3D vision-based bolt loosening assessment using photogrammetry, deep neural networks, and 3D point-cloud processing

Xiao Pan, T.Y. Yang[*]

Department of Civil Engineering, The University of British Columbia, Vancouver, Canada

ARTICLE INFO

ABSTRACT

Structural bolts are essential structural elements. Detection of structural bolt loosening is of great importance to provide earlier warnings of structural damages and prevent catastrophic system-level collapse. Most existing studies about bolt loosening assessment were built in 2D computer vision, where the assessment may be restricted based on the camera views. In this paper, a novel 3D vision-based methodology is proposed for autonomous bolt loosening assessment. First, a 3D point cloud of bolted connection is created using readily available 2D images. Second, a new convolutional neural network (CNN)-based method is developed to localize structural bolts in the 3D point cloud. Further, a 3D point cloud processing algorithm is developed to recognize and quantify bolt loosening. Parameter studies were conducted to investigate the effectiveness of the proposed pipeline. Finally, a real-world implementation has been conducted to quantify bolt loosening on a steel column base connection with bolts. The results indicate that the proposed bolt loosening assessment methodology can effectively localize and quantify bolt loosening at high accuracy and low cost.

1. Introduction

Civil engineering structures such as buildings and bridges deteriorate continuously due to aging effects and environmental impacts such as earthquakes. Damage inspection of these structures is of vital importance to maintain their functionalities. Earlier damage identification before natural disasters can greatly alleviate or prevent the catastrophic failure in the event of natural disasters. Rapid post-disaster inspection of these structures provides crucial information to building owners or decision makers to make informed decisions. Traditional structural maintenance consists of regular human inspections on site, which are slow and heavily dependent on the proper training and experience of inspectors. Data obtained by visual inspection are manually analyzed and documented by the inspectors or engineers, which is inherently biased and can be inconsistent from time to time. This may result in false conclusion and generate erroneous evaluation reports. On the other hand, manual inspection procedures are usually unsafe to the inspectors and inefficient, because the civil structures are relatively large and can be constructed in a harsh environment. Besides, some portions of large or complex infrastructures are practically impossible to be inspected by humans. These issues result in a very high long-term maintenance cost and become even more challenging when rapid assessment of infrastructure on a regional scale is demanded for decision-makers right after the event of natural hazards such as earthquakes. Within this context, there is a compelling need to provide efficient, accurate and economical automated structural health monitoring (SHM) methods to replace traditional manual inspection methods, using novel sensing technologies and data processing algorithms, which are more reliable, consistent, and less dependent on

environmental conditions [1,2].

Vibration-based SHM methods have been established in recent decades to enhance the efficiency of damage inspections. In general, vibration-based SHM methods typically rely on contact type sensors to measure global structural response for damage detection from the estimated structural modal properties. The bolt loosening assessment using contact sensors has been attempted in recent years (e. g., [3–5]). Several limitations can be observed. The performance of many contact sensors are generally affected by environmental conditions, such as temperature and humidity [6,7]. These methods require dedicated experts for high-precision instrumentation, and a professional postprocessing software to compensate the environmental influences [8,9]. Moreover, as for bolt loosening assessment, these approaches could recognize the change of the behavior in loosened bolted connections, but is unable to localize which bolts are loosened exactly [10].

Recent years have witnessed the success of computer vision-based (or vision-based) SHM, as a novel, accurate, efficient and economical complementary approach to vibration-based methods, for structural damage detection [11,12]. On one hand, vision systems have been demonstrated for structural vibration measurement [13,14]. On the other hand, the vision-based methods are designed to automate and enhance the efficiency of the visual inspection of civil structures by humans. Vision-based damage detection methods have been investigated for different structural types, such as reinforced concrete (RC) structures [15–17], steel structures [18–20], masonry structures [21,22]. These studies have demonstrated vision-based methods can be effectively used to analyze various types of structural damages based on 2D images. Despite of the promising results, the damage detection methods in these studies were developed in the area of 2D computer vision. The evaluation outcomes are relatively sensitive to the camera locations and viewing angles where the photos were taken.

In recent years, vision-based bolt loosening detection has been attempted. Most of the exist studies are developed in 2D computer vision, which can be categorized into front view-based detection and side view-based detection methods as explained below.

Front view-based bolt loosening detection typically consists of localization of bolts in an image captured from the front view (i.e., camera plane parallel to the bolt head), followed by quantification of bolt loosening rotation angle. For example, Kong & Li [23], adopted image registration method to estimate for bolt loosening angle. Besides, many existing front view-based detection methods such as [24,25], employed Hough Transform (HT) algorithm [26] to segment bolt edges and measure the rotation angle. Additionally, Huynh et al. [27] investigated a combined CNN and HT method to localize the bolts and quantify the bolt rotation, and demonstrated its applicability in a bridge connection. Further, Ta & Kim [28] enhanced the similar approach to detect bolt loosening and corrosion concurrently. Despite of the progress made, their methods bear several limitations coming from the HT algorithm. To wit: 1) The HT algorithm employed in these studies may not detect lines and circles of bolt edges very well, when the environment becomes more complicated (e.g., presence of washers, light reflection on the bolts, shadows, effects of background noise in a relatively low light condition. 2) The HT algorithm is applied to two static images to extract edge lines of bolts. However, the measurement range is only up to 60° for hexagon-shaped bolts due to their geometry. This constraint may not be satisfied in real-world scenarios when structures are experiencing severe shaking (due to major earthquake) and the bolts get loosened by a relatively large angle above 60°. To relax this constraint, Zhao et al. [29] applied a dual-class single shot detector to localize the bolt and a specific symbol on the bolt simultaneously, where the bounding box locations are used to quantify bolt rotation angle. In this study, the dual-class single shot detector is trained to detect a specific symbol on the bolt. However, different types of bolts may have different symbols. Without excessive training for many different types of bolts, the detector is directly applicable to other types of bolts. Moreover, this method may not be very accurate due to bounding box jittering, which is known as a common issue in many object detection methods. Further, Pan & Yang [30], developed a combined CNN-based detection-tracking approach to automatically localize and continuously track the bolt loosening rotation. Although the study has achieved real-time speed and high accuracy, the KLT tracker [31] employed in this study relies on the existence of visual distinct features (i.e., surface texture information) to be effectively applied. In rare situations, if there exist minimal local distinct features of bolts, it requires additional artificial marks to be added for the algorithm to perform well.

On the other hand, side view-based detection methods localize bolts and quantify bolt loosening as the exposed shank length in images captured from the side view. In this case, the exposed shank length along its longitudinal direction is estimated. In recent years, side view-based bolt loosening detection methods have been established, such as a combined support vector machine and HT method [10,32,33]. These studies were more focused on qualitative evaluation by localizing the bolts in an image and assigning "loosened" or "non-loosened" label to the detected bolts, which may not perform well on bolts loosened by a small amount. This is because a lightly loosened bolt is very similar to a tightened bolt in an RGB image, which may lead to false classification by the vision algorithms. Further, studies on quantification of bolt loosening were conducted. Zhang et al. [34] employed the Faster R–CNN to localize tight and loose bolts, and quantify bolt loosening. While this method shows the capability of localizing and quantify the bolt loosening length, the quantification process is based on 2D images, where the outcomes are sensitive to camera locations and viewing angle. More recently, Zhang & Yuen [35] proposed a bolt loosening quantification method using an orientation-aware bounding box approach, which can address multiple orientations of bolt assemblies in the image. The method uses the aspect ratio of the loosened bolts as the looseness quantification metric. However, the experiments in this study were conducted on bird-eye view images, which means bolts loosened by a small amount may not be effectively captured. Yang et al. [36] trained multiple YOLO detectors to identify loosened bolts. In their study, artificial marks were manually added to the surface of bolts and nuts. Identification of bolt loosening is based on the relative movement of marks on the bolt and nut when the bolt gets loosened. In practical scenarios, adding marks for all the bolts requires excessive human efforts, and becomes time-consuming and difficult particularly for large-scale civil structures with many bolted connections.

In summary, despite of the promising results and progress in bolt loosening detection achieved by the existing studies, the front view-based and side view-based methods as aforementioned were built in the domain of 2D computer vision, and have their respective limitations as previously described. In vision-based structural damage detection, there is a compelling need and trend to develop more

comprehensive damage quantification methods in 3D space. For this purpose, 3D computer vision can be considered as an economical and highly accurate method. Existing investigation of 3D vision methods in structural damage quantification remains in the infancy stage. As for bolt loosening assessment, a very recent study [37] employed a stereo vision-based method to quantify bolt loosening. In this research, an RGB-D camera was employed to acquire RGB images and their corresponding depth maps. A Mask RCNN [38] detector was applied to segment the bolts in an image. The depth map generated by the stereo camera was used to quantify the exposed shank length of bolts. However, the depth map generated by a single pair of stereo cameras typically may contain a large amount of error due to noise and suboptimal background lighting conditions, particularly in a complex environment. Without further optimization, the stereo depth map may be highly noisy and inaccurate, which can greatly limit their capability in real-world applications.

To address these limitations, in this paper, a novel 3D vision-based methodology is developed to automatically detect and quantify bolt loosening at high accuracy and low cost. The proposed pipeline consists of 1) robust, accurate and economical vision-based 3D reconstruction procedures to generate a 3D dense point cloud of the bolted connection, 2) a multi-view object detection method to extract bolts from the 3D point cloud, and 3) a 3D point cloud processing method to quantify bolt loosening. Parameter studies have been conducted to assess the effectiveness of the proposed pipeline. Finally, a real-world implementation of the proposed methodology has been made to a structural column base connection. The results indicate that the proposed methodology can reliably identify and quantify bolt loosening at high accuracy of ~0.5 mm. This is particularly useful to achieve the early identification of the bolts which just get loosened by a small amount, which is more difficult to be detected by the existing 2D vision methods using 2D images. To the
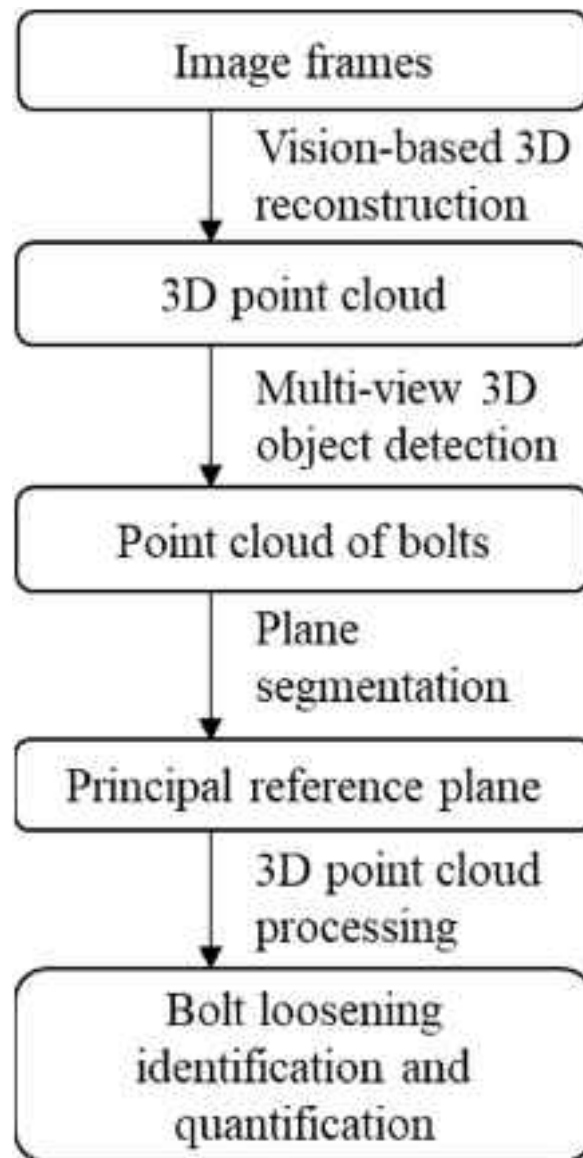

Fig. 1. 3D vision-based bolt loosening assessment procedures.

best of the author's knowledge, the proposed methodology is the first-of-its-kind using 3D vision-based reconstruction, 3D object detection methods, and 3D point cloud processing in a single pipeline for autonomous structural bolt loosening identification and quantification. Existing studies about bolt loosening assessment were developed in 2D computer vision, which rely on 2D images captured at a particular viewing angle. Hence, their evaluation outcomes are sensitive to camera locations. As a comparison, the proposed method incorporates many camera views, and provides more accurate, robust, and comprehensive evaluation outcome of bolt loosening in 3D space. In addition, the proposed methodology yields high accuracy at a much lower cost than the use of high-precision terrestrial laser scanners.

2. Methodology

In this section, the 3D vision-based bolt loosening identification and quantification methodology is presented. It consists of vision-based 3D reconstruction using 3D images, vision-based 3D localization of bolts, and bolt loosening quantification using 3D point cloud processing (Fig. 1). First, 2D image sources are collected by common RGB cameras such as webcams, smartphones cameras, or unmanned aerial vehicles (UAVs). To illustrate the pipeline, a simple structural bolted connection setup with tight and loosened bolts is selected. Other similar experimental setups have been used to examine various vision-based bolt loosening detection methods in recent years (e.g., [10,32–34]; [35]).

2.1. Vision-based 3D reconstruction

The proposed method first involves 3D reconstruction of the structural bolted component. The 3D reconstruction pipeline requires the input images of the structural bolted device to be collected from multiple views. For this purpose, the image collection can be conveniently achieved by using a camera to take photos or record videos. Video recording at high enough resolution (e.g., 1080p, 4 K) and high enough frame rates (e.g., 30 or higher) is usually used to provide abundant camera views at high quality and efficiency for 3D reconstruction. On the other hand, photos of high enough resolution can also be directly collected for 3D reconstruction, without recording videos. The camera can be any consumer-grade cameras which offers reasonably high quality and image resolution (e.g., 1080p, 4 K, or higher), such as smartphone cameras, or economical webcams. In the field applications, a drone or unmanned ground robots equipped with an appropriate RGB camera can be used for efficient data collection. After the data collection, 3D reconstruction procedures are applied to the collected images. It should be noted that, in order to reconstruct a 3D object accurately, it is required to collect a sufficient number of images at various camera views. There are generally three criteria for camera view selection as summarized in the existing literature [39,40] including: 1) the images collected should cover the entire object of interest, 2) a minimum of 50% image overlap between two adjacent views, and 3) images should be sharp and have a sufficient resolution. The 3D reconstruction procedures have 4 essential steps, as shown in Fig. 2.

In step 1, image-to-image association is established from input images to determine a scene graph. During this step, key feature points of images are detected by scale-invariant feature transform [41]. Next, the images that share similar features will be connected by the vocabulary tree (VocTree) algorithm [42]. After the aforementioned steps, a scene graph can be obtained where all the image pair connections are established.

In step 2, the structure-from-motion algorithm is implemented to estimate the sparse cloud. In this case, the KD-tree algorithm [43] is first implemented to search the closest matched features in the image pairs. The camera matrices for any image pair can be determined using epipolar geometry [39]. Further, the estimated camera matrices for all image pairs can be utilized to determine a sparse point using the triangulation methods [44]. Finally, the camera parameters and the initial point cloud estimated are finetuned
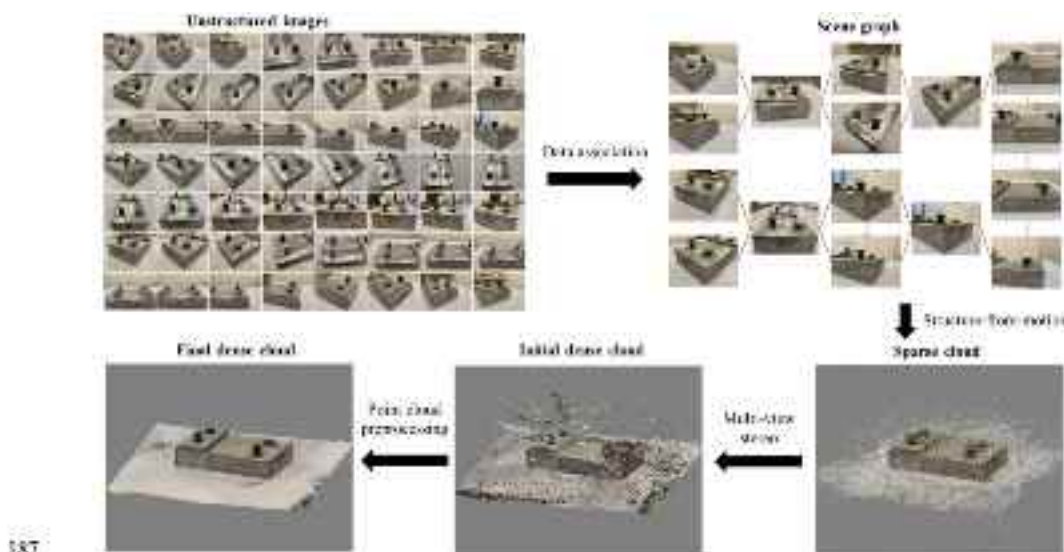

Fig. 2. Vision-based 3D reconstruction.

by the bundle adjustment (BA) method [45] to minimize the reprojection error. In the end, a refined sparse cloud is obtained for further processing.

In step 3, a multi-view stereo algorithm is implemented to generate a dense 3D point cloud using the sparse 3D point cloud and the images from the scene graph. During this process, the depth value at every pixel location of the selected images is computed using the semi-global matching method [46]. The dense point cloud determined has sub-pixel accuracy [46].

In step 4, the dense point cloud preprocessing is implemented to reduce noise. It is necessary to reduce the effects of noise before the bolt loosening quantification procedures to alleviate the effects of these outlier points. In this study, two denoising techniques including the statistical outlier removal [47] and radius-based outlier removal algorithms are considered. Detailed description and implementation of the noise reduction techniques can be referred to the Open3D, which is an open-source library for 3D data processing [48].

2.2. Automatic 3D bolt loosening quantification

Bolt loosening inspection procedures consist of vision-based 3D bolt localization, followed by the quantification of loosened bolts using 3D point cloud processing.

2.2.1. 3D localization of structural bolts

This section describes the proposed vision method for automatic 3D bolt localization in the 3D point cloud. This step is crucial for two reasons. 1) It provides regions of interest to allow the bolt quantification procedures to be effectively applied. 2) It eliminates the need in processing the entire point cloud in the subsequent quantification steps, thus greatly enhancing the computational efficiency.

Many popular existing point cloud processing algorithms such as PointPillar [49], are generally trained to process organized LiDAR point clouds, and cannot be directly used to process unorganized photogrammetry point cloud generated using vision-based 3D reconstruction. Meanwhile, the PointPillar algorithm does not have satisfactory localization accuracy. As a result, other parts of point clouds (e.g., noise, washers, steel plates) may be wrongly classified as point clouds of bolts by these algorithms, leading to a failure in identifying loosened bolts. Lastly, as these algorithms are generally very computationally demanding, they are usually used to process relatively sparse point clouds (as 3D object detection in autonomous driving), which may hamper their deployment to dense point cloud processing.

For these reasons, this paper proposes a more accurate, and simple yet effective 3D bolt localization based on multi-view CNN-based vision object detector. The generated 3D point cloud is automatically rendered using virtual camera views, leading to multiple 2D images. The virtual camera path can be set based on the ground reference plane defined during 3D reconstruction. After that, a pretrained YOLOv3-tiny detector for bolt localization is adopted from [30] to recognize bolts in the rendered images. The detected bounding boxes are then fused to achieve the 3D localization of bolts, using multiple virtual camera views, as shown in Fig. 3. For illustration purpose, Fig. 3 shows the implementation of a two-view detection. More than two views can be used.

2.2.2. Structural bolt loosening quantification

This section presents the procedures to quantify bolt loosening, which is defined as the distance between the bottom of the bolt head and the supporting surface underneath the bolt, as illustrated in Fig. 4. A sub-cloud convex hull (SCCH) method is proposed for automated quantification of bolt loosening, which is defined as the exposed bolt shank length. The detailed implementation procedures are described herein:

- After the 3D localization of bolts, the direction of the longitudinal axis of the bolt is first determined. This is achieved by plane segmentation algorithm, MSAC (Torr & Zisserman, 2000). During this step, the first principal plane will be identified as the bolt head plane. This plane is deemed as a reference plane for bolt loosening quantification. The normal vector of the reference plane will be retrieved, which represent the longitudinal direction of the bolt.
- A series of sub-cloud sampling is implemented (with a relatively small step) along the direction of the reference plane normal, for each point cloud within each 3D boundary box. Within the sub-cloud, the points will be projected to the reference plane. Further, the convex hull of the projected points belonging to each cuboid boundary will be determined, respectively. The area of each convex hull will be calculated. The convex hull represents the estimated cross section.
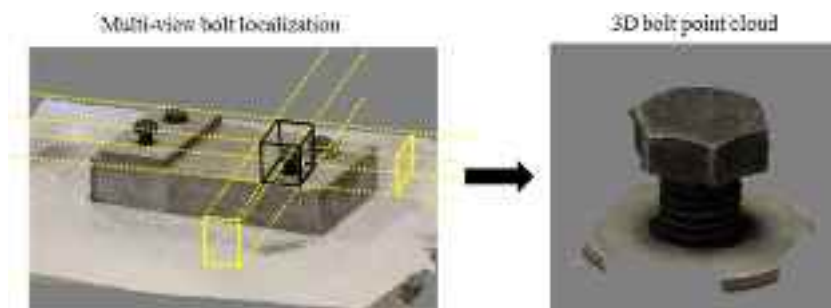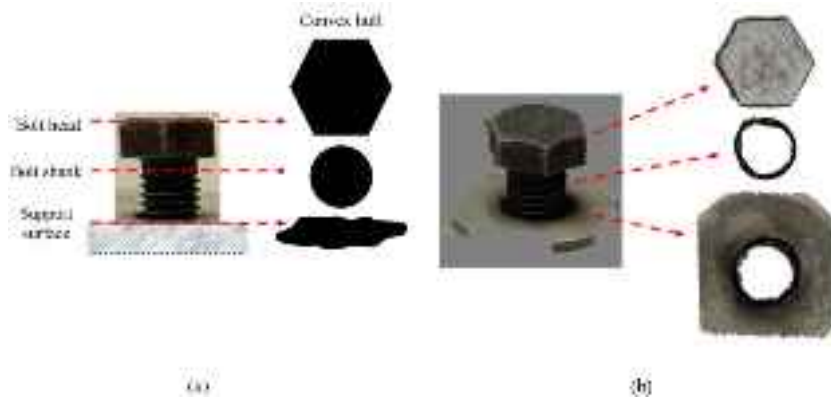

Fig. 3. 3D bolt localization.

Fig. 4. Cross section of bolt: (a) idealized; (b) from the 3D point cloud.

- The above sampling processes are repeated. The region of each convex hull will be consistently checked for each sub-cloud. The bolt loosening length can be estimated considering the two following criteria. a) If a bolt is tight, there will be a relatively consistent convex hull area at the beginning (i.e., bolt head region), and then a sudden increase of the convex hull when the sampling reaches the supporting element below the bolt. b) If a bolt is loosened, there will be a decrease of the convex hull region, when the sampling crosses the boundary between the bolt head and the exposed shank (due to loosening). Using the information described above, the bolt loosening can be recognized and quantified.

The pseudocode for the quantification procedures is provided in Table 1. Overall, compared to the traditional 2D vision-based methods, the proposed 3D vision-based localization and quantification method is fully automated during the image and 3D point cloud data processing, without human intervention to identify a reference plane to identify bolt loosening.

Table 1
Bolt loosening quantification algorithm.

% Input
Require: Vision-based object detector for bolt localization
Require: Step size for point cloud sampling
Require: The point cloud of the bolted component
% Search for reference planes
Determine 2D rendered images using virtual cameras from the 3D point cloud
Apply the pretrained CNN (i.e., YOLOv3-tiny in this study) to localize bolts
Concatenate the predicted bounding boxes to generate the 3D localization boundary
Crop and save the point clouds within all the 3D boundary
% Search for reference planes
For each point cloud within each 3D boundary
Reference plane = Apply MASC algorithm to detect the principal plane
Record plane normal
End for
% Quantification of bolt loosening length
For each point cloud within each 3D boundary
% Stepping along the direction of the normal vector of the reference plane
While the current step is within the range of the 3D boundary:
Sample a sub-cloud using a small predefined step size
Project all sub-cloud points on the reference plane
Determine and record the convex hull information of the projected points
End while
End for
% Output: Bolt loosening quantification
Detect changes in convex hull (cross section):
Bolt looseness = Distance between the bottom of the bolt head and the supporting surface underneath the bolt

3. Experiments and parameter studies

To investigate the effectiveness and potential limitations of the proposed methodology, experiments and parameter studies were carried out under different lighting and noise conditions. First, a benchmark model is generated using 114 images of 4 K resolution collected by a consumer grade RGB camera at a distance between 0.2 m and 1.5 m to the bolted connection, under a normal lighting condition. The selection of distance is dependent on the image resolution and the required 3D reconstruction accuracy. In general, if the image resolution is lower or the required 3D reconstruction accuracy is higher, the distance should be smaller, and vice versa. The zoom-in or zoom-out function is not used to maintain the original quality of image frames. The proposed pipeline is applied to the 3D point cloud to quantify bolt loosening, which is compared with the manual measurement. The result show that the estimated length (i. e., 14.8 mm) is very close to the ground truth value, 15 mm, as measured by a Fowler digital caliber (with the measurement accuracy of 0.02 mm). This provides a benchmark scenario for comparisons in the parameter studies that follow. The number of images used in the parameter studies stays constant, i.e., 114. During the parameter studies, the original collected images were augmented with the effects as described in Section 3.1 and Section 3.2. These images will be used to reconstruct the 3D point cloud in the respective scenario. In the end, the performance of the proposed method will be evaluated on each 3D point cloud. There exist many different performance metrics such as the root mean square error (RMSE) and mean absolute deviation (MAD) in machine learning applications (e.g., [50, 51]). In this study, the MAV metric is selected. Then, the error percentage is defined as follows,

$$Error = \frac{|Ground\ truth - Prediction|}{Ground\ truth} \times 100\% \tag{1}$$

where the ground truth value is measured by a Fowler digital caliber with a precision of 0.02 mm, and the prediction value is estimated by the proposed method.

3.1. Effects of lighting conditions

Different lighting conditions can occur in the real world. Cameras produce different pixel values of an image of the same object under different lighting conditions. Besides, the image noise is affected by the lighting condition for modern cameras in general. For example, the RGB images captured by modern cameras typically contain more noise in low light condition, than images captured by the same camera under a reasonably good lighting condition. As a result, the lighting condition can have a big impact on the 3D reconstruction result. For this reason, this section presents a parameter study on three different lighting conditions. Fig. 5 shows sample images under three different lighting conditions. Evidently, the images captured in low light condition contain much more noise.

Fig. 6 shows the 3D point clouds generated under the three lighting conditions. The 3D point clouds are well reconstructed in moderate and high light condition as shown in Fig. 6 (a) and (b), while contain more missing regions and noise in a low light condition as presented in Fig. 6 (c). Nevertheless, the proposed pipeline can be effectively applied to all three cases to identify bolt loosening. This is because the use of convex hull in the proposed pipeline can omit the small local missing parts, while still providing an effective measurement based on the reconstructed regions. In addition, Table 2 presents a summary of quantification results under different lighting conditions. It has been shown that the proposed methodology performs well in both moderate light and high light condition, while the accuracy decreases in the low light condition with an error of ~2 mm. Overall, the accuracy is deemed satisfactory in all the lighting conditions investigated.

3.2. Effects of image noise

Images captured by modern cameras inevitably contain noise due to different factors. In this section, the effect of three different types of noise (i.e., detector noise, salt and pepper noise, and speckle noise), and compound noise (i.e., the combination of the three basic noise types), is assessed.

3.2.1. Description of the detector noise

The detector noise is a common type of noise that appears in almost all images, owing to the discrete nature of radiation. This type of noise can be generally modeled with a Gaussian distribution with zero mean and a specified dispersion (i.e., standard deviation or variance). The pixel value in an image is corrupted by adding the noise value sampled from the distribution to its original value. Three variance values of the Gaussian distribution are investigated including 0.01, 0.0225 and 0.04.
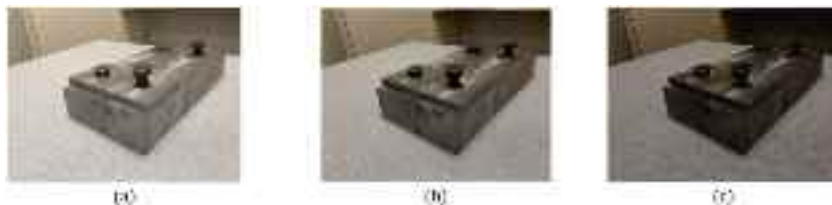

Fig. 5. Sample images acquired under 3 different lighting conditions: (a) high light, (b) moderate light, (c) low light.

Fig. 6. 3D point clouds generated based on images under three lighting conditions: (a) high light, (b) moderate light, (c) low light.

Table 2
Bolt loosening quantification under three different lighting conditions.

Scenarios	Estimated values [mm]	Error
Low lighting	13.2	12%
Moderate lighting	14.7	2%
High lighting	14.8	1.3%

3.2.2. Description of the salt and pepper noise

The salt and pepper noise a type of data drop-out noise, possibly due to errors in the data transmission. If a pixel value is affected by this type of noise, it will be either set to the maximum value or zero. Other unaffected pixels remain unchanged. In this section, the noise ratio is defined as the percentage of affected pixels. These affected pixels will be randomly selected from the image where half of them is assigned to zero and half of them is assigned to the maximum value. Three noise ratios are examined including 0.01, 0.025 and 0.05.

3.2.3. Description of the speckle noise

The speckle noise is another type of data drop-out noise, which is attributed to imaging sensors affected by different environmental conditions during image acquisition. This noise is a multiplicative noise, which can be described by $J = I + n*I$, where J refers to the corrupted image pixel, I is the original image pixel, n is sampled from a uniform distribution with zero mean and a predefined variance. Three variance values of the uniform distribution are investigated including 0.01, 0.0225 and 0.04.
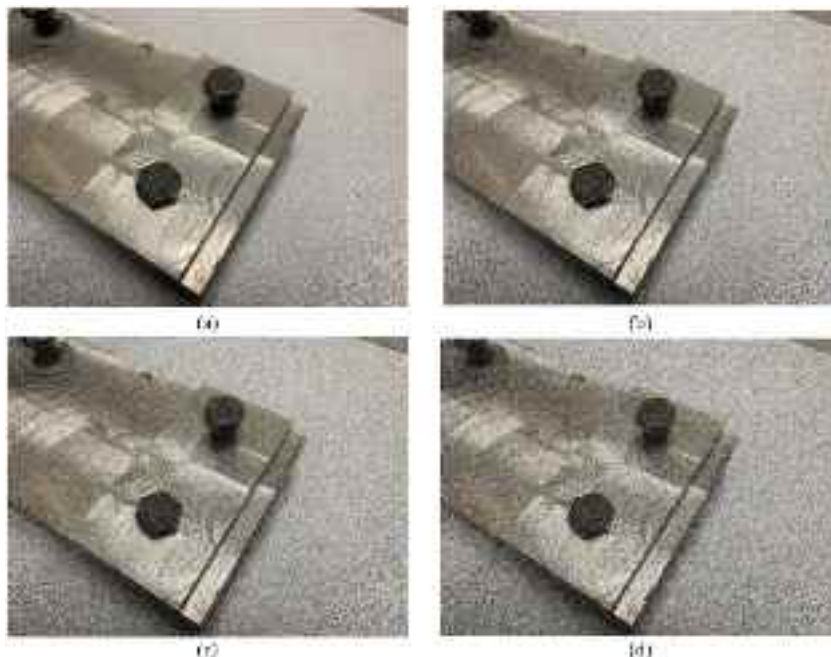

Fig. 7. Image of bolts: (a) without manually added noise, (b) with low noise, (c) with moderate noise, (d) with high noise.

3.2.4. Description of the compound noise

In this section, the compound noise is defined as the combination of the Gaussian noise, salt and pepper noise, and speckle noise. The magnitude of compound is determined by summing up the former three types of noise at three different levels, respectively. Therefore, the three levels of the compound noise include: low (Gaussian variance = 0.01, Salt and pepper noise ratio = 0.01, Speckle noise variance = 0.01), moderate (Gaussian variance = 0.0225, Salt and pepper noise ratio = 0.025, Speckle noise variance = 0.0225), high (Gaussian variance = 0.04, Salt and pepper noise ratio = 0.05, Speckle noise variance = 0.04).

3.2.5. Summary and discussion about the noise effects

For illustration, Fig. 7 presents a comparison of sample images under the three levels of compound noise described in Section 3.2.4. It is shown that images are greatly corrupted, particular in Fig. 7 (c) and (d).

Fig. 8 shows a close-up comparison of the 3D point cloud generated using the images corrupted by three levels of compound noise, respectively. In addition, Table 3 summarizes bolt loosening quantification for the selected bolt under different noise scenarios. Overall, the proposed pipeline identifies all the loosened bolts, and is relatively robust against different types of noise at all levels. The worst-case scenario corresponds the compound noise at the highest level, where error is the maximum among all the cases. This is because the added noise smoothens the transition region from the bolt cap to the bolt shank in the 3D point cloud. In the 3D point cloud obtained using original images without added noise, the transition from the bolt cap to the bolt shank is abrupt and clear. In the 3D point cloud obtained using corrupted images, the transition from the bolt cap to the bolt shank is contaminated by noise, and becomes smoother. In general, as the noise level increases, the quantification accuracy decreases, which applies to all the noise types investigated. The maximum error of 2.8 mm is observed under the highest level of the compound noise.

4. Implementation and results

To further demonstrate the effectiveness of the proposed methodology, the proposed 3D vision-based bolt loosening length quantification pipeline is implemented on a real engineering scenario where a structural steel column is connected to the ground with bolts. First, a 4 K video was recorded by a consumer grade drone for about 20 s. Then, about 137 images were extracted from the video. Next, the 3D reconstruction procedures were applied to the images to determine a dense 3D point cloud. The vision-based 3D reconstruction procedures were first implemented to generate the initial dense point cloud in Meshroom and Metashape with the Python API. Next, the statistical outlier removal and radius outlier removal algorithms were established in Open3D to significantly reduce the effect of outliers. The final reconstructed 3D point cloud is shown in Fig. 9. The 3D point cloud of the bolted connection is successfully reconstructed. Once the 3D point cloud is obtained, the multi-view 3D bolt localization method, as described in Section 2.2, which consists of multi-view CNN-based bolt localization, was applied to identify bolts from the 3D point cloud. In this experiment, the two bolts were successfully identified by detector in both views, as shown in Fig. 9.

Further, the proposed SCCH method was applied to recognize and quantify the bolt loosening. For the loosened bolt, the ground truth bolt loosening length was set up at 16 mm (with a help of the digital caliber), while the estimated value is 15.77 mm. For the non-loosened bolt, the estimated value is 0.12 mm. This shows that the quantification error for both two bolts is less than 0.5 mm compared to the ground truth values. Therefore, this demonstrated that the proposed 3D vision-based bolt loosening quantification method can be effectively used to identify lightly loosened bolts with the exposed shank length as low as ~0.5 mm. This is particularly useful to deal with the bolts which just get loosened by a small amount, which is difficult to be recognized by the existing 2D vision methods using 2D images captured at only a particular viewing angle. Besides, the proposed pipeline is fully automated during the image and point cloud data processing steps, without human intervention to select a particular reference angle or plane, which is more superior to the existing side view-based 2D vision methods. Overall, the implementation demonstrates the applicability and effectiveness of the proposed bolt loosening inspection pipeline. When integrated with drones or other ground robots, the proposed pipeline can make the inspection process more efficient and convenient for typical bolted connections in the field.

5. Conclusions

Structural bolts are widely utilized to connect structural components. The clamping forces generated in bolts are highly affected by bolt loosening. Identification of bolt loosening is of vital importance to maintain the structural integrity and performance. This paper presents a 3D vision-based bolt loosening quantification pipeline, which aims to address the limitations of existing side view-based quantification methods built upon 2D computer vision. The proposed method consists of vision-based 3D reconstruction, a newly proposed multi-view detection method for bolt localization, and a newly developed SCCH point cloud processing algorithm to recognize and quantify bolt loosening. By localizing the bolt regions of interest, the multi-view 3D bolt localization method significantly reduces the size of point clouds to be further processed by the SCCH method. The SCCH method leverages the robustness of CNN-based object detection and point cloud processing techniques to provide an automated and accurate bolt loosening quantification. Several conclusions and novelties of the paper are summarized herein: 1) To the best of the author's knowledge, the proposed methodology is the first-of-its-kind to develop an end-to-end 3D vision-based pipeline to quantify bolt loosening length using 3D vision and 3D point cloud processing algorithms. 2) The proposed pipeline achieves the high accuracy (~0.5 mm) at low cost (with consumer grade cameras). 3) The proposed pipeline does not rely on the manual selection of a viewing angle, as compared to the existing 2D vision-based methods which require 2D images captured at a specific viewing angle. Overall, this paper has demonstrated the proposed method can be effectively applied to automate the bolt loosening inspection process under reasonable lighting conditions, which has the potential to improve the level of autonomy and efficiency, as well as lowering the inspection cost at large scale when integrated with inspection robots.

Fig. 8. Close-up view of the 3D point cloud: (a) from the original collected images without added noise, (b) from the images with the highest compound noise.

Table 3
Bolt loosening quantification under different noise scenarios.

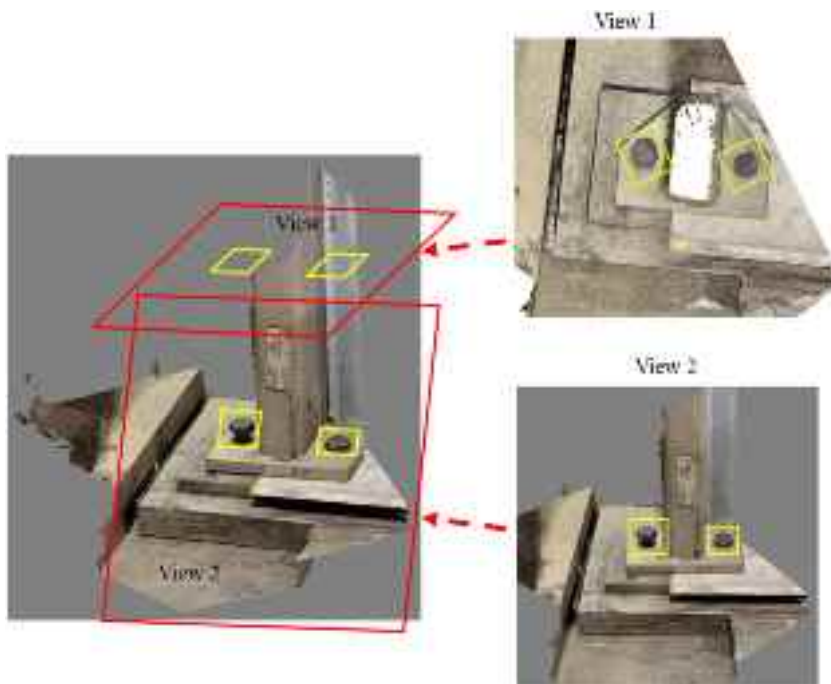| Scenarios	Noise level	Estimated values [mm]	Error
Original image with no added noise	None	14.8	1.3%
Detector noise	Low	14.2	5.3%
	Moderate	13.6	9.3%
	High	12.9	14.0%
Salt and pepper noise	Low	14.2	5.3%
	Moderate	14.1	6.0%
	High	13.7	8.7%
Speckle noise	Low	14.3	4.7%
	Moderate	13.3	11.3%
	High	13.2	12.0%
Compound noise	Low	13.9	7.3%
	Moderate	13.3	11.3%
	High	12.2	18.7%

Fig. 9. Structural bolt localization in the 3D point cloud of the steel column base connection.

In addition, limitations of the present study and further research are also discussed. a) The 3D vision methods are developed for bolt loosening assessment in the present study. The concept of 3D vision-based methods should be further extended to identify and quantify other types of structural damages such as concrete cracks and spalling. b) The drone used in this study was manually controlled. In the future, autonomous image collection algorithm will be developed and validated for drones, to further enhance the degree of autonomy during the data collection process.

Author statement

Xiao Pan: conceptualization, methodology, computational analysis, experimental validation, original draft preparation, manuscript review. T.Y. Yang: manuscript review, supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

References

[1] O.S. Salawu, Detection of structural damage through changes in frequency: a review, Eng. Struct. 19 (9) (1997) 718–723.
[2] Y. An, E. Chatzi, S.H. Sim, S. Laflamme, B. Blachowski, J. Ou, Recent progress and future trends on damage identification methods for bridge structures, Struct. Control Health Monit. 26 (10) (2019) e2416.
[3] J. Yang, F.K. Chang, Detection of bolt loosening in C– C composite thermal protection panels: II. Experimental verification, Smart Mater. Struct. 15 (2) (2006) 591.
[4] T. Wang, G. Song, S. Liu, Y. Li, H. Xiao, Review of bolted connection monitoring, Int. J. Distributed Sens. Netw. 9 (12) (2013), 871213.
[5] E. Sevillano, R. Sun, R. Perera, Damage detection based on power dissipation measured with PZT sensors through the combi- nation of electro-mechanical impedances and guided waves, Sensors 16 (5) (2016) 639.
[6] Y. Xia, B. Chen, S. Weng, Y.Q. Ni, Y.L. Xu, Temperature effect on vibration properties of civil structures: a literature review and case studies, J. Civ. Struct. Health Monit. 2 (1) (2012) 29–46.
[7] J. Li, J. Deng, W. Xie, Damage detection with streamlined structural health monitoring data, Sensors 15 (4) (2015) 8832–8851.
[8] T.C. Huynh, J.T. Kim, Quantification of temperature effect on impedance monitoring via PZT interface for prestressed tendon anchorage, Smart Mater. Struct. 26 (12) (2017), 125004.
[9] T.C. Huynh, J.T. Kim, RBFN-based temperature compensation method for impedance monitoring in prestressed tendon anchorage, Struct. Control Health Monit. 25 (6) (2018) e2173.
[10] L. Ramana, W. Choi, Y.J. Cha, Fully automated vision-based loosened bolt detection using the Viola–Jones algorithm, Struct. Health Monit. 18 (2) (2019) 422–434.
[11] B.F. Spencer Jr., V. Hoskere, Y. Narazaki, Advances in computer vision-based civil infrastructure inspection and monitoring, Engineering 5 (2) (2019) 199–222.
[12] X. Pan, Three-dimensional Vision-Based Structural Damage Detection and Loss Estimation–Towards More Rapid and Comprehensive Assessment, Doctoral dissertation, University of British Columbia, 2022, https://doi.org/10.14288/1.0422384.
[13] D. Tan, J. Li, H. Hao, Z. Nie, Target-free vision-based approach for modal identification of a simply-supported bridge, Eng. Struct. 279 (2023), 115586.
[14] X. Pan, T.Y. Yang, Y. Xiao, H. Yao, H. Adeli, Vision-based real-time structural vibration measurement through interactive deep-learning-based detection and tracking methods, Eng. Struct. 281 (2023), 115676.
[15] X. Liang, Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization, Comput. Aided Civ. Infrastruct. Eng. 34 (5) (2019) 415–430.
[16] X. Pan, T.Y. Yang, Postdisaster image-based damage detection and repair cost estimation of reinforced concrete buildings using dual convolutional neural networks, Comput. Aided Civ. Infrastruct. Eng. 35 (5) (2020) 495–510.
[17] S. Tavasoli, X. Pan, T.Y. Yang, Real-time autonomous indoor navigation and vision-based damage assessment of reinforced concrete structures using low-cost nano aerial vehicles, J. Build. Eng. (2023), 106193.
[18] C.M. Yeum, S.J. Dyke, Vision-based automated crack detection for bridge inspection, Comput. Aided Civ. Infrastruct. Eng. 30 (10) (2015) 759–770.
[19] X. Kong, J. Li, Image registration-based bolt loosening detection of steel joints, Sensors 18 (4) (2018) 1000.
[20] X. Pan, T.Y. Yang, 3D Vision-based Out-of-plane Displacement Quantification for Steel Plate Structures Using Structure-from-motion, Deep Learning, and Point-cloud Processing, Computer-Aided Civil and Infrastructure Engineering, 2022.
[21] N. Wang, Q. Zhao, S. Li, X. Zhao, P. Zhao, Damage classification for masonry historic structures using convolutional neural networks based on still images, Comput. Aided Civ. Infrastruct. Eng. 33 (12) (2018) 1073–1089.
[22] N. Wang, X. Zhao, P. Zhao, Y. Zhang, Z. Zou, J. Ou, Automatic damage detection of historic masonry buildings based on mobile deep learning, Autom. ConStruct. 103 (2019) 53–66.
[23] X. Kong, J. Li, Vision-based fatigue crack detection of steel structures using video feature tracking, Comput. Aided Civ. Infrastruct. Eng. 33 (9) (2018) 783–799.
[24] J.H. Park, T.C. Huynh, S.H. Choi, J.T. Kim, Vision-based technique for bolt-loosening detection in wind turbine tower, Wind Struct. 21 (6) (2015) 709–726.
[25] J.H. Park, T. Kim, J. Kim, Image-based bolt-loosening detection technique of bolt joint in steel bridges, in: 6th International Conference on Advances in Experimental Structural Engineering 11th International Workshop on Advanced Smart Materials and Smart Structures Technology, 2015, pp. 1–2. Champaign, IL.
[26] P.E. Hart, R.O. Duda, Use of the Hough transformation to detect lines and curves in pictures, Commun. ACM 15 (1) (1972) 11–15.
[27] T.C. Huynh, J.H. Park, H.J. Jung, J.T. Kim, Quasi- autonomous bolt-loosening detection method using vision-based deep learning and image processing, Autom. ConStruct. 105 (2019), 102844.

[28] Q.B. Ta, J.T. Kim, Monitoring of corroded and loosened bolts in steel structures via deep learning and hough transforms, Sensors 20 (23) (2020) 6888.

[29] X. Zhao, Y. Zhang, N. Wang, Bolt loosening angle detection technology using deep learning, Struct. Control Health Monit. 26 (1) (2019), e2292.

[30] X. Pan, T.Y. Yang, Image-based monitoring of bolt loosening through deep-learning-based integrated detection and tracking, Comput. Aided Civ. Infrastruct. Eng. (2022) 1–16.

[31] C. Tomasi, T. Kanade, Detection and tracking of point, Int. J. Comput. Vis. 9 (1991) 137–154.

[32] Y.J. Cha, K. You, W. Choi, Vision-based detection of loosened bolts using the Hough transform and support vector machines, Autom. ConStruct. 71 (2016) 181–188.

[33] L. Ramana, W. Choi, Y.J. Cha, Automated vision-based loosened bolt detection using the cascade detector, in: C. Walber, E. Wee Sit, P. Walter, S. Seidlitz (Eds.), Sensors and Instrumentation, 5, Springer, 2017, pp. 23–28.

[34] Y. Zhang, X. Sun, K.J. Loh, W. Su, Z. Xue, X. Zhao, Autonomous bolt loosening detection using deep learning, Struct. Health Monit. 19 (1) (2020) 105–122.

[35] Y. Zhang, K.V. Yuen, Bolt damage identification based on orientation-aware center point estimation network, Struct. Health Monit. 21 (2) (2022) 438–450.

[36] X. Yang, Y. Gao, C. Fang, Y. Zheng, W. Wang, Deep learning-based bolt loosening detection for wind turbine towers, Struct. Control Health Monit. 29 (6) (2022) e2943.

[37] H. Sohn, J. Chung, Detection and quantification of bolt loosening using RGB-D camera and Mask R-CNN, Smart Struct. Syst. Int. J. 27 (5) (2021) 783–793.

[38] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969. Venice, Italy.

[39] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge university press, 2003.

[40] Y.F. Liu, X. Nie, J.S. Fan, X.G. Liu, Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction, Comput. Aided Civ. Infrastruct. Eng. 35 (5) (2020) 511–529.

[41] D.G. Lowe, Distinctive image features from scale-invariant key- points, Int. J. Comput. Vis. 60 (2) (2004) 91–110.

[42] D. Nister, H. Stewenius, Scalable recognition with a vocabulary tree, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) 2, Ieee, 2006, June, pp. 2161–2168.

[43] M. Muja, D.G. Lowe, Fast approximate nearest neighbors with automatic algorithm configuration, VISAPP (1) 2 (331-340) (2009) 2.

[44] R.I. Hartley, P. Sturm, Triangulation, Comput. Vis. Image Understand. 68 (2) (1997) 146–157.

[45] B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon, Bundle adjustment—a modern synthesis, in: International Workshop on Vision Algorithms, Springer, Berlin, Heidelberg, 1999, September, pp. 298–372.

[46] H. Hirschmuller, Stereo processing by semiglobal matching and mutual information, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2007) 328–341.

[47] A.C. Carrilho, M. Galo, R.C. Santos, Statistical outlier detection method for airborne lidar data, in: International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2018.

[48] Q.Y. Zhou, J. Park, V. Koltun, Open3D: A Modern Library for 3D Data Processing, 2018 *arXiv preprint arXiv:1801.09847*.

[49] A.H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, O. Beijbom, PointPillars: fast encoders for object detection from point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12697–12705.

[50] A. Botchkarev, A new typology design of performance metrics to measure errors in machine learning regression algorithms, Interdiscipl. J. Inf. Knowl. Manag. 14 (2019) 45–76.

[51] M.Z. Naser, A.H. Alavi, Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences, Architect. Struct. Construct. (2021) 1–19.