Concrete Crack Quantification Using Voxel-Based Reconstruction and Bayesian Data Fusion

Chaobo Zhang , *Member, IEEE*, Maziar Jamshidi , Chih-Chen Chang , Xiaojun Liang, *Member, IEEE*, Zhiwen Chen , *Member, IEEE*, and Weihua Gui

Abstract—Concrete cracks are one of the most apparent indicators for possible structural deterioration and need to be periodically inspected. However, for current image-based automated crack inspection techniques, accurate and detailed crack quantification and assessment remain a challenging task. Most of these techniques require high-quality input images, which may be difficult to ensure in practice. Besides, simply merging crack detections from multiple images to generate a large crack map may result in an inaccurate outcome for crack severity assessment. In this article, a novel crack quantification framework is proposed to identify complete crack geometric properties utilizing a set of unordered inspection images. To realize this, cracks in images are detected by an instance segmentation convolutional neural network. Subsequently, the crack segmentations from multiple separate images are systematically aggregated through voxel-based reconstruction and Bayesian data fusion. This framework outputs a crack model that can retrieve accurate geometric properties of each crack segment by recognizing the crack's inherent branching patterns. The capability and performance of the proposed crack quantification framework are validated on cracked concrete specimens in a laboratory setting. Also, a field test on a cracked concrete wall was carried out using images captured by a UAV to demonstrate the efficacy of the proposed framework in practical conditions.

Index Terms—Bayesian data fusion, crack quantification, instance segmentation network, voxel-based reconstruction.

Chaobo Zhang and Xiaojun Liang are with the Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: zhangchb@pcl.ac.cn; liangxj@pcl.ac.cn).

Maziar Jamshidi is with the Department of Civil Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: maziar.jamshidi@ucalgary.ca).

Chih-Chen Chang is with the Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: cechang@ust.hk).

Zhiwen Chen and Weihua Gui are with the Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen 518066, China, and also with the School of Automation, Central South University, Changsha 410017, China (e-mail: zhiwen.chen@csu.edu.cn; gwh@csu.edu.cn).

I. INTRODUCTION

CONCRETE structures need to be periodically inspected to assess their current functional state, predict their future condition, and make informed decisions on their maintenance and rehabilitation. One important inspection item for concrete structures is to investigate whether cracks have appeared or propagated in their critical components [1]. According to structural inspection manuals, a crack is defined as a linear fracture in concrete and its type, size, orientation, and location need to be recorded to describe its seriousness [2], [3]. Conventionally, manual visual inspection has been the primary method for gathering information about cracks. It is however reckoned that manual visual inspection could not only be labor-demanding and time-consuming but also lead to rather subjective and inaccurate results [4]. Hence, there is a need for more reliable, and automated methods for detecting and characterizing cracks for concrete structures.

Owing to the prevalence of low-cost and high-quality digital cameras, image-based techniques are vastly incorporated into crack inspection tasks. In an image-based technique, in order to obtain geometrical properties (e.g., width and length) of cracks, pixels that form the cracks need to be extracted with high precision [5]. This process is called segmentation and has been broadly developed and utilized for crack inspection. The crack segmentation output of an inspection image is a 2-D binary mask where each pixel is labeled as either part of a crack or not. Existing crack segmentation techniques can be divided into two groups: methods that are based on manually-selected features and those recent methods that utilize deep learning for feature extraction. Techniques in the first group detect crack pixels by applying a threshold or a machine learning classifier to handcrafted image features [6]. The accuracy of these techniques, however, depends on the quality of the features that are manually selected. Therefore, their performance may be hampered for actual images taken under varying environmental conditions [7]. On the other hand, techniques based on deep learning could extract high-level features directly from raw images by using large-scale datasets to train convolutional neural networks

(CNNs) [8]. Researchers in several studies (e.g., [9], [10]) have utilized the power of CNNs for crack segmentation by treating cracks as distinct objects in inspection images. Results of those studies demonstrated that CNN-based approaches are less sensitive and dependent on image conditions than the traditional crack segmentation methods in the first group.

The next stage is to extract the physical properties of the identified cracks from the crack masks. To transform the pixel domain data into the physical domain, a typical approach is to employ the pinhole camera model where the camera must be directed perpendicularly to the concrete surface at a measured distance [11], [12]. This method requires special instruments operating under controlled conditions. Also, they are usually applicable to concrete surfaces with specific geometric shapes. These conditions could be difficult to meet when images are captured by a nonstationary camera moving at varying distances from structural surfaces. Images taken from the camera mounted on an unmanned aerial vehicle are examples of such situations. To eliminate the need for special camera settings, Valença et al. [13], [14] utilized auxiliary equipment, such as predefined pattern boards and terrestrial laser scanners to perform image rectification and calculate the spatial resolution for crack measurement. In another study, Jahanshahi et al. [15] reconstructed 3-D models from images to carry out perspective correction and measure cracks on flat concrete surfaces. Similarly, Liu et al. [16], [17] proposed a combination of manual feature-based crack segmentation and 3-D scene reconstruction to locate the 3-D positions of the edges of thin cracks. However, their method simply stitches segmentations from multiple images to assess a crack's severity. Such an indiscriminate combination of the segmentation information without a systematic way to correct for false detections may render the crack quantification results unreliable. Besides, in inspection videos or images taken by a robotic platform, it is likely that the same scene is captured from many different view angles and distances generating a large volume of unordered and complex overlapping visual data [18]. In such a situation, the shared information among the entire collected image set can be leveraged to filter false segmentations.

In this article, a novel crack quantification framework is proposed that retrieves detailed crack properties from a set of unordered inspection images by incorporating CNN-based segmentation, voxel-based reconstruction and Bayesian data fusion. Voxels are volumetric cubes arranged in a 3-D grid and are analogous to pixels in 2-D images. Constructing crack models in the form of voxels rather than a cluster of point clouds makes it possible to apply image processing techniques in 3-D space and significantly reduces the data storage size of crack information [19]. In this article, unlike the previous 3-D reconstruction-based studies, the branching patterns of cracks are also recognized in the quantification process. Given the geometric complexity of a crack pattern, which usually consists of several segments connected at branching points, each crack segment should be analyzed individually to get its length, width, and orientation [20], [21]. Moreover, in the proposed framework, to increase the precision and robustness of the overall quantification process, Bayesian data fusion is utilized to systematically aggregate crack masks from multiple images. Finally, to understand a crack mode

and severity, a crucial piece of information is its relation with the underlying structural component. Therefore, in a comprehensive assessment framework, it is necessary to associate the locations and properties (e.g., orientation) of identified cracks with the structural components to be able to interpret their significance [22]. In this article, a procedure is laid out that employs as-designed models of the inspected structure to assign detected cracks to the underlying members with any shapes and types.

The following are the main contributions of the proposed article.

1) A new framework is laid out to retrieve detailed crack properties from a set of inspection images. Compared to the existing methods, the proposed technique can perceive branching patterns of cracks in 3-D voxel space. It can also associate the detected crack properties with the underlying structural components which is a crucial piece of information when it comes to crack severity assessment.
2) A crack instance segmentation CNN is developed which can predict crack mask quality. The predicted quality score plays a key role in the process of accurate crack width quantification.
3) A Bayesian data fusion technique is proposed whereby crack segmentation outcomes from multiple images are systematically aggregated. The novelty of the proposed method is that accurate segmentations can favorably skew the final outcome. Application of this technique enhances the quantification accuracy of size, orientation, and location of each crack segment.

The rest of the article is organized as follows. Details of the research methodology are described in Section II. In Section III, experimental results are presented to validate the capability of the proposed techniques. Finally, Section IV concludes this article.

II. METHODOLOGY

The proposed crack quantification technique takes a set of inspection images and the corresponding crack segmentation results as input and outputs a voxel-based crack representation model with detailed crack properties. As shown in Fig. 1, the whole process can be divided into four stages.

1) Crack segmentation in acquired images.
2) Voxel-based crack reconstruction
3) Bayesian data fusion.
4) Crack measurement and visualization.

Details of each stage are described in the following.

A. Crack Segmentation

To extract the image pixels pertinent to a target crack, a defect instance segmentation architecture called DIS-YOLO [23] was adopted. It was shown that DIS-YOLO outperforms the state-of-the-art models in crack instance segmentation. In this network, a class-specific confidence score (object confidence × class probability) from YOLOv3 [24] is used to give out predictions. This score reflects the confidence of the network in detecting the existence of cracks but does not reflect the quality of the predicted masks very well, and such a misalignment tends
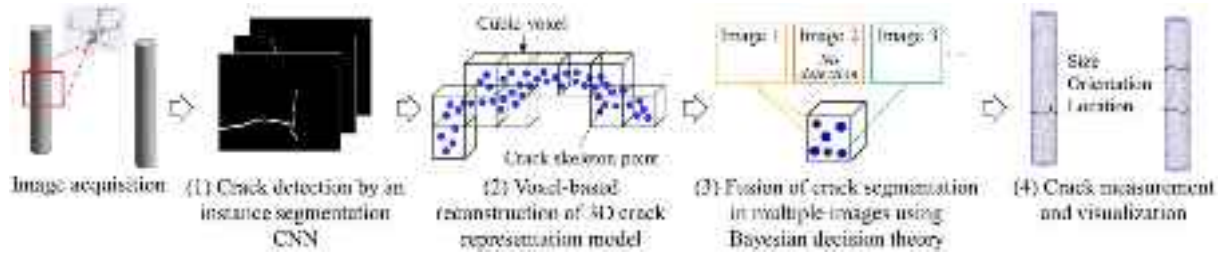
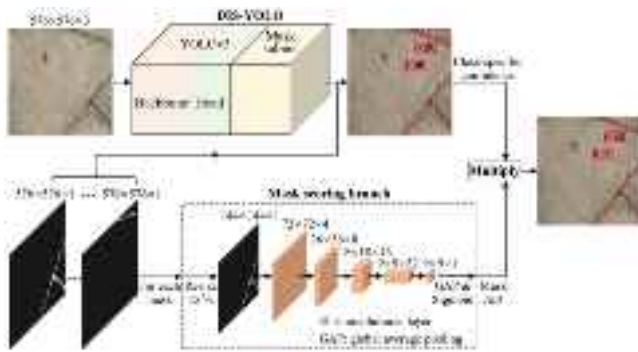Fig. 1.   Overview of the proposed crack quantification framework.

Fig. 2.   Modified DIS-YOLO with mask scoring branch.

to degrade the evaluation results [25]. To tackle this problem, this study modifies DIS-YOLO architecture by adding an additional mask scoring branch to learn the quality of the predicted masks (see Fig. 2). The intersection-over-union (IoU) between the predicted and ground-truth masks indicates the mask quality. The mask scoring branch takes the predicted masks from DIS-YOLO as input and outputs a mask IoU score for each prediction. The product of the mask IoU and the corresponding class-specific confidence gives the final crack segmentation confidence. The architecture design of the mask scoring branch contains five separate 3×3 convolutional layers with a stride of 2, followed by one global average pooling layer and a sigmoid activation function.

Training and testing were performed on a PC running the Ubuntu 17.04 operating system with an Intel Core i7-7700 CPU and a GeForce GTX 1060 6GB GPU. The model was trained with a publicly available concrete crack dataset, which contains 571 crack images with pixel resolutions ranging between 266×236 and 1400×884. The dataset is randomly split into training, validation, and testing subsets with a ratio of 7:1:2. The whole network is trained end-to-end with a loss described as follows:

$$\text{Total loss} = \text{Confidence loss} + \text{Classification loss}$$
$$+\text{Localization loss} + \text{Mask subnet loss}$$
$$+\text{Mask scoring loss}. \tag{1}$$

The confidence, classification, localization, and mask subnet losses belong to DIS-YOLO network [23]. The last term is added to the total loss as a result of the addition of the mask scoring branch to the network. Since the predicted mask IoU score needs to regress to a specific real value, sum-squared loss is used to penalize the errors from the mask scoring branch. The network
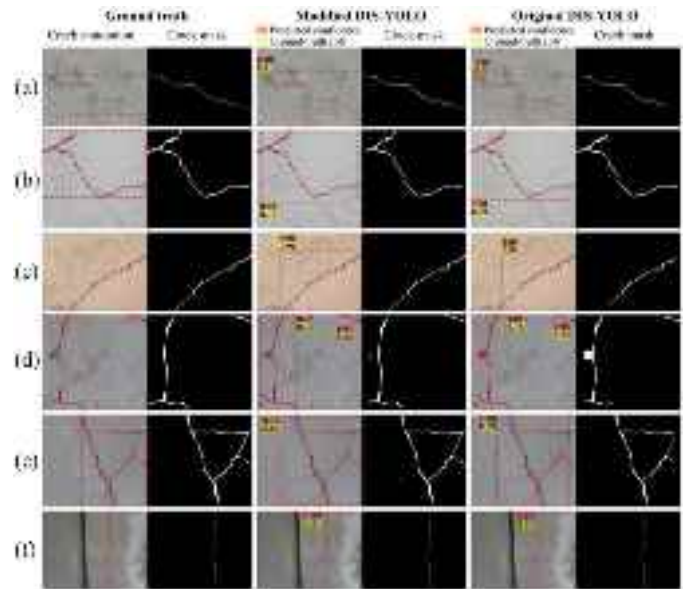

Fig. 3.   Comparison of test samples between the original and modified DIS-YOLO.

was trained using the Adam optimizer [26] with a batch size of 2 and a weight decay of 0.0001. To speed up the training process, all layers in the DIS-YOLO were initialized with the pretrained weights from [23] and only the network heads (YOLOv3's head subnet, the mask subnet and mask scoring branch in Fig. 2) were fine-tuned for 40 epochs with a learning rate of 10^{-4}, and 10^{-5} for every 20 epochs. During testing, a final crack segmentation confidence threshold of 0.25 was used to obtain the predicted results for each image.

Compared with the original DIS-YOLO, the added mask scoring branch could increase the testing accuracy of crack segmentation by 3.2% and achieve an average precision (AP) equal to 87.3% for the mask-level IoU metric of 0.5. To provide a more comprehensive evaluation, in addition to the standard instance segmentation evaluation metric AP, the semantic segmentation evaluation metric IoU is used. Test results show that the IoU of the modified DIS-YOLO with mask scoring branch is 72.5%, which is 6.3% higher than that of the original DIS-YOLO. However, this accuracy gain came at the cost of an approximately 20% slower test speed (i.e., about 8 FPS when tested on an NVIDIA GeForce GTX 1060 GPU). Fig. 3 shows test examples for both the original and modified DIS-YOLO. In this figure, the ground-truth IoU of each crack mask is reported under its corresponding predicted segmentation confidence. The
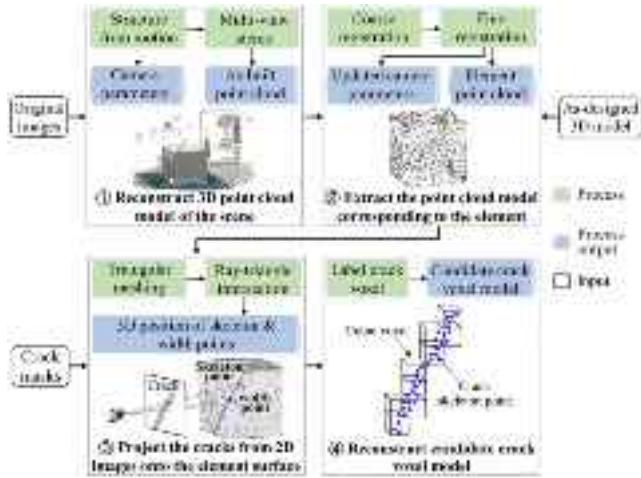
Fig. 4. Reconstruction process of crack voxel model.

Fig. 5. Illustration of reconstructed crack voxel model.

illustrated examples show that the modified DIS-YOLO can predict crack bounding boxes and masks more accurately compared to the original DIS-YOLO. This indicates that additional training information from the mask scoring branch is beneficial for both object detection and mask prediction in DIS-YOLO. Also, the added mask scoring branch provides a crack segmentation confidence that can well reflect the quality of the predicted mask, as shown for all cracks in Fig. 3. The predicted segmentation confidences of the modified DIS-YOLO show slightly larger errors for some very thin and complex cracks [see Fig. 3(e) and (f)], which are still well below those of the original DIS-YOLO.

B. Voxel-Based Reconstruction

Reconstructing 3-D objects from images is a classic computer vision problem that has been a subject of research for decades. Among various image-based 3-D reconstruction techniques, multiview stereo (MVS) that reconstructs the 3-D scene based on stereo correspondence is the most successful method in terms of robustness and number of applications [27]. The MVS algorithms can be classified into four general groups according to their output scene representations which can be depth maps, point clouds, voxel spaces and meshes. Among them, the voxel representation that reconstructs the 3-D scene with a regular grid structure provides great flexibility for model manipulation [28]. Such a voxel output is suitable for performing 3-D volumetric analysis such as morphological operations, but may not be suitable for accurate crack width measurement since its accuracy is limited by the resolution of the voxel grid. In this article, a novel and robust voxel-based crack reconstruction framework is proposed which can produce a crack model where each voxel encodes accurate crack width information from multiple 2-D images. This process is accomplished through the four steps illustrated in the flowchart shown in Fig. 4.

First, a dense 3-D point cloud of the scene is generated using the structure from motion (SFM) and MVS techniques. The SFM process adopts VisualSFM [29] to compute the camera parameters by detecting and matching feature points from multiple images using the scale-invariant feature transform [30]. Once
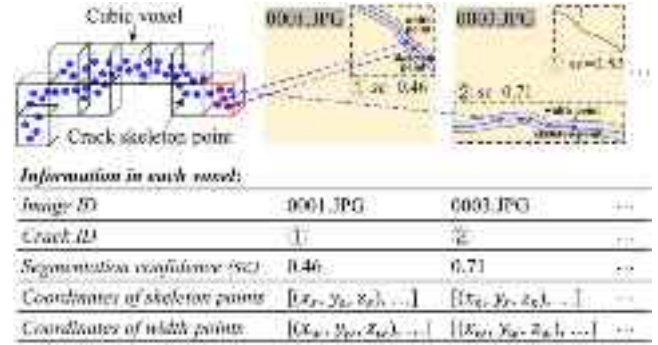
the camera parameters are known, patch-based MVS [31] is performed to generate a dense 3-D point cloud model of the scene. The output point cloud is then scaled to real size using a uniform scale factor obtained from the known dimensions of an object in the scene.

The next step is to extract the element point cloud from the reconstructed scene. This task could be accomplished by 3-D registration of the as-built data with the as-designed model [32]. The as-built data is the reconstructed point cloud from the previous step, which presents the current appearance of a scene. The as-designed model represents the corresponding 3-D model of a scene that is usually stored in CAD or BIM engines. The registration process between the as-built point cloud and the as-designed model consists of a coarse registration for rough alignment followed by a fine registration. The purpose of registration is to find the association between the as-built data with its corresponding as-designed 3-D model object. In this article, the coarse registration is performed manually using fiducial markers with known locations on the as-designed 3-D model [18]. The fine registration is performed using a procedure based on the Iterative closest point proposed by Glira et al. [33]. It should be noted that, for both registration steps, only the translation and rotation transformation are computed to update the camera parameters so as to keep the reconstructed point cloud model in real size.

The third step is to locate the crack points in 3-D space by projecting the crack from 2-D images onto the element surface. To do so, the extracted element point cloud from the previous step is first converted to a triangular mesh surface model. With the known camera parameters, crack skeleton and width points, as the key points defining the shape of the crack, can be projected from the image onto the element's triangular mesh surface. A crack skeleton point is obtained by running a morphological thinning algorithm on the crack segmentation masks, and the closest crack edge point to each skeleton point corresponds to its width point (see Fig. 5) [12].

Finally, the crack voxel model is reconstructed by building a volumetric space and labeling voxels containing any skeleton points as candidate crack voxels. As illustrated in Fig. 5, each voxel of the reconstructed crack model encodes the crack information from multiple images. Specifically, each voxel stores five pieces of information.

1) Image IDs indicating the images containing the skeleton points of the voxel.
2) Crack IDs indicating the crack on the image that the skeleton points of the voxel belong to.
3) Crack segmentation confidence for each crack ID.
4) 3-D coordinates of the skeleton points (x_s, y_s, z_s).
5) 3-D coordinates of the corresponding width points (x_w, y_w, z_w).

Given this information, in the next stage, it is inferred whether a candidate crack voxel belongs to an actual crack or it is a false detection.

C. Data Aggregation

1) Tracking Voxels in Multiple Images: Each voxel can have projections in multiple images. Although for each candidate crack voxel, there is at least one image with a detected crack in its corresponding region, the projected regions of the crack voxel on other images may have no crack detected by the segmentation network. By utilizing the shared information among all contributing images, the segmentation outcomes from multiple individual images can be fused to increase the accuracy and robustness of the reconstructed crack voxel model. To realize such data aggregation, a candidate crack voxel first needs to be tracked in multiple visible images. With the estimated camera parameters calculated in the 3-D reconstruction process, the location of a candidate crack voxel in each image can be calculated by projecting the voxel onto the image. Here, to maintain a consistent projected region in each image, the sphere projection model introduced by Yeum *et al.* [18] is adopted. In each image, the smallest bounding box containing the projected sphere is taken as the projected image region of the voxel.

Due to uncertainty in the crack's location and the camera's view angles, there might be images that fail to show some voxels clearly and should be excluded in the data fusion process. For instance, the candidate crack voxel may be too far away from the optical center of the image (i.e., crack is not detectable) or it is occluded by other objects in the scene. To tackle this problem, a visibility checking algorithm is proposed in Algorithm 1 based on the reference image selection and photo-consistency measure. The reference image for a candidate crack voxel is defined as the image with the best crack visibility. By comparing the photo-consistency of the projected region in the reference image with the corresponding regions in other images, their visibility can be evaluated. In this article, the projected image regions are converted to grayscale and resized to a constant size of $n \times n$ (n is set to 30), and the normalized cross-correlation (NCC) coefficient [27] shown in the following equation is adopted to measure the photo-consistency, (2) shown at the bottom of this page.

In (2), $\text{IR}(p_x, p_y)$ and $I(p_x, p_y)$ denote the intensity values at pixel location (p_x, p_y) of the projected region in reference image IR and the investigated image I, respectively, and $\overline{\text{IR}}$ and $\bar{I}$ denote their corresponding mean intensity values. In the proposed algorithm, if the NCC coefficient is less than a given threshold ρ, it indicates the content in the image being investigated is dissimilar to the reference image and should be labeled as invisible.

For each candidate voxel, to initiate the algorithm, a list of potential reference images is generated by searching for the images that satisfy the following constraints: they must include the projected region of the voxel; they must provide crack skeleton points for the voxel; and the angle between the normal of the voxel (taken as the surface normal of the closest triangular mesh face element) and the direction from the voxel to the camera's optical center is smaller than $90°$. The potential reference images in the list are then sorted in descending order according to their crack segmentation confidence scores. Afterward, these images are analyzed sequentially to find a reference image that has more than γ (equal to 3) number of visible images as shown in Algorithm 1. This constraint guarantees that the reference image itself is not occluded by any obstacles. Here, a preset threshold $\rho = 0.3$ for NCC is selected as the visibility criterion, however, the NCC is not invariant to rotation and may lead to a low similarity score for two images captured from different view angles. To address this issue, the projected region of an image being checked against a potential reference image is sequentially rotated with a $30°$ interval around its center. After each rotation, the NCC coefficient is calculated to check whether the visibility is larger than the threshold. In the end, if the algorithm could successfully find a reference image, that image and all the visible images are stored and linked to the evaluated candidate crack voxel, otherwise, the voxel will be removed from the model.

2) Bayesian Data Fusion: For a set of candidate crack voxels that are successfully tracked in multiple images, a rigorous approach is needed to determine whether they belong to an actual crack or not. Previously, Chen *et al.* [34] proposed a data fusion approach based on Bayesian inference that provides a systematic framework for decision making. In this article, this promising approach was further improved by including the segmentation confidence of the network, which reflects the reliability of the network in finding a crack object and generating a high-quality mask (see Section II-A). As a general rule, the closer the value of segmentation confidence gets to 1, the more likely it is that the segmentation outcome is correct and accurate. In the proposed data fusion process, the segmentation outcomes are divided into two categories of low confidence (LC) and high confidence (HC) detections using a limiting value λ. Thus, for a candidate crack voxel v that contains n visible images, the outcome y can be presented as the set below

$$y = \{(i, Y_i) \,|\, i \in \{1, \ldots, n\}, \, Y_i \in \{\text{NO}, \text{LC}, \text{HC}\}\}. \quad (3)$$

$$\text{NCC}(\text{IR}, I) = \frac{\sum_{p_x=1}^{n} \sum_{p_y=1}^{n} \left\{ \left[\text{IR}(p_x, p_y) - \overline{\text{IR}} \right] \cdot \left[I(p_x, p_y) - \bar{I} \right] \right\}}{\sqrt{\sum_{p_x=1}^{n} \sum_{p_y=1}^{n} \left[\text{IR}(p_x, p_y) - \overline{\text{IR}} \right]^2} \cdot \sqrt{\sum_{p_x=1}^{n} \sum_{p_y=1}^{n} \left[I(p_x, p_y) - \bar{I} \right]^2}} \quad (2)$$

Algorithm 1: Pseudocode for Tracking Reference and Visible Images.

Input: V: Candidate crack voxel model
$\quad\quad\quad$ v: A single voxel in V
$\quad\quad\quad$ I: A set of all inspection images
$\quad\quad\quad$ *SR:* Sorted reference image list for v
Output: IR: Reference image for v
$\quad\quad\quad\quad$ IV list of visible images for v

1: $\quad$ IR$\leftarrow\emptyset$
2: $\quad$ **for** each $r \in$ SR **do**
3: $\quad\quad$ $r_p\leftarrow$projected region of v on image r
4: $\quad\quad$ **for** each $i \in I$ **do**
5: $\quad\quad\quad$ $i_p\leftarrow$projected region of v on image i
6: $\quad\quad\quad$ **if**
$\quad\quad\quad\quad$ $\max\{\text{NCC}[r_p, \text{rotate}(i_p, \theta = 270, \ldots, 330)]\} \geq$
$\quad\quad\quad\quad$ ρ **do**
7: $\quad\quad\quad\quad$ add i to IV
8: $\quad\quad\quad$ **end if**
9: $\quad\quad$ **end for**
10: $\quad\quad$ **if** length $(IV) \geq \gamma$ **do**
11: $\quad\quad\quad$ $IR \leftarrow r$
12: $\quad\quad\quad$ **return** IR, IV; **break**
13: $\quad\quad$ **end if**
14: $\quad$ **end for**
15: $\quad$ **if** $IR = \emptyset$ **do**
16: $\quad\quad$ remove v from V
17: $\quad$ **end if**

Note that "NO" denotes the case that no crack segmentation is identified according to the confidence threshold specified in Section II-A. The goal is to decide whether voxel v represents an actual crack (i.e., AC) or not, given the segmentation outcomes from multiple visible images. Assuming that the segmentation outcomes from different images are independent, for each candidate voxel, the posterior odds ratio can then be obtained as

$$\frac{P\left(\text{AC}|y\right)}{P\left(\overline{\text{AC}}|y\right)} = \prod_{i=1}^{n} \frac{P\left(Y_i|\text{AC}\right)}{P\left(Y_i|\overline{\text{AC}}\right)} \cdot \frac{P\left(\text{AC}\right)}{P\left(\overline{\text{AC}}\right)} \quad\quad (4)$$

where $P(\text{AC})$ denotes a prior probability of a voxel containing an actual crack; $P(\overline{\text{AC}})$ denotes the probability of a voxel containing no crack; the conditional probabilities $P(Y_i|\text{AC})$ and $P(Y_i\overline{\text{AC}})$, respectively, denote the probability of occurrence of a segmentation outcome $Y_i \in \{\text{NO}, \text{LC}, \text{HC}\}$ given the presence and absence of an actual crack. The first ratio on the right-hand side, also called the likelihood ratio, depends on the accuracy of the network in detecting crack voxels and can be obtained from the testing stage of the network. The second ratio on the right-hand side represents the prior odds which can be estimated from site investigation. If the calculated posterior odds ratio is larger than 1 for a given segmentation outcome y, this implies that the candidate voxel under consideration is more likely an actual crack voxel. The predefined parameters could be obtained from the *in situ* inspection data. For instance, if the target is to inspect cracks on the concrete piers of a certain highway bridge*, a prior* crack inspection test on a few bridge piers could
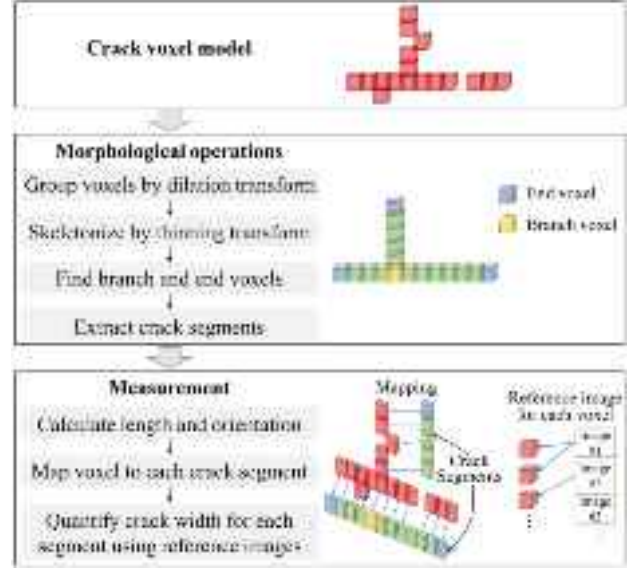

Fig. 6. Crack properties retrieval procedure.

be performed to estimate the parameters with the manually inspected or labeled ground-truth crack.

D. Crack Properties Retrieval

The above steps give out a filtered crack voxel model wherein each voxel contains the crack width information from multiple images and a reference image. To retrieve crack properties, such as width, length, orientation, and location, a 3-D volumetric image processing framework is proposed, which is summarized in Fig. 6. To extract the crack segments, a four-step 3-D morphological operation is sequentially performed. The first step is to perform a dilation transform using a spherical structuring element to group the crack voxels that belong to each complete crack [2]. Next, the centerline of each crack (i.e., crack skeleton voxel) is extracted through a 3-D medial axis thinning operation [35]. In this step, some outliers can be removed by eliminating cracks that have skeleton lengths below a preset size. Afterward, the branch and end voxels of a crack can be found through morphological analysis. In the end, the crack segments of a crack are extracted, and any two crack segments that share the same branch voxel and have a similar crack direction at the shared branch voxel are merged.

In the next stage, the length, orientation, and width of each crack segment can be measured. The length is approximated as the product of voxel resolution and the number of skeleton voxels in each segment. The orientation of each crack segment is measured by finding the angle of crack direction which is the vector connecting two ends relative to the underlying structural component's major axis (e.g., vertical direction for columns) [36]. To measure crack width, rather than relying on the voxel size, the width information from the reference image of each crack voxel, described in Section II-C1, is utilized. This is because among all images that contribute to skeleton points in a voxel, the reference image most likely has the best image quality and the most accurate crack segmentation. It should be noted that

crack width is defined as twice the distance between the skeleton point and the corresponding width point, therefore the minimum detectable crack width on 2-D images is three pixels.

The crack voxels are grouped for each complete crack in the above morphological operations. Hence, they need to be assigned to each crack segment for retrieving each segment's width properties. Also, there are always some errors in the crack segmentation and the 3-D reconstruction process, which cause slight discrepancies between the locations of the reconstructed crack voxels and the locations of the crack segments from morphological operations. To address this problem, a mapping process is performed to map the closest crack skeleton voxel to each reconstructed crack voxel based on the nearest neighbor search method [37] (see Fig. 6). Subsequently, the maximum and average crack widths in each skeleton voxel can be measured using the crack width information from the reference image of the matched crack voxel. There may exist a few skeleton voxels that have more than one or no matched crack voxel. This does not affect the maximum and average width properties of the crack segment because in calculating those qualities all crack voxels grouped as one segment are considered. Finally, the skeleton voxels of each crack segment can be visualized on the as-designed 3-D model to reveal the crack's location relative to the structural component.

III. EXPERIMENTAL VALIDATION

In this section, the proposed crack quantification framework was validated using a set of cracked concrete members. The experiment is to evaluate the accuracy of crack voxel model reconstruction and the crack quantification using images taken under various conditions, including different working distances, viewpoints, and crack sizes.

A. Implementation Details

For the test, two concrete specimens made of engineered cementitious composites (ECC) were used. They were underwent tensile tests to measure the tensile stress-strain relationship of the ECC, which resulted in transverse cracks being developed on their surfaces. The specimens have cylindrical shapes with diameters of around 150 mm and lengths of about 900 mm. For validation of the crack quantification technique, as shown in Fig. 7, these specimens were placed approximately 800 mm apart. The concrete specimens can be regarded as small-scale circular bridge piers and the experimental setup could mimic the on-site crack inspection of bridge piers surface. A total of 18 images with a resolution of 4032×3024 pixels were taken using a smartphone camera with a fixed focal length of 4.2 mm and a pixel pitch of 1.4 μm. These images were taken from different view angles at a distance between 0.5 and 2 m away from the specimens, which leads to a maximum spatial resolution of 6 pixels/mm.

In order to evaluate the performance of the modified DIS-YOLO in practical conditions, the model trained on the public crack dataset in Section II-A was adopted to segment cracks on the acquired images. Due to limited computer memory, the resolution of the input images used for training a deep CNN
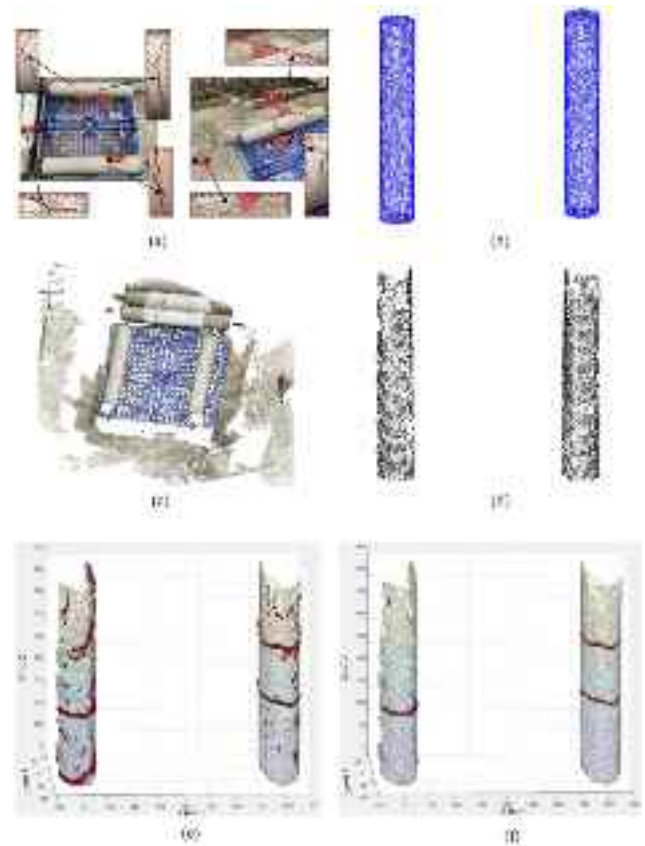

Fig. 7.   Key steps in the crack voxel model reconstruction process.

is usually selected at a much lower resolution than the original inspection images. However, to ensure a good crack segmentation result on large images, the inspection image was scanned using a fixed-size sliding window at 50% overlap. The window size was set to 672×504 pixels which is close to the training image size of the network (576×576 pixels). The outputs from all sliding windows of an inspection image were then merged by examining whether any two crack segmentation masks have overlapping crack pixels. In such a case, a new segmentation confidence was assigned to the merged crack that is the average of the confidence scores of the overlapping cracks.

To reconstruct the crack voxel model, the voxel size needs to be predefined. A large voxel size may decrease the quantification accuracy of crack length and orientation due to the growing geometric error from bigger voxels [28]. While a small voxel size can give an unclear projected image region that may cause errors in tracking visible images and the subsequent data fusion process. In this article, the voxel size was set to be 5 mm that is a reasonable value to achieve a balance between these two effects. The specimen's real length was used to compute the scale factor and convert the reconstructed 3-D point cloud model to real size. To extract the element point cloud, an as-designed 3-D model of the two specimens was created in BIM format using Autodesk Revit. After preparing these input data and manually performing the coarse registration, all the remaining processes for crack voxel model reconstruction are done automatically.

TABLE I
ACCURACY OF CRACK VOXEL MODEL BEFORE AND AFTER
BAYESIAN DATA FUSION

Model	Precision	Recall	F1-score
Candidate crack voxel model	0.301	0.745	0.429
After Bayesian data fusion ($\lambda=0.6$)	0.836	0.709	0.767
After Bayesian data fusion ($\lambda=1.0$)	0.859	0.640	0.733

To determine the likelihood and prior odds ratios for Bayesian data fusion described in Section II-C2, a test was performed in advance with the limiting parameter λ set to a median value of 0.60 and the ground truth labeling of cracks on other tested specimens. Based on the segmentation outcome of the network on a separate test dataset, the conditional probabilities $P(\text{HC}|\text{AC})$, $P(\text{LC}|\text{AC})$ and $P(\text{NO}|\text{AC})$ in the numerator of likelihood ratio were computed to be 0.315, 0.451, and 0.234 respectively. Similarly, the terms $P(\text{HC}|\overline{\text{AC}})$, $P(\text{LC}|\overline{\text{AC}})$, and $P(\text{NO}|\overline{\text{AC}})$ in the denominator of likelihood ratio were calculated as 0.017, 0.053, and 0.930, respectively. These conditional probabilities are fixed values and depend on the accuracy of the segmentation network. The prior odds ratio $P(\text{AC})/P(\overline{\text{AC}})$ was estimated to be around 0.050, which means in general for each 5-mm voxel on the specimen's surface, it is twenty times more likely to observe no crack.

B. Results and Discussion

The reconstruction accuracy of the crack voxel model before and after Bayesian data fusion, as well as the crack measurement accuracy, were evaluated to investigate the overall performance of the proposed crack quantification framework. Fig. 7 shows the key steps in the crack voxel model reconstruction. Fig. 7(a) displays samples with the predicted bounding boxes and their pertinent masks for each crack instance. The developed crack segmentation network could successfully identify cracks on images, however, the results still contain some false positives from the background. This phenomenon commonly happens in practice due to the unavoidable discrepancy of image content between the image dataset used for training of the network and the site-specific inspection images.

Fig. 7(b) and (c) shows the as-designed and the as-built point cloud model, respectively. These two point cloud models were registered to extract the as-built element point cloud in Fig. 7(d). Then, following the procedure described in Section II-B, the crack skeleton and width points extracted from the crack masks were projected onto the element surface to form the candidate crack voxel model as shown in Fig. 7(e). In this process, most of the false positives from the background that have no projections on the element surface were eliminated. However, those false positives that lie on the element surface on the inspection image would still produce false crack voxels, such as the falsely detected steel chain in Fig. 7(a).

Fig. 7(f) shows the crack voxel model after Bayesian data fusion. It can be seen that a significant number of false voxels have been removed while the true crack voxels are retained by the proposed Bayesian data fusion method. Table I compares
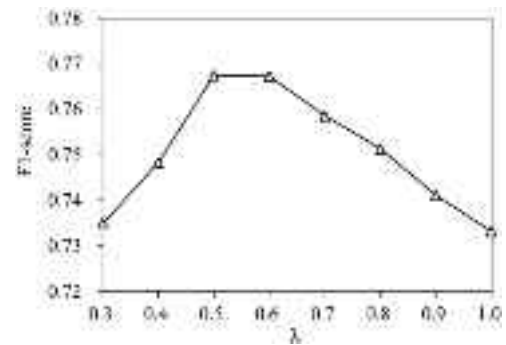

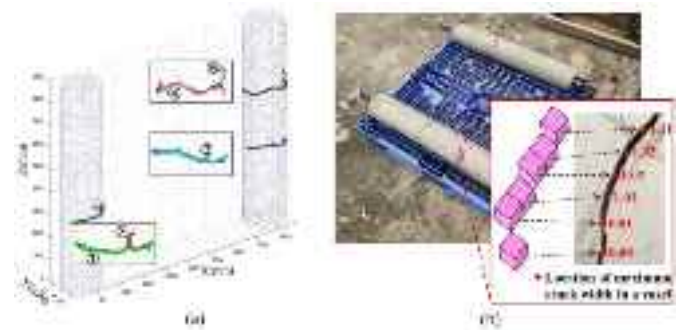Fig. 8.   Bayesian data fusion with different limiting parameter λ.

Fig. 9.   Evaluation process of the crack quantification accuracy.

the precision, recall and F1-score between the candidate crack voxel model and the voxel model after Bayesian data fusion. It is observed that the F1-score of the crack voxel model after Bayesian data fusion is 76.7%, which is 33.8% higher than the candidate voxel model. The recall of these two models is comparable but the precision of the crack voxel model after Bayesian data fusion is 2.78 times higher than the candidate voxel model. This reflects the fact that the data fusion technique could successfully filter out the false-positive detections without losing true positives in the process.

For Bayesian data fusion, the limiting parameter λ is a user-defined hyperparameter as described in Section II-C2. The crack voxel reconstruction accuracy under different values of λ was analyzed and summarized in Fig. 8. It can be seen that the F1-score could achieve the highest value between 0.5 and 0.6, and would be decreasing outside this range. When λ is equal to 1.0, the proposed data fusion approach is the same as the standard Bayesian model [34] which is unaware of the importance of different segmentation confidences. This extreme case could achieve an F1-score of 73.3%, which is 3.4% lower than the case with $\lambda= 0.6$ (see Table I). The above results demonstrate that the performance of Bayesian decision making can be improved by including the confidence of the crack segmentation network.

Next, the proposed crack quantification procedure outlined in Fig. 6 was performed on the crack voxel model after Bayesian data fusion to extract the width, length and orientation of each crack segment. Note that in this article, any complete crack with a total skeleton length below 5 voxels is considered an outlier and is removed from the model. Fig. 9(a) displays the reconstructed

TABLE II
QUANTIFICATION RESULTS OF CRACK LENGTH AND ORIENTATION

Crack segment number	Length (mm)			Orientation (°)		
	Prediction	Ground truth	Error	Prediction	Ground truth	Error
1	205	225	-20	90.00	86.27	3.73
2	30	15	15	19.47	45.00	-25.53
3	180	205	-25	88.05	86.31	1.74
4	215	225	-10	82.29	83.93	-1.64
5	55	35	20	8.05	9.46	-1.41

TABLE III
QUANTIFICATION RESULTS OF CRACK WIDTH

| Crack segment number	Location number	Predicted crack width (mm)	Measured crack width (mm)	Error (mm)
1	1	1.18	1.21	-0.03
	2	0.97	0.80	0.17
2	3	0.68	0.59	0.09
	4	0.74	0.69	0.05
3	5	1.43	1.36	0.07
	6	1.03	0.95	0.08
4	7	1.02	0.93	0.09
	8	0.78	0.74	0.04
5	9	0.88	0.80	0.09
	10	0.59	0.51	0.08

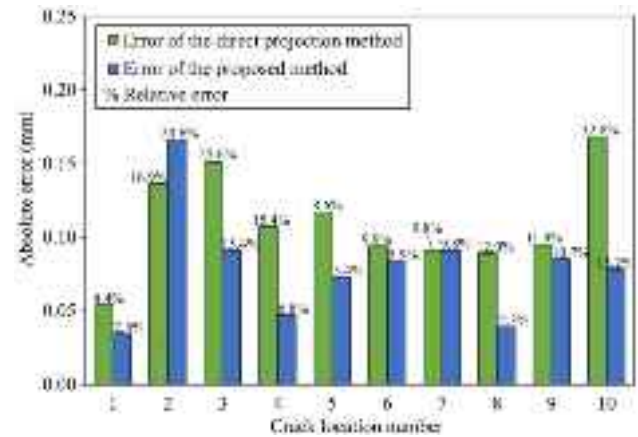Fig. 10. Example of predicting crack width at location number 5 using direct projection method.

Fig. 11. Comparison of crack width accuracy between the proposed method and the direct projection method.

crack segment model together with the as-designed point cloud model to visualize the crack's location on the element, where each color represents a separate crack segment. Table II gives the quantification results of crack length and orientation. It is seen that the errors of crack segments' length are in the order of 10–25 mm. For the orientation of the crack segment, the errors are less than 5° except for the very short segment number 2. These results are acceptable as longer crack segments are more critical than shorter ones considering the crack severity assessment.

To quantitatively evaluate the accuracy of crack width, the predicted maximum widths in several crack voxels are compared with the measured crack widths at the same locations. The location of the maximum crack width in each voxel could be obtained from its corresponding reference image. Fig. 9(b) illustrates a reference image with the corresponding voxels and maximum crack width locations. The reference image provides crack width information for crack voxels in the middle part of crack number 4 in Fig. 9(a). Table III gives the comparison of crack width at ten selected locations. Results show that most locations have an error of less than 0.1 mm, and when the real crack width is larger than 0.9 mm, the relative error can drop to below 10%. It should be noted that the minimum detectable crack width in this experiment is about 0.5 mm when assuming a crack width of 3 pixels on an image and a working distance of 500 mm. The limitation of minimum detectable crack width and the quantification error on narrow cracks could be mitigated by increasing the pixel resolution of the crack width by using a camera with a larger focal length or taking photos at a shorter working distance. Nonetheless, the proposed crack quantification technique is successfully validated with the capability of retrieving various crack properties including the width, length,
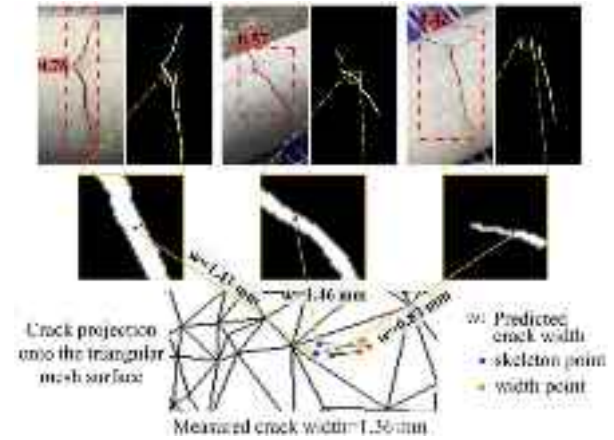
orientation, and location of each crack segment in a complete crack.

As a comparative study, the crack width is also measured using direct projection method [17]. In this method, crack width is retrieved by projecting its width points from 2-D images back onto the 3-D triangular surface model reconstructed using SFM and MVS techniques. The procedure of this method is similar to the first three steps of the voxel reconstruction process described in Section II-B. Therefore, similar to the proposed framework, it can be utilized to measure cracks on uneven or curved structural surfaces using images captured under uncontrolled conditions. In Fig. 10, prediction of the crack width at location number 5 using the direct projection method is presented as an illustrative example. Here, for the same crack location detected in multiple images, the average crack width is used. It can be seen that the crack with the highest segmentation confidence, i.e., from the reference image, can provide more accurate measurement, while directly averaging the crack width from multiple images may introduce greater errors from the low-confidence crack segmentations. Fig. 11 compares the crack width error between the proposed framework and the direct projection method at the locations given in Table III. The results show that the proposed
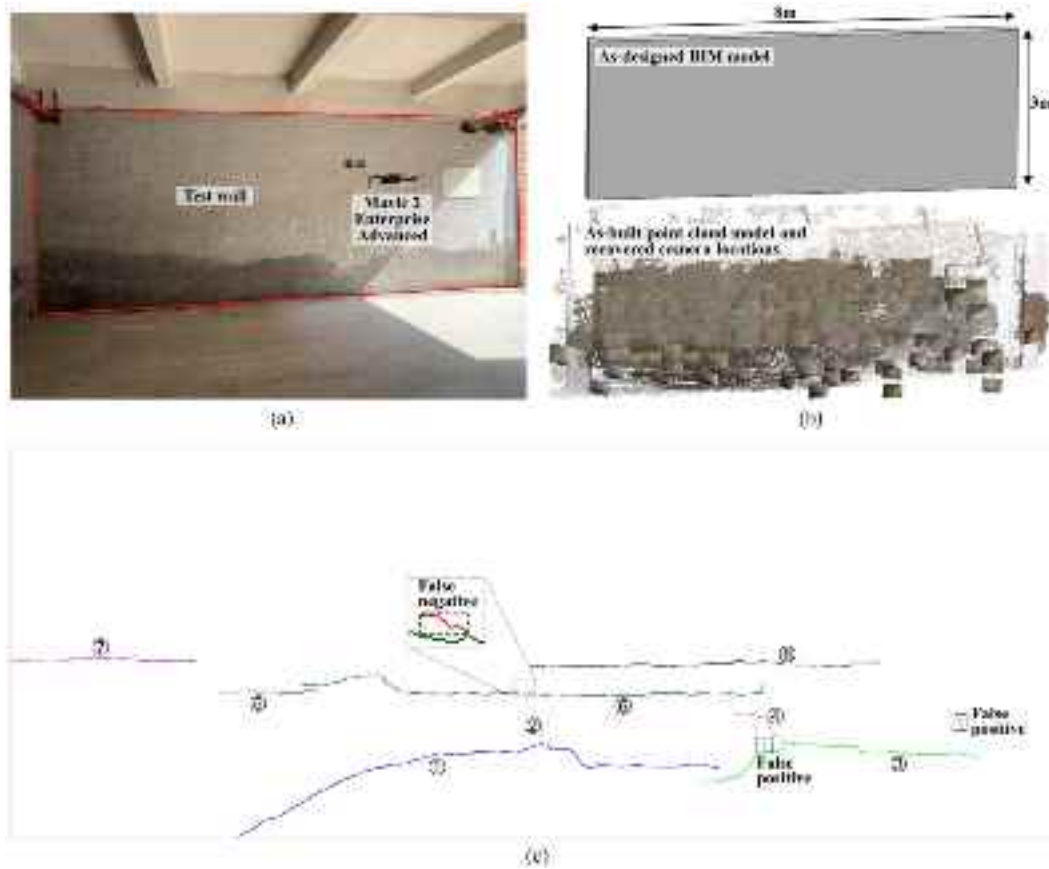
Fig. 12. Field test of a concrete wall and its reconstructed crack segment model.

method on average offers a 4.7% lower relative error compared to the direct projection method. Particularly, for the narrow cracks at locations number 3 and number 10, the advantage of the proposed method is more obvious where the relative error is decreased by more than 10%. It should be noted that in the direct projection method, all detected cracks are projected onto the triangular surface model. Therefore, false segmentations can easily generate false crack location and width information in 3-D space, and thus further undermine the accuracy of the overall crack measurement. Moreover, with the 3-D crack points generated as the output of the direct projection method, it would be difficult to apply image processing techniques to calculate the length and orientation of each crack segment.

C. Field Test

A field test was conducted on a concrete wall of a factory room, as shown in Fig. 12(a). The wall's dimensions are 8 m × 3 m, and it has several transverse cracks on its surface. Photos of the wall were taken using the commercial UAV "Mavic 2 Enterprise Advanced" equipped with a high-resolution camera. Since the GPS signal for precise indoor navigation is weak, manual control of the UAV was adopted throughout the experiment. A total of 96 images were taken with a resolution of 8000 × 6000 pixels at a working distance of approximately between 0.5 and 1.5 m. The focal length and pixel pitch of the camera are fixed at 4 mm and 0.8 μm, respectively. Under this experimental setting, the maximum spatial resolution and minimum detectable crack width can be calculated as 10 pixels/mm and 0.3 mm, respectively.

Fig. 12(b) shows the as-designed 3-D model and the reconstructed point cloud model of the wall. Similar to most concrete structures in service, the surface of the wall has dirt, holes, or other inevitable imperfections, which usually provide enough distinctive feature points for accurate point cloud reconstruction. The final crack segment model is illustrated in Fig. 12(c). This model is obtained following the implementation details of crack segmentation, crack voxel model reconstruction and Bayesian data fusion described in Section III-A. It should be noted that the conditional probabilities and prior odds ratio in the Bayesian data fusion process were obtained through a prior crack inspection test on another wall in the same factory room. When the limiting parameter λ was set to 0.6, the terms $P(\text{HC}|\text{AC})$, $P(\text{LC}|\text{AC})$, $P(\text{NO}|\text{AC})$, $P(\text{HC}|\overline{\text{AC}})$, $P(\text{LC}|\overline{\text{AC}})$, $P(\text{NO}|\overline{\text{AC}})$, and $P(\text{AC})/P(\overline{\text{AC}})$ was estimated to be 0.357, 0.419, 0.224, 0.020, 0.042, 0.938, and 0.031, respectively. The results in Fig. 12(c) show that major cracks could be recognized successfully. However, a couple of very short cracks were detected as false positives or negatives.

In Table IV, the length, width and orientation of a total of eight detected crack segments are presented. It can be seen that the errors in the predicted lengths and orientations for all

TABLE IV
CRACK QUANTIFICATION RESULTS OF THE FIELD TEST

Crack segment number	Length (mm)			Orientation (°)			Width (mm)			
	Prediction	Ground truth	Error	Prediction	Ground truth	Error	Location number	Prediction	Ground truth	Error
1	3705	3700	5	81.84	81.84	0.00	1	1.36	1.31	0.05
							2	2.25	2.18	0.07
2	245	255	-10	19.89	24.08	-4.19	3	0.48	0.42	0.06
							4	0.39	0.29	0.10
3	2225	2205	20	85.12	85.17	-0.05	5	0.75	0.71	0.04
							6	1.20	1.13	0.07
4	465	465	0	56.31	57.15	-0.85	7	0.81	0.87	-0.06
							8	0.50	0.46	0.04
5	760	740	20	88.85	89.22	-0.36	9	0.58	0.53	0.05
							10	0.42	0.38	0.04
6	3595	3590	5	89.67	89.76	-0.08	11	0.52	0.57	-0.05
							12	1.17	1.20	-0.03
7	1420	1415	5	89.59	89.59	0.00	13	0.43	0.37	0.07
							14	0.58	0.61	-0.03
8	2670	2675	-5	89.46	89.46	0.00	15	0.79	0.73	0.06
							16	1.13	1.11	0.02

crack segments are less than 20 mm and 5°, respectively. Also in Table IV, the predicted maximum widths for two locations of each crack segment are compared with the measured crack widths at the same locations. In terms of the accuracy of crack width prediction, the measurement error of most crack locations, including the narrow cracks at locations 4, 10, and 13, can be controlled within 0.1 mm. The proposed method could offer an error of less than 10% for cracks with widths greater than 0.5 mm. This indicates that a promising width measurement accuracy could be achieved when the crack width is more than 5 pixels on an image. The above results demonstrate that the proposed crack quantification framework has a consistent performance for both laboratory and field conditions.

IV. CONCLUSION

In this article, a new crack assessment and 3-D visualization framework is proposed that identifies crack segments and quantifies their various properties using a set of unordered inspection images. The proposed technique begins with an instance segmentation network that generates segmentation masks for the cracks that appeared in all images. The proposed network can output a crack segmentation confidence score for each prediction, which is consistent with the IoU between the predicted and the ground-truth mask. Testing results on the public concrete crack dataset showed that the developed network could achieve the instance segmentation accuracy of AP = 87.3% using the mask-level IoU metric of 0.5, and the semantic segmentation accuracy of IoU = 72.5%. Next, the crack segmentation results of multiple images are aggregated through the 3-D reconstruction and Bayesian data fusion techniques. Specifically, in the proposed data fusion approach, rather than treating crack segmentation outcomes equally, segmentations with higher confidence scores can potentially improve the final result. The above steps result in a voxel-based crack representation model where each voxel encodes crack width information derived from a reference image. Following a proposed image tracking algorithm, for each crack voxel, a reference image is found that has the best quality and most accurate crack segmentation among all visible images. Finally, a crack quantification procedure is proposed to retrieve the detailed crack properties for each crack segment given the inherently complex and branching pattern of cracks in 3-D space.

The capability and performance of the proposed techniques were validated on cracked concrete specimens using the images acquired from different locations and view angles. Results show that the proposed Bayesian data fusion approach can significantly improve the accuracy of crack voxel model reconstruction and achieve an F1-score of 0.767, which is 33.8% higher than the method without data fusion and 3.4% higher than the standard Bayesian data fusion approach. It is also illustrated that the obtained crack voxel model can be used to accurately quantify the length and orientation of relatively long crack segments in a complete crack. By quantifying the crack width on reference images, a promising accuracy for crack widths is achieved as well. Compared with the direct projection method where crack width measurements across multiple images are simply averaged, the proposed method could provide a 4.7% lower average error in terms of crack width. In order to further verify the robustness and applicability of the proposed framework, a field test was conducted on a cracked concrete wall with images taken by a commercial UAV. Results show that the calculated crack length and orientation match well with the actual measured values. And for cracks wider than 0.5 mm, the relative crack width error was less than 10%. These results demonstrate the efficacy of the proposed framework to quantify cracks in the field environment.

The results of the proposed crack quantification framework promote its possible adoption in automated inspection systems where often a large volume of unordered and complex visual inspection data is collected and processed. Moreover, compressing the crack information of collected images in the form of voxel models can significantly reduce the need to retain and store a massive amount of inspection images over the lifetime of a piece of infrastructure. In the future, more field tests on different concrete structures, specifically those with very narrow cracks, should be performed to validate and further improve the robustness and practicality of the proposed framework.

REFERENCES

[1] T. Chisholm, R. Lins, and S. Givigi, "FPGA-based design for real-time crack detection based on particle filter," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5703–5711, Sep. 2020.

[2] L. Wu, S. Mokhtari, A. Nazef, B. Nam, and H. B. Yun, "Improvement of crack-detection accuracy using a novel crack defragmentation technique in image-based road assessment," *J. Comput. Civil. Eng.*, vol. 30, no. 1, Jan. 2016, Art. no. 04014118.

[3] R. G. Lins and S. N. Givigi, "Automatic crack detection and measurement based on image analysis," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 3, pp. 583–590, Mar. 2016.

[4] L. Guo, R. Li, and B. Jiang, "A cascade broad neural network for concrete structural crack damage automated classification," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2737–2742, Apr. 2021.

[5] H. Chen and H. Lin, "An effective hybrid atrous convolutional network for pixel-level crack detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, Apr. 2021.

[6] T. Nishikawa, J. Yoshida, T. Sugiyama, and Y. Fujino, "Concrete crack detection by multiple sequential image filtering," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 27, no. 1, pp. 29–47, Jan. 2012.

[7] Y. J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, May 2017.

[8] S. Zhou and W. Song, "Deep learning-based roadway crack classification using laser-scanned range images: A comparative study on hyperparameter selection," *Autom. Construction*, vol. 114, Jun. 2020, Art. no. 103171.

[9] W. Choi and Y. J. Cha, "SDDNet: Real-time crack segmentation," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 8016–8025, Sep. 2020.

[10] W. Jiang, M. Liu, Y. Peng, L. Wu, and Y. Wang, "HDCB-Net: A neural network with the hybrid dilated convolution for pixel-level crack detection on concrete bridges," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5485–5494, Aug. 2021.

[11] S. Jiang and J. Zhang, "Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 6, pp. 549–564, Jun. 2020.

[12] K. Jang, Y. K. An, B. Kim, and S. Cho, "Automated crack evaluation of a high-rise bridge pier using a ring-type climbing robot," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 1, pp. 14–29, Jan. 2021.

[13] J. Valença, I. Puente, E. Júlio, H. González-Jorge, and P. Arias-Sánchez, "Assessment of cracks on concrete bridges using image processing supported by laser scanning survey," *Construction Build. Mater.*, vol. 146, pp. 668–678, Aug. 2017.

[14] J. Valença and E. Julio, "MCrack-Dam: The scale-up of a method to assess cracks on concrete dams by image processing. The case study of itaipu dam, at the Brazil–Paraguay border," *J. Civil Struct. Health Monit.*, vol. 8, no. 5, pp. 857–866, Oct. 2018.

[15] M. R. Jahanshahi, S. F. Masri, C. W. Padgett, and G. S. Sukhatme, "An innovative methodology for detection and quantification of cracks through incorporation of depth perception," *Mach. Vis. Appl.*, vol. 24, no. 2, pp. 227–241, Feb. 2013.

[16] Y. F. Liu, S. Cho, B. F. Spencer Jr, and J. S. Fan, "Concrete crack assessment using digital image processing and 3D scene reconstruction," *J. Comput. Civil Eng.*, vol. 30, no. 1, Jan. 2016, Art. no. 04014124.

[17] Y. F. Liu, X. Nie, J. S. Fan, and X. G. Liu, "Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 5, pp. 511–529, May 2020.

[18] C. M. Yeum, J. Choi, and S. J. Dyke, "Autonomous image localization for visual inspection of civil infrastructure," *Smart Mater. Struct.*, vol. 26, no. 3, Feb. 2017, Art. no. 035051.

[19] S. Cho, S. Park, G. Cha, and T. Oh, "Development of image processing for crack detection on concrete structures through terrestrial laser scanning associated with the octree structure," *Appl. Sci.*, vol. 8, no. 12, pp. 1–19, Nov. 2018.

[20] T. Yamaguchi and S. Hashimoto, "Practical image measurement of crack width for real concrete structure," *Electron. Commun. Jpn.*, vol. 92, no. 10, pp. 1–12, Oct. 2009.

[21] R. S. Adhikari, O. Moselhi, and A. Bagchi, "Image-based retrieval of concrete crack properties for bridge inspection," *Autom. Construction*, vol. 39, pp. 180–194, Apr. 2014.

[22] 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, Apr. 2019.

[23] C. Zhang, C. C. Chang, and M. Jamshidi, "Simultaneous pixel-level concrete defect detection and grouping using a fully convolutional model," *Struct. Health Monit.*, vol. 20, no. 4, pp. 2199–2215, Jan. 2021.

[24] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, pp. 1–6, 2018.

[25] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++: Better real-time instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 1108–1121, Feb. 2022.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Rep.*, 2015, pp. 1–15.

[27] Y. Furukawa and C. Hernández, "Multi-view stereo: A tutorial," *Found. Trends Comput. Graph. Vis.*, vol. 9, no. 1/2, pp. 1–148, 2015.

[28] J. Taraben and G. Morgenthal, "Methods for the automated assignment and comparison of building damage geometries," *Adv. Eng. Informat.*, vol. 47, Jan. 2021, Art. no. 101186.

[29] C. Wu, "Towards linear-time incremental structure from motion," in *Proc. Int. Conf. 3D Vis.*, 2013, pp. 127–134.

[30] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[31] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1362–1376, Aug. 2010.

[32] F. Bosché, "Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction," *Adv. Eng. Informat.*, vol. 24, no. 1, pp. 107–118, Jan. 2010.

[33] P. Glira, N. Pfeifer, C. Briese, and C. Ressl, "A correspondence framework for ALS strip adjustments based on variants of the ICP algorithm," *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2015, no. 4, pp. 275–289, Aug. 2015.

[34] F. C. Chen, M. R. Jahanshahi, R. T. Wu, and C. Joffe, "A texture-based video processing methodology using bayesian data fusion for autonomous crack detection on metallic surfaces," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 4, pp. 271–287, Apr. 2017.

[35] T. C. Lee, R. L. Kashyap, and C. N. Chu, "Building skeleton models via 3-D medial surface axis thinning algorithms," *CVGIP, Graph. Models Image Process.*, vol. 56, no. 6, pp. 462–478, Nov. 1994.

[36] Z. Zhu, S. German, and I. Brilakis, "Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation," *Autom. Construction*, vol. 20, no. 7, pp. 874–883, Nov. 2011.

[37] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Softw.*, vol. 3, no. 3, pp. 209–226, Sep. 1977.

Chaobo Zhang (Member, IEEE) received the master's and Ph.D. degrees in civil engineering from Hong Kong University of Science and Technology, Hong Kong, in 2017 and 2020, respectively.

He is currently a Postdoctoral Fellow with the Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen, China. His research interests include intelligent sensing, big data analysis, deep learning and computer vision for industrial applications.

Maziar Jamshidi received the M.Sc. degree in structural engineering from the Sharif University of Technology, Tehran, Iran, in 2012, and the Ph.D. degree in civil engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2018.

He is currently a Postdoctoral Associate with the Civil Engineering Department, University of Calgary, Calgary, AB, Canada. He has also been working in the field of structural control with a focus on regenerative damping systems and energy harvesting. His research interests include civil structural health monitoring, including the application of deep learning for visual inspection and vibration-based damage identification of infrastructure systems.

Chih-chen Chang received the master's and Doctoral degrees in aeronautics and astronautics from Purdue University, West Lafayette, IN, USA, in 1989 and 1993, respectively.

He is currently a Professor Emeritus with the Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong. His research interests included condition monitoring, control and assessment of large scale structures, innovative signal and data processing for civil engineering applications.

Zhiwen Chen (Member, IEEE) received the B.S. degree in electronic information science and technology and the M.S. degree in electronic information and technology from Central South University, Changsha, China, in 2008 and 2012, respectively, and the Ph.D. degree in electrical engineering and information technology from the University of Duisburg-Essen, Duisburg, Germany, in 2016.

He is currently an Associated Professor with Central South University. His research interests include model-based and data-driven fault diagnosis and health monitoring, data analytics, etc.

Xiaojun Liang (Member, IEEE) received the B.Eng. degree in engineering mechanics and aerospace from Tsinghua University, Beijing, China, in 2012, and the Ph.D. degree in mechanical engineering and applied mechanics from the University of Pennsylvania, Philadelphia, PA, USA, in 2017.

From 2017 to 2020, he was a Senior R&D Engineer with intelligent robot division of JD Group. Since 2021, he is currently an Assistant Research Fellow with Peng Cheng Laboratory, Shenzhen, China. His research interests include mechanical modeling and analysis, industrial intelligent system, industry control optimization based on mechanism and data fusion, etc.

Weihua Gui received the B.S. degree in electrical engineering and the M.S. degree in control engineering from Central South University, Changsha, China, in 1976 and 1981, respectively.

From 1986 to 1988, he was a Visiting Scholar with the University GH Duisburg, Duisburg, Germany. Since 1991, he has been a Full Professor with Central South University, Changsha, China. Since 2013, he has been an Academician with the Chinese Academy of Engineering, Beijing, China. His current research interests include modeling and optimal control of complex industrial processes, distributed robust control, and fault diagnosis.