensor accuracy is presented along with surface area accuracy tests and an exemplary inspection of a bridge pillar column.

Key words: road design; horizontal alignment; general transition curves.

1. Introduction

Periodic inspections of a bridge structure are the main source of information regarding its condition during the operation and maintenance phase. Their correct execution guarantees the safety of the bridge structure and, most importantly, to its users. Over the years, they were performed manually by specialized bridge inspectors and were labour-intensive. However, some problems with bridge inspections have been identified. Kong and Frangopoulos showed that preventing minor defects from turning into major ones can reduce maintenance costs up to 65% [1]. A study by Phares et al. proves that 60% of inspectors did not photograph defects such as cracks at the bottom of the inspected bridge [2]. This shows that the maintenance decision-making process, which is critical from the point of view of structure safety, is based on error and information loss to which the inspection process is prone. The answer to that could be the integration of Building Information Modelling technology that is being introduced to almost every aspect of transport infrastructure [3] as BIM focuses on preserving information about a structure through its life cycle.

The most notable work in this field was done by the researchers involved in the development of the SeeBridge (Semantic Enrichment Engine for Bridges) project [4]. They published many research papers describing Information Delivery Manual and Model View Definition for a BIM-backed bridge inspection [5], defect information integration into BIM models of reinforced concrete bridges [6], defect detection with Artificial Intelligence [7, 8], and as-is BIMs acquisition [9, 10]. However, they were not the only researchers working in this area. Other examples include the work of Isailović et al., where an approach to the detection and modelling of spalling defects into IFC based BIMs was presented [11]; Shim et al. described a BIM-based Bridge Management System for cable-stayed bridges [12, 13]; and later the next generation of BMS utilising Digital Twining was proposed [14, 15]. All these researchers utilised or proposed the use of a 3D reconstruction to obtain data for modelling structural defects.

The most popular 3D reconstruction techniques used for this are laser scanning and photogrammetry. They are well known in the construction industry [16, 17]. However, the work presented by Popescu et al. shows that there are some alternatives to these methods. They proved that a scanner equipped with RGB-D cameras is also viable for a bridge inspection [18]. This type of camera streams not only three standard channels: Red, Green and Blue (RGB), but also provides a fourth channel – Depth, i.e. distance from the camera focal point (Fig. 1). Moreover, handheld RGB-D camera modules are becoming increasingly popular nowadays. This is mainly because they provide a means to reconstruct scenes at an affordable cost even in real time [19]. These advantages of RGB-D cameras make them ideal for capturing 3D information needed in as-is BIM enrichment. Yet in

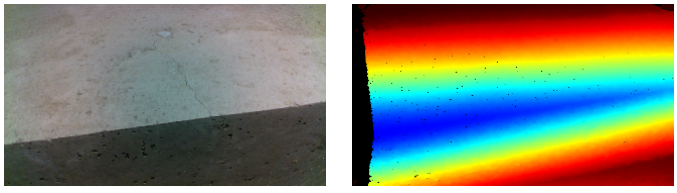

Fig. 1. Data provided by RGB-D camera; RGB image (left) and depth map (right)

*e-mail: bartosz.wojcik@polsl.pl

Manuscript submitted 2020-12-07, revised 2021-02-12, initially accepted for publication 2021-03-31, published in June 2021

the literature and in the reviews on the topic of 3D reconstruction, they are not recognised with reference to the previously mentioned topics or construction in general [16, 17, 20–22].

These cameras attracted a lot of interest when Microsoft presented the first Kinect with Xbox 360 premiere. It was so successful that it was often used in many different areas. However, the main problem of the first two devices from the Kinect family was the size of the sensor and the need for an external power supply. This led to the development of a new version that could be more flexible. The latest Azure Kinect is significantly smaller and can be powered through a USB cable, thus it can work with a tablet or smartphone. Most importantly, it was not developed for the console. It is sold as a Development Kit meant to be integrated into advanced computer vision tools.

However, Microsoft was not the only company to realize how much potential RGB-D cameras have. For example, Intel also started working in this area. They presented a RealSense line, with devices like D415, D435/D435i, D455, and L515. Another good example would be OAK-D, which is included in OpenCV AI KIT that amassed over 1.3 million USD on Kickstarter in 20 minutes [23].

As stated before, the first two Kinects were built to track human motion, as they were used as an input to control the Xbox console. This area is still being explored by researchers [24]. However, the RGB-D cameras have found application in a wider scope, mainly in machine vision as they are great modules to integrate into robotic systems. For example, Hong et al. integrated Intel's RealSense D435 camera into an online autonomous object-packing system [25].

Furthermore, there are also some examples of RGB-D camera applications in the construction industry. Jahanshahi et al. described RGB-D-based pavement defect detection and quantification [26]. Roca et al. mounted an RGB-D camera on UAV; this enabled them to perform geometric control of the building façade [27]. Bellés and Pla used two depth sensors to reconstruct sewer manholes [28] Abdelbarr et al. utilized multiple cameras to monitor the displacement field of a structure [29]. Xu et al. presented an approach for modelling and problem solving of building defects using point clouds and enhanced case-based reasoning, which also featured an RGB-D sensor as a data acquisition device [30]. Nahangi et al. performed a comparison of depth sensors with a laser scanner to evaluate their application for pipe radius extraction [31]. Beckman et al. combined deep-learning object detection with a depth sensor to not only detect but also quantify the spalling defects [32].

However, the examples above used the first or second version of Microsoft Kinect sensors, which were released in 2010 and 2013, respectively. As far as it is understandable in the case of research conducted a few years later, it is hard to consider these devices novel nowadays. This shows that in the context of the construction industry it is harder to come by studies that use the current generation of depth sensors. These are smaller and have lower power consumption, which makes them truly mobile. One example would be the work of Kim et al. However, they fused an RGB-D with a DSLR camera to enable crack measurements as the resolution of an RGB sensor in Intel RealSense D435 was too low for this task [33].

In this work, we demonstrate that utilising the latest generation of an RGB-D camera coupled with a tablet device is viable to aid bridge visual inspection. We describe a complete solution that would provide a means to extract the surface defect features as well as enable the creation representation ready to be embedded into the BIM model. Moreover, due to low costs and no changes in bridge inspectors' routine, our solution could be deployed rapidly.

The rest of this paper is organized as follows. Section 2 introduces the depth sensor that was used in this study. Next, the proposed solution facilitating the measurement of the surface area of the defects on the planar concrete surface and creating its representation ready for inclusion in the BIM model in the IFC standard is described in Section 3. Finally, Section 4 presents the results of the tests and a case study conducted to assess our solution. Lastly, Section 5 contains a discussion, and conclusions are drawn in Section 6.

2. The used sensor

Intel RealSense D435i Depth Camera was used for data acquisition in this study. An excerpt from its technical specification listing the basic features of the modules is shown in Table 1. The device utilises an active infrared stereovision to reconstruct depth. This means that it is obtained from two IR cameras placed at a certain distance from each other, similarly to how human vision works. Additionally, as it is an active method, a dotted pattern is projected by an IR laser emitter on the reconstructed scene. These dots, which are invisible to the naked eye, provide additional features that help the matching algorithm find a solution.

Apart from these infrared imagers, the colour camera is also present, providing additional RGB information that can be associated with depth data. Moreover, this device has a built-in Internal Measuring Unit (IMU), which provides timestamped readings from the accelerometer (3 Degrees of Freedom) and the gyroscope (also 3 Degrees of Freedom). This IMU can be used for enchanting tracking of the device path through the scene.

The process of depth reconstruction occurs in Intel RealSense Vision Processor D4, which is a part of the depth module. Thus, depth is streamed directly from the camera with other video streams. All this data is streamed via a USB cable which also supplies power to the device. Small dimensions combined with the fact that this device can work with smartphones, tablets, and laptops make data acquisition quite easy.

Additionally, the depth data provided by D435i can be easily converted into an organized point cloud, although due to the difference in the field of view of depth and colour frames, only part of this point cloud is available with RGB data. However, it has to be emphasised that this sensor is not meant for end-users, rather for OEM developers or researchers. Additionally, it was not developed with accurate measurements in mind, but rather focuses on mobility and real-time scene reconstruction. According to available documentation, D435i depth cameras should have less than $\pm 2\%$ error for depth values and spatial noise up to 2 meters [34].

Table 1
Intel RealSense Depth Camera D435i Specification [34]

Dimensions [mm] (length × depth × height)	90 × 25 × 25
Depth technology	Active IR Stereo
Depth FOV [°] (horizontal / vertical / diagonal)	86 / 57 / 94
Max Depth Output resolution [px] (horizontal × vertical)	1280 × 720
Max Depth Frame Rate [fps]	90
IR Imagers Baseline [mm]	50
Left/Right IR Imagers FOV [°] (horizontal / vertical / diagonal)	91.2 / 65.5 / 100.6
Left/Right IR Imagers resolution [px] (horizontal × vertical)	1280 × 800
Max Left/Right IR Imagers Frame Rate [fps]	90
IR Imagers Distortion [%]	≤ 1.5
IR Projector FOV [°] (horizontal / vertical / diagonal)	100.4 / 69/ 110.4
Optical Power [mW] (average / peak)	350 / 440
Laser Wavelength at 20°C [nm]	850 ± 10
Colour Camera FOV [°] (horizontal / vertical / diagonal)	69.4 / 42.5 / 77
Colour Camera resolution [px] (horizontal × vertical)	1920 × 1080
Max Colour Camera Frame Rate [fps]	30
Colour Camera Distortion [%]	≤ 1.5
Internal Measuring Unit Degrees of Freedom	6
Acceleration Range [g]	±4
Accelerometer Sample Rate [Hz]	62.5, 250
Gyroscope Range [°/s]	±1000
Gyroscope Sample Rate [Hz]	200, 400

3. A proposed solution

In many systems supporting infrastructure maintenance, photos are still the main source of information about inspected structures. This is because images are not only easy to acquire, as nowadays almost everybody carries a digital camera around, but also because humans process visual data better than any other. However, despite these advantages, images cannot provide the exact information needed (Table 2), such as the length, surface area, or volume. Therefore, they are only used for qualitative assessment.

This gap could be filled by RGB-D cameras by adding a fourth channel with depth which would enable the measurement of defects and thus provide a means to easily quantify damage without significant changes to inspection routines. Inspectors would still have to take photos of defects; the only difference would be a different type of camera used. Moreover, knowing this relationship between the photo and depth provides the opportunity to create a textured 3D defect representation that would allow to embed it into the BIM model in the way described by Hüthwohl et al. [6].

Table 2
Types and main properties of the surface defects collected during the visual inspection [6, 21, 22]

Surface defect	Measurable defect property
Spalling	Area and depth
	Volume
Scaling	Area and depth
	Volume
Efflorescence	Area
Delamination	Area
Organic contamination	Area
Inorganic contamination	Area
Crack	Width

In this study. we combine the previously described sensor with a tablet device to propose a low-cost RGB-D-based solution as an indirect visual examination method for defect surface area measurements, like efflorescence, delamination, organic/inorganic contamination, and in some cases also scaling. Additionally, our method would enable the creation of results ready to be embedded into the BIM model.

The scheme of the proposed solution can be divided into the following steps:

- RGB-D data acquisition;
- Image processing;
- Point cloud processing;
- Surface reconstruction and area measurements;
- Texture mapping;
- IFC model enrichment.

3.1. Image processing. A semi-automatic GraphCut-based algorithm was used [35]. It is an implementation of a lazy snapping interactive image segmentation tool presented by Li et al. [36], and enables the removal of the image background with minimal input, as it performs a series of graph cuts based on paths defined by the user, to create a binary mask that separates objects in the foreground.

As the results were not always satisfactory, the initial mask was refined by an active contour model, also called snakes, described by Kass et al. [37] It is an automatic iterative region-growing model that takes a seed, in this case the output of background removal, and grows or shrinks it to fill the region in an image.

This approach proved to be enough to segment out defects from the concrete surface in the colour frames captured by an RGB-D camera. The example of the original image and the mask obtained as a result of image processing are shown in Fig. 2.

Fig. 2. Image processing: colour frame (left) and corresponding binary mask (right)

3.2. Point cloud processing. Having the defect already segmented out of colour frames, it is already possible to combine it with depth data to measure its relative surface area. However, due to the noisy nature of raw depth (Fig. 3 left), it would yield poor results, thus it has to be processed to enhance it.

Fig. 3. Point cloud processing: original point cloud (left), point cloud projected on to a plane fitted with RANSAC (middle), and point cloud representing the segmented defect (right)

As mentioned before, the depth frame can be easily transformed into a point cloud, with the only condition that the camera intrinsics are known. Since the depth stream is internally mapped to the colour one, a pin-hole camera model is used for the RGB sensor. That simplifies the transformation, so the 3D point coordinates are being calculated in the following way

$$\begin{cases} x = \dfrac{d(x_p - pp_x)}{fx}, \\ y = \dfrac{d(y_p - pp_y)}{fy}, \\ z = d, \end{cases} \tag{1}$$

where:
x, y, z – 3D point coordinates,
x_p, y_p – pixel coordinates,
pp_x, pp_y – principal point coordinates in image coordination system,
f_x, f_y – focal length of the image plane as a multiple of pixel width/height,
d – pixel depth.

This transformation facilitates processing a point cloud, instead of the depth image. Then a plane fitting Random Sample Consensus model can be used. RANSAC, described by Fischler and Bolles [38], proved to be efficient in the iterative estimation of mathematical model parameters in noisy data. In our case, the model is described by a linear plane equation

$$Ax + By + Cz + D = 0, \tag{2}$$

where A, B, C, D – plane coefficients.

Additionally, for fitting a plane model with RANSAC, a distance threshold value can be provided as an input, this parameter defines the tolerance while selecting inliers and outliers in the data, and because of that, it can significantly influence the model coefficient calculations.

Having this in mind, model tolerance was set to $\pm 2\%$ of the mean depth value that was calculated through the depth frame, including only valid depth pixels. This specific value was chosen in accordance with the noise levels stated in the datasheet [34].

After the plane model is computed, every point in a point cloud that has the correct depth value assigned is projected to the fitted plane, resulting in an edited point cloud (Fig. 3 middle).

Moreover, the point clouds acquired by the depth camera are organized point clouds, i.e. points arranged in rows and columns instead of a list as in unorganized point clouds. Indexes of these rows and columns correspond to the pixels in the colour/depth frame. Because of that, it is possible to extract points associated with surface defects, using the binary mask acquired in the image segmentation process (Fig. 3 right).

3.3. Surface reconstruction and area measurement. Surface reconstruction is a process in which the surface of an object is approximated from points contained in a cloud. As the cloud provided by the depth camera is organised, this task is easier and the method described by Holz and Behnke [39] can be applied.

Acquiring this reconstructed mesh facilitates measuring the surface area of the segmented defect as a sum of all triangular faces (Fig. 4).

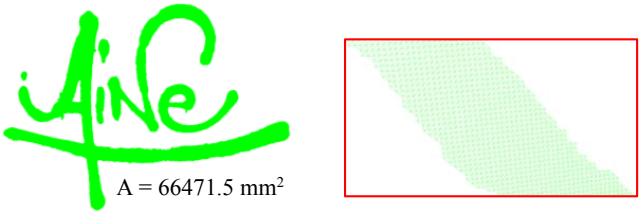

Fig. 4. Results of organized fast meshing (left), close-up (right)

The mesh presented in Fig. 4 composes of over 1.8 million triangles, which when saved into Wavefront.obj file format has a size of over 52 MB. Geometric representation of this size is impractical, as it contains too many faces for planar geometry. As there is no reason for this triangle count, mesh decimation has to be performed with the algorithm presented by Garland and Heckbert [40]. It is capable of producing high fidelity simplified polygonal representations using quadric error metrics and iterative pair contractions. This enabled the decimation of the initially acquired mesh (Fig. 4 left), resulting in a reduction in the triangle count to 275 faces (Fig. 5 top left), which corresponded to a decrease in file size to less than 12 KB.

However, it has to be emphasised that this decimation simplifies the shape of the mesh, altering its original surface area, thus it has to be measured on the initial mesh and stored for future usage.

Fig. 5. Decimated mesh (top left), mesh mapped texture (top right), and textured mesh (bottom)

3.4. Texture mapping. To create a realistic defect representation ready to be embedded into the BIM model, a connection between the colour frame and the decimated mesh has to be established. This process is called texture mapping and can be achieved by projecting 3D points to 2D image coordinates, the texture map created in this process defines the way that the 2D image is applied to 3D mesh (Fig. 5). The projection itself can be done with a formula derived from the previously presented image-to-point transformation (1):

$$\begin{cases} x_p = f_x \dfrac{x}{z} - pp_x \,, \\ y_p = f_y \dfrac{y}{z} - pp_y \,, \end{cases} \tag{3}$$

where:
x, y, z – 3D point coordinates ,
x_p, y_p – pixel coordinate,
pp_x, pp_y – principal point coordinates in the image coordination system,
f_x, f_y – focal length of the image plane as a multiple of the pixel width/height.

The textured mesh (Fig. 5 bottom) can be exported to a file that contains mesh geometry and texture map information, another file that contains the material information needed to connect it to the latter file with the texture, which is just an RGB frame. Additionally, a text file containing the measured surface area of the initial mesh can be created.

3.5. IFC model enrichment. The integration of defect information into BIM models of reinforced concrete bridges is already described by Hüthwohl et al. [6]. They presented a way of modelling bridge defects into IFC files, using *IfcSurfaceFeature* class with a textured mesh as a geometric representation and a set of predefined properties attached to it in the form of *IfcPropertySet*. The same defects are then grouped into *IfcElementAssembly* that are then aggregated into any *IfcProduct* subclass representing a structural element.

4. The validation of the solution

To evaluate the proposed solution, a series of tests were conducted. First, the depth quality assessment was done, then the accuracy tests of surface area measurement followed by a case study to test real-life applicability were performed.

4.1. Depth quality. To enable depth quality assessment, Intel provides Depth Quality Tool with the published whitepaper that describes the test methodology and measured parameters [41]. To assure that the sensor is properly calibrated and calculate the depth error for different distances, the depth quality was tested according to it. A vertical surface was used as a target, the camera was placed perpendicularly on a tripod at a distance of 300, 500, 750, and 1000 mm. Ground truth values were measured by a laser rangefinder. For each position, 100 frames were recorded, with 80% Field of View being analysed.

The depth quality was assessed with the help of the following parameters: Fill Rate, Z Accuracy, and spatial noise. While Fill Rate determines the ratio of pixels containing a valid depth value (non-zero) to all pixels, Z Accuracy provides the information on how close the depth values are to the measured ground truth (distance from the planar target to the camera depth origin). The RMS Error (Root Mean Square Error) or spatial noise measures how noisy the depth values are assuming that the planar target is observed.

The latter is additionally divided into Plane Fit RMS Error and Subpixel RMS Error, where the subpixel error is independent of the distance from the target. The mean values of these parameters calculated from 100 frames are presented in Table 3.

Table 3
Depth quality assessment results (100 frames, 80% FoV)

Distance [mm]	Mean Fill Rate [%]	Mean Z Accuracy [%]	Mean Plane Fit RMS Error [%]	Mean Subpixel RMS Error [px]
350	99.124	–1.568	0.270	0.181
500	98.867	–0.973	0.521	0.255
750	99.762	–0.718	0.613	0.194
1000	99.796	–0.304	0.785	0.185

As seen in the results of the depth quality assessment, despite visible noise, the sensor approximates the depth with an accuracy better than stated in the datasheet. However, this alone contributes to the surface area measurement error in a quadric way, as the area measured on a fitted plane will differ from the expected value with a squared ratio of depth to ground truth

$$A' = \left(\frac{d}{gt}\right)^2 \cdot A \,, \tag{4}$$

where
A, A' – original and measured area ,
d – depth provided by the sensor ,
gt – ground truth.

4.2. Accuracy of the measured surface area. To assess the accuracy of the solution surface area measurement, the final test was performed. The modified setup for depth quality assessment was used for that task, this time a printed target was placed on the wall. This target consisted of a black rectangle, 120 mm high and 65 mm wide resulting in a surface area of 7800 mm^2, printed on a white sheet of paper. Again, for each position, 350, 500, 750, and 1000 mm, 100 frames per distance were captured and analysed. The target is seen from 500 mm, at each stage of image segmentation and plane fitting is shown in Fig. 6.

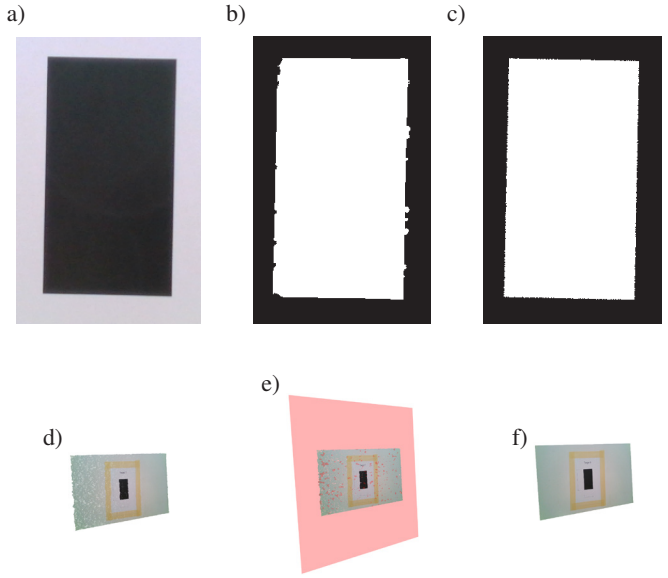

Fig. 6. Image segmentation and plane fitting: a) RGB image of the measured target, b) initial mask created with Lazy Snapping, c) final mask refined with Active Contours, d) point cloud created from raw depth, e) plane model fitted to the point cloud with RANSAC, f) processed point cloud

Minimal, maximal, and mean measured areas with Relative Standard Deviation (RSD), Mean Average Error (MAE), and Root Mean Square Error (RMSE) for each camera position are presented in Table 4.

As can be seen in the presented data (Table 4), a low RSD indicates that the measured values are tightly packed around the mean. This means that the measurements are repeatable. However, if we consider Z accuracy error (Table 3) from the depth quality analysis, the accuracy should be increasing as its value changed from over –1.5% for a distance of 350 mm to almost –0.3% for a distance of 1000 mm. If we approximate the expected errors of the Z accuracy, we can see (Table 5) that errors due to other influences are rising in the opposite direction. One of the reasons seems to be the fact that as the distance increases, the portion of the FoV taken by the marker decreases, as its size is fixed through the measurements.

In conclusion, for now, the presented solution can measure the surface area of defects with over 95% accuracy, but this can be enchanted with RGB-D cameras with a better depth accuracy and certain measures taken during data acquisition, like filling as much frame as possible with the measured area.

Table 4
Results of surface area measurement tests

Distance [mm]	Max [mm^2]	Min [mm^2]	Mean [mm^2]	RSD [%]	MAE [%]	RMSE [%]
350	7532.77	7516.04	7523.81	0.0393	–3.541	–3.541
500	7573.99	7534.46	7560.70	0.107	–3.068	–3.070
750	7561.72	7428.46	7476.27	0.291	–4.151	–4.160
1000	7558.81	7431.04	7483.28	0.364	–4.061	–4.075

Table 5
Approximated influences of various factors on the surface area measurements

Distance [mm]	Approximated error due to Z accuracy influence [%]	Approximated error due to other influences [%]
350	–3.111	–0.430
500	–1.936	–1.134
750	–1.435	–2.725
1000	–0.614	–3.461

4.3. Case study. To test the real-life applicability of the proposed concept, a final case study was performed in the form of two inspections in the field. For this purpose, one of the columns of the box girder bridge pillar was selected, as it already featured graffiti defects, and the as-built BIM model of the substructure was assembled (Fig. 7). The inspections were conducted in the summer of 2019 and 2020. Intel RealSense D435i Depth Camera and Microsoft Surface Go device were used for data acquisition.

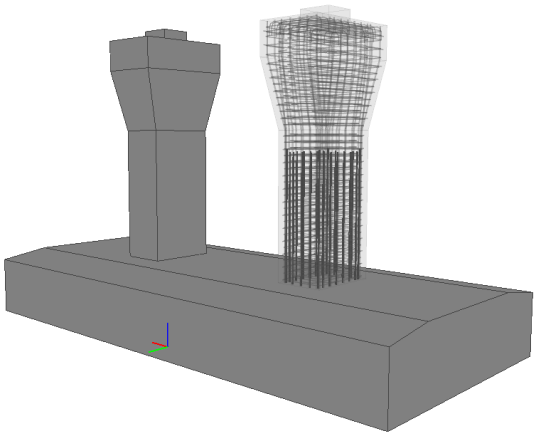

Fig. 7. BIM model of the bridge substructure

Capturing surface defects with a combination of these devices proved to be viable. However, some problems regarding large defects were encountered, as it was harder to cover the whole surface with a single frame. However, the main problem of this solution proved to be the localization of the defect in the

BIM model. In the case of graffiti, the location is not critical and can be approximated, because a correct assignment to the surface of the structural element is enough. Nevertheless, this problem was resolved with the help of a previously acquired textured mesh, obtained by photogrammetry techniques. The use of this 3D representation enabled us to align the features visible on the colour frames with the same features present on a textured mesh, even after the appearance of the surfaces changed in time (Fig. 8).

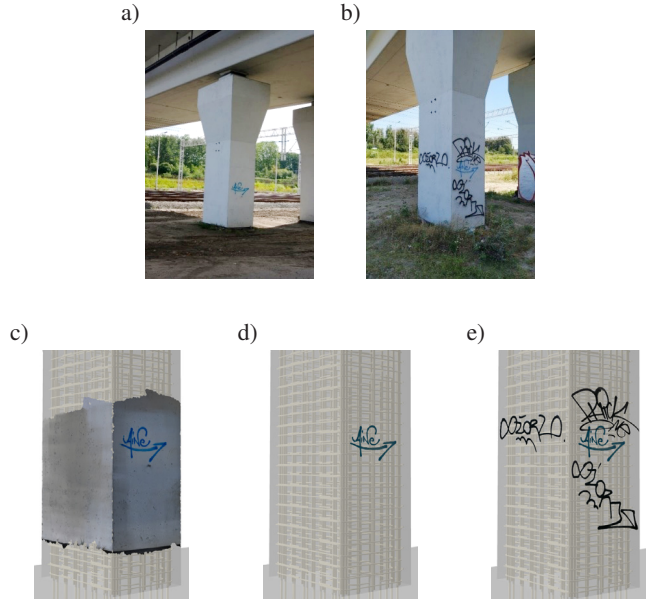

Fig. 8. Case study: a) selected bridge pier in spring 2019, b) selected bridge pier in summer 2020, c) photogrammetric textured mesh used for defect localization, d) results of the first inspection embedded into BIM model, e) updated model with results of the second inspection

5. Discussion

In the article, a solution that uses a low-cost RGB-D camera and a tablet to enchant the inspection of planar concrete surfaces was presented. With this approach, we managed to obtain over 95% of accuracy in surface area measurement, which is more than satisfactory in the light of regular infrastructure inspections commonly done with the naked eye. The proposed method is also suitable for documenting various types of defects (Fig. 9) and their data are also collected with ease and without significant changes in the routine of bridge inspectors as they still need to take a photo of the defects, but with a different camera.

However, one of the biggest problems occurring during the conducted research was defect localization. It was discovered that it was not possible to arbitrarily assign an accurate localization of defects on the surface of the element in the BIM model. This problem was solved with the help of a 3D representation of the column, previously acquired by photogrammetry. However, this type of solution is not entirely satisfactory as it would require a complete scan of the structure, which is a very labour-intensive task [42].

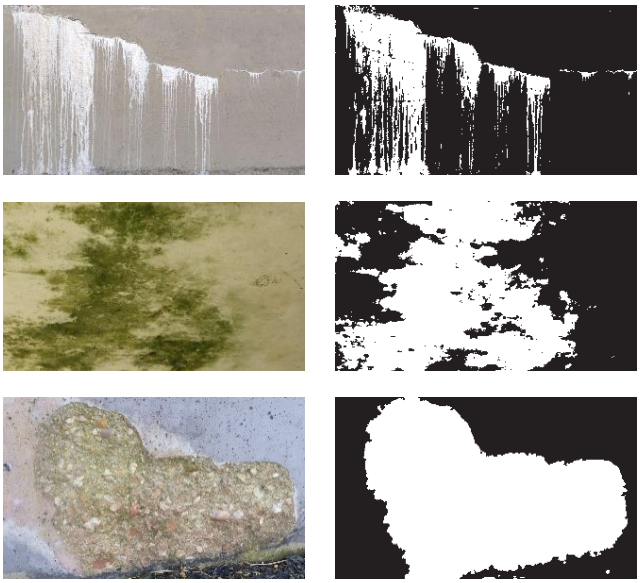

Fig. 9. Images of different surface defects (left column) retrieved from the dataset presented in [7] and the result of the segmentation with algorithm proposed in Section 0 (right column). From the top: efflorescence, vegetation, scaling, rust stains

It is also worth noting that the proposed solution can be further improved with the use of AI methods, as they were successfully implemented previously for defect detection [7,8,11]. It can also benefit from using better camera sensors as proved by Kim et al. [33].

6. Conclusions and future work

We believe that our novel solution can greatly improve the infrastructure inspection and maintenance process, as it is well known that the currently used tools are imperfect and the whole process is prone to loss of information. It is especially apparent when we still lack tools to fully harness Building Information Modelling potential in this area. With the use of the presented methods, work on the inspection of infrastructure can be more precise, faster, and safer. This can be achieved due to the contactless and instant form of taking measurements without exposing the inspector to approaching unsafe areas within the structure. The management of the structure, as well as individual defects, can also be improved thanks to the use of the open IFC format, which does not restrict management units to using proprietary commercial solutions and allow for the free exchange of data between parties.

It is also a step towards even further developing modern infrastructure inspection techniques in future works. The IFC format is under constant development and is shaped according to its users' needs, including infrastructure managers. Further restrictions on the inspector's exposure to danger can be obtained with the increasingly common use of UAV vehicles. Moreover, thanks to the use of fast and accurate AI methods like, e.g. Mask R-CNN [43] together with better camera sensors already present in modern RGB-D cameras, the described methods can

serve as a cornerstone to modernize the standards for conducting infrastructure inspections.

Acknowledgments. Work presented in this study was supported by *European Union* through the *European Social Fund* as a part of a *Silesian University of Technology as a Centre of Modern Education based on research and innovation* project, number of grant agreement: POWR.03.05.00 00.z098/17-00.

References

[1] J.S. Kong and D.M. Frangopol, "Life-Cycle Reliability-Based Maintenance Cost Optimization of Deteriorating Structures with Emphasis on Bridges", *J. Struct. Eng.* 129(6), 818–828 (2003).

[2] B.M. Phares, G.A. Washer, D.D. Rolander, B.A. Graybeal, and M. Moore, "Routine Highway Bridge Inspection Condition Documentation Accuracy and Reliability", *J. Bridg. Eng.* 9(4), 403–413 (2004).

[3] A. Costin, A. Adibfar, H. Hu, and S.S. Chen, "Building Information Modeling (BIM) for transportation infrastructure – Literature review, applications, challenges, and recommendations", *Autom. Constr.* 94, 257–281 (2018).

[4] "SeeBridge". [Online]. Available: https://seebridge.net.technion.ac.il/.

[5] R. Sacks *et al.*, "SeeBridge as next generation bridge inspection: Overview, Information Delivery Manual and Model View Definition", *Autom. Constr.* 90, 134–145 (2018).

[6] P. Hüthwohl, I. Brilakis, A. Borrmann, and R. Sacks, "Integrating RC Bridge Defect Information into BIM Models", *J. Comput. Civ. Eng.* 32(3), (2018).

[7] P. Hüthwohl and I. Brilakis, "Detecting healthy concrete surfaces", *Adv. Eng. Informatics* 37, 150–162 (2018).

[8] P. Hüthwohl, R. Lu, and I. Brilakis, "Multi-classifier for reinforced concrete bridge defects", *Autom. Constr.* 105, 102824 (2019).

[9] R. Lu, I. Brilakis and C. R. Middleton, "Detection of Structural Components in Point Clouds of Existing RC Bridges", *Comput. Civ. Infrastruct. Eng.* 34(3), 191–212 (2019).

[10] R. Lu and I. Brilakis, "Digital twinning of existing reinforced concrete bridges from labelled point clusters", *Autom. Constr.* 105, 102837 (2019).

[11] D. Isailović, V. Stojanovic, M. Trapp, R. Richter, R. Hajdin, and J. Döllner, "Bridge damage: Detection, IFC-based semantic enrichment and visualization", *Autom. Constr.* 112, 103088 (2020).

[12] C.S. Shim, H. Kang, N.S. Dang, and D. Lee, "Development of BIM-based bridge maintenance system for cable-stayed bridges", *Smart Struct. Syst.* 20(6), 697–708 (2017).

[13] N.S. Dang and C.S. Shim, "BIM authoring for an image-based bridge maintenance system of existing cable-supported bridges", *IOP Conf. Ser. Earth Environ. Sci.* 143(1), 012032 (2018).

[14] S. Dang, H. Kang, S. Lon, and S. Changsu, "3D Digital Twin Models for Bridge Maintenance", *10th Int. Conf. Short Mediu. Span Bridg.,* 2018, pp. 73.1–73.9.

[15] C.S. Shim, N.S. Dang, S. Lon, and C.H. Jeon, "Development of a bridge maintenance system for prestressed concrete bridges using 3D digital twin model", *Struct. Infrastruct. Eng.* 15(10), 1319–1332 (2019).

[16] Z. Ma and S. Liu, "A review of 3D reconstruction techniques in civil engineering and their applications", *Adv. Eng. Informatics* 37, 163–174 (2018).

[17] Q. Wang and M.K. Kim, "Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018", *Adv. Eng. Informatics* 39, 306–319 (2019).

[18] C. Popescu, B. Täljsten, T. Blanksvärd, and L. Elfgren, "3D reconstruction of existing concrete bridges using optical methods", *Struct. Infrastruct. Eng.* 15(7), 912–924 (2019).

[19] S. Izadi *et al.*, "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera", in *Proceedings of the 24th annual ACM symposium on User interface software and technology – UIST '11*, 2011, p. 559.

[20] J. Hoła, J. Bień, Ł. Sadowski, and K. Schabowicz, "Non-destructive and semi-destructive diagnostics of concrete structures in assessment of their durability", *Bull. Polish Acad. Sci. Tech. Sci.* 63(1), 87–96 (2015).

[21] J. Bień, T. Kamiński, and M. Kużawa, "Taxonomy of non-destructive field tests of bridge materials and structures", *Arch. Civ. Mech. Eng.* 19(4), 1353–1367 (2019).

[22] J. Bień, M. Kużawa, and T. Kamiński, "Strategies and tools for the monitoring of concrete bridges", *Struct. Concr.* 21(4), 1227–1239 (2020).

[23] "OpenCV AI Kit". [Online]. Available: https://www.kickstarter.com/projects/opencv/opencv-ai-kit.

[24] B. Liu, H. Cai, Z. Ju, and H. Liu, "RGB-D sensing based human action and interaction analysis: A survey", *Pattern Recognit.* 94, 1–12 (2019).

[25] Y.-D. Hong, Y.-J. Kim, and K.-B. Lee, "Smart Pack: Online Autonomous Object-Packing System Using RGB-D Sensor Data", *Sensors* 20(16), 4448 (2020).

[26] M.R. Jahanshahi, F. Jazizadeh, S.F. Masri, and B. Becerik-Gerber, "Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor", *J. Comput. Civ. Eng.* 27(6), 743–754 (2013).

[27] D. Roca, S. Lagüela, L. Díaz-Vilariño, J. Armesto, and P. Arias, "Low-cost aerial unit for outdoor inspection of building façades", *Autom. Constr.* 36, 128–135 (2013).

[28] C. Bellés and F. Pla, "A Kinect-Based System for 3D Reconstruction of Sewer Manholes", *Comput. Civ. Infrastruct. Eng.* 30(11), 906–917 (Nov. 2015).

[29] M. Abdelbarr, Y.L. Chen, M.R. Jahanshahi, S.F. Masri, W.M. Shen, and U.A. Qidwai, "3D dynamic displacement-field measurement for structural health monitoring using inexpensive RGB-D based sensor", *Smart Mater. Struct.* 26(12) (2017).

[30] Z. Xu, S. Li, H. Li, and Q. Li, "Modeling and problem solving of building defects using point clouds and enhanced case-based reasoning", *Automation in Construction* 96(February), 40–54 (2018).

[31] M. Nahangi, T. Czerniawski, C.T. Haas, and S. Walbridge, "Pipe radius estimation using Kinect range cameras", *Autom. Constr.* 99 (March 2017), 197–205 (2019).

[32] G.H. Beckman, D. Polyzois, and Y.J. Cha, "Deep learning-based automatic volumetric damage quantification using depth camera", *Automation in Construction* 99(November 2018), 114–124 (2019).

[33] H. Kim, S. Lee, E. Ahn, M. Shin, and S.-H. Sim, "Crack identification method for concrete structures considering angle of view using RGB-D camera-based sensor fusion", *Struct. Heal. Monit.*, 1–13 (2020).

[34] Intel, "Intel® RealSense™ Camera D400 series Product Family Datasheet" (2019).

[35] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': interactive foreground extraction using iterated graph cuts", in *ACM SIGGRAPH 2004 Papers on – SIGGRAPH '04*, 2004, p. 309.

[36] Y. Li, J. Sun, C.K. Tang, and H.Y. Shum, "Lazy snapping", *ACM SIGGRAPH 2004 Pap. SIGGRAPH 2004*, 303–308 (2004).

[37] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models", *Int. J. Comput. Vis.* 1(4), 321–331 (Jan. 1988).

[38] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", *Commun. ACM* 24(6), 381–395 (1981).

[39] D. Holz and S. Behnke, "Fast Range Image Segmentation and Smoothing Using Approximate Surface Reconstruction and Region Growing", 2013, 61–73.

[40] M. Garland and P.S. Heckbert, "Surface simplification using quadric error metrics", in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques – SIGGRAPH '97*, 1997, 209–216.

[41] Intel, "Intel® RealSense™ Camera: Depth testing methodology", 2018.

[42] B. Wójcik and M. Żarski, "Asesment of state-of-the-art methods for bridge inspection: case study", *Arch. Civ. Eng.* 66(4), 343–362 (2020).

[43] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN", *IEEE Trans. Pattern Anal. Mach. Intell.* 42(2), 386–397 (2017).