# VISION-BASED APPROACHES FOR QUANTIFYING CRACKS IN CONCRETE STRUCTURES

P. Shokri [1,*], M. Shahbazi [2], D. Lichti [3], J. Nielsen [1]

[1] Dept. of Electrical Engineering, University of Calgary, Calgary, T2N 1N4 Canada – (parnia.shokri1, nielsenj) @ucalgary.ca,
[2] Centre de géomatique du Québec, Saguenay, G7H 1Z6 Canada – mshahbazi@cgq.qc.ca
[3] Dept. of Geomatics Engineering, University of Calgary, Calgary, T2N 1N4 Canada – ddlichti@ucalgary.ca

**Technical Commission II**

**KEY WORDS:** Stereo Vision, Deep Learning, 3D Reconstruction, Semantic Segmentation, Concrete Cracks

**ABSTRACT:**

In this paper, a combination of photogrammetric, computer-vision, and deep-learning approaches are proposed for accurate detection and quantification of cracks from the images of concrete structures. In particular, a semantic segmentation approach using UNet is applied, which is trained on a customized dataset of real-world images. Then, two photogrammetric methods are assessed for reconstructing the full figure of the cracks from stereo images. One approach is based on detecting the dominant structural plane surrounding the crack and projecting the crack pixels to this 3D plane. The second approach is based on matching the crack pixels across two images. To be able to perform the 3D reconstructions accurately, a rigorous calibration of the intrinsic calibration parameters of the cameras is performed. The relative orientation parameters between the stereo cameras are also determined in the calibration procedure. Extensive experiments are performed to evaluate each phase of this detection-and-quantification workflow. In general, cracks can be detected with an average precision of 87.48% and recall of 87.45%. They can be reconstructed in 3D with an accuracy as high as 0.05 mm.

## 1. INTRODUCTION

In order to maintain infrastructure in a healthy status, its conditions must be inspected regularly (Lattanzi and Miller, 2017). In general, the task of inspection is comprised of two components: visual inspection and metric inspection (Jahanshahi and Masri, 2013; Koch et al., 2014). While visual inspection refers to visually examining condition of an asset and identifying its damage, metric inspection refers to accurately measuring the facts that may have an impact on the health condition of the asset. In particular, optical and image-based approaches have gained popularity for structural health monitoring (Kerle et al., 2019) due various reasons: 1) they are contact-less and non-destructive methods; 2) the equipment (mainly cameras) for this purpose is inexpensive; 3) the full figure of the structure's damage or deformation can be provided as a large field of view can be covered by cameras; 4) techniques of computer vision, deep learning and machine learning can be used to automate the inspection tasks; 5) integrating these solutions with robotic systems allows real-time collection and processing of information (Jahanshahi et al., 2013; Lee et al., 2019; Mohan and Poobal, 2018).

In this paper, we propose a flow of image-based approaches for automatic detection and quantification of the full figures of cracks in concrete structures. Figure 1 shows two samples of images captured from concrete sidewalks and building façades. The camera used in this study is a commercial stereo camera, the ZED system (Stereo Labs, USA), as shown in Figure 2. The cameras have sensors of 2208x1242 pixels and are equipped with wide-angle lenses with a focal length of 2.8 mm. As such, the effect of lens distortions can be significant in the quality of stereo reconstruction. Therefore, the first specific objective of this study is accurately calibrating the low-cost, wide-angle, stereo ZED camera not only to eliminate the distortions but also to determine the relative orientation parameters between the two cameras.

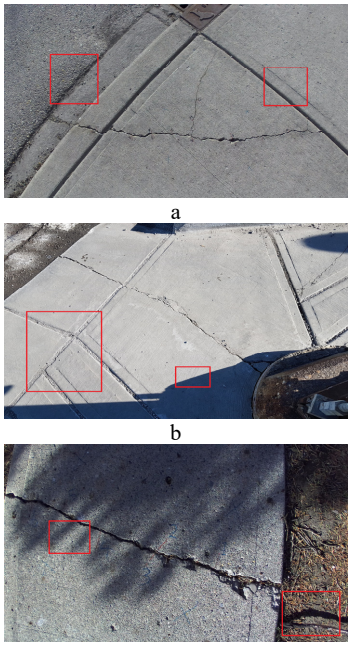
Figure 1. Samples of images with cracks on concrete structures


Figure 2. ZED stereo camera (Image source: stereolabs.com)

Detecting cracks in concrete structures is generally a very challenging task because the structure surface is covered by all types of crack-like features, e.g., concrete joints, artificial patterns, and shadow outlines (Figure 3). Therefore, the second specific objective of this study is to implement an automated image-based crack detection solution. Finally, if one can

* Corresponding author

reconstruct the cracks in three dimensions (3D), the status of the crack can be assessed. For instance, cracks longer or wider than a certain threshold can be identified for immediate maintenance. Therefore, the third specific objective of this study is 3D reconstructing the full figure of cracks with sub-millimeter level of accuracy. The task of 3D reconstruction from images is a well-studied topic. Several photogrammetric approaches, either based on stereo vision or structure from motion (SfM), exist to complete this task. In the case of quantifying cracks, the "scale" of 3D reconstruction is of paramount importance. The main reason is that only a few millimeters of error can affect our conclusion as to whether the crack represents considerable damage or not. A stereo rig of cameras provides the advantage of defining the object scale without the need for external observations if the baseline of the stereo cameras is accurately determined. In the case of SfM, one needs to access an external observation (e.g., reference distances between tie points) to define the scale (Jahanshahi and Masri, 2013; Kim et al., 2017). Another solution to determine the scale is assuming that the crack is lying on a planar surface whose 3D model relative to the camera is known. This assumption is often correct as concrete structures are piece-wise planar objects (Shan et al., 2016).



a

b

c

Figure 3. Examples of concrete structures with features that visually resemble cracks; a) concrete joints; b) artificial patterns; c) outlines of shadows from vegetations

## 2. METHODOLOGY

### 2.1 Camera Calibration

Although the manufacturer (Stereo Labs) provides the intrinsic calibration parameters of the two cameras as well as their mounting parameters, we found out that these parameters need considerable improvement to yield millimeter-level of accuracy in 3D reconstruction. Equation (1) describes the model used by the factory for calibrating the cameras.

$$x_{ij} - (-f_{x_m} x'' + C_{x_m}) = 0$$
$$y_{ij} - (f_{y_m} y'' + C_{y_m}) = 0$$

where

$$\begin{bmatrix} u_{ij} & v_{ij} & w_{ij} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \mathbf{R}_o^j & -\mathbf{R}_o^j \mathrm{C}_j \end{bmatrix} \tilde{\mathrm{X}}_i$$

$m$: camera index $\in \{0,1\}$

$i$ : tie-point index

$j$ : image index

$o$ : denoting object coordinate system

$$y''_{ij} = y'_{ij}\,(1 + K_{1_m} r_{ij}^{\,2} + K_{2_m} r_{ij}^{\,4} + K_{3_m} r_{ij}^{\,6}) + 2P_{2_m} x'_{ij} y'_{ij} + P_{1_m}\left(r_{ij}^{\,2} + 2\,y'^{\,2}_{ij}\right)$$

$$x''_{ij} = x'_{ij}\,(1 + K_{1_m} r_{ij}^{\,2} + K_{2_m} r_{ij}^{\,4} + K_{3_m} r_{ij}^{\,6}) + 2P_{1_m} x'_{ij} y'_{ij} + P_{2_m}\left(r_{ij}^{\,2} + 2\,x'^{\,2}_{ij}\right)$$

$$x'_{ij} = u_{ij} \,/\, w_{ij}$$

$$y'_{ij} = v_{ij} \,/\, w_{ij}$$

$$\mathrm{r}^2_{ij} = x'^{\,2}_{ij} + y'^{\,2}_{ij}$$

$K_{1_m}, K_{2_m}, K_{3_m}$ : radial lens distortion coefficients

$P_{1_m}, P_{2_m}$ : tangential lens distortion coefficients

(1)

In Equation (1), $\mathbf{R}$ is the rotation matrix from the object to the camera coordinate system, C is the position of the image perspective center in the object coordinate system, $\tilde{\mathrm{X}}$ is the homogeneous coordinates of the object point, and $(x, y)$ are the observations of the point in the image. ($C_x$, $C_y$, $f_x$, $f_y$) are the interior orientation parameters. These equations were integrated into a self-calibrating free-network bundle adjustment (Shahbazi et al., 2017) for calibrating the cameras. Relative orientation (RO) stability constraints were also added to the self-calibrating bundle adjustment. For an image pair, $k$, the relative rotations $\mathbf{R}_l^r$ and translations $\mathrm{r}_l^r$ between the left camera and right camera can be denoted as follows, where $l$ denotes the left camera and $r$ denotes the right camera in the stereo rig.

$$\mathbf{R}_r^l = \mathbf{R}_o^{l_k} (\mathbf{R}_o^{r_k})^{\mathrm{T}}$$
$$\mathrm{r}_r^l = \mathbf{R}_o^{l_k} (\mathrm{r}_{r_k}^o - \mathrm{r}_{l_k}^o)$$

(2)

This equation must hold for all image pairs. That is, for a different image pair, $h$, the stability of the RO parameters obliges Equation (3). Despite being in the form of translation vectors, these equations involve rotation matrices too; i.e., they constrain both the lever arm offsets and the boresight angles.

$$\mathbf{R}_o^{r_h} (\mathrm{r}_{l_h}^o - \mathrm{r}_{r_h}^o) = \mathbf{R}_o^{r_k} (\mathrm{r}_{l_k}^o - \mathrm{r}_{r_k}^o)$$
$$\mathbf{R}_o^{l_h} (\mathrm{r}_{r_h}^o - \mathrm{r}_{l_h}^o) = \mathbf{R}_o^{c_k} (\mathrm{r}_{r_k}^o - \mathrm{r}_{l_k}^o)$$

(3)

Our experiments showed that the model of Equation (1) was insufficient to accurately model all the distortions in the stereo images of the ZED camera. As such, in another calibration attempt, the collinearity equations of the perspective projection model were augmented with five radial lens distortion terms as well as affine sensor distortion terms (Equations (4) and (5)).

$$x_{ij} - c_{x_m} + \delta_{x_{ij}} + f_m \frac{u_{ij}}{w_{ij}} = 0$$
$$y_{ij} - c_{y_m} + \delta_{y_{ij}} + f_m \frac{v_{ij}}{w_{ij}} = 0$$

(4)

The interior orientation parameters of the cameras include $(c_x, c_y, f)$, and $(\delta_x, \delta_y)$ are the distortion corrections that are modeled as follows.

$$\begin{aligned}\delta_{x_{ij}} &= (x_{ij} - c_{x_m})(k_{1_m} r_{ij}^2 + k_{2_m} r_{ij}^4 + k_{3_m} r_{ij}^6 + k_{4_m} r_{ij}^8 + k_{5_m} r_{ij}^{10}) \\ &\quad + p_{1_m}(r_{ij}^2 + 2(x_{ij} - c_{x_m})^2) + 2p_{2_m}(x_{ij} - c_{x_m})(y_{ij} - c_{y_m}) \\ &\quad + s_{1_m}(x_{ij} - c_{x_m}) + s_{2_m}(y_{ij} - c_{y_m}) \\ \delta_{y_{ij}} &= (y_{ij} - c_{y_m})(k_{1_m} r_{ij}^2 + k_{2_m} r_{ij}^4 + k_{3_m} r_{ij}^6 + k_{4_m} r_{ij}^8 + k_{5_m} r_{ij}^{10}) \\ &\quad + p_{2_m}(r_{ij}^2 + 2(y_{ij} - c_{y_m})^2) + 2p_{1_m}(x_{ij} - c_{x_m})(y_{ij} - c_{y_m})\end{aligned} \quad (5)$$

where:

$r_{ij} = \sqrt{(x_{ij} - c_{x_m})^2 + (y_{ij} - c_{y_m})^2}$

$k_{1_m}, k_{2_m}, k_{3_m}, k_{4_m}, k_{5_m}$ : radial lens distortion parameters

$p_{1_m}, p_{2_m}$ : decentering lens distortion parameters

$s_{1_m}, s_{2_m}$ : sensor affine distortion parameters

However, unmodelled systematic errors in the residuals along the x-direction could still be observed with this calibration model. Therefore, inspired by the models presented in (Lichti et al., 2015), a quartic polynomial as a function of x (Equation (6)) was added to the distortion terms in the x-direction while considering only three radial lens distortion coefficients. This model significantly reduced the systematic errors in the residuals. Further discussions around this subject will be provided in Section 3.

$$\delta_{x_{ij}}^{poly} = s_{3_m}(x_{ij} - c_{x_m})^2 + s_{4_m}(x_{ij} - c_{x_m})^3 + s_{5_m}(x_{ij} - c_{x_m})^4 \quad (6)$$

### 2.2 Crack Detection

In this paper, a semantic segmentation approach based on convolutional neural networks is used for detecting cracks from RGB images. In particular, the UNet network (Ronneberger et al., 2015), is applied. The architecture of UNet includes two paths. The first path is the encoder that captures the global context of the image. The second path is the decoder, which allows localizing the object precisely. Since cracks are fine objects with a semantic context in a concrete background, this network architecture is effective for crack segmentation.

In order to train this network, an adequate number of annotated images is required. There are a few datasets publicly available for crack segmentation. Still, these datasets do not represent the challenging scenarios in real life and only consist of very close-range and simple images. Besides, studies of crack identification are conventionally applied to low-resolution images up to 227x227 pixels (Dung and Anh, 2019). To address this data-adequacy challenge, a new dataset is collected and labeled. The dataset consists of 670 challenging RGB images and their corresponding semantic segmentation labels. For further enhancement, this dataset is merged with a public dataset available on Mendeley Data (Özgenel, 2019), which includes 458 simpler images. The overall dataset is resized to 512x512 pixels for our experiments. This dimension of training images is considered high-resolution compared to the existing models in (Neff et al., 2017) and (Dung and Anh, 2019), which use 128x128 and 227x227 image sizes only. As the images in the training set are resized to 512x512 pixels, some information is lost in high-detailed scenes. As a result, before resizing the images, for every training image, a few patches of size 512x512 pixels are manually sampled. These patches are then added to the training set as well. Adding these patches to the training set improves the segmentation method, specifically for detecting smaller cracks. Randomly, 80% of our original dataset (80% of 670 images), as well as 80% of the Mendeley dataset (80% of 458 images), are used to form the training set, resulting in a total of 902 images. Then, the corresponding patches of each training image from our share of the original data are also added to the dataset. For testing the performance of our segmentation approach, an independent set of 226 images (with no patches) is used. This testing set consists of the remaining 20% of our original dataset and the remaining 20% of the Mendeley dataset.

In order to reduce the risk of overfitting, data augmentation techniques should be used (Krizhevsky et al., 2012). In this paper, to augment the training dataset, affine transformations including random rotations from -45° to 45°, scaling from 0.8 to 1.2, and shear from -15° to 15° are used. Moreover, a random four-point perspective transformation is applied. As a result of this data augmentation, the training dataset is doubled in size.

The input/output layers of the original UNet network are adjusted for high-resolution input/output, i.e., 512x512 pixels. The UNet model is trained for 220 epochs with a batch size of 4. In each epoch, the network is iterated through all the training images. Adam Optimizer is used as the update rule with the learning rate of 1e-4, beta-1 of 0.9, and beta-2 of 0.999. All the other parameters of the network are similar to the original work of Ronneberger et al. (2015).

It is worth mentioning that for preparing the ground truth of our dataset, an annotation algorithm is developed following the work of Jahanshahi et al. (2013) based on morphological opening and closing operations. This approach makes the annotation task semi-automatic, i.e., the user must only clean up the false positive crack pixels detected by the algorithm instead of manually labeling the complete figure of the crack.

### 2.3 3D Reconstruction of Cracks

Two photogrammetric approaches are assessed to reconstruct the detected cracks in 3D. The first 3D reconstruction approach assumes that the cracks are lying on a planar object. To implement this approach, first, salient features from the stereo images are detected using Harris corner detector (Derpanis, 2004). Their descriptors are extracted using SURF approach (Bay et al., 2008), and they are robustly matched across the two images given the knowledge of the relative orientation parameters (obtained through the calibration). Then, the calibration parameters (including the RO parameters) of the stereo cameras are used in a simple spatial intersection process to determine the 3D coordinates of these tie points. Next, using Random Sample Consensus (RanSaC), the dominant plane, which best fits the tie points immediately surrounding the detected crack, is determined. Finally, the collinearity equations for crack pixels observed in the left image along with additional constraints to enforce the crack points to lie on the dominant plane are solved (with zero degrees of freedom) to determine the 3D coordinates of each point (Förstner and Wrobel, 2016). The main advantage of this approach is its speed since images do not need to be undistorted and rectified. Moreover, the crack needs to be detected only in the left image. Our experiments show that a crack averagely consisting of 50,000 pixels can be reconstructed in 0.06 seconds (CPU usage only). The main drawback of this approach is that the depth of the cracks can no longer be estimated.

In the second approach, no assumption about the shape of the concrete structure is made. Instead, the detected pixels from the left and right images are matched to each other. This method is more time-consuming since we need to not only perform the deep-learning-based detection on both images but also must complete the matching process for all the crack pixels. To facilitate matching, images are first undistorted and then stereo-rectified using the method of Fusiello et al. (2000). As such corresponding epipolar lines become parallel, and the search space for matching becomes very small. Basically, for each crack point from the left image, the corresponding crack point is one of the detected pixels that lie on their corresponding epipolar line in the right image. Therefore, matching can be done faster and with fewer errors. The main advantage of this approach is that, for larger cracks, the depth of the cracks can also be estimated. On the other hand, it takes a longer time for rectifications, and it is prone to matching errors.

## 3. EXPERIMENTS AND RESULTS

### 3.1 Calibrations

In order to self-calibrate this stereo camera system, a total of 92 image pairs were captured from the multi-depth, multi-resolution test-field of Figure 4.

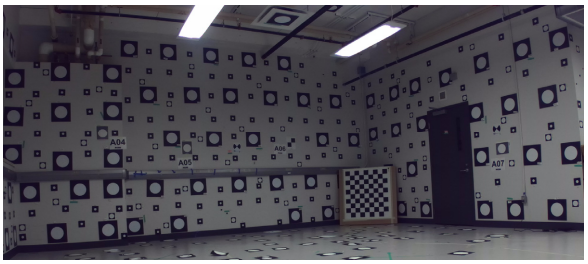


Figure 4. The calibration test-field

The test-field included 132 targets, and a total of 16760 observations were made in the calibration images. An independent set of 25 image pairs were also captured to check the accuracy of the calibrations. The network configuration for the calibration and the test images are shown in Figure 5 and Figure 6, respectively. When calibrating the camera with the model of Equation (5), an unmodelled error-pattern was noticed in the residuals. This pattern specifically existed in the residuals along the x-direction, and a high correlation between them and the x-coordinates was observed (Figure 7). A similar pattern, with residuals of larger magnitudes, was also observed when using the calibration model suggested by the manufacturer (Figure 8). As discussed in Section 2.1, the model of systematic errors in the x-direction was augmented with a quartic polynomial. As a result, the unmodelled systematic errors in the residuals were almost removed (Figure 9). The root-mean-square (RMS) and standard deviation (StD) of the magnitude of the residuals were also reduced to 0.05, 0.04 pixels from 0.08, and 0.04 pixels, respectively. In the calibration process, the relative orientation parameters between the two cameras were also estimated with a precision of 0.67 mm for the baseline vector and 0.017 degrees for the relative rotations. In the check images, the RMS of the residuals was 0.12 pixels using our calibration approach. Using the manufacturer parameters resulted in an RMS error of 1.18 pixels. The distribution of the residuals is shown in Figure 10. The RMS of the errors of 3D stereo reconstruction with our calibration parameters was only 0.9 mm while it was 25.2 mm with the manufacturer's parameters. These accuracies are measured based on our pre-knowledge of the exact size of the checkerboard pattern (Figure 10).

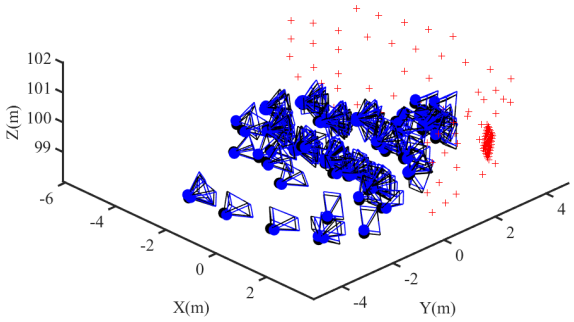



Figure 5. Network of calibration image pairs

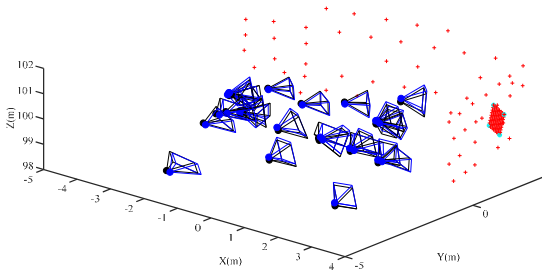



Figure 6. Network of check image pairs

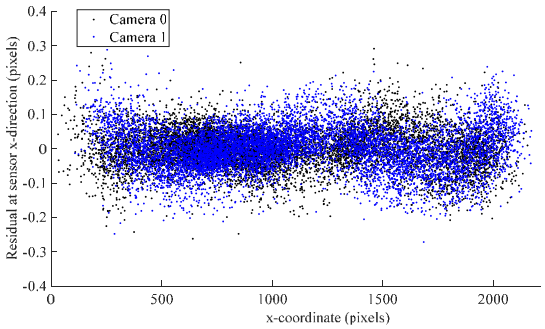
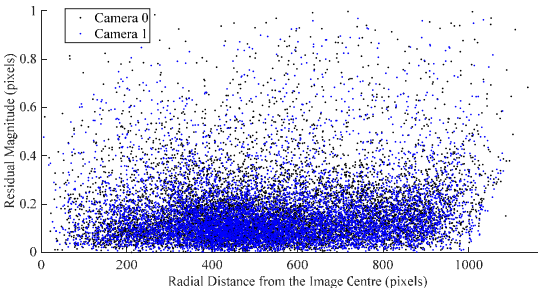
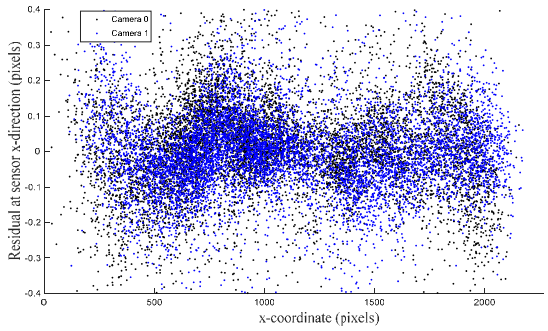



Figure 7. Unmodelled systematic errors in the residuals along the x-direction using the photogrammetric model including five radial lens distortion coefficients as described Equation (5)




a




b

Figure 8. Unmodelled systematic errors in the residuals using the manufacturer's calibration model as described in Equation (1); a) magnitude of the residuals; b) x-component of the residuals

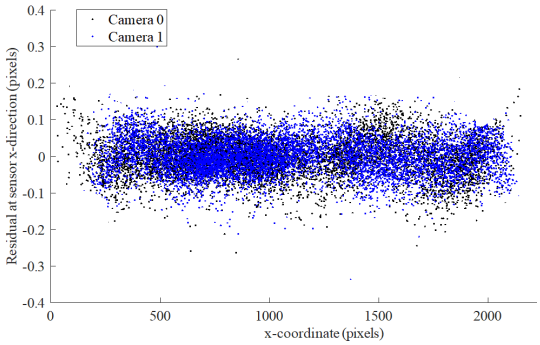
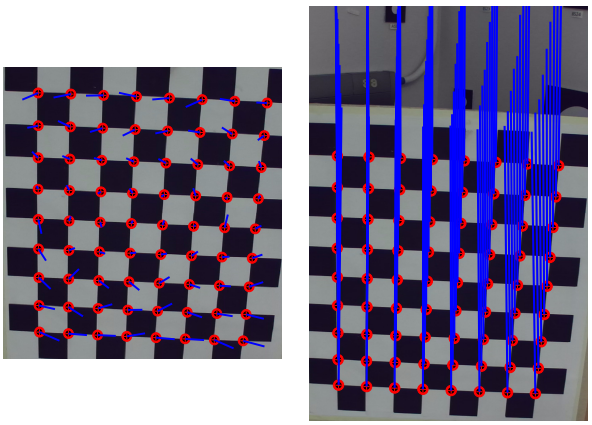



Figure 9. Residuals after adding the quartic distortion terms described in Equation (6)



a b

Figure 10. Residuals from the check data magnified with a factor of 500; a) our calibration; b) manufacturer's calibration

### 3.2 Crack Segmentation

In order to assess the performance of our UNet-based segmentation approach on the testing set, the average precision, recall, F1-score, and the Intersection over Union (IoU) measures were calculated. These results are shown in Table 1. The average F1-score was calculated by taking the mean of the F1-scores computed for each test image. The required time for detection on a test image in average is 60 milliseconds using a machine equipped with a GeForce GTX 1080 Ti graphical processing unit.

| Precision | Recall | F1 | IoU |
|---|---|---|---|
| 87.48 | 87.45 | 85.71 | 58.31 |

Table 1. Performance of UNet for crack segmentation

One of the most recent studies in the field of concrete crack segmentation (Dung and Anh, 2019) claims to outperform the state-of-the-art by gaining a maximum F1-score of 89.3% on very close-range images with a size of 227x227 pixels, in which cracks occupy a considerable area in each image. By achieving an F1-score of 85.71% using UNet, our proposed approach compares well to the state-the-art, with the benefit of being able to segment finer cracks with more challenging backgrounds captured in larger-size images. A few results of the test dataset are shown in Figure 15. As can be seen in this figure, the network successfully distinguishes between the crack-like features, such as concrete seams, and the real cracks even in complex scenarios. It is worth mentioning that although the test images were resized to 512x512 pixels to be fed to the segmentation network, the output masks were resized back to their original size for comparison with the ground-truth labels and measuring the performance variables of Table 1.

### 3.3 Crack quantification

To assess the performance of the suggested 3D reconstruction approaches, 10 stereo pairs were captured, as shown in Figure 16. The images were taken from 1 to 2 meters away from sidewalks. The average spatial resolution in these images was 0.914 mm. These images were fed to the segmentation network, and their cracks were detected. To collect some ground-truth data about the 3D structure of the cracks, few checkmarks were painted on the ground along the crack's length and width. The actual distances between these marks were manually measured with a caliper. In 10 pairs, a total of 47 ground-truth distances were collected. The checkmarks were distributed evenly along the cracks. Figure 11 shows a sample image with the checkmarks. Figure 12 displays a sample stereo pair after rectification; it can be seen than non-linearities due to distortions were well adjusted. Figure 13 presents the full extent of the crack in this stereo pair, which was reconstructed in 3D using the planar approach (the matching approach results in a very similar model).

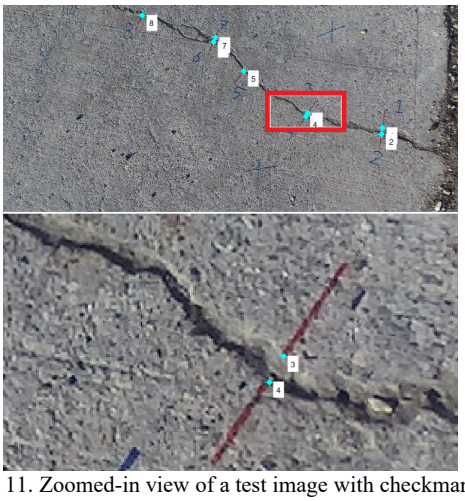
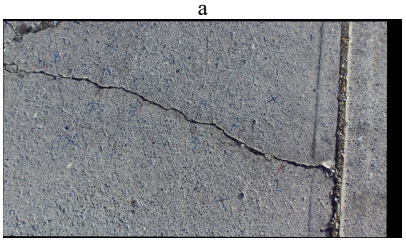


Figure 11. Zoomed-in view of a test image with checkmarks for assessing the accuracy of 3D reconstructions

The proposed approach, based on the planar assumption, resulted in reconstruction errors with an RMS of 1.23 mm and StD of 1.27 mm. The second approach, based on stereo matching and spatial intersection, resulted in errors with an RMS of 0.86 mm and StD of 0.72 mm. This approach resulted in a slightly higher 3D reconstruction accuracy. In both methods, the accuracy on some checkmarks was as high as 0.05 mm. The reconstruction accuracies from both methods were reasonable considering the average spatial resolution of 0.91 mm and the limited precision of checkmark detection in the images. We also noticed that the ZED camera always caused a sort of blur in the right edge of all left images and the left edge of all right images, as shown in Figure 14. Since many of our checkmarks were observed close to image edges, this blurring issue could have also affected the check accuracy.

a



b

Figure 12. Sample rectified stereo images (the original images belong to the 7th row of Figure 16)

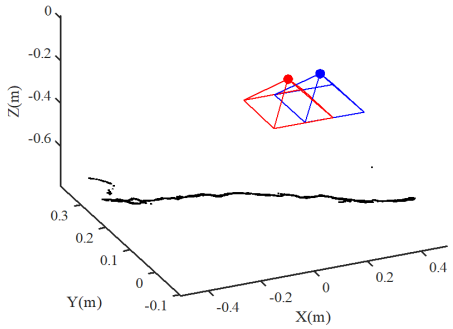
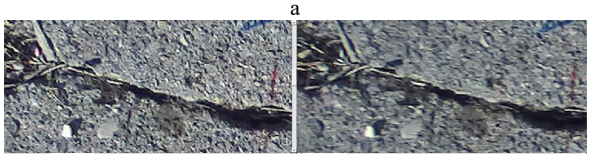



Figure 13. 3D model of a crack that was detected from the images of Figure 12

a



b

Figure 14. The camera generates a sort of blurring effect on the edges of the images; a) right edges of both images, where the left image is blurred; b) left edges of both images, where the right image is blurred

## 4. CONCLUSIONS

In this paper, a complete image-based workflow for identifying and characterizing cracks in concrete structures was proposed and evaluated. The crack detection approach was based on convolutional neural networks. Crack quantification was performed using photogrammetric approaches. The proposed approaches were applied to the stereo images captured by a commercial camera, ZED System. This wide-angle camera needed to be calibrated with additional parameters to yield acceptable accuracies. It was shown that our proposed calibration approach improved the accuracy by 96% compared to the manufacturer's calibrations. The proposed detection approach was able to segment cracks with a precision of 87.48% and F1-score of 85.71%. Despite the state-of-the-art, in which similar accuracies are achieved for detecting simple cracks from simple backgrounds in low-resolution images, our approach was tested on high-resolution images with complex cracks and backgrounds. In terms of 3D reconstructing the cracks, an RMS error of 0.86 mm was observed, which was reasonable given the spatial resolution of our test images and the limited precision of our checkmarks.

## ACKNOWLEDGEMENTS

This research was supported by multiple funding resources including NSERC Discovery Grant, Mitacs Accelerate, Canada's Tri-Council Agencies New Frontiers in Research Fund, and NSERC College and Community Applied Research and Development Grant. This study was also sponsored by Industrial SkyWorks Inc. (ON, Canada).

Figure 15. Samples of test images with the results of crack segmentation; left column shows the RGB images and right column shows the detection results

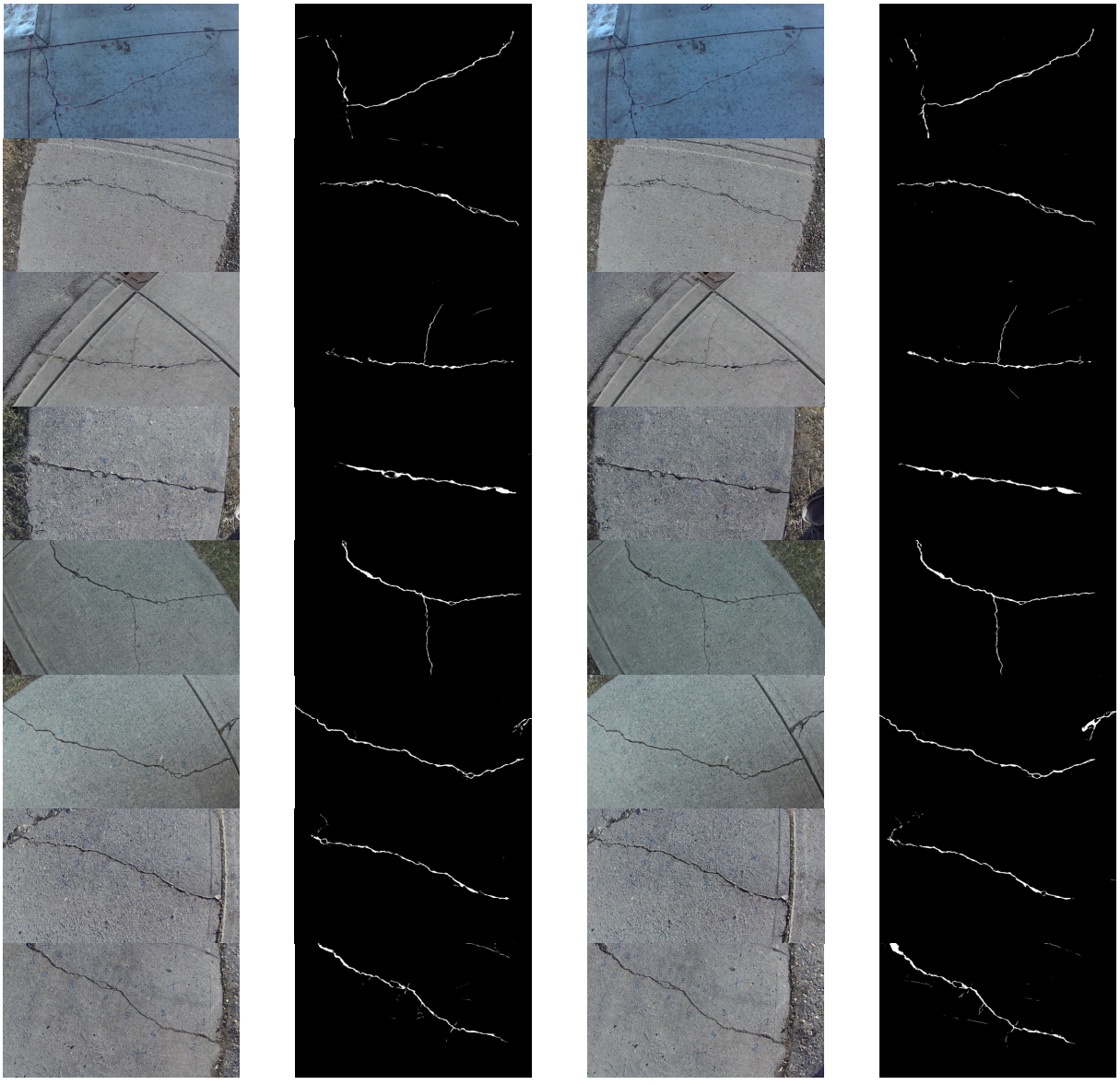


Figure 16. Test images for 3D reconstruction of the cracks; from left to right, the columns correspond to the left image of the stereo pair, the detected crack in the left image, the right image and the detected crack in the right image.

## REFERENCES

Bay H., Ess A., Tuytelaars T., Gool L.V.. 2008. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3), 346-359.

Derpanis, K., 2004. The Harris Corner Detector. Report, York University, ON, Canada.

Dung, C.V., Anh, L.D., 2019. Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, 99, 52-58.

Förstner, W., Wrobel, B.P., 2016. *Photogrammetric Computer Vision*. Springer Nature, Cham.

Fusiello, A., Trucco, E., Verri, A., 2000. Compact algorithm for rectification of stereo pairs. *Machine vision and applications*, 12(1), 16-22.

Jahanshahi, M.R., Masri, S.F., 2013. A new methodology for non-contact accurate crack width measurement through photogrammetry for automated structural safety evaluation. *Smart Materials and Structures*, 22(3), 035019.

Jahanshahi, M.R., Masri, S.F., Padgett, C.W., Sukhatme, G.S., 2013. An innovative methodology for detection and quantification of cracks through incorporation of depth perception. *Machine vision and applications*, 24(2), 227-241.

Kerle, N., Nex, F., Gerke, M., Duarte, D., Vetrivel, A., 2019. UAV-based structural damage mapping: A review. *ISPRS International Journal of Geo-Information*, 9(1), 14.

Kim, H., Lee, J., Ahn, E., Cho, S., Shin, M., Sim, S.H., 2017. Concrete crack identification using a UAV incorporating hybrid image processing. *Sensors*, 17(9), 2052.

Koch, C., Paal, S., Rashidi, A., Zhu, Z., König, M., Brilakis, I., 2014. Achievements and challenges in machine vision-based inspection of large concrete structures. *Advances in Structural Engineering*, 17(3), 303-318.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097-1105.

Lattanzi, D., Miller, G., 2017. Review of robotic infrastructure *inspection systems. Journal of Infrastructure Systems*, 23(3), 04017004.

Lee, Donghan, Kim, J., Lee, Daewoo, 2019. Robust Concrete Crack Detection Using Deep Learning-Based Semantic Segmentation. *International Journal of Aeronautical and Space Sciences*, 20(1), 287-299.

Lichti, D.D., Sharma, G.B., Kuntze, G., Mund, B., Beveridge, J.E., Ronsky, J.L., 2015. Rigorous geometric self-calibrating bundle adjustment for a dual fluoroscopic imaging system. *IEEE transactions on medical imaging*, 34(2), 589-598.

Mohan, A., Poobal, S., 2018. Crack detection using image processing: A critical review and analysis. *Alexandria Engineering Journal*, 57(2), 787-798.

Neff, T., Payer, C., Štern, D., Urschler, M., 2017. Generative Adversarial Network based Synthesis for Supervised Medical Image Segmentation. In *Proceedings of OAGM and ARW Joint Workshop*, 140-145.
Özgenel, Ç.F., 2019. Concrete Crack Segmentation Dataset. https://data.mendeley.com/datasets/jwsn7tfbrp/1 (3 Apr 2019).

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of International Conference on Medical image computing and computer-assisted intervention*, 234-241.

Shahbazi, M., Sohn, G., Théau, J., Ménard, P., 2017. Robust structure-from-motion computation: application to open-pit mine surveying from unmanned aerial images. *Journal of Unmanned Vehicle Systems*, 5(4), 126-145.

Shan, B., Zheng, S., Ou, J., 2016. A stereovision-based crack width detection approach for concrete surface assessment. *KSCE Journal of Civil Engineering*, 20(2), 803-812.

Shi, Y., Cui, L., Qi, Z., Meng, F., Chen, Z., 2016. Automatic road crack detection using random structured forests. *IEEE Transactions on Intelligent Transportation Systems*, 17(12), 3434-3445.