d Collection

Doctoral Dissertation

CONCRETE CRACK EVALUATION FOR CIVIL INFRASTRUCTURE USING COMPUTER VISION AND DEEP LEARNING

Hyunjun Kim

Department of Urban and Environmental Engineering
(Urban Infrastructure Engineering)

Graduate School of UNIST

2020

CONCRETE CRACK EVALUATION FOR CIVIL INFRASTRUCTURE USING COMPUTER VISION AND DEEP LEARNING

Hyunjun Kim

Department of Urban and Environmental Engineering
(Urban Infrastructure Engineering)
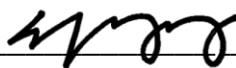
Graduate School of UNIST

Concrete Crack Evaluation for Civil Infrastructure Using Computer Vision and Deep Learning

A dissertation

submitted to the Graduate School of UNIST

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

Hyunjun Kim

January 7, 2020

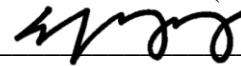Approved by

_____

Advisor

Myoungsu Shin

Concrete Crack Evaluation for Civil Infrastructure
Using Computer Vision and Deep Learning

Hyunjun Kim

This certifies that the dissertation of Hyunjun Kim is approved.

January 7, 2020

(signature)

_____
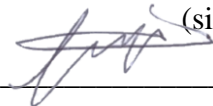
Committee Chair: Myoungsu Shin

(signature)
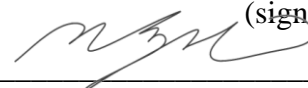
_____

Committee Member: Sung-Han Sim

(signature)

_____

Committee Member: Young-Joo Lee

(signature)

_____

Committee Member: Yun-Kyu An

(signature)

_____

Committee Member: Soojin Cho

(signature)

_____

Committee Member: Sukhoon Pyo

ABSTRACT

Concrete Crack Evaluation for Civil Infrastructure Using Computer Vision and

Deep Learning

by

Hyunjun Kim

Doctor of Philosophy in Engineering

Ulsan National Institute of Science and Technology

Surface cracks of civil infrastructure are one of the important indicators for structural durability and integrity.   Concrete cracks are typically investigated by manual visual observation on the surface, which is intrinsically subjective because it highly depends on the experience of inspectors.   Furthermore, manual visual inspection is time-consuming, expensive, and often unsafe when inaccessible structural components need to be assessed.   Computer vision-based approach is recognized as a promising alternative that can automatically extract crack information from images captured by the digital camera.   As texts and cracks are similar in terms of consisting distinguishable lines and curves, image binarization developed for text detection can be appropriate for crack identification purposes.   However, although image binarization is useful to separate cracks and backgrounds, the crack assessment is difficult to standardize owing to the high dependence of binarization parameters determined by users. Another critical challenge in digital image processing for crack detection is to automatically distinguish cracks from an image containing actual cracks and crack-like noise patterns (e.g., stains, holes, dark shadows, and lumps), which are often seen on the surface of concrete

structures.   In addition, a tailored camera system and the corresponding strategy are necessary to effectively address the practical issues in terms of the skewed angle and the process of the sequential crack images for efficient measurement.   This research develops a computer vision-based approach in conjunction with deep learning for accurate crack evaluation of for civil infrastructure.   The main contribution of the proposed approach can be summarized as follows: (1) a deep learning-based approach for crack detection, (2) a hybrid image processing for crack quantification, and (3) camera systems for the practical issues on civil infrastructure in terms of a skewed angle problem and an efficient measurement with the sequential crack images.   The proposed research allows accurate crack evaluation to provide a proper maintenance strategy for civil infrastructure in practice.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1   INTRODUCTION

Civil infrastructure is subjected to various loadings during their design life, such as environmental load, self-weight, and service load.   These kinds of loadings can induce structural damage and even failure, potentially resulting in social and economic losses.   Thus, most industrialized countries perform a regular inspection of civil infrastructure systems to evaluate structural soundness and provide the associated countermeasure.   Particularly in the case of concrete structures, surface cracks are one of the common items for the inspection process, because cracks often indicate structural damage and the corresponding problem with concrete durability [1-3].   Indeed, as cracks are an important indicator of structural health, the monitoring of surface cracks is considered as an essential process for structural maintenance.

Surface cracks of the concrete structure are typically investigated by manual visual inspection of the surface.   The crack information observed by inspectors, including direction, size, length, and width, are tracked over time to assess the current condition and anticipate the crack growth, assisting with maintenance plans.   Manual visual observation is the most common way in practice for monitoring surface cracks of the concrete structure to evaluate whether the crack growth over time would degrade structural safety.   However, manual visual observation is costly, labor intensive, and even inaccurate, because the inspection results inevitably depend on the personal opinion.

A wide variety of techniques have been introduced, including digital image processing, nondestructive evaluation and testing [4-7], dynamic property-based algorithm [8,9], and new crack sensor [10], overcoming the issues of manual visual inspection.   Digital image processing is considered as a powerful alternative to manual visual inspection, which can automatically extract crack information from images taken by digital cameras.   Digital image correlation (DIC) is also able to be utilized for crack detection purposes [11,12] by comparing

surface images before and after crack initiation. However, applying DIC is often impractical particularly when taking the reference image without cracks is unavailable. For reference-free crack assessment, Dare *et al*. [13] developed a route-finder algorithm that identifies a crack based on the connection of two given points provided by a user. The semi-automatic approach is found to be inefficient, when a number of crack images need to be processed. More recent studies have been devoted to presenting automatic methods with minimizing human intervention. Edge detection is also applied to crack detection, as the boundary between crack and background pixels can be found as edges [14,15]. Because the edges are often disconnected and undetected, additional post-processing is necessary to carefully perform a complete crack detection. Another approach is to calculate the difference in pixels associated with cracks and backgrounds, which are generally dark and bright, respectively. Image binarization algorithms are capable of transforming the dark and bright pixels in a grayscale image into a binary image that contains only black and white information. Although image binarization has been developed primarily for text detection [16-20], it has a strong potential to be utilized for crack identification purposes, because texts and cracks have essentially similar in terms of consisting distinguishable lines and curves [21,22]. Mathematical morphology, in conjunction with image binarization, is adopted for enhancing identification performance, transforming noisy crack pixels to realistic crack shapes [23-25]. Jahanshahi *et al*. [26] summarized more information with respect to crack detection methods based on digital image processing.

Among the diverse crack detection methods, image binarization is seen to have a strong potential for extracting crack information from images effectively. In the binarization process, pixels in a grayscale image with higher pixel values than a specified threshold are marked as one (white) in the converted binary image, while those with lower values become zero (black). Each binarization technique has its own scheme to calculate the threshold, generally based on

the statistical properties including the mean and standard deviation in pixel values. Thus, crack identification performance inevitably depends on the selected binarization method and the associated parameters.

On the other hand, deep learning consists of a cascade of multiple layers, which has been recently introduced as a powerful tool for identifying crack information [27-30]. Surface images of the concrete structure, labeled as either a cracked or an intact surface, are used for training a classification model based on convolution neural network (CNN) [31]. In the validation stage, the trained classification model is applied to new surface images. Previous studies that employed deep learning have successfully detected cracked regions; however, the classification in the presence of crack-like noncracks, which are unavoidable in real-world applications, was not fully studied. It is important to accurately detect and filter possible noncrack objects in concrete surface images. However, this problem has rarely been discussed in literature.

Furthermore, a tailored camera system and the corresponding strategy are necessary to perform accurate measurement of cracks on civil infrastructure, considering the following issues: (1) the skewed angle of view of the camera with respect to the concrete surface and (2) the absence of efficient crack identification strategy to fully extract crack information on the sequential images. Cracks on two-dimensional images may not be vertically aligned with the concrete surface because of the skewed angle of view of the camera. Thus, a three-dimensional (3D) space is required, particularly when taking images from a distance or from unmanned aerial vehicles (UAVs), because it is difficult to conveniently align the camera with the target surface. In addition, although previous studies that employed UAV can provide multiple crack images, crack measurements are difficult to efficiently conduct with a large number of the sequential crack images.

This research presents a computer vision-based crack evaluation strategy using deep

learning. The following three features are mainly addressed to conduct effective and automatic crack evaluation for providing a proper maintenance of civil infrastructure: (1) a deep learning-based approach for crack detection, (2) a hybrid image processing for crack quantification, and (3) camera systems for the practical issues on civil infrastructure in terms of a skewed angle problem and an efficient measurement with the sequential crack images.

CHAPTER 2　　BACKGROUND

2.1　　Computer Vision Techniques

Computer vision techniques have been recognized as a promising tool in the field of civil engineering for improved inspection and monitoring of civil infrastructure.　Images and videos taken from digital cameras are processed by computer vision techniques, potentially addressing the challenges in manual visual inspection in terms of time, cost, and safety. Automated damage detection is typically performed to identify concrete cracks that are one of the important indicators for structural soundness [13-15,21-30].　In addition, concrete spalling [32,33], fatigue cracks in steel [34], steel corrosion [35,36], and asphalt defects [37,38] are also considered to detect visual defects on civil engineering structures for inspection purposes.　As the localization of the obtained surface damages on civil infrastructure is a key step, a variety of computer vision techniques have been proposed for structural component recognition [39-41].　Here, automated classification of piers, slabs, and background is conducted on each digital image.　Furthermore, the monitoring of civil infrastructure based on computer vision techniques is employed to measure physical quantities, such as strains [11,12,42,43] and displacements [44,45].　In contrast the conventional approaches using wired or wireless contact sensors [46,47], computer vision techniques can provide the advantages of non-contact methods.　Advances in computer vision-based approaches for the inspection and monitoring of civil engineering structures are summarized [48].

In the following sections, a wide variety of computer vision techniques have been summarized for automated damage detection in civil infrastructure, particularly for crack identification purposes.　Beginning with the conventional image binarization methods, machine learning and the recent works of deep learning are arranged chronologically.

2.2 Image Binarization

Image binarization methods are used to convert a grayscale image into a binary image.   In the case of color images, pixel values of the original image must be turned into grayscale values ranging from zero (black) to 255 (white) by calculating a weighted sum of their red, green, and blue components.   In the binarization process, each pixel in the grayscale image is examined by comparing its pixel value with a threshold: a higher pixel value than the threshold leads to a pixel value of one (white) in the binary image, and a lower value results in a pixel value of zero (black), as can be seen in the example in Fig. 2.1.   For example, when the average pixel value is used as a threshold, the pixel value is higher than the corresponding threshold in A, and thus the binarization result is one (white), whereas the result is zero (black) in B, because the pixel value is lower than the threshold.   Binarization methods can be distinguished by the means they use to determine the threshold [16-20].   These involve the binarization parameters, including (1) the window in which the threshold calculation is conducted (see the blue and red boxes in Fig. 2.1) and (2) the sensitivity that controls the contributions of the statistical parameters of the pixel values to the threshold calculation.   After image binarization, the resulting image is then binarized with the pixel values zero and one, as shown in Fig. 2.1.

Grayscale Image

123	120	125	119	122	120	117
115	118	113	129	120	76	71
124	118	128	124	119	77	118
121	[A] 121	125	120	80	[B] 79	113
70	75	76	71	78	128	115
127	125	123	125	128	122	121

[A]
Pixel Value: 121
Threshold: 106
Binary Result: 1

[B]
Pixel Value: 79
Threshold: 101
Binary Result: 0

Binary Image

1	1	1	1	1	1	1
1	1	1	1	1	0	0
1	1	1	1	1	0	1
1	1	1	1	0	0	1
0	0	0	0	0	1	1
1	1	1	1	1	1	1
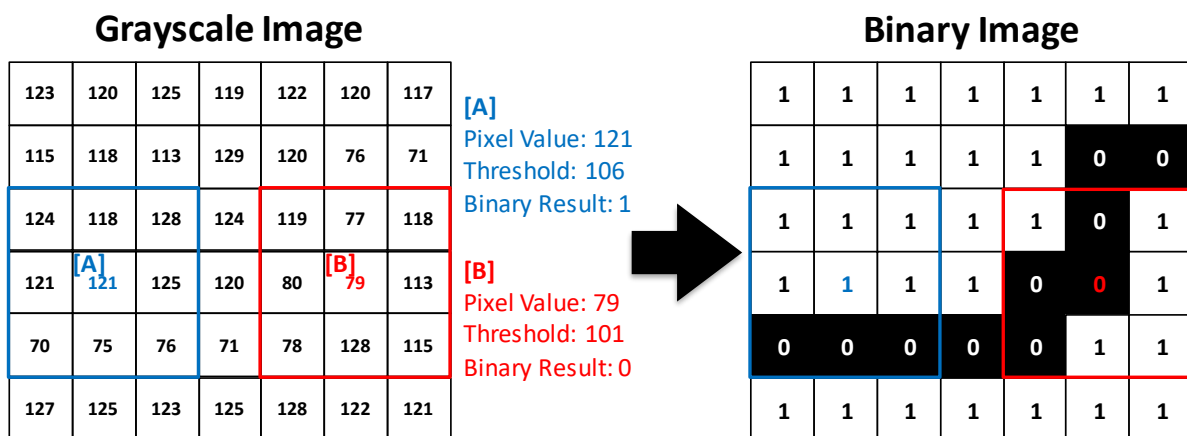
Fig. 2.1 Schematic demonstration of image binarization using 3 × 3 windows (thresholds are

6

selected manually for the demonstration).

A number of image binarization methods have been developed to acquire binary images more effectively and accurately, particularly for text identification purposes, such as Bernsen's, Niblack's, Sauvola's, and Wolf's methods, and the NICK method [16-20].   Bernsen [16] proposed Eq. 2.1 to calculate a threshold by considering maximum and minimum intensities of a selected window.

$$T_{Bernsen} = \frac{Z_{max} - Z_{min}}{2}$$
Eq. 2.1

where Z_{max} and Z_{min} are the maximum and minimum values in the pixel histogram of each window.   When $Z_{max} - Z_{min}$ evaluated in a window is less than a prescribed value, the full image is used for determining the threshold instead.   This method is generally useful for distinguishing a specific object from the background in the case of high-contrast images.

Niblack [17] considered the local mean and standard deviation of pixel values in a window:

$$T_{Niblack} = m + k \times s$$
Eq. 2.2

where k is the sensitivity, and m and s are the mean and standard deviation in a selected window, respectively.   While Niblack's method is simple and straightforward to use, its performance can be significantly degraded when the image background is noisy owing to its high dependency on the standard deviation.

Sauvola and Pietikäinen [18] modified Niblack's method to mitigate its sensitivity to the standard deviation by normalizing the standard deviation by a factor R, the dynamic range, as in Eq. 2.3.

$$T_{Sauvola} = m \times \left\{ 1 - k \times \left(1 - \frac{s}{R} \right) \right\}$$
Eq. 2.3

Sauvola's method is known to be effective for searching texts from noisy backgrounds as a

result of considering the dynamic range, unless the pixel-value difference between text and non-text is small.

To address the deficiency of Sauvola's method, Wolf and Jolion [19] normalized the contrast and mean in the equation for computing a threshold, as follows:

$$T_{Wolf} = (1-k) \times m + k \times M + k \times \frac{s}{R} \times (m-M) \qquad \text{Eq. 2.4}$$

where M is the minimum pixel value of the entire grayscale image.   In Wolf's method, the dark colors can be separated effectively from the backgrounds, because this method considers the minimum pixel value of the entire image in deciding a threshold.

Khurshid *et al*. [20] developed the NICK method based on Niblack's method by adding the mean square to the variance as follows:

$$T_{NICK} = m + k\sqrt{B + m^2} \qquad \text{Eq. 2.5}$$

where B is the variance.   This method shifts the threshold down by adding the mean square to the variance to delete background noises in the source image.   As shown, each image binarization method has its own equation to determine the threshold based on statistical properties of grayscale pixels in each window.   Thus, the image binarization results vary with the image binarization method used as well as the selected binarization parameters.

The primary goal of applying image binarization methods to the field of civil engineering is to accurately identify crack information such as widths and locations [21,22].   For this purpose, the sensitivity and window size of a binarization method must be properly determined to localize the exact crack pixels from the grayscale image.   For example, the pixel values of backgrounds in the grayscale image are higher than the pixel values of crack elements; thus, a large window size increases the threshold by including more background pixels, resulting in overestimated crack information in the binary image, as shown in Fig. 2.2.   Furthermore, sensitivity is a governing variable in the threshold equations of the binarization methods,

possibly leading to incorrect results, if an inappropriate value is assigned, as can be seen in Fig. 2.3. Thus, measured crack locations and widths can be inaccurate depending on the selected window and the sensitivity, while optimal binarization parameters tailored to crack identification have not been studied to date.

(a)   (b)   (c)
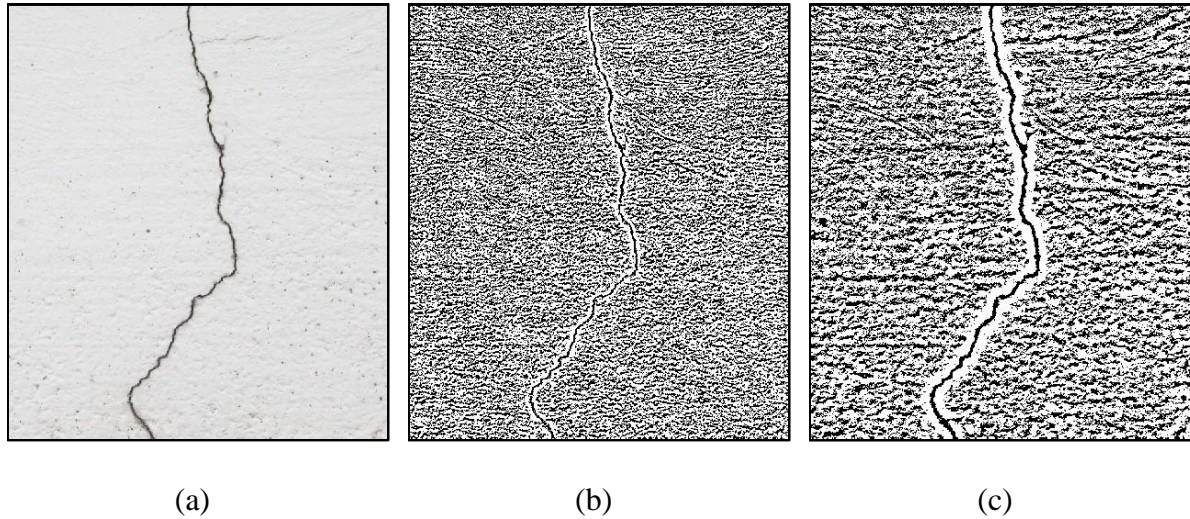
Fig. 2.2 Image binarization using Niblack's method with different window sizes: (a) original image, (b) 30 × 30 window, and (c) 100 × 100 window.

(a)   (b)   (c)
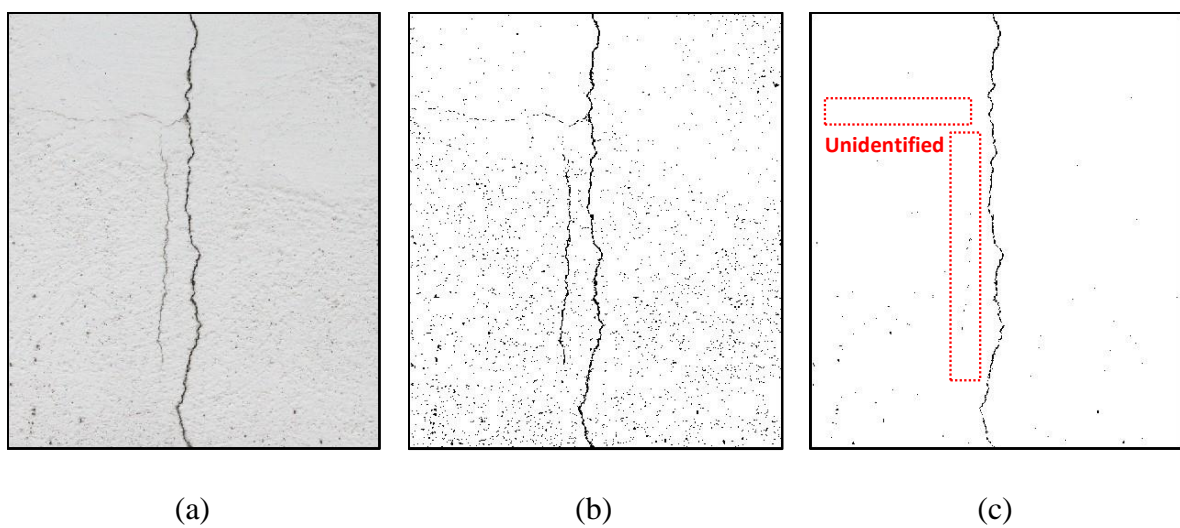
Fig. 2.3 Image binarization using Sauvola's method with different sensitivities: (a) original image, (b) sensitivity of 0.1, and (c) sensitivity of 0.5.

2.3 Machine Learning

Machine learning is recognized as a robust and automated tool in a wide variety of civil engineering applications. In particular, supervised learning, a type of machine learning, can be utilized to resolve crack recognition issues in conjunction with computer vision-based techniques. The combined approach typically involves identifying the unique characteristics of cracks and noncracks in the training set, which are further used in classification methods such as random forests [49] and support vector machines (SVMs) [50]. The trained model is subsequently applied to new images that are not utilized in the training set. For example, Csurka *et al.* [51] proposed a bag-of-words (BoW) model for the natural image classification of objects such as cars, phones, trees, and books. This process consists of three stages as follows: (1) feature extraction, (2) visual vocabulary construction, and (3) classification.

Feature extraction, a process of determining the unique characteristics of an image, is a vital part of object identification using image processing. The geometric patterns (e.g., eccentricity and number of pixels in each pixel group) and statistical properties of pixel intensities (e.g., mean and standard deviation) have been typically selected as important features to classify cracks and noncracks and thereby generate a classification model [52-55]. Furthermore, modern feature extraction algorithms adopted in the field of computer vision can be employed to detect the salient features of cracks to enable accurate identification [56-58]. In particular, speeded-up robust features (SURF) [58], one of the most widely used local feature detectors, has a proven performance in terms of computational time [59]. In contrast to Csurka *et al.* [51], who used scale-invariant feature transform (SIFT) [57] for feature extraction, SURF has a strong potential in terms of computational efficiency and high performance. SURF, which is designed to obtain distinctive features from digital images, consists of two main procedures: (1) interest point detection and (2) interest point description. To detect the interest points on elements such as blobs, corners, and edges, the determinant of the Hessian

matrix is used as a measure for evaluating the local change around each pixel. After the interest points are obtained, Haar wavelet responses are calculated within a circular neighborhood; an orientation is then assigned to each point using these responses. A square region is subsequently generated along the obtained orientation to address the image rotations. A feature vector with 64 elements is finally computed using the Haar wavelet responses in both the horizontal and vertical directions in 4×4 sub-regions.

The feature vectors of all the interest points are used to generate a visual word that serves as a representative, small image segment to demonstrate features such as color, shape, and surface texture. An image contains various interest points and the corresponding feature vectors; therefore, it is necessary to determine the characteristic features of cracks and noncracks to efficiently handle the large volume of images in the training stage. k-means clustering [60], which is a popular method for cluster analysis, is introduced to determine the representative clusters, in which the mean values of the feature vectors are the visual words. The results of the k-means clustering (i.e., visual words) are then grouped, and this group is called visual vocabulary or the bag of features.

To categorize the visual vocabulary through k-means clustering, Csurka *et al.* [51] used SVM, which is one of the most common classification algorithms owing to its robustness, computational efficiency, and resistance to over-fitting. When two different sets (i.e., cracks and noncracks) of images are trained for the classification, a visual vocabulary should be first generated from all the images using k-means clustering. Subsequently, the frequency of occurrence of the visual words in the vocabulary is calculated for each category. The obtained feature histograms are then inputted to the SVM to construct the classification model. Among the various SVM classifiers (e.g., linear, quadratic, cubic, and Gaussian), the linear SVM classifier, which is the most widely used. Although user-defined empirical thresholds are unnecessary in these methods, crack-like noncracks that share similar geometry and colors with

cracks still remain undistinguishable.   For an effective classification, advanced features need

to be extracted from cracks and noncracks to generate a robust classification model.

2.4   Deep Learning

Deep learning, which is inspired by the activities of the human brain, has recently been

recognized as a powerful method in a variety of research fields, including natural language

processing, speech recognition, and computer vision.   In particular for the field of computer

vision, various CNNs have been proposed for image classification purposes, including AlexNet

[61], VGGNet [62], ZFNet [63], GoogLeNet [64], and ResNet [65].   Among them, AlexNet

presented by Krizhevsky *et al.* [61] successfully classified natural images into 1,000 categories.

In contrast to the conventional machine learning, the architecture of AlexNet is a hierarchical

structure, having five convolutional layers and three fully connected layers.   Each

convolutional layer handles an input image having different kernels and the corresponding

sizes.   Furthermore, AlexNet is equipped with rectified linear units (ReLUs) and max pooling

between the convolutional layers to enhance the classification performance in terms of the

computational time and accuracy.   After passing through the convolutional layers, the output

will go through three fully connected layers with the softmax activation function to identify the

class of the image, such as animal, car, fruit, or vegetable.   Fig. 2.4 shows the overall process

of the deep learning and machine learning approaches, modified from the study by Zheng *et al.*

[66].   Note that the CNN-based method directly uses global features for the classification,

whereas the SURF-based method uses visual words clustered from local features.   A typical

method of applying CNN is to employ a scanning window, in which the input images are

divided into a number of sub-images with a fixed resolution, as shown in Fig. 2.4 [27,67].

The sub-images are manually categorized as either a cracked surface or as an intact surface to

build the classification model, which is utilized to localize the location of crack pixels with

additional binarization process [68,69]. However, when the objects are located at the edges in each sub-image, the identification is difficult to perform correctly. To properly handle the issue, although the results of CNN with the decision-making based on Naive Bayes [70] or infrared thermography images [71] shows strong potential, the scanning window is found to be still inefficient in that the intact surface, which takes up a majority of an image, has the highest influence in the training.

In addition to image classification, advanced technologies have been developed for single-object localization and multiple-object detection by combining possible object locations with a CNN, such as R-CNN [72], SPPNet [73], Fast R-CNN [74], Faster R-CNN [75], SSD [76], R-FCN [77], and YOLO [78]. As an alternative to the scanning window, these models can be employed to automatically detect multiple damages on digital images taken from a reinforced concrete [79], a tunnel [80], a road [81], an aircraft [82], and a sewer pipe [83,84]. Although multiple damages can be successfully recognized and localized by the previous models, manual annotation of damage types and the corresponding bounding boxes in each image for the ground truth is computationally inefficient. In addition, additional post processing is required to localize the location of crack pixels in each of the detected regions [85].

Recently, a variety of semantic segmentation models have been developed to perform pixel-to-pixel object masking, such as FCN [86], U-Net [87], SegNet [88], PSPNet [89], and DeepLab [90]. The semantic segmentation models are successfully utilized to categorize the pixels with respect to multiple damage types, such as cracks, leakage, spalling, and efflorescence [91,92]. Furthermore, the models optimized for identifying crack pixels are also presented utilizing FCN [93-95], U-Net [96], and SegNet [97,98]. Although various models have been proposed with advanced techniques to improve the identification accuracy [99-103], manual annotation of the ground truth in the pixel-level for the training stage is computationally inefficient. In addition, the loss of spatial resolution due to a number of

convolution and pooling layers is still remained as a critical factor. Furthermore, crack

identification from images that contain crack-like noncrack objects have received little

attention, despite this case being quite common in practice.
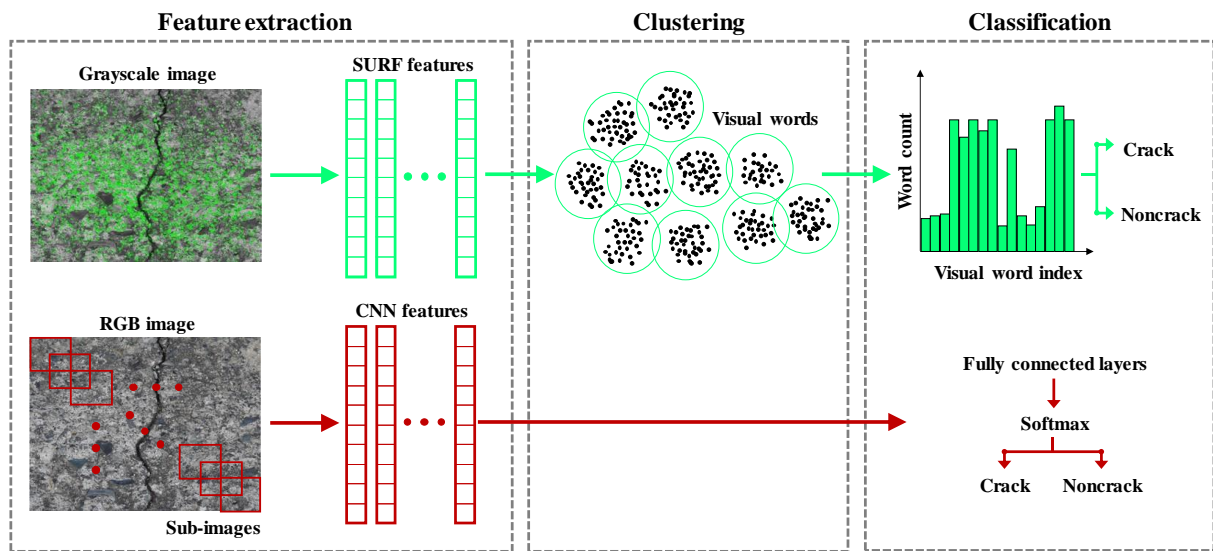
Fig. 2.4 Schematic of SURF-based and CNN-based methods (modified from the study by Zheng *et al*. [66]).

2.5   Summary

In this chapter, the applications of image binarization method, machine learning, and deep

learning on crack identification purposes were summarized. While these kinds of computer

vision-based approaches can provide efficient and effective inspection of surface cracks,

several issues needed to be carefully addressed for civil infrastructure: (1) robust and automated

crack detection and quantification on realistic surface images that contain crack-like noncracks

and (2) a tailored camera system and the corresponding strategy to perform accurate crack

evaluation, even with the skewed angle and the sequential images. Although image

binarization had a strong potential to effectively extract crack pixels on surface images, the

crack assessment was difficult to standardize owing the high dependence of binarization

parameters. Deep learning, state-of-the-art technique for semantic segmentation, was able to perform automated crack detection from the training process; however, the classification accuracy highly depended on the manually labeled ground truth in the pixel-level. Thus, a robust classification model based on deep learning in the region-level, in conjunction with the image binarization with associated optimal parameters can be a viable solution for robust and automated crack detection and quantification on realistic surface images. Furthermore, the practical issues in the real-world implementation, including the skewed angle problem and the absence of efficient crack measurements for the sequential images, can be properly solved by a tailored camera system and the associated strategy for accurate crack evaluation.

In the following chapters, concrete crack evaluation for civil infrastructure are presented using computer vision and deep learning. The proposed approaches consist of three main parts: (1) a deep learning-based crack detection, (2) a hybrid processing for crack quantification, and (3) camera systems for civil infrastructure. Here, CHAPTER 3 and CHAPTER 4 are related to robust and automated crack detection and quantification on realistic surface images; subsequently, CHAPTER 5 focuses on a tailored camera system and the associated strategy for accurate crack evaluation with the real-world implementation.

CHAPTER 3   DEEP LEARNING-BASED CRACK DETECTION

This study proposes a deep learning-based approach for automated crack detection purposes, consisting of two main processes: (1) generation of the crack candidate regions (CCRs) and (2) CNN-based classifications [104].   To properly address realistic surface images, the concept of CCRs that can be actual cracks or crack-like noncracks is employed by the image binarization method.   Subsequently, CNN features are extracted from all the CCRs to construct a robust classification model.   The obtained classification model is finally applied to new images to evaluate the classification performance.

3.1   Crack Candidate Region (CCR)

The proposed approach is employed for detecting cracks in concrete surface images that may contain crack and/or crack-like noncrack objects.   For this purpose, the proposed framework is designed to initially select crack candidates on realistic surface images.   The selected crack candidates constitute the CCRs, which are further utilized in building and applying the classification model.

The crack candidates, which represent both actual crack and crack-like noncrack objects, are extracted from a concrete surface image for effective classification.   The crack elements are typically represented by dark colors, which can be simply extracted using image binarization methods.   In the image binarization approach, all the pixels are converted into zero (black) or one (white) based on a threshold calculated using the statistical properties, such as pixel intensities and user-defined parameters such as sensitivity and window size.   Among the various image binarization methods [16-20] available for detecting the CCRs, Sauvola's binarization is utilized in the present study owing to its high performance in noisy and high-contrast images [105,106], as shown in Eq. 2.3.   Note that the sensitivity controls the

16

contribution of the statistical properties, and the window represents a rectangular box in which the threshold of each pixel is calculated. In contrast to other methods that directly employ the standard deviation, Sauvola's binarization makes it possible to amplify the contribution of the standard deviation in an adaptive manner by a factor of R, making it effective with noisy and high-contrast images. The image binarization finally returns the crack and noncrack objects marked as black in the binary images. Most of the obtained objects appear to be clearly noncracks because of noisy surface textures, which can be removed based on their geometric patterns such as the eccentricity and the number of pixels in each pixel group, as shown in Eq. 3.1.

$$e > e_{threshold}$$
$$A > A_{threshold}$$

Eq. 3.1

where e and A are the eccentricity and the number of pixels of a pixel group in the binary image, respectively. Thus, the computational efficiency can be improved by filtering the unnecessary noisy objects. Finally, the smallest rectangles containing crack candidates, are marked in the original image, as shown in Fig. 3.1. Note that the CCR may contain either a true crack or a crack-like noncrack object. This implies that if only Sauvola's binarization is applied to an input image without further deep learning-based classification, all the CCRs are considered as cracks, even if some of them are noncracks (0% accuracy for true negative).
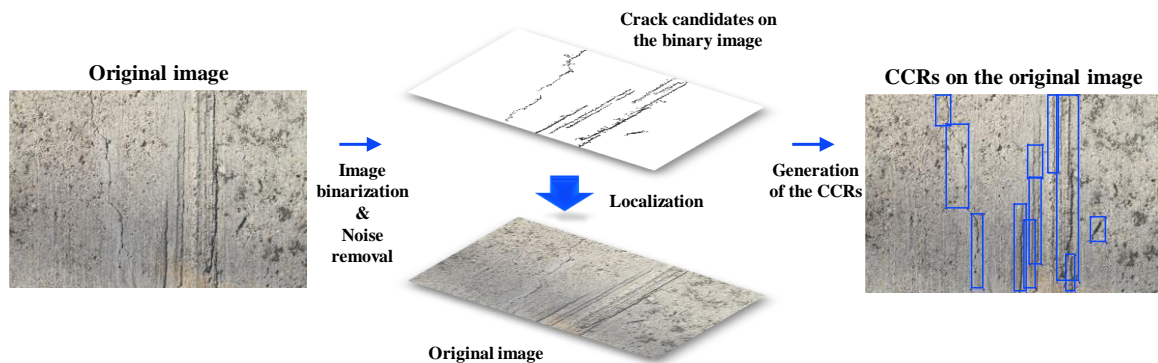
Fig. 3.1 Generation of the CCRs in the entire image.

The advantages of the CCRs in the proposed framework can be summarized as follows:

(1) The application of the CCRs is tailored to the classification of actual cracks and crack-like noncrack objects.   Previous studies on the use of deep learning for crack detection need to train the intact surface, which takes up a majority of an image. However, the CCRs enable constructing a robust classification model trained with only cracks and crack-like noncracks based on the binary information.

(2) The computational efficiency can be enhanced because only the selected CCRs are used in the training and testing stages.   Manual annotation of the ground truth for the training stage is computationally inefficient in the previous studies.   In contrast, considering that the image background, which does not contain possible crack or noncrack objects, occupies a major portion of the concrete surface image, excluding the background can significantly reduce the computational burden.

3.2   Classification Models

To construct the classification models, SURF and CNN features are obtained from the CCRs in the proposed approach to compare their classification performances.   In the SURF-based method, a grayscale image is used to extract the local features.   Here, a concrete surface image typically contains a large number of local features because of the noisy surface texture, affecting the classification of the cracks and noncracks.   Because the important features are largely located on crack-like shapes (either actual cracks or noncracks), the binary information of the CCRs is supported to preferentially select the SURF features on the crack segments, whereas most of the noisy SURF features on the concrete surface are filtered out, as shown in Fig. 3.2.   In contrast to the SURF-based method, the CNN-based method resizes the RGB

image to a fixed resolution of 227×227×3 for the input size in the employed CNN architecture.

Note that the input size of AlexNet is introduced in the proposed approach.
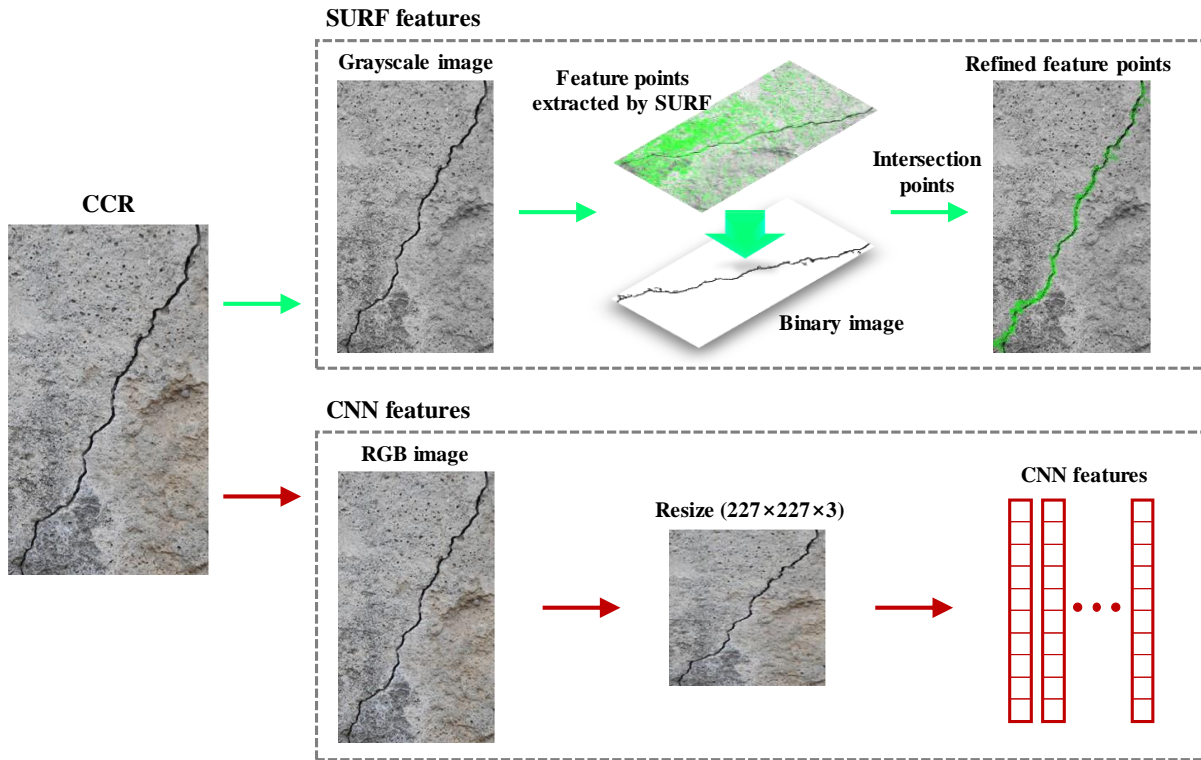
Fig. 3.2 Feature extraction process of SURF and CNN.

The classification models of the SURF-based and CNN-based methods are constructed using the CCRs obtained from the concrete surface images.   From the features obtained using SURF, the visual words that contain representative, small image segments are generated using k-means clustering.   Subsequently, the obtained visual words are grouped to create a visual vocabulary.   Here, the frequency of occurrence of the visual words in each category (i.e., cracks and noncracks) is calculated, from which the classification model is obtained using the linear SVM classifier.   The trained model can be used to categorize new CCRs.   Fig. 3.3 shows the schematic of the overall process of the proposed approach.

Training stage

Set of crack and noncrack images	

⇩ Image binarization & Noise removal

Generation of the CCRs

⇩ SURF   ⇩ CNN

SURF-based classifier   CNN-based classifier

Validation stage

New images

⇩ Image binarization & Noise removal

Generation of the CCRs

⇩   ⇩

Crack / Noncrack   Crack / Noncrack

SURF-based classification   CNN-based classification
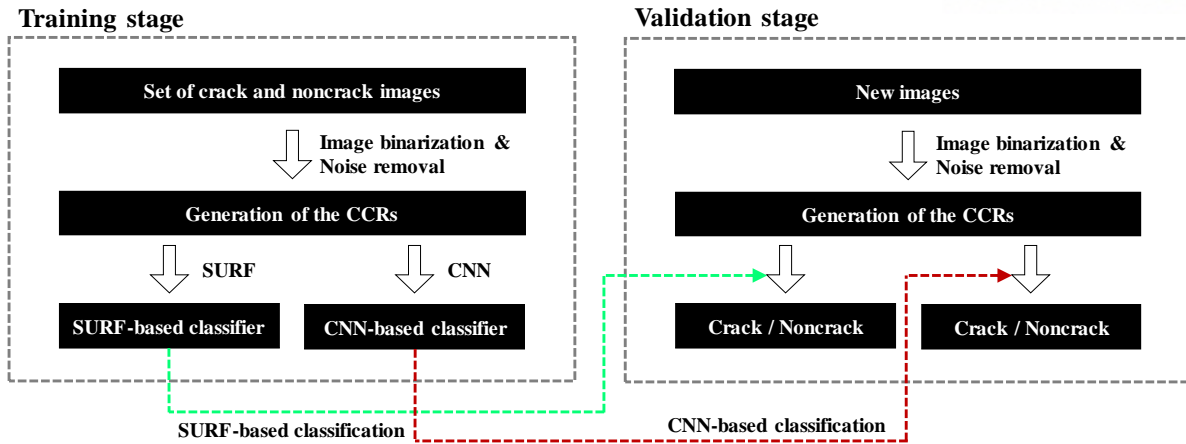
Fig. 3.3 Flow chart of the proposed approach for concrete crack detection.

3.3 Experimental Validation

3.3.1 Experimental Setup

The proposed crack detection approach is evaluated to demonstrate its performance using surface images obtained from concrete structures. The image binarization is applied to 487 images captured using digital cameras (see Table 3.1) to extract the CCRs including cracks and noncracks. The user-defined parameters of the image binarization are selected as 0.07 and 131 for the sensitivity and the window size, respectively [105,106]. In addition, the thresholds of the noise object removal are selected as 0.9 and 5,000 for the eccentricity and the number of pixels in each pixel group, respectively. Finally, 3,186 CCRs are generated, which consist of 527 actual cracks and 2,659 noncracks. To obtain a robust classification model, the image set is collected from various concrete surfaces under different working distances between the camera and the concrete surface, and under different illuminance conditions. Fig. 3.4 shows typical sample images taken from the set. The images contain noncracks such as dark shadows, stains flowing down from the top, dust, and protruding lumps generated from the casts, which are generally found in concrete structures. Furthermore, these kinds of crack-like noncracks are found to be similar to cracks in terms of geometry (e.g., long and thin) and

20

color (both are dark).   Note that the image database also includes branched cracks, spalling,

and various orientations of cracks.

Table 3.1 Specifications of used cameras.

	EOS-1D X	COOLPIX 900s
Manufacture	Canon	Nikon
Image resolution	17.9 M pixels	15.9 M pixels
Focal length	100 mm	4.3-357mm
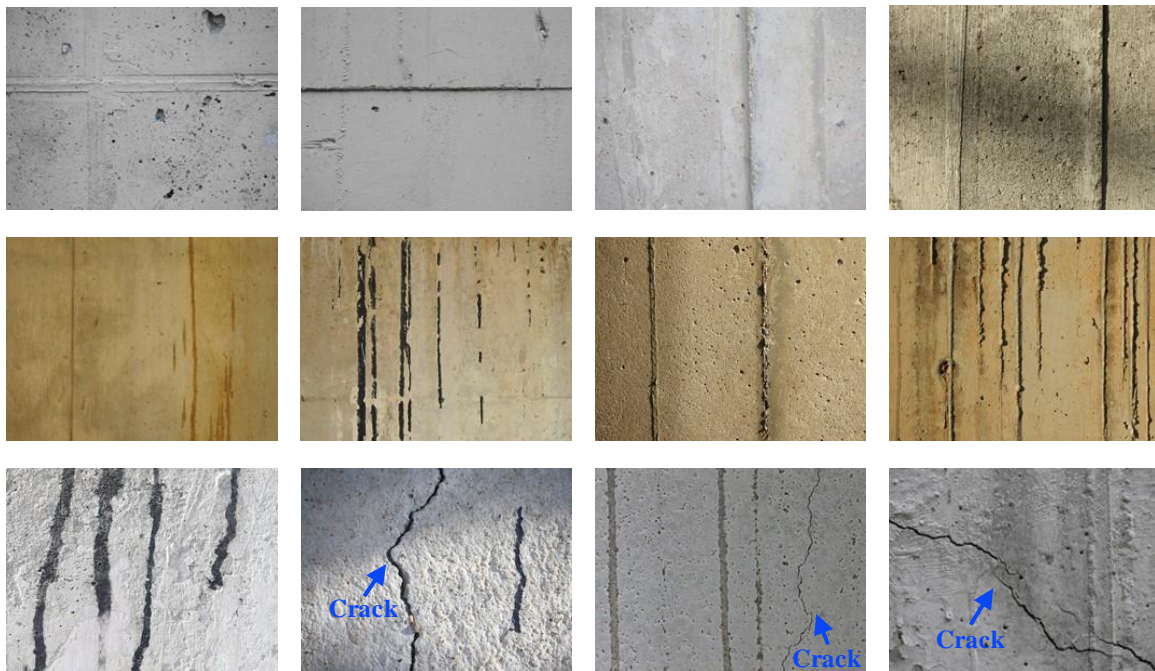
Fig. 3.4 Sample images of concrete surfaces used for experimental validation.

3.3.2   Performance Comparison

The classification models of the SURF-based and CNN-based methods are implemented using

MATLAB.   To evaluate the classification performances with respect to the size of CCRs, six

sets (i.e., 100, 200, 500, 1,000, 2,000, and 3,000) of CCRs are constructed from 3,186 CCRs.

In the feature extraction stage, SURF and CNN features are obtained by following the procedure of the proposed approach, as shown in Fig. 3.2. To obtain the classification model of the SURF-based method, three cases with different sizes of visual words (i.e., 100, 500, and 1,000) are considered in the *k*-means clustering. In addition, three cases with different minibatch sizes (i.e., 50, 100, and 200) are selected for the CNN-based method. With regard to the computational environment, a PC with an Intel Core i7-7700 processor clocked at 3.60 GHz and with 16 GB of RAM was employed. Moreover, a dedicated GPU (NVIDIA GeForce GTX 1080) was used.

Fig. 3.5 shows the typical classification results. Both the SURF-based and CNN-based methods successfully categorize the CCRs in the sample images as either a crack or as a noncrack, as indicated by the blue and red boxes. Note that only a few representative CCRs are shown for effective demonstration.

The trained classification models of the SURF-based and CNN-based methods are compared to quantitatively evaluate the performances. A 10-fold cross validation is conducted for each CCR set (i.e., 100, 200, 500, 1,000, 2,000, and 3,000). Fig. 3.6 shows the results of the SURF-based method with three different visual words (i.e., SURF-100, SURF-500, and SURF-1000) and those of the CNN-based method with three different minibatch sizes (i.e., CNN-50, CNN-100, and CNN-200). Here, the following five performance metrics are selected to compare the models:

- *precision*: TP / (TP+FP)

- *recall*: TP / (TP+FN)

- *F1 score*: 2 × (*precision* × *recall*) / (*precision* + *recall*)

- *accuracy*: (TP+TN) / (TP+FP+FN+TN)

- *computational time* in the training stage
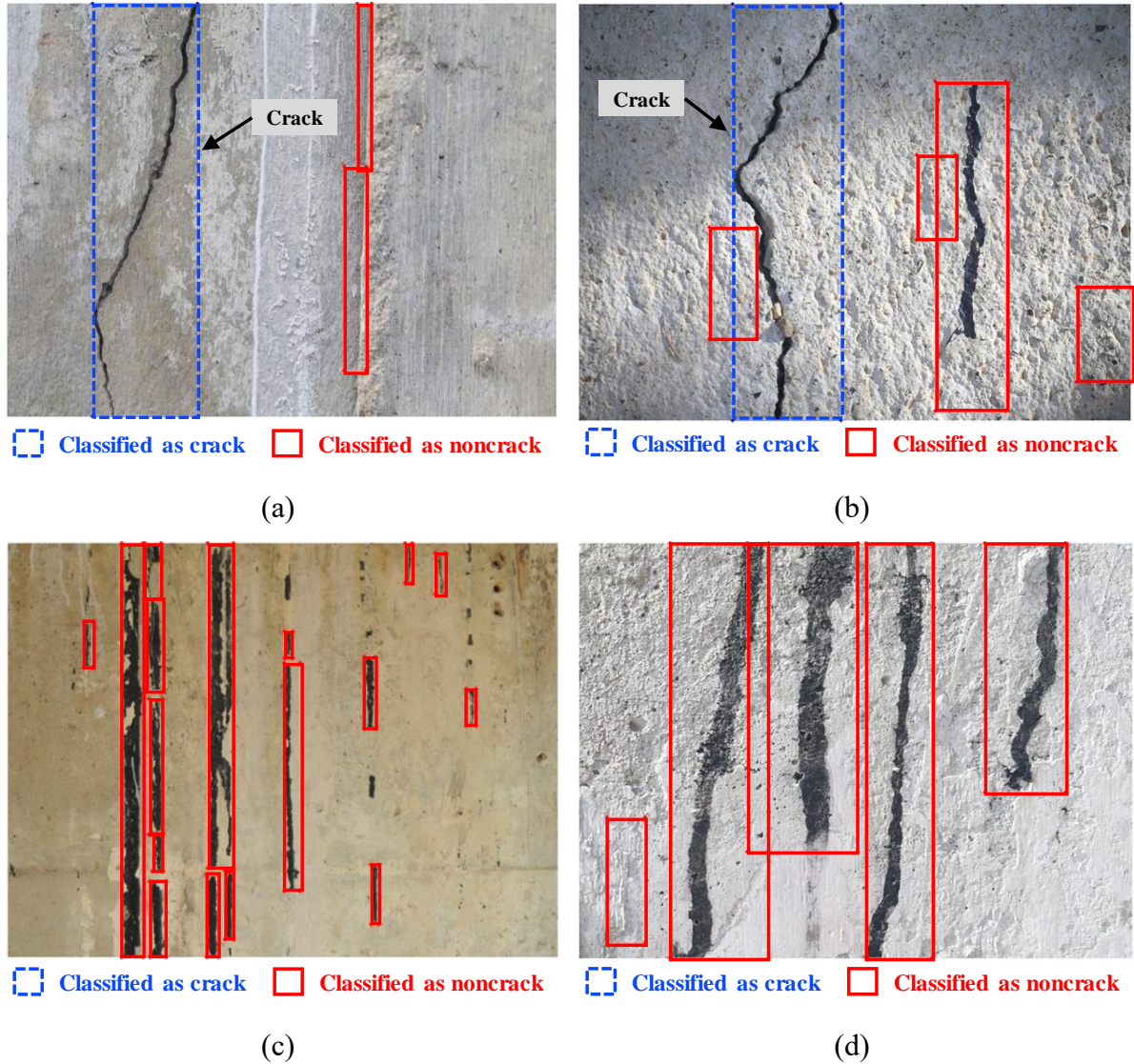
Fig. 3.5 Typical classification results of cracks and noncracks from the CCRs (both the SURF-based and CNN-based methods correctly classify the CCRs).

TP, FP, FN, and TN denote true positive, false positive, false negative, and true negative, respectively. As shown in Fig. 3.6b, the recall values corresponding to the SURF-based and CNN-based methods exhibit increasing trends with respect to the number of CCRs. However, the recall value of the SURF-based method decreases when the largest size of the CCRs is employed (i.e., 3,000) because of overfitting. In terms of the precision, as shown in Fig. 3.6a, the precision of the CNN-based method is higher than that of the SURF-based method, and is

reflected in the high F1 score (Fig. 3.6c) and accuracy (Fig. 3.6d).   In particular, the F1 score and the accuracy of CNN-50 significantly increase higher than those of the SURF-based method when 3,000 CCRs are used in the training.   Thus, when a sufficient minibatch size is used, CNN is observed to exhibit consistently high performance metrics.   In addition, the computational time for generating each classification model exhibits increasing trends in accordance with the number of CCRs, as shown in Fig. 3.6e.   Although the CNN-based method is slightly better than the SURF-based method, it is difficult to directly compare them because the SURF-based and CNN-based methods are implemented on different processing units of CPU and GPU, respectively.
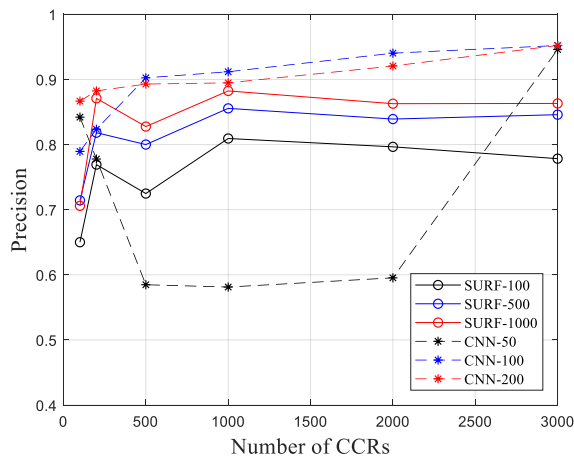
(a)
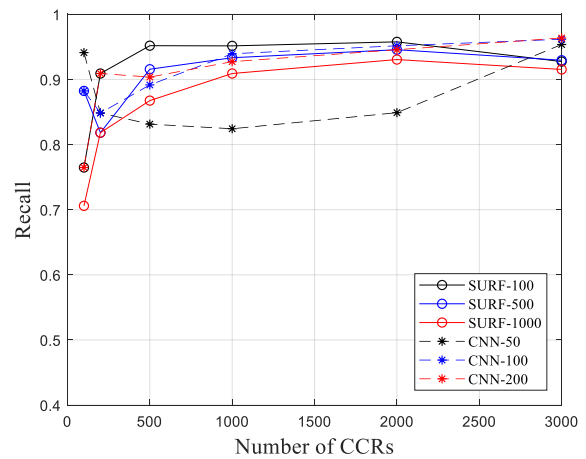
(b)
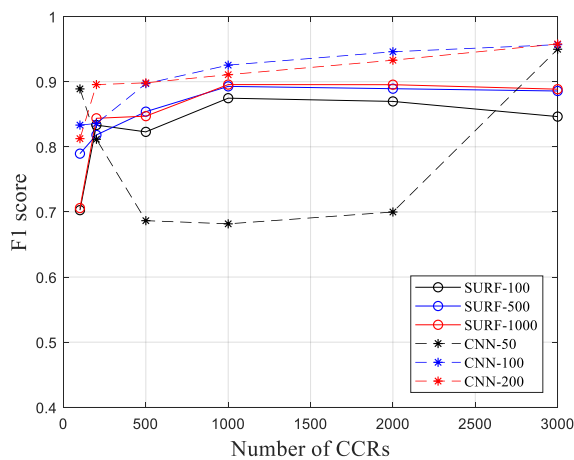
(c)
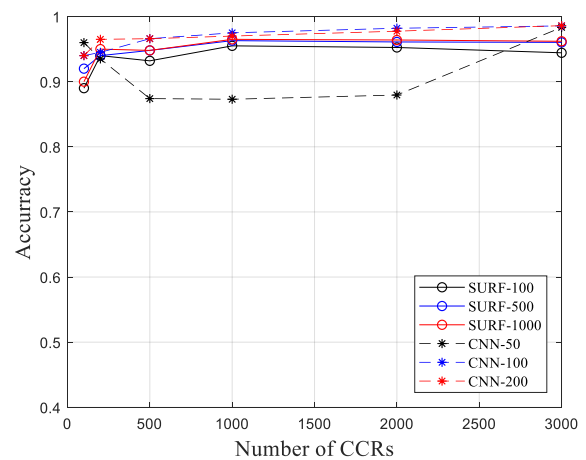
(d)
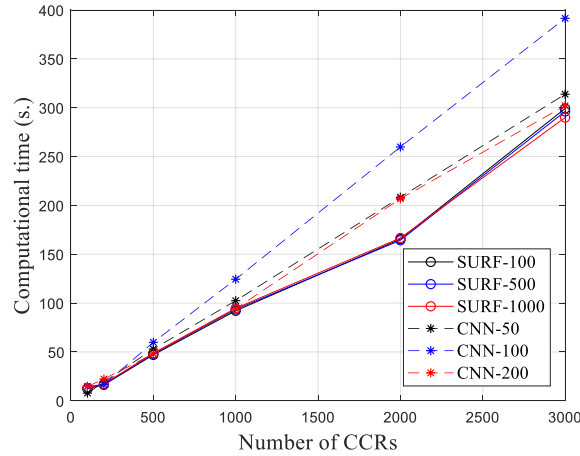
(e)

Fig. 3.6 Comparison of the SURF-based and CNN-based methods in terms of: (a) *precision*, (b) *recall*, (c) *F1 score*, (d) *accuracy*, and (e) *computational time.*

The classification models of the SURF-based and CNN-based methods can be compared for specific CCR cases to qualitatively understand their characteristics.   In particular, SURF-1000 and CNN-200 are used to categorize the CCRs in concrete surface images that are not used in the training stage.   Fig. 3.7 shows the classification results for the four cases.   Note that cases 1, 2, 3, and 4 represent dark stains flowing down from the top; protruding lumps generated between the casts; cement leaking from the cast; and surface cracks, respectively. As shown in Fig. 3.7b, d, f, and h, CNN-200 correctly classifies all the CCRs in the four cases as either a crack or as a noncrack, as indicated in the blue and red boxes, respectively.   In particular, the crack-like noncracks in cases 1, 2, and 3 that share similar geometry and colors with those of cracks are successfully detected as noncracks.   Furthermore, the cracks with small widths are accurately recognized in case 4.   In contrast to the CNN-based method, false positives and negatives are found in case of the SURF-based method (see Fig. 3.7a, c, e, and g).   These examples show that the overall performance of the CNN-based method is better than that of the SURF-based method.   Nevertheless, for the images used in this study, both the SURF-based and CNN-based methods correctly classify cracks and noncracks in most

25

cases.

(a)

(b)

(c)

(d)

(e)

(f)

(g)　　　　　　　　　　　　　　　　(h)
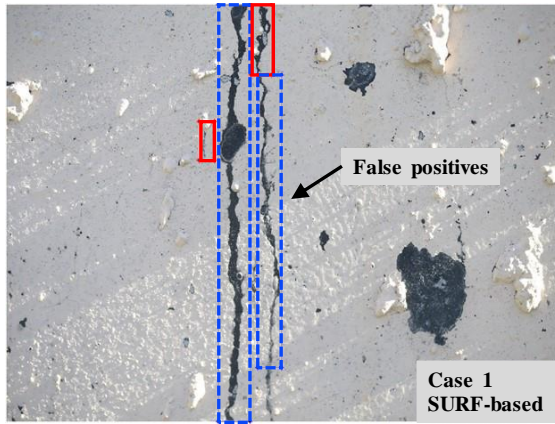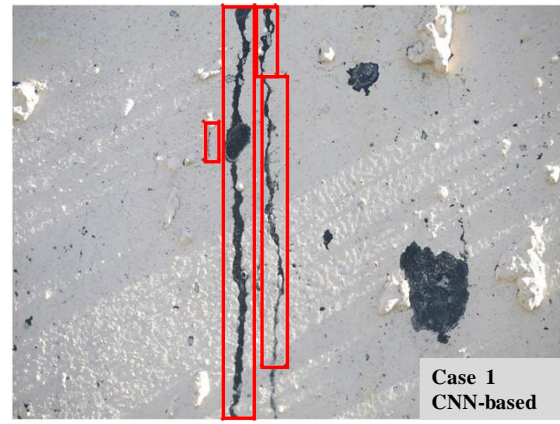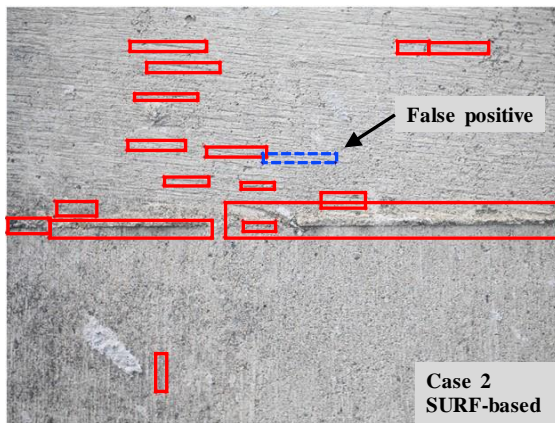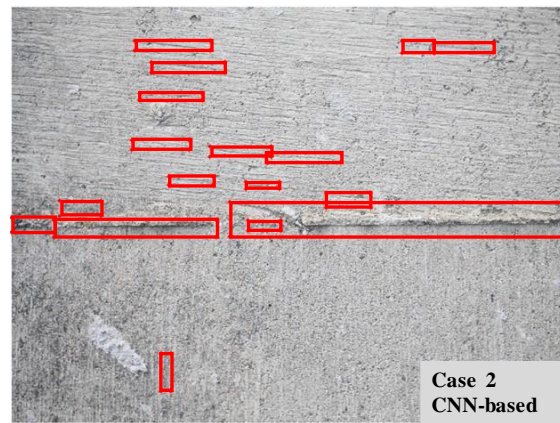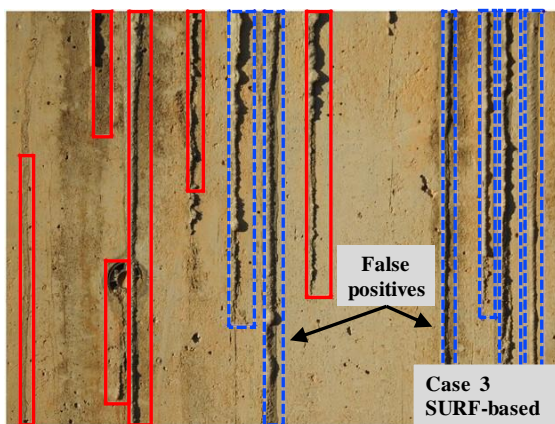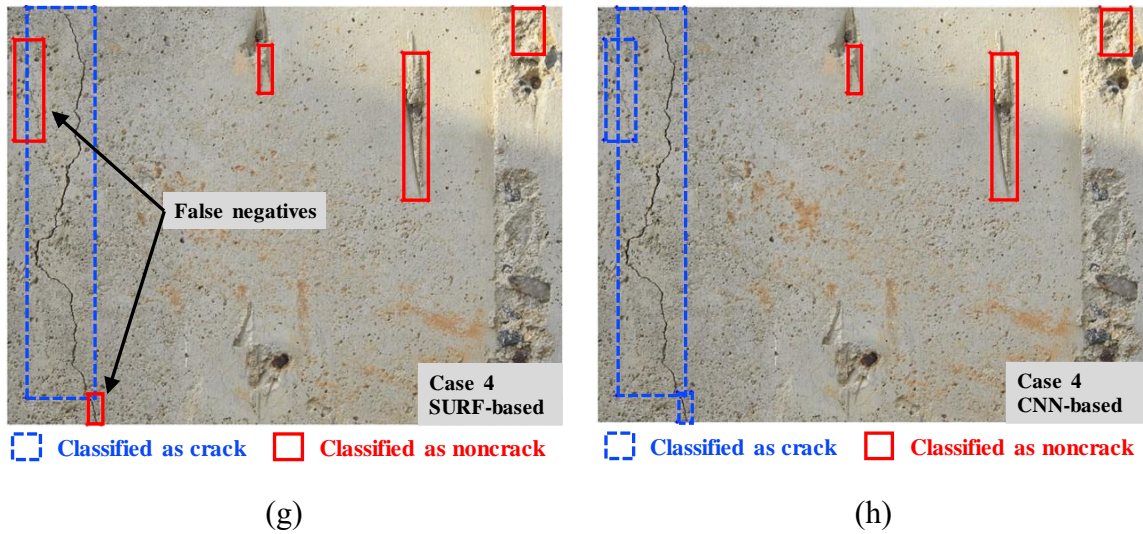
Fig. 3.7 Classification of cracks and noncracks from the CCRs: (a) case 1 with the SURF-based method, (b) case 1 with the CNN-based method, (c) case 2 with the SURF-based method, (d) case 2 with the CNN-based method, (e) case 3 with the SURF-based method, (f) case 3 with the CNN-based method, (g) case 4 with the SURF-based method, and (h) case 4 with the CNN-based method.

Although the classification performance of the CNN-based method is better in classifying actual cracks and crack-like noncrack objects, some of the CCRs could be successfully categorized only using the SURF-based method.   As shown in Fig. 3.8, both the SURF-based and CNN-based methods yield false negatives; however, the CNN-based method has an additional false detection from the lump on the concrete surface.   Thus, the local features extracted using the SURF can in some instances correctly classify the CCRs that were incorrectly categorized using the CNN-based method.   Hence, the combined use of deep neural networks and SVM classifiers with local/global features is found to have a potential to improve the classification performance.
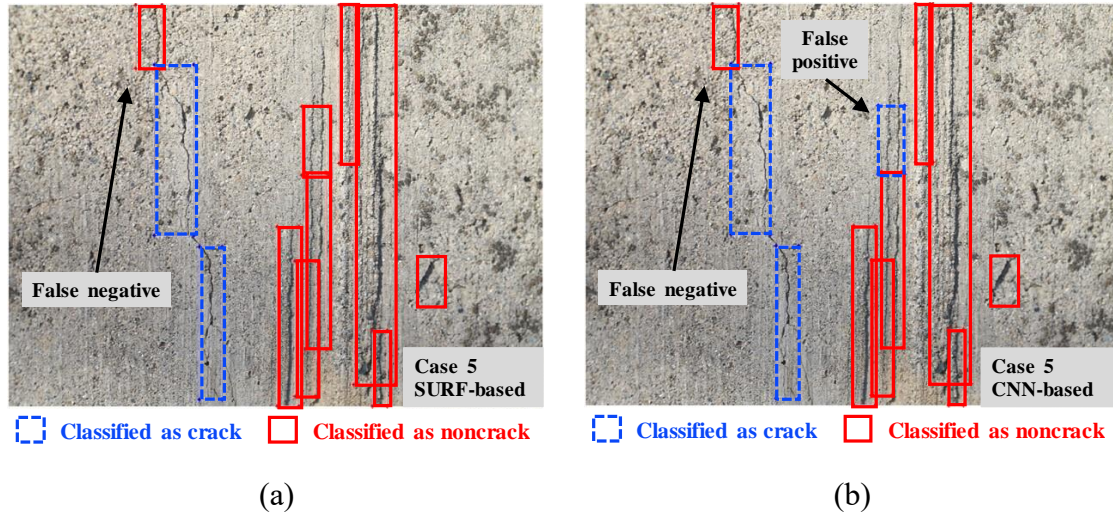
Fig. 3.8 Classification of cracks and noncracks from the CCRs: (a) case 5 with the SURF-based method and (b) case 5 with the CNN-based method.

To clearly show the advantage of the proposed crack detection, a comparative analysis has been conducted for three different classification models of previous studies and the proposed approach. Model A represents a classical classification constructed with k-means clustering and SVM. Widely used features for training in the literature [52-55] are selected, including geometric patterns and statistical properties of crack and crack-like noncrack objects on concrete surface images. Model B is constructed based on CNN with the scanning window [27,67-71]. Model C built with CNN represents the proposed approach. All the number of CCRs in the training set are constant in each model (i.e., 527 cracks and 2,659 intact surfaces or crack-like noncracks), and the parameters correspond to the highest performance shown in Fig. 3.6 are selected here. In the validation stage, a 10-fold cross validation is conducted, in which all the classification models are applied to the CCRs containing largely cracks and crack-like noncracks. The training configuration for the three models are summarized in Table 3.2.

The validation results in Table 3.2 clearly shows the efficacy of the proposed approach. The low performance metrics of Model A reveals that the geometric patterns and statistical properties are inadequate features to distinguish cracks and crack-like noncracks. In addition,

without using crack-like noncracks results in poor classification results in Model B.   As such, the CNN features trained with cracks and crack-like noncracks are the critical enablers for successful crack detection.

Table 3.2 Comparison of classification models with CCRs containing largely cracks and crack-like noncracks.

		Classification model A	Classification model B	Classification model C*
Training configuration	Features	Geometric patterns and statistical	CNN features	CNN features
	Classification model	SVM	CNN	CNN
	Training data	Cracks and crack-like noncracks	Cracks and intact surfaces	Cracks and crack-like noncracks
Validation results	*Precision*	0.51	0.24	0.94
	Recall	0.49	1.00	0.96
	F1 score	0.50	0.38	0.95
	Accuracy	0.84	0.47	0.98

*proposed approach

3.4   Summary

This research proposed a deep learning approach to determine the existence and location of cracks in concrete surface images containing possible crack-like noncrack objects.   The main contribution of this study was to propose a classification framework based on the CCRs for detecting cracks in the presence of noncrack objects that share similar image characteristics (i.e., shape and color).   In the training stage, concrete surface images with cracks and noncracks were prepared, from which CCRs were automatically extracted using image binarization.   After the CCRs were generated, the SURF-based and CNN-based methods were applied to the CCRs to extract the important features of the cracks and noncracks, which were subsequently used to construct classification models.   The obtained classification models were validated using concrete surface images that were not part of the training set.   The experimental results confirmed that the proposed framework could successfully detect both cracks and crack-like noncracks using CCRs.   Furthermore, the CNN-based method was found to be more accurate and efficient than the SURF-based method for crack detection purposes.   The experimental results can be summarized as follows:

(1) Cracks and noncrack objects were effectively extracted and categorized from concrete surface images using the proposed crack detection framework based on the extracted CCRs.

(2) The overall performance of the CNN-based method was better than that of the SURF-based method in most cases.   The *Precision* and *F1 score* were higher for the CNN-based method provided that sufficiently large minibatch sizes and CCR set sizes were used.   The *Recall* and *accuracy* of the CNN-based and SURF-based methods were largely the same.

(3) In some cases, the SURF-based method was able to classify CCRs that were incorrectly classified using the CNN-based method. Combining deep neural networks and SVM classifiers with local/global features could enable improved classification performance compared to using each method separately.

(4) By introducing various crack-like noncracks in the form of CCRs in the training, the proposed framework would enable accurate detection of cracks from concrete surface images in the presence of noncrack objects.

The proposed deep learning learning-based approach has a strong potential for automated crack detection of concrete structures.

CHAPTER 4   HYBRID PROCESSING FOR CRACK QUANTIFICATION

As the binarization parameters provided by operators are the governing factors in practice, crack quantification from the image binarization is difficult to standardize.   To systematically address the issue, this chapter describes the proposed strategy to determine appropriate image binarization methods with optimal binarization parameters [105].   Furthermore, a hybrid image processing strategy is developed in conjunction with the UAV-based system, considering a combination of different sets of binarization parameters for accurately extracting the crack width while minimizing loss of length [106].

4.1   Optimal Binarization Parameters

The comparative analysis of the binarization methods with the associated optimal parameters is conducted to demonstrate their performance.   Although a set of the optimal window size and sensitivity are determined, the binarization results of each method are inevitably different despite using the same crack image, because each method has a different scheme for calculating the threshold.   To investigate the performance of the image binarization methods, the following criteria are used as follows: (1) measurement accuracy of the crack width, (2) identified crack length, and (3) computation time.

A set of different types of crack images are prepared and analyzed for each binarization method.   The image set accounts for various concrete textures and crack locations, widths, and patterns.   All color images in the set are converted to grayscale; the image binarization methods are applied to generate binary images.   Subsequently, the crack widths in pixels are calculated by counting the number of crack pixels at specific locations at which the crack widths are desired.   Finally, the actual crack width in millimeters can be obtained using the following camera pinhole model.

$$W_r = D_p W_p = \frac{10 D_w}{P_c L_f} W_p \qquad \text{Eq. 4.1}$$

where W_r is the real crack width in millimeters, D_p is the resolution of the imaging device, W_p is the obtained crack width in pixels, D_w is the working distance in millimeters, P_c is the pixels per centimeter of the used camera sensor, and L_f is the focal length of the camera in millimeters. The converted crack widths are compared with the reference widths resulted by an optical microscope to analyze the estimation accuracy. Note that the optical microscope that provides crack observations in a small region is inappropriate for investigating overall crack patterns. Assuming that the reference by the microscope is a true width, the measurement error in image binarization is defined by

$$Error = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(\frac{d_c - d_m}{d_m}\right)_i^2} \qquad \text{Eq. 4.2}$$

where d_c is the calculated crack widths using image binarization methods, d_m is the measured crack width using an optical microscope, and n is the number of specified locations for comparing crack widths. In addition, the accuracy in terms of the identified crack length is evaluated through comparison with all cracks, each of which is verified visually. The following index R_{Length} is defined to quantify the performance in terms of the crack length as

$$R_{Length} = \frac{L_{Est}}{L_{Total}} \qquad \text{Eq. 4.3}$$

where L_{Total} is the total crack length in an image for reference, and L_{Est} is the estimated crack length using the image binarization methods. Finally, the calculation time for processing an image binarization method is examined to compare the computation load for each method.

The optimal binarization parameters of each method are to be determined from the average values of calculated errors from the diverse images in the image pool. Note that the pixel resolution of the used digital camera and the working distance, which are also critical factors,

are not included in this study, because all of the crack images are captured using the same digital camera, and the working distances in the experiment are kept within a small range and thus do not significantly affect the binarization results.

4.2   Hybrid Image Processing

A hybrid image processing strategy is presented to utilize a combination of different sets of binarization parameters for accurately extracting the crack width while minimizing loss of length.   The proposed hybrid image processing strategy comprises two stages: (1) image pre-processing to prepare the image for further analysis and (2) crack width estimation using the hybrid approach.

The image pre-processing is conducted, which stage consists of two steps: (1) image undistortion and (2) conversion from color image to grayscale image.   As shown in Fig. 4.1a, the lens may produce a distorted image, from which crack width estimation can be seriously impaired.   To calibrate this image [107], a black and white checker board is captured using the camera with different angles and distances, to estimate the intrinsic and extrinsic parameters. After determining the camera parameters, the image taken by that camera is undistorted, as shown in Fig. 4.1b.   Subsequently, the calibrated image is converted to grayscale, as the image colors are unnecessary for identifying cracks.

(a)   (b)

Fig. 4.1 Illustrative example for the image calibration algorithm: (a) image distortion resulted from wide-angle lens and (b) image undistortion using image calibration.

Hybrid image processing is applied to the pre-processed image to determine crack width and length accurately. As stated earlier, a set of binarization parameters of sensitivity and window size is difficult to estimate crack width and length simultaneously. Thus, the hybrid approach employs two sets of binarization parameters, each of which provides the least error in width and length estimations, respectively. Let P_w and P_l designate these two sets:

P_w: optimal parameters minimizing estimation errors in crack width

P_l: optimal parameters minimizing estimation errors in crack length

P_w and P_l are then separately employed to generate two binary images using an image binarization method. P_l inevitably results in a higher threshold than that of P_w to convert more pixels to crack elements.

The binary images are subsequently processed using the steps for skeleton and edge detection and crack width calculation. Following width estimation, the obtained width information is recorded in each location of the skeleton pixels. The sets of skeleton pixels and their related crack width are defined as follows:

S_w: set of skeleton pixels obtained using P_w

S_l: set of skeleton pixels obtained using P_l

$w(P, S)$: crack width at location S obtained using P

where P is either P_w or P_l, and S is a set of skeleton pixels. Selecting P_w to produce the

accurate crack width of $w(P_w, S_w)$ results in more unidentified crack elements than P_l. Thus, S_w is a subset of S_l, because the obtained thresholds of P_l are greater than those of P_w. However, the calculated widths obtained using P_l (i.e., $w(P_l, S_l)$) are overestimated owing to the high thresholds.

The final crack widths are a combination of $w(P_w, S_w)$ and $w(P_l, S_l\text{-}S_w)$, which are the crack widths using P_w at S_w and P_l at $S_l\text{-}S_w$, respectively. The overestimated crack width $w(P_l, S_l\text{-}S_w)$ has to be corrected, which enables the simultaneous generation of accurate crack width and length. The calibration for the overestimated $w(P_l, S_l\text{-}S_w)$ can be performed by utilizing the ratio of $w(P_w, S_w)$ and $w(P_l, S_w)$, as defined in Eq. 4.4:

$$\alpha = \frac{1}{N} \sum_{i=1}^{N} \frac{w_i\left(P_w, S_w\right)}{w_i\left(P_l, S_w\right)}$$

Eq. 4.4

where α is the calibration factor, N is the number of skeleton pixels in S_w, and w_i is the crack width at the ith skeleton pixel. The calibration factor α is then multiplied by $w(P_l, S_l\text{-}S_w)$ to correct the overestimation. The crack widths of $w(P_w, S_w)$ and $\alpha w(P_l, S_l\text{-}S_w)$ are combined to provide complete width and length information in S_l. The overall procedure of the proposed hybrid image processing strategy is summarized in Fig. 4.2.

Crack image and associated distance

Step 1: Image pre-processing

Image undistortion

Conversion to grayscale

Step 2: Crack width estimation

Crack width estimation using P_w

Crack width estimation using P_l

Width calibration for $S_{l\text{-}w}$

Crack width in S_w

Crack width in $S_{l\text{-}w}$

Combined crack width in S_w & $S_{l\text{-}w}$
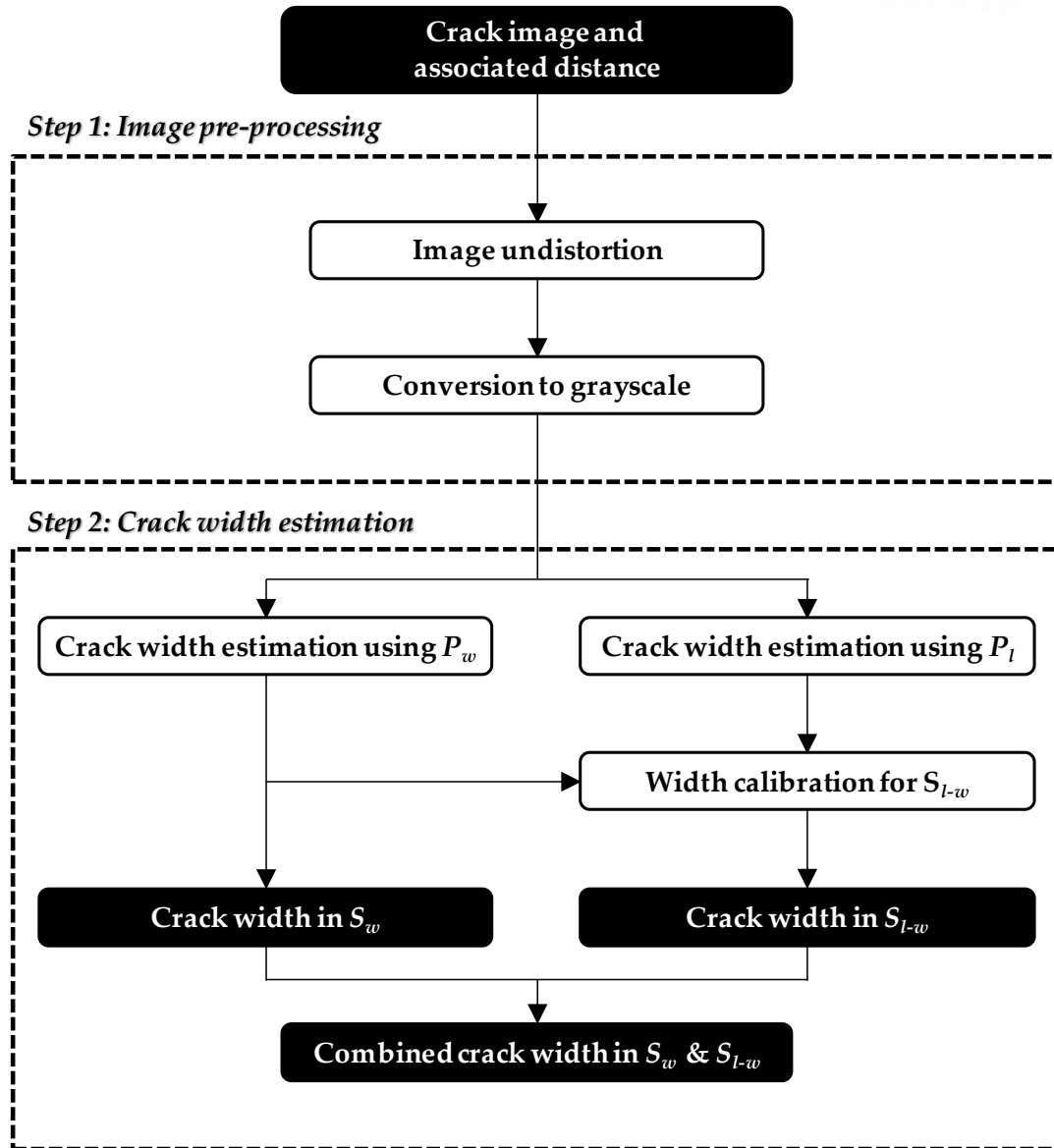
Fig. 4.2 Schematic outline of the hybrid image processing strategy.

4.3   Experimental Validation

4.3.1   Comparative Analysis

A parametric analysis of each binarization method is performed to determine the optimal parameters for crack monitoring.   A total of 21 images with different conditions, including concrete textures, illuminances, crack sizes, lengths, widths, patterns, and directions, are captured using the digital camera, as shown in Fig. 4.3.   These cracks are mostly caused by

shrinkage, creep, and repeated external loads.   The obtained images are processed using the

image binarization methods to compare their crack widths and the references resulted by the

optical microscope.   Note that the color targets are attached to the concrete surface as shown

in Fig. 4.4 to identify the specific locations where the references are measured.   At least seven

points for each case are employed to yield consistent width estimation results.

Specifications of the digital camera and the optical microscope used are summarized in

Table 4.1.   The working distances between the camera and the concrete surface in the

experiment are taken between 200 and 350 mm, which result in pixel resolutions of 0.0142 to

0.0243 mm/pixel, respectively.   Thus, the crack widths that range from 0.11 to 1.73 mm in

this experiment can be estimated reasonably.

Fig. 4.3 Example crack images in different conditions.

Table 4.1 Specifications of used digital camera and optical microscope.

	Digital Camera	Optical Microscope
Manufacturer	CANON	MICRO B2B
Model	EOS-1DX	MSP-2000
Image Resolution	18.1 Mega Pixels	2.0 Mega Pixels
Focal Length	100 mm	0 - 70 mm
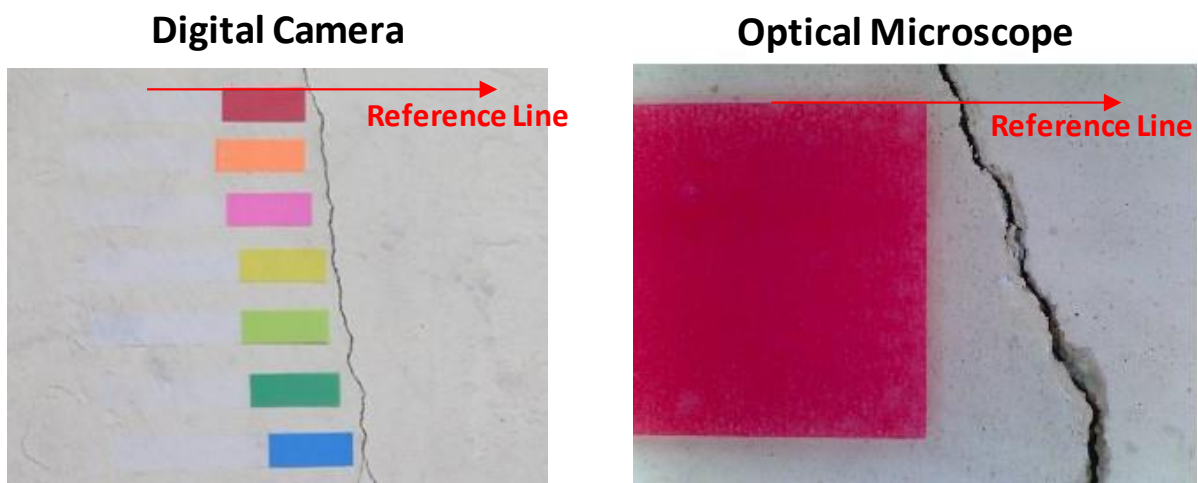
Fig. 4.4 Comparison of crack widths using digital camera and optical microscope.

The average errors are calculated using a wide range of sensitivities and window sizes to determine the optimal parameters in aspects of the width accuracy. The results of each method are presented in Fig. 4.5, which contains the black and white areas that represent the maximum and minimum errors, respectively. The blue circles in Fig. 4.5 represent all combinations of binarization parameter that produce width estimation errors less than 10%. While neither the sensitivity nor the window size is seen to have any significant trends in Bernsen's method, the sensitivity is a governing factor in the other methods (i.e., Niblack's, Sauvola's, and Wolf's methods, and the NICK method). The optimal binarization parameters of each method are selected inside of the white areas based on the minimum error of measured

crack widths, as summarized in Table 4.2.   As these binarization parameters are valid only for

the experimental setup in this section, this procedure should be repeated if different camera,

working distance, and target crack width are employed.
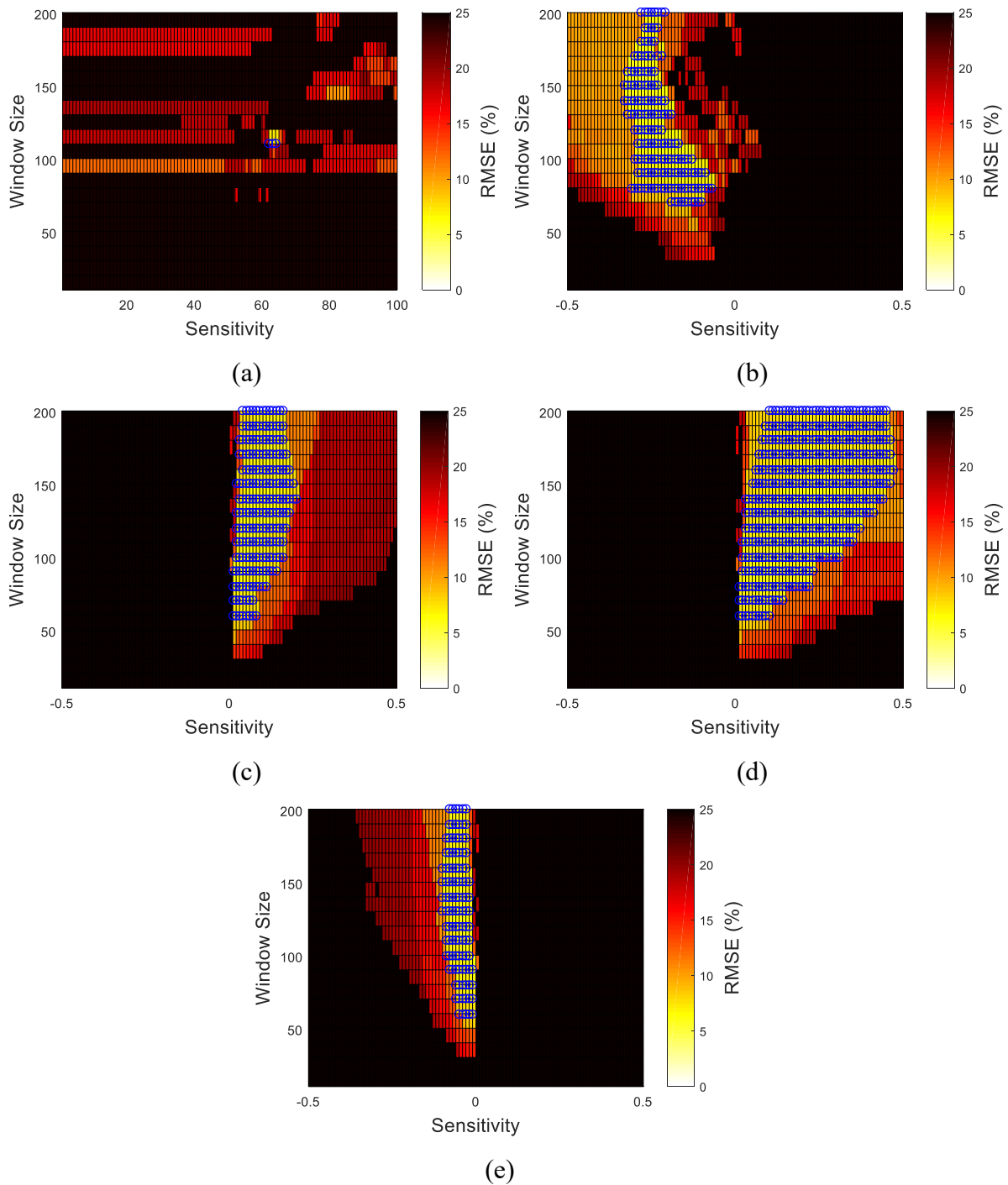
Fig. 4.5 Errors in measured crack widths in the image binarization methods: (a) Bernsen, (b)

Niblack, (c) Sauvola, (d) Wolf, and (e) NICK methods.

Table 4.2 Optimal parameters of the image binarization methods.

Method	Sensitivity	Window Size	Error (%)
Bernsen	62–64	111	7.19
Niblack	−0.26	141	5.96
Sauvola	0.07	131	5.98
Wolf	0.27	131	6.18
NICK	−0.05	121	5.96

To investigate the performances of the image binarization methods with the associated optimal parameters, the binarization results are compared utilizing new crack images. The Bernsen's, Niblack's, Sauvola's, and Wolf's methods, and the NICK method with the optimal binarization parameters determined in Table 4.2 are applied to crack images for crack quantification. As shown in Fig. 4.6, six crack images are prepared, of which the first two contain cracks similar to the previous process, while the other four have different conditions, such as concrete crack pattern, illuminance, and surface texture. Note that the binarization result in each method and their corresponding optimal parameters in Fig. 4.6a, is representatively presented in Fig. 4.7.
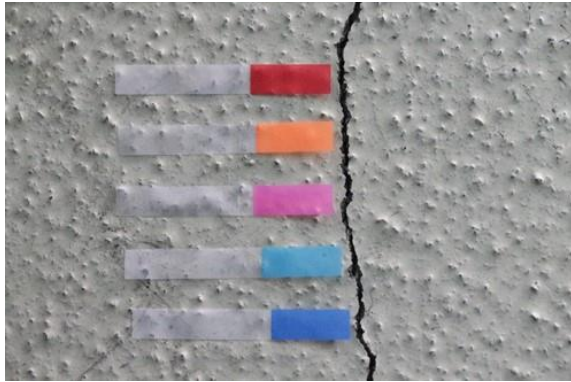
(a)

(b)

(c)

(d)

(e)

(f)

Fig. 4.6 Crack images used for comparative analysis: (a) Case 1, (b) Case 2, (c) Case 3, (d) Case 4, (e) Case 5, and (f) Case 6.

| (a) | (b) | (c) | (d) | (e) | (f) |
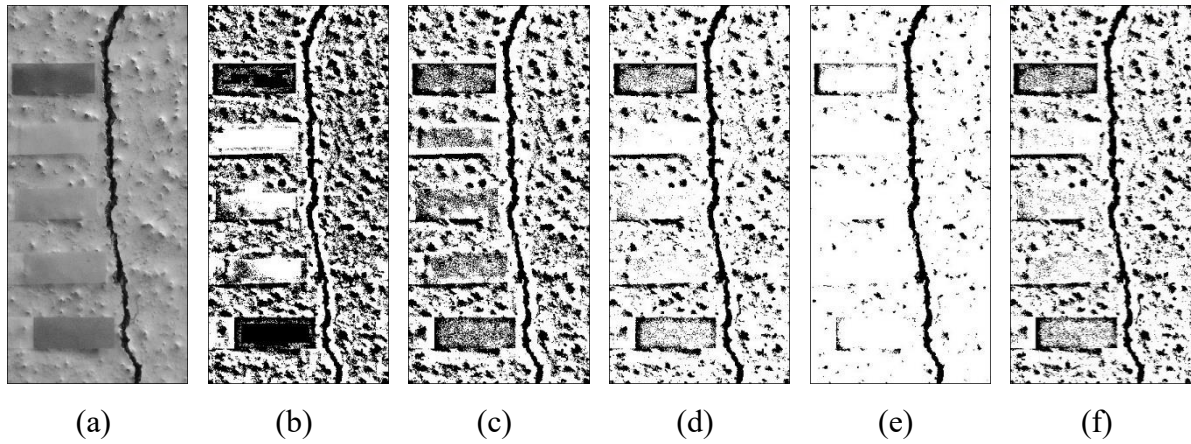
Fig. 4.7 Binarization results of each method with their optimal parameters in Case 1: (a) original grayscale image, (b) Bernsen, (c) Niblack, (d) Sauvola, (e) Wolf, and (f) NICK binarization results.

The aforementioned criteria are utilized to investigate the performance of each method, as presented in Fig. 4.8.   All the image binarization methods and the associated optimal parameters are seen to work reasonably for identifying crack widths, because the estimation errors are less than 11% in all the cases, as shown in Fig. 4.8a.   Thus, the optimal parameters resulted by the proposed approach work well for crack images used in the image pool, as well as new images.   Subsequently, the identified crack length is evaluated and is also graphically represented in Fig. 4.8b.   Most binarization methods can search crack length with high accuracy; however, Wolf's method is seen to be inappropriate for measuring full lengths, particularly for the small cracks in Fig. 4.6c and f.   This result will be analyzed later using the calculated threshold of each method.   Last, the computation time represents the total time required for processing the selected crack image, also analyzed in Fig. 4.8c.   Bernsen's method is observed to require much more computation time than the others, because this approach involves additional histogram calculation in determining threshold values.
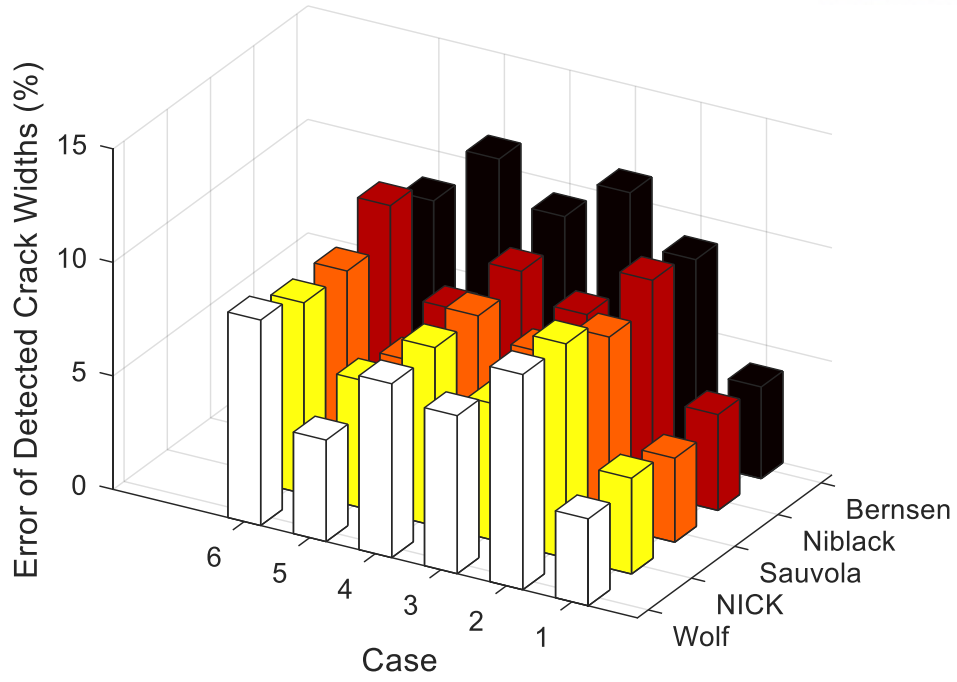
(a)
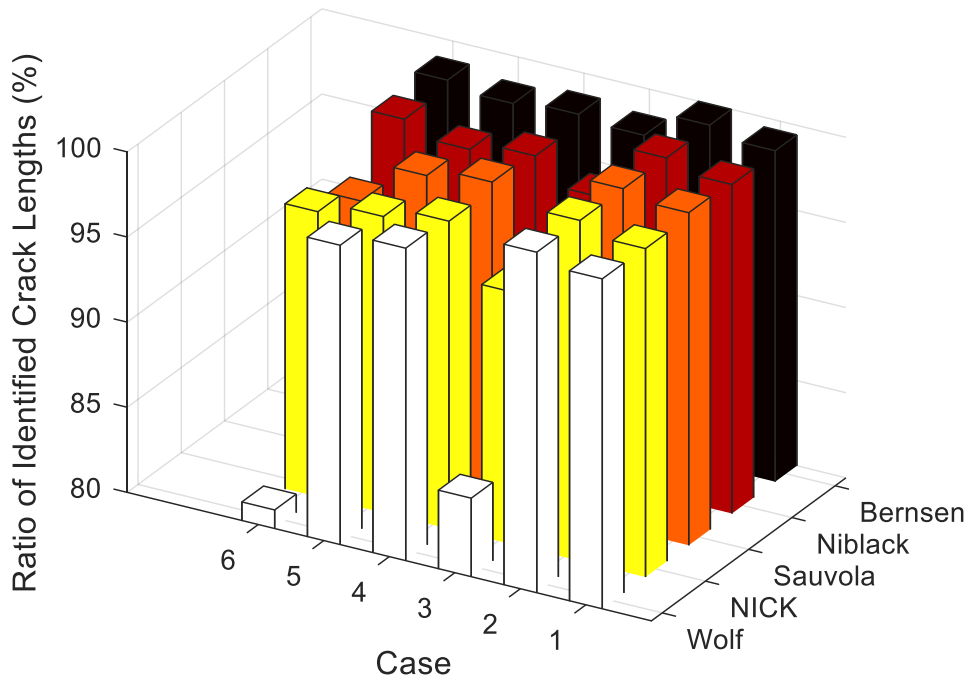
(b)
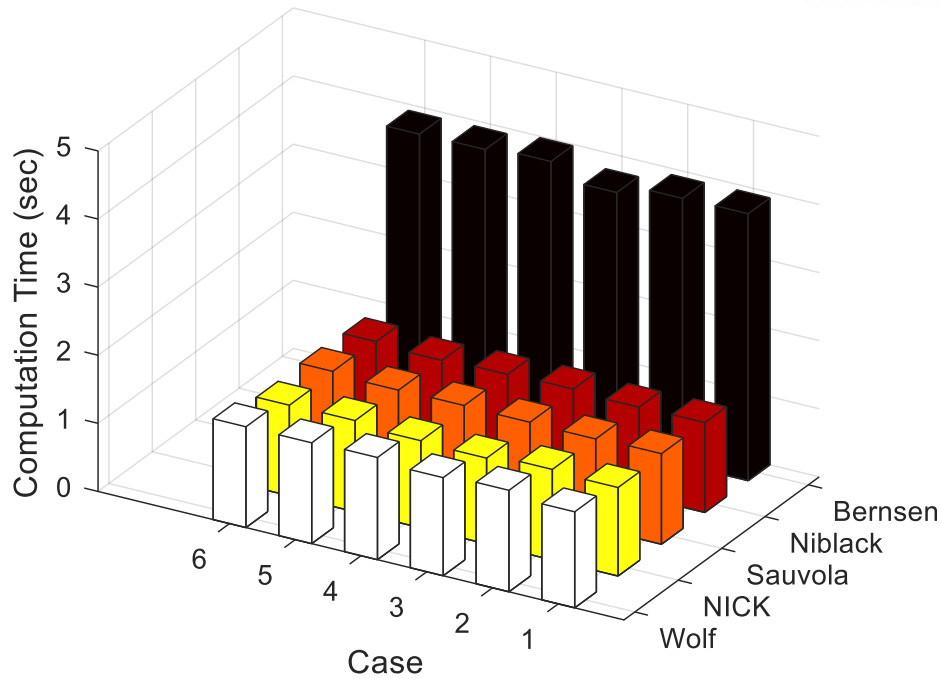
(c)

Fig. 4.8 Experimental results: (a) accuracy of measured crack widths, (b) ratio of identified crack lengths, and (c) computation time.

To further interpret the performance differences in Fig. 4.8, the thresholds calculated from each method are compared with the grayscale pixel values of cracks. Note that Bernsen's method is not included in this analysis, because the pixel-valued threshold is not used in the binarization process, in contrast to the other methods. The "grayscale" in the legend of Fig. 4.9 represents the grayscale pixel values of the original image on the red lines between points A and B. In addition, the distance between two black circles represents the reference width determined by the optical microscope. As shown in Fig. 4.9a, all the methods are observed to estimate the crack width accurately. As Wolf's method is designed to be sensitive to the standard deviation of the pixel values, as in Eq. 2.4, the thresholds in noncrack pixels, which have relatively smaller variations compared with the crack regions, are lower than those of other methods. Because of this characteristic of Wolf's method, it is difficult to determine the

crack width accurately for unclear cracks with small standard deviations, as shown in Fig. 4.9b. However, as a result of the low threshold, Wolf's method is advantageous in that false positives resulted by dust, aggregates, and holes are much fewer than those of the other methods. Thus, Wolf's method is considered to be efficient for identifying clear cracks, also resulting in a low chance of false positive crack detections, whereas Niblack, Sauvola, and NICK are appropriate solutions, when a crack image contains various types of clear and unclear cracks. The combination of Wolf's method and the other methods (i.e., Niblack, Sauvola, and NICK) can be potentially used to improve the results with accurate crack widths while reducing false positive crack detections. Note that because in this study, the recognizable cracks are selected to evaluate the measurement accuracy of the calculated crack width, the errors of Wolf's method are almost similar to those of the other methods in Fig. 4.9a.

(a)

Original image

Reference width

(b)
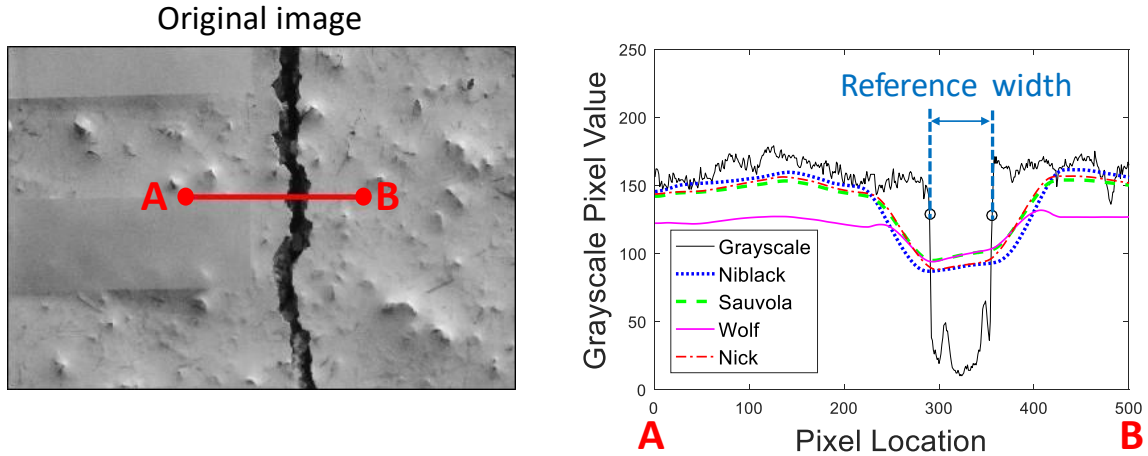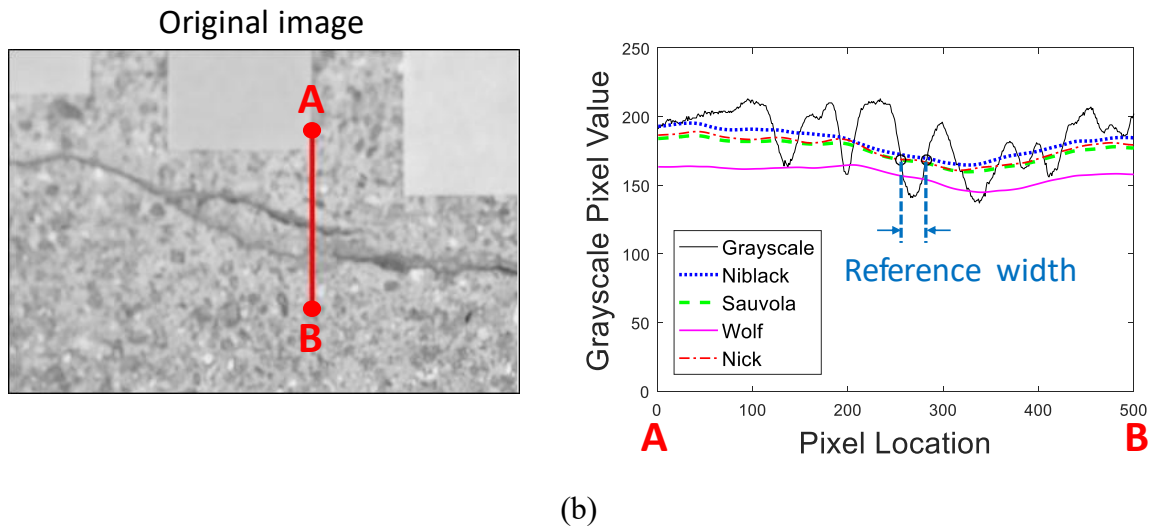
Fig. 4.9 Comparison between thresholds of the image binarization methods: (a) Case 1 and (b) Case 3.

4.3.2   Hybrid Image Processing

In this section, a hybrid image processing strategy described in Section 4.2 is evaluated in terms of accurate crack quantification.   The proposed hybrid image processing is able to measure crack widths accurately with minimizing the loss of crack length information.   To enable this, Sauvola's method is employed here because of its high performance in noisy and high-contrast images, as shown in the previous section.   The proposed approach is further implemented on an UAV-based prototype for the experimental validation on concrete structures.

Recent advances in UAV technologies have developed high-mobility and low-cost UAVs, rapidly broadening their real-world civil engineering application [108-114].   For example, aerial images taken by UAVs can be utilized to construct 3D structural models [115-118], evaluate road conditions [119-121], and conduct traffic surveillance and management [122-124].   Furthermore, the application of UAVs in conjunction with computer vision has also shown great potential to overcome the disadvantages of visual inspection for concrete crack monitoring [125-129].   UAV enables the taking of images in proximity to surface cracks in civil engineering structures, facilitating better crack identification results.
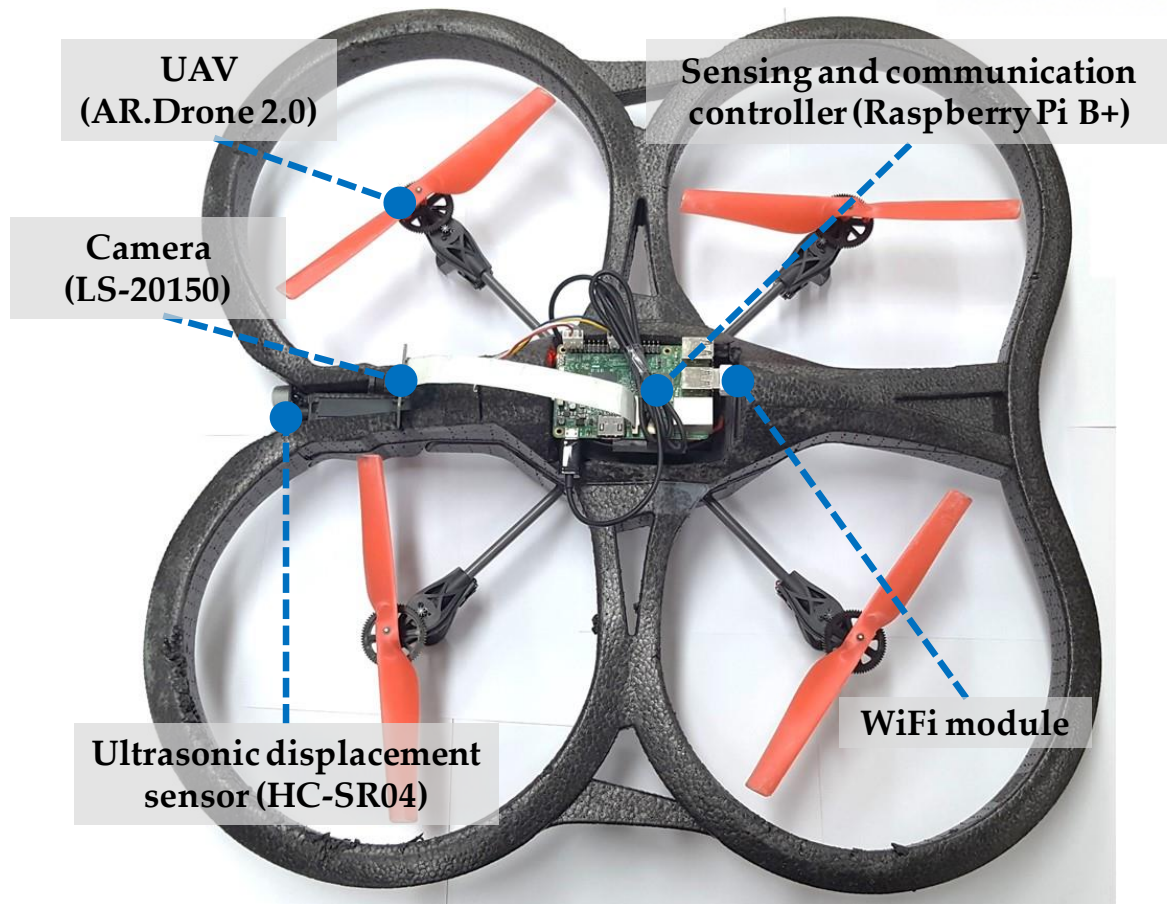
Fig. 4.10 UAV-based system for crack information acquisition.

The UAV-based prototype employed in this study is designed to effectively acquire necessary data for crack identification. The prototype is designed based on an off-the-shelf quadcopter, Parrot AR.Drone 2.0, because of its high mobility, cost-effectiveness, and convenient control interface using a smartphone. The proposed UAV is equipped with four essential components: a single-board computer, a camera, an ultrasonic displacement sensor, and a WiFi module, as shown in Fig. 4.10 and Table 4.3.

Table 4.3 System component.

Component	Model	Specification
UAV	Parrot AR.Drone 2.0	Dimensions: 58 cm × 13 cm × 58 cm Weight: 1.8 kg
Single-board computer	Raspberry Pi B+	CPU: 700 MHz single-core Weight: 45 g
Camera	LS-20150	Resolution: 2592 pixels × 1944 pixels Weight: 10.3 g
Ultrasonic displacement sensor	HC-SR04	Measurable distance: 2 cm – 4 m Weight: 8.5 g

Raspberry Pi B+, a low-cost low-power single-board computer running Linux, is employed to control sensing and communication. The Raspberry Pi B+ is interfaced with the camera, the displacement sensor, and the USB WiFi module. The Raspberry Pi B+ takes images using the camera, measures the working distance between the camera and the concrete structure, and is controlled by and sends data to a remote computer using the WiFi module. The USB WiFi module mounted on the Raspberry Pi provides wireless connection between the UAV-based system and the operator's computer through a WiFi router. Remote access to the Raspberry Pi of the UAV-based system allows operators to acquire image and distance information when desired and to wirelessly transmit the acquired data. The operator can monitor the video being taken by the camera, and instantly acquire image and distance data that are wirelessly transmitted to the operator's computer.

The camera module (LS-20150) and the ultrasonic displacement sensor (HC-SR04) provide crack images and the corresponding working distances, which are required to determine crack sizes. In previous studies [125-129], quantitative assessment of cracks was ineffective or unavailable because measured distance information was not obtained. The camera module has a maximum resolution of five million square pixels, which is adequate for

crack image acquisition, despite its light weight of 10.3 g.   The small focal ratio (F-number) of the camera module enables the highest shutter speed; thus, any effect of the movement and vibration of the UAV on the crack images is minimized.   The obtained crack images can be blurred because of the intrinsic vibration and movement of the UAV, and thus the image blur is an important issue that has to be addressed.   The image blur is closely related to the exposure time of the camera shutter when capturing images, and can be alleviated by increasing the shutter speed, resulting in low brightness.   Thus, an optimal shutter speed has to be selected considering the trade-off; a shutter speed of 1/1000 s for the LS-20150 camera module is sufficiently fast to produce bright and clear images in most cases.

All the components of the proposed system are selected to be low-cost and lightweight. The total weight of the sensing and communication components (i.e., Raspberry Pi with the camera, the ultrasonic displacement sensor, and the WiFi module) is approximately 60 g, which does not significantly affect the flight of the UAV.   To further reduce the weight, the sensing and communication components are designed to share the UAV's battery.   The power consumption of the Raspberry Pi is approximately 2 W, which is significantly less than that of the UAV (70 W).

A field testing is conducted to demonstrate the validity of the proposed hybrid image processing in conjunction with the UAV.   First, parametric analysis is conducted to determine the two sets of optimal parameters, P_w and P_l.   Subsequently, crack quantification by the proposed hybrid image processing is performed using crack images obtained from the UAV-based system.

In the parametric analysis, 20 crack images with different crack widths, surface textures, sizes, lengths, and directions were obtained to address various concrete conditions.   The collected image pool is processed by Sauvola's method with a wide range of sensitivity and window size.   Then, crack width and length information are measured.   Fig. 4.11 shows a

typical image included in the image pool.   An optical microscope is then utilized here to

calculate reference widths at the specific locations, where the color targets are attached, as

shown in Fig. 4.11.

Two cost functions are defined to determine the optimal parameters in terms of crack

width and length.   The first cost function J$_w$, for optimal width, is defined as,

$$\mathrm{J}_w = \frac{1}{N_p N_t} \sum_{i=1}^{N_p} \sqrt{\sum_{j=1}^{N_t} \left(\frac{w_e - w_m}{w_m} \right)^2_{ij}}$$
<div align="right">Eq. 4.5</div>

where N_p is the total number of images in the pool, N_t is the number of color targets, w_e is the

estimated crack width from Sauvola's method, and w_m is the width measured by the optical

microscope.   The second cost function J$_l$, for optimal length, is defined as,

$$\mathrm{J}_l = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\left| \frac{l_e - l_t}{l_t} \right| + \frac{C_d}{C_t} \right)_i$$
<div align="right">Eq. 4.6</div>

where l_e is the estimated crack length from Sauvola's method, and l_t is the total length verified

visually in the grayscale image. C_d and C_t are the numbers of measured crack and total pixels

in the entire binary image, respectively.   The second term in Eq. 4.6 prevents all the pixels

from being converted into cracks.

The cost function values of each crack information set (i.e., crack width and length) are

investigated with respect to the binarization parameters, as shown in Fig. 4.12, to determine

the two sets of optimal parameters.   From the results, the sensitivity is observed as a governing

factor rather than the window size in both cost functions.   The lowest cost function values in

each case, marked as the blue circles, are selected to determine the optimal parameters

summarized in Table 4.4.

Raspberry Pi camera   **Optical microscope**
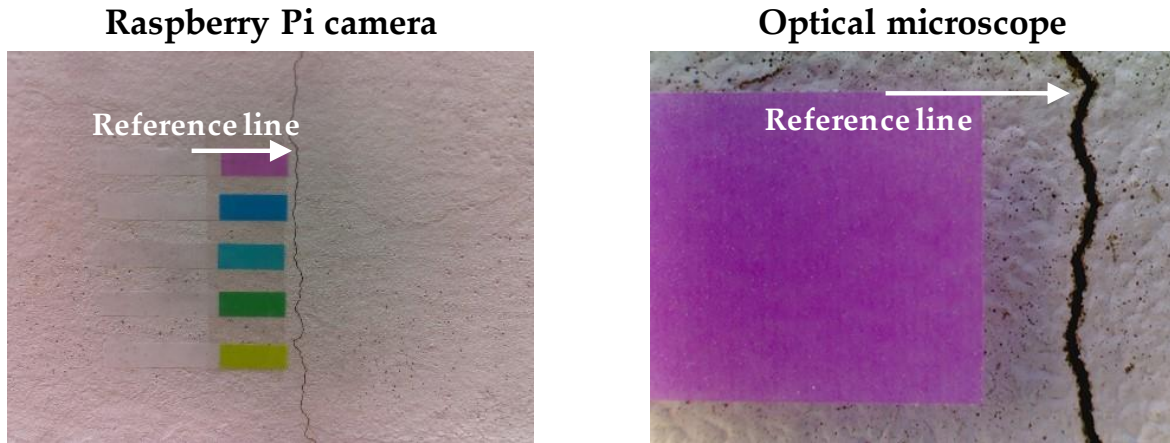
Fig. 4.11 Comparison of measured crack widths to references.

(a)   **(b)**
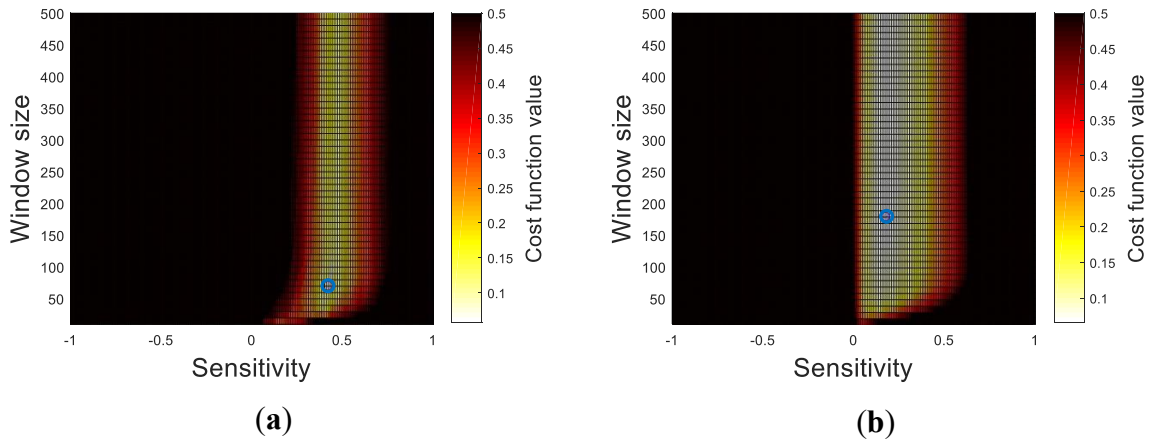
Fig. 4.12 Cost functions in terms of window size and sensitivity: **(a)** J_w and **(b)** J_l.

Table 4.4 Optimal crack width and length parameters.

	Sensitivity	Window size	Cost function
P_w	0.42	70	0.057
P_l	0.18	180	0.065

A field testing is conducted on a concrete wall of the gymnasium building in the UNIST campus (see Fig. 4.13).   The concrete wall has diverse crack sizes and shapes owing to shrinkage, creep, and external loads.   The UAV-based system acquires the crack images and

the corresponding working distances utilizing the Raspberry Pi camera and the displacement

sensor, while flying in front of the concrete wall.   Note that the crack widths are also measured

by the optical microscope as the reference, comparing with the UAV-based crack system.

Fig. 4.13 Experimental validation using concrete wall.

The crack quantification based on the hybrid image processing is applied to the obtained

images.   To validate the performance of the proposed hybrid strategy, the binarization results

are compared with those of Sauvola's method with default parameters adopted from [18].   As

shown in Fig. 4.14, the proposed hybrid approach outperforms the default parameters.   Note

that the black objects on the bottom side of the captured images are the part of the UAV.

Quantitative comparisons of crack widths and lengths are conducted at a total of 15 points in

three crack regions as presented in Table 4.5 and Table 4.6.   With the default values, cracks

with widths less than 0.25 mm were typically unidentified or underestimated.   In contrast, the

hybrid method measured all range of crack widths reliably, because small cracks unidentified

by Sauvola's method using P_w can also be identified and calibrated accurately.

Cracks thinner than about 0.1 mm, which can be seen with the naked eye, are not found

even with the hybrid image processing as shown in Fig. 4.14. Although the measurement accuracy of digital image processing is not as good as the manual visual observation, its efficiency in terms of inspection time is critical particularly when a number of crack images are to be processed. The accuracy-related issue can be resolved by appropriate camera and lens that can identify cracks thicker than the minimum width of interest for maintenance purposes.

(a)

(b)

(c)

(d)

(e)

(f)
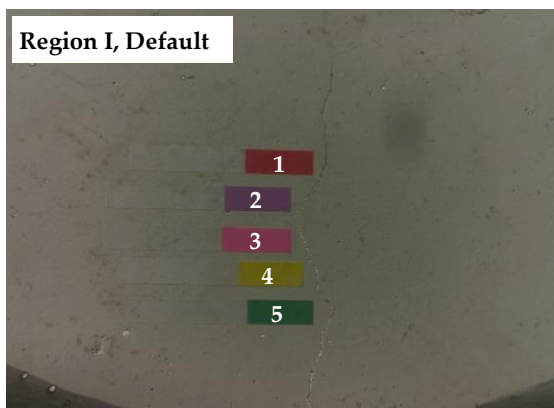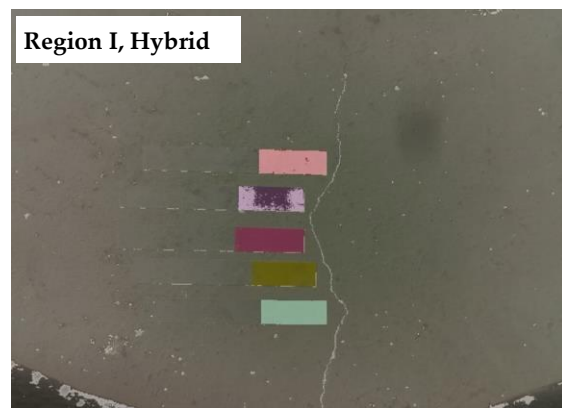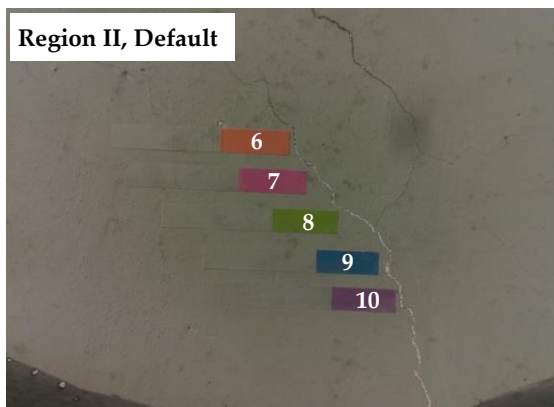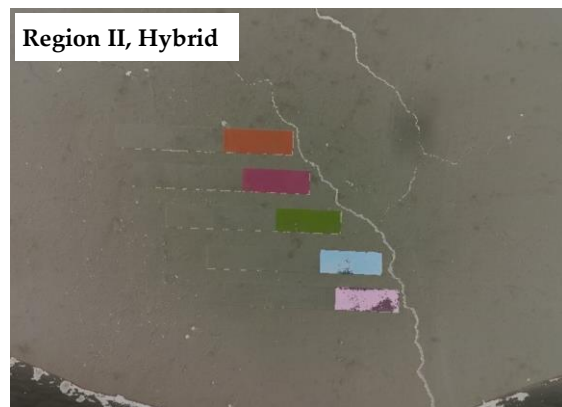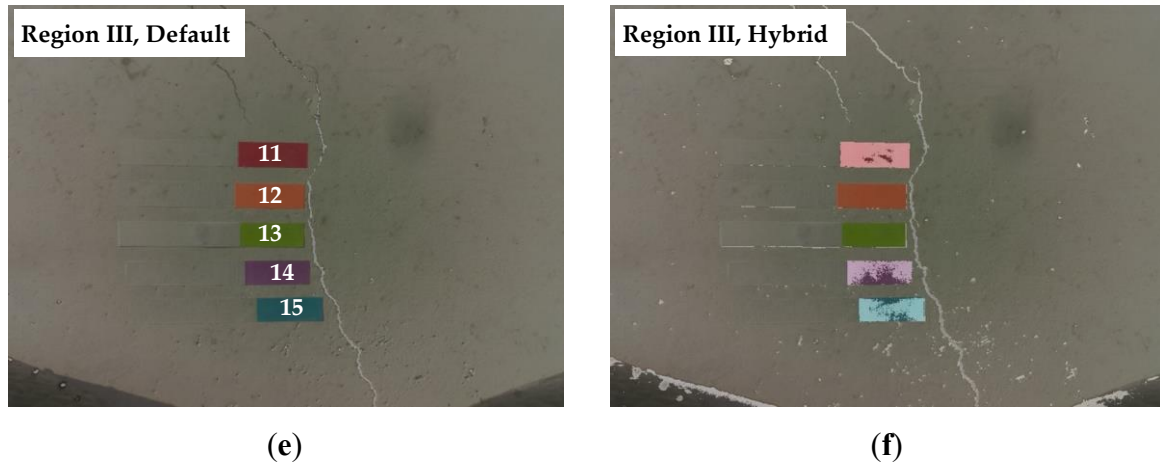
Fig. 4.14 Crack identification results: (a) region I, Sauvola's method with default parameter, (b) region I, hybrid method, (c) region II, Sauvola's method with default parameter, (d) region II, hybrid method, (e) region III, Sauvola's method with default parameter, and (f) region III, hybrid method.

Table 4.5 Comparison of obtained crack widths.

Region	Location	Crack width calculation (mm)		
		Default (difference)	Hybrid (difference)	Microscope
	1	N/A*	0.14 (0.02)	0.12
	2	N/A*	0.14 (0.02)	0.12
I	3	0.15 (-0.07)	0.20 (-0.02)	0.22
	4	0.15 (-0.08)	0.20 (-0.03)	0.23
	5	N/A*	0.13 (-0.01)	0.14
	6	N/A*	0.22 (0.03)	0.19
	7	0.20 (-0.03)	0.25 (0.02)	0.23
II	8	0.30 (-0.02)	0.30 (-0.02)	0.32
	9	0.25 (0.01)	0.25 (0.01)	0.24
	10	0.35 (-0.04)	0.40 (0.01)	0.39

	11	N/A*	0.22 (0.03)	0.19
	12	0.49 (-0.04)	0.49 (-0.04)	0.53
III	13	0.49 (-0.01)	0.49 (-0.01)	0.50
	14	0.59 (0.04)	0.59 (0.04)	0.55
	15	0.59 (0.04)	0.59 (0.04)	0.55

* The crack is unidentified.

Table 4.6 Comparison of obtained crack lengths.

Region	Total crack length calculation (mm)		
	Default (error)	Hybrid (error)	Manual
I	37.49(52.3%)	72.86(7.3%)	78.57
II	79.18(42.0%)	128.75(5.7%)	136.50
III	95.01(18.8%)	115.99(0.9%)	117.02

4.4　Summary

This research presented a framework for accurate crack quantification using the hybrid image processing strategy in conjunction with UAV.　A prototype of the UAV-based system was designed by a Raspberry Pi connected to a camera, an ultrasonic displacement sensor, and a WiFi module.　The Raspberry Pi controlled sensing and wireless communication, providing crack images with associated distances on demand.　The obtained information was subsequently processed by the hybrid image processing method using two sets of optimal parameters P_w and P_l, to accurately measure crack widths while minimizing loss of crack lengths.　The results of experimental evaluation can be summarized as follows:

(1) While the crack widths less than 0.25 mm were typically unidentified or underestimated in case of the default values, the proposed hybrid method measured all ranges of crack widths reliably.

(2) The maximum length estimation errors were 7.3 % and 52.3 % for the hybrid method and Sauvola's binarization with the default parameters, respectively, proving significant performance improvement by the hybrid method.

Consequently, the results of experimental evaluation on a concrete wall show that the proposed UAV and hybrid image processing-based crack identification strategy effectively and reliably quantifies cracks.

CHAPTER 5   CAMERA SYSTEMS FOR CIVIL INFRASTRUCTURE

This chapter develops a couple of camera systems with the associated strategies to properly address the practical issues on civil infrastructure, including a skewed angle problem and an efficient measurement for the sequential crack images.   Previous chapters have shown a potential of computer vision-based techniques with deep learning for concrete crack detection and quantification; however, advanced camera systems are necessary to handle the skewed angle issue and efficiently extract all the crack information on the sequential images.   For these purposes, two camera systems are presented for accurate crack evaluation, as follows: (1) RGB-D camera-based system and (2) camera slider-based system.   The RGB-D camera-based system is specially designed to provide technical improvements for crack identification using digital image processing: a robust depth estimation considering the skewed angle problem and a high measurement resolution when images are captured from a distance.   Furthermore, another camera slider-based system is proposed to efficiently process the sequential crack images for crack measurements.

5.1   RGB-D Camera-based System

A camera system is developed in this study to simultaneously acquire a high-resolution crack image and the corresponding depth information.   The proposed camera system (Fig. 5.1) is designed to combine a commercial RGB-D camera with a high-resolution digital camera, improving the measurement resolution.   The high-resolution crack image captured using the digital camera (with a focal length of 500 mm) enables crack identification from a distance; this cannot be achieved by the RGB-D camera.   Furthermore, as both cameras are perfectly aligned and bolted to share a part of the angle of view, a stereo calibration between the cameras is needed only once regardless of the position of the tripod.   Table 5.1 lists the detailed

specifications of the system.
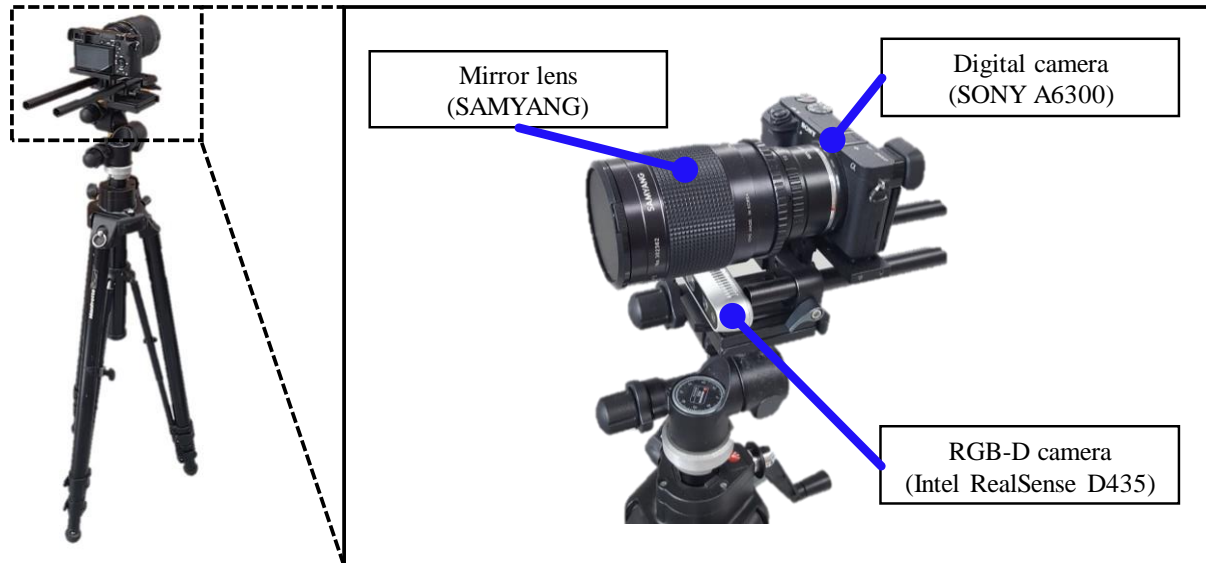
Fig. 5.1 Camera system in the proposed approach.

Table 5.1 Specification for system components.

Component	Model	Specification
Digital camera	SONY A6300	Image resolution: 6000 × 4000 pixels Weight: 404 g
Lens	SAMYANG 500mm F8.0 MC Mirror	Focal length: 500 mm F-number: f/8.0 Weight: 330 g
RGB-D camera	Intel RealSense D435	Image resolution: 1280 × 720 pixels Depth range: 0.2–10 m Weight: 72 g

The RGB-D camera (Intel RealSense D435), located under the digital camera, is used to obtain depth as well as RGB information.   The RGB-D camera is based on the active infrared (IR) stereo vision technology, wherein non-visible static IR patterns are projected using an embedded IR projector; subsequently, the patterns captured using the two embedded IR cameras are utilized for stereo camera calibration to calculate the depth value in each pixel.

Although this camera itself can provide RGB and depth information, the depth values can be sensitive to environmental conditions, such as sunlight, because the wavelength of the projected IR patterns falls in the range of wavelength of sunlight. Furthermore, it is difficult to identify cracks from a distance because of the low-resolution of the image (approximately 1 megapixels) captured using a camera with a short focal length lens. To improve the measurement resolution, a high-resolution crack image captured using the digital camera with a relatively long focal length lens is used instead of the low-resolution crack image. The digital camera (SONY A6300) is equipped with a mirror lens (SAMYANG 500 mm F8.0 MC Mirror) to capture high-resolution crack images of concrete structures. The mirror lens is selected instead of a zoom lens, owing to its low weight, portability, and cost-effectiveness. Furthermore, the mirror lens has a relatively long focal length of 500 mm, allowing to capture high-resolution crack images from a farther distance.

With the camera system, a crack identification strategy based on a sensor fusion algorithm is proposed. The main objective is to accurately measure crack information regardless of the angle of view by combining the commercial RGB-D and high-resolution digital cameras using a sensor fusion algorithm. The proposed crack identification method consists of two steps: (1) approximation of the concrete surface to a plane in a 3D space, and (2) crack width calculation based on a coordinate transformation.

An approximate plane corresponding to the concrete surface is determined for a robust and accurate estimation of the depth data. Although the RGB-D camera itself can provide RGB and depth information, the obtained depth is sensitive to environmental conditions. In particular, in the case of outdoor environments, the noise in the depth data can be significant, because solar radiation degrades the identification performance of RGB-D cameras. As the use of such noisy depth measurement decreases the accuracy, a plane model that approximates the concrete surface is employed for a robust estimation of the depth data. Note that this study

focuses on the plane surfaces of concrete structures to investigate the skewed angle of view of the camera system.

To construct the approximate plane corresponding to the concrete surface, random sample consensus (RANSAC) [130] is employed. A point cloud, which comprises image coordinates in the x- and y-axes and depth in the z-axis of the RGB-D camera, is generated from the region of interest (ROI) on the concrete surface, from which a plane model is employed for RANSAC.

$$ax + by + cz + d = 0 \qquad \text{Eq. 5.1}$$

where a, b, c, and d are the constants of the plane equation. The approximate plane considered as a consensus model is constructed using the inliers in the point cloud, in which the depth on the predetermined plane can be expressed as follows:

$$Z(x, y) = \frac{-a_o x - b_o y - d_o}{c_o} \qquad \text{Eq. 5.2}$$

where $Z(x,y)$ is the depth on the approximate plane in terms of the image coordinates in the x- and y-axes of the RGB-D camera; and a_o, b_o, c_o, and d_o are the constants in the approximate plane equation. As the depth calculated from the approximate plane of the concrete surface is provided for each pixel, a robust and accurate estimation can be conducted with minimal noise in the depth information. Fig. 5.2 shows a graphical description of the process of constructing the approximate plane.

A sensor fusion method, which combines the RGB-D and high-resolution digital cameras, is proposed to improve the crack measurement resolution. Although the RGB-D camera can provide RGB and depth information for crack identification, the low-resolution of the image and the relatively short focal length are the governing factors that need to be carefully addressed for an effective crack inspection from a distance. To overcome these issues, a high-resolution crack image taken using the digital camera (with a 500 mm focal length lens) is used instead

of the low-resolution crack image taken using the RGB-D camera based on a coordinate transformation. The crack identification based on the sensor fusion method is done in two steps: (1) detection of crack pixels and (2) width calculation based on the coordinate transformation.

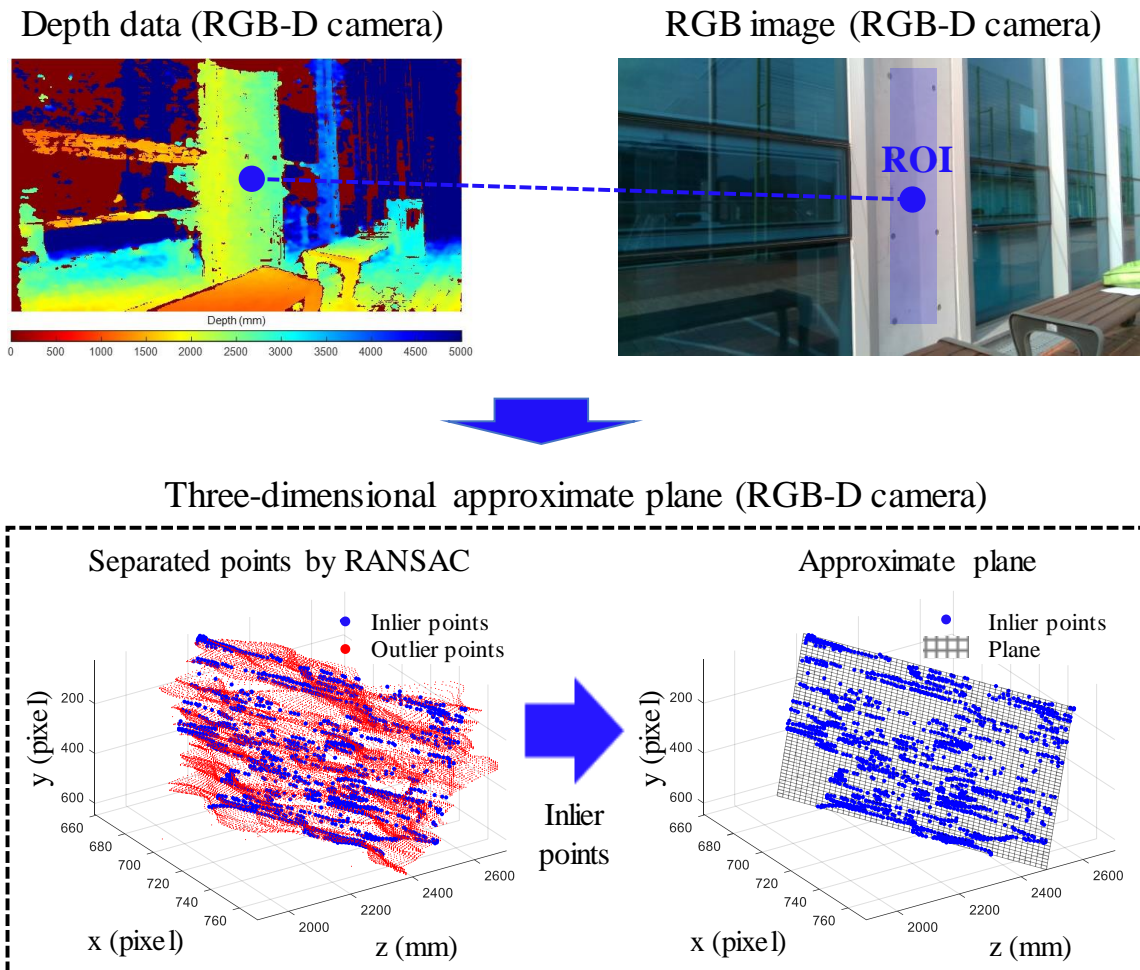Fig. 5.2 Step 1: Plane approximation of concrete surface.

The proposed frameworks for crack detection and quantification described in CHAPTER 3 and CHAPTER 4 are applied to the RGB images captured using the digital camera to identify the locations of crack pixels. The crack pixels on the high-resolution crack image are transferred onto the approximate plane of the RGB-D camera based on a coordinate

transformation to accurately calculate the crack widths. Although the digital camera provides a high-resolution crack image, the lack of depth information in each pixel decreases the crack measurement accuracy when the camera is not vertically aligned with the concrete surface. To obtain depth values in each crack pixel, the calibration patterns in a checkerboard are captured using both the digital and RGB-D cameras; subsequently, a coordinate transformation is modeled using two linear equations in the x- and y-axes to map the crack pixels. The corresponding crack pixels in the image coordinate of the RGB-D camera are applied to the approximate plane in Eq. 5.2 to calculate the depth of each pixel. The 3D points of the crack pixels, which comprise image coordinates in the x- and y-axes and depth in the z-axis, are further mapped onto the camera coordinate of the RGB-D camera utilizing intrinsic parameters.

$$sf \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & skew & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X(x_c, y_c) \\ Y(x_c, y_c) \\ Z(x_c, y_c) \end{bmatrix} \qquad \text{Eq. 5.3}$$

where x_c and y_c represent the location of the crack pixels in terms of the image coordinates in the x- and y-axes of the RGB-D camera, respectively; $skew$ is the skew coefficient; sf is the scale factor; f_x and f_y are the focal lengths; and c_x and c_y are the principal points. Here, the skew coefficient is zero, and $Z(x_c, y_c)$ is the depth value on the approximate plane in Eq. 5.2 corresponding to the crack pixel. $X(x_c, y_c)$ and $Y(x_c, y_c)$ are the locations of the crack pixels in terms of the camera coordinate of the RGB-D camera, which can be derived from Eq. 5.3 as follows:

$$X(x_c, y_c) = \frac{(x_c - c_x)Z(x_c, y_c)}{f_x} \qquad \text{Eq. 5.4}$$

$$Y(x_c, y_c) = \frac{(y_c - c_y)Z(x_c, y_c)}{f_y} \qquad \text{Eq. 5.5}$$

As the depth $Z(x_c, y_c)$ of the crack pixels in the camera coordinate of the RGB-D camera is

known from the approximate plane, the 3D coordinates of each crack pixel can be determined.

The crack width is then calculated using the 3D coordinates of the crack pixels. The skeleton and edges of the crack pixels in the binary image are transferred onto the camera coordinate of the RGB-D camera. Subsequently, the crack direction in each skeleton pixel is calculated based on its adjacent pixels, from which two edge pixels nearest to the skeleton pixel are used to calculate the crack width. Fig. 5.3 shows the schematic of the width calculation based on the coordinate transformation.
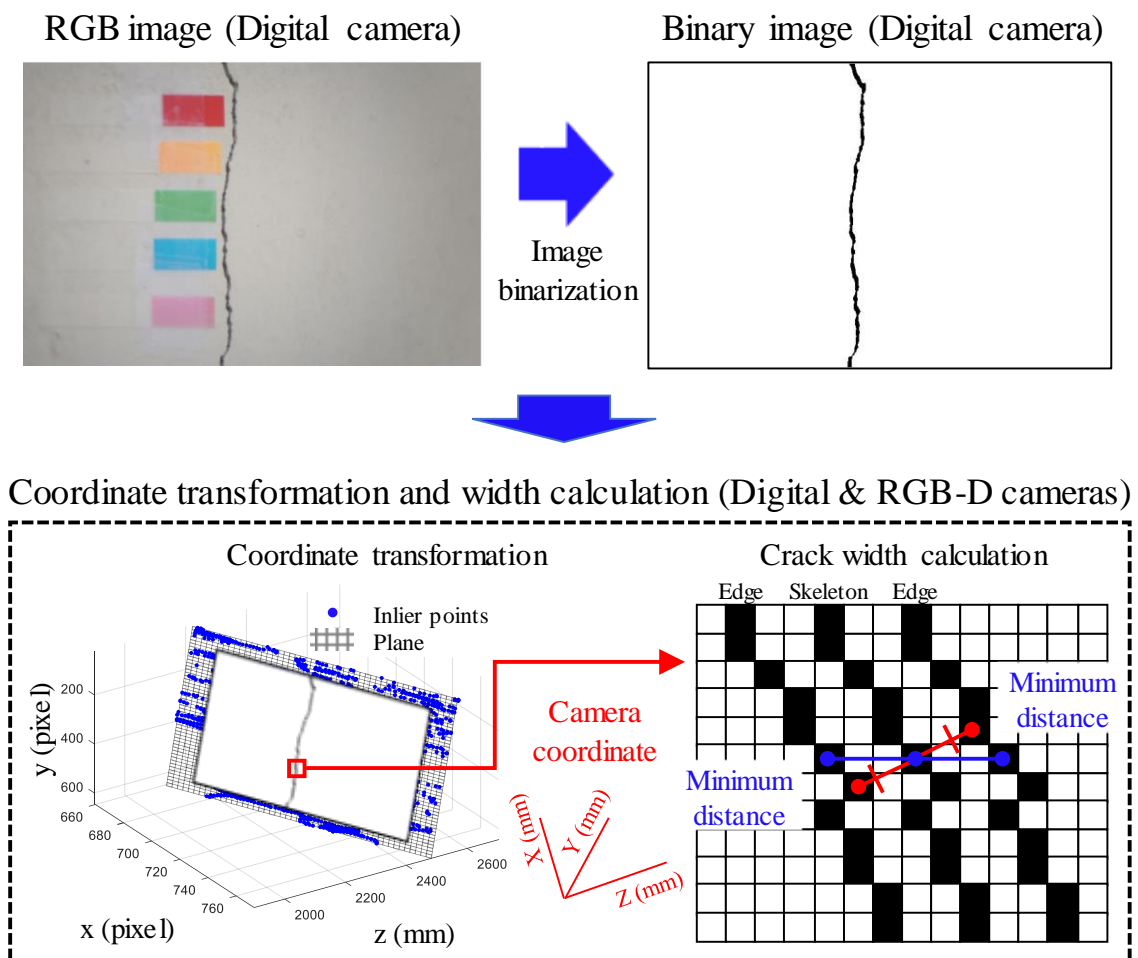
Fig. 5.3 Step 2: Crack width calculation based on a coordinate transformation.

5.2 Camera Slider-based System

Another camera slider-based system is presented to sequentially capture crack images on the surface of concrete structures.   The proposed camera system, as shown in Fig. 5.4, is designed by combining a commercial automatic slider and a high-resolution digital camera with a macro lens for improving the measurement resolution.   Thus, the digital camera implemented on the automatic slider can efficiently provide a high-resolution set of the sequential crack images. The detailed specifications of the camera system are listed in Table 5.2.
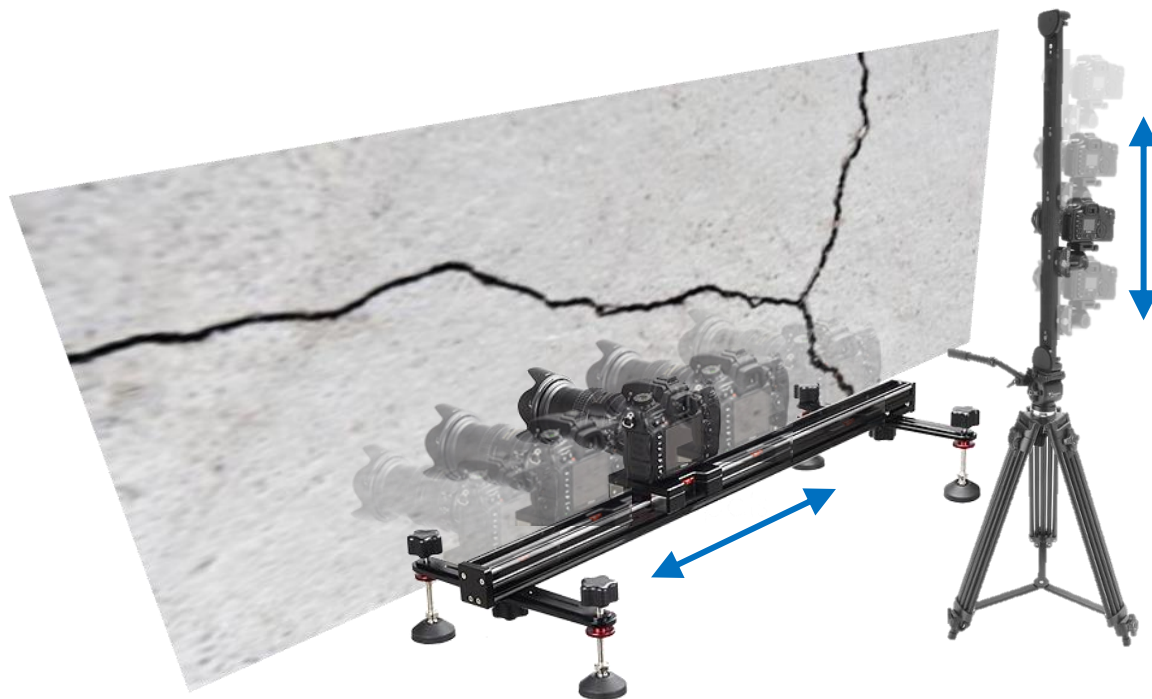
Hardware configuration for scanning surface cracks
• High-resolution digital camera
• Macro lens for a close shot
• Camera slider for automatic movement
• Control interface using a smartphone

Fig. 5.4 Camera system in the proposed approach.

Table 5.2 Specification for system components.

Component	Model	Specification
Automatic slider	KONOVA Slider K5	Maximum velocity: 0.1 m/s Minimum velocity: 0.003 m/s
Digital camera	SONY A7R2	Image resolution: 7952×5304 pixels Weight: 625 g
Lens	SONY SEL50M28	F-number: F2.8 Weight: 236 g

The proposed camera system is optimized to provide the sequential images for fully identifying microcracks. According to American Concrete Institute ACI 224R-90 [131], tolerable crack widths under various exposure conditions are less than 0.4 mm. To precisely measure such microcracks, the digital camera (SONY A7R2) is equipped with a macro lens (SONY SEL50M28) to take close-up crack images with a high-resolution. Furthermore, as the Bluetooth connectivity provides wireless connection between the operator's smartphone and the digital camera mounted on the automatic slider, the proposed camera system allows operators to acquire the sequential crack images when desired and to wirelessly transmit the obtained data.

With the sequential crack images obtained from the proposed camera system, a tailored crack identification strategy is presented for efficient crack measurements. Although a panorama image constructed by stitching the captured crack images allows crack measurements, it is inefficient because a combination of high-resolution images cause memory problems. To properly handle this issue, a distributed crack identification strategy based on image registration techniques is developed. The proposed approach consists of two steps: (1) estimation of a registration information between the sequential crack images and (2) the distributed crack identification to calculate crack information on each image.

To obtain the registration information between the sequential crack images, the SURF features are utilized.   SURF is designed to detect the interest points on elements of the image such as blobs, corners, and edges, of which Haar wavelet responses are calculated within a circular neighborhood to assign an orientation; subsequently, a square region is generated along the obtained orientation to address the image rotations.   A feature vector with 64 elements is finally computed using the Haar wavelet responses in both the horizontal and vertical directions in 4 × 4 sub-regions.   Because a concrete surface image typically contains a large number of the SURF features on the noisy surface texture, such features in each image can be utilized to detect correspondences between the sequential crack images for calculating a 2-D geometric transformation (Fig. 5.5).   Here, the pairwise distance between the feature vectors is computed to find the matching features, from which a homography matrix (projective transformation) can be estimated as follows:

$$w \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
Eq. 5.6

where x' and y' represent the location of the matching features in terms of the image coordinates in the x- and y-axes of the first image, respectively; x and y represent the location of the matching features in terms of the image coordinates in the x- and y-axes of the second image, respectively; H is the homography matrix between the sequential crack images.   However, because the mismatched features can decrease the accuracy of the projective transformation, RANSAC [130] is employed here.   Among the extracted matching features, a random subset is selected to estimate the homography matrix between the sequential crack images.   This process is repeated when the approximate matrix is finally considered as a consensus model that contains a high number of inliers.   For example, when the inliers in Fig. 5.6 are used to estimate the homography matrix using Eq. 5.6, two outliers can be successfully removed.

All the projective transformations are finally converted with respect to the first image; subsequently, the minimum and maximum output limits over all transformations are calculated to determine the area without overlap between the sequential crack images.

A distributed crack identification is performed in each crack image based on the obtained registration information between the sequential images. As the concept of the proposed approach is to calculate crack information in each image instead of a full panorama image, efficient crack identification can be conducted. For this purpose, all the projective transformations with respect to the first image are applied to each image, respectively, applying geometric transformations. The proposed frameworks for crack detection and quantification described in CHAPTER 3 and CHAPTER 4 are subsequently employed for identifying the locations of crack pixels. In the case of the conventional method, a full panorama image constructed by stitching the sequential crack images is directly utilized to calculate crack information, resulting in computational inefficiency. However, the proposed approach can efficiently obtain crack information from each image without overlap between the sequential crack images. The distributed crack identification strategy is graphically compared with the crack detection on a panorama image, as shown in Fig. 5.7.
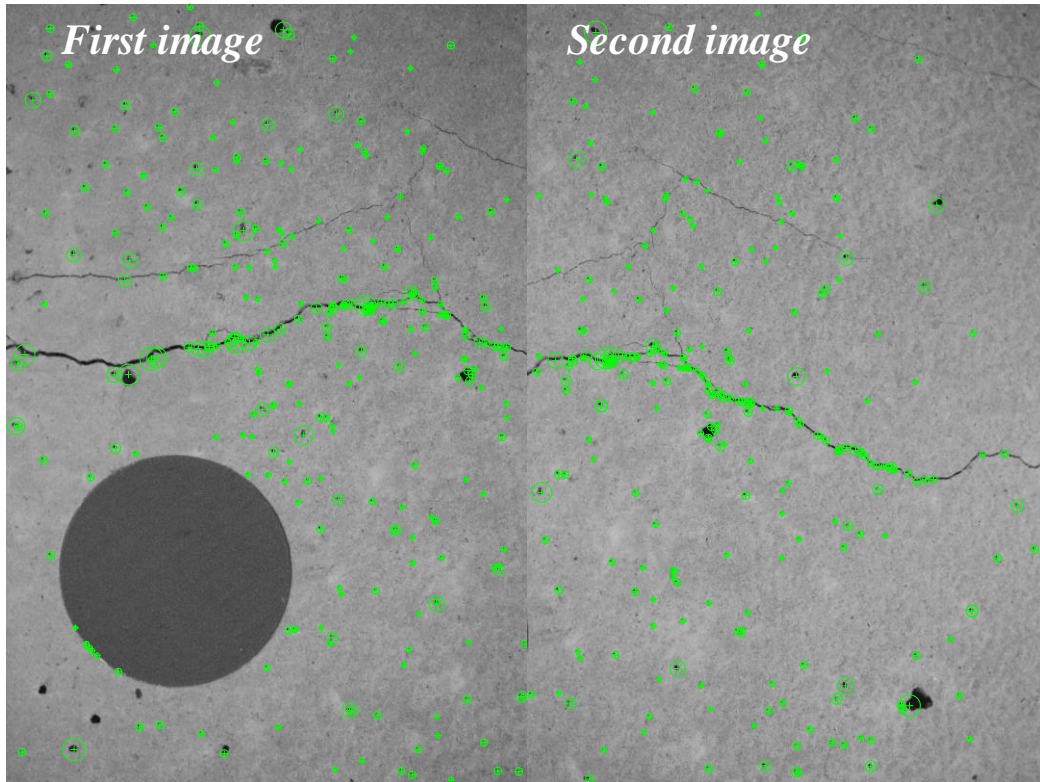
Fig. 5.5 Detection of SURF features on the sequential crack images.
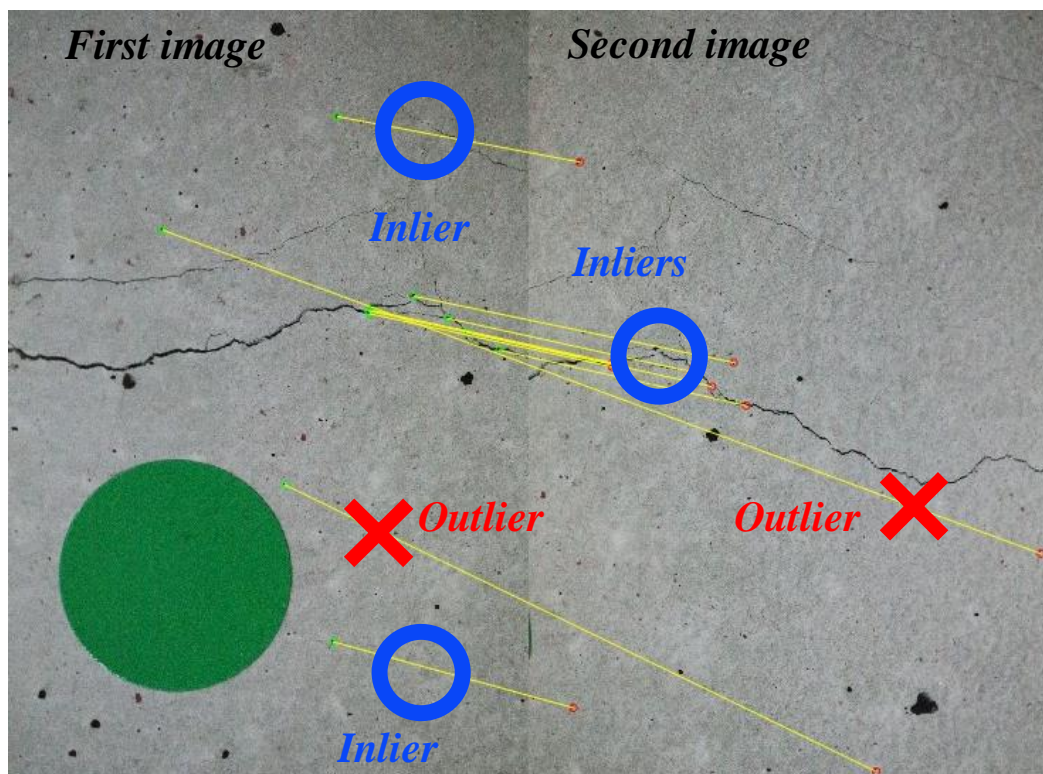
Fig. 5.6 RANSAC-based outlier detection for robust estimation of projective transformations.

69
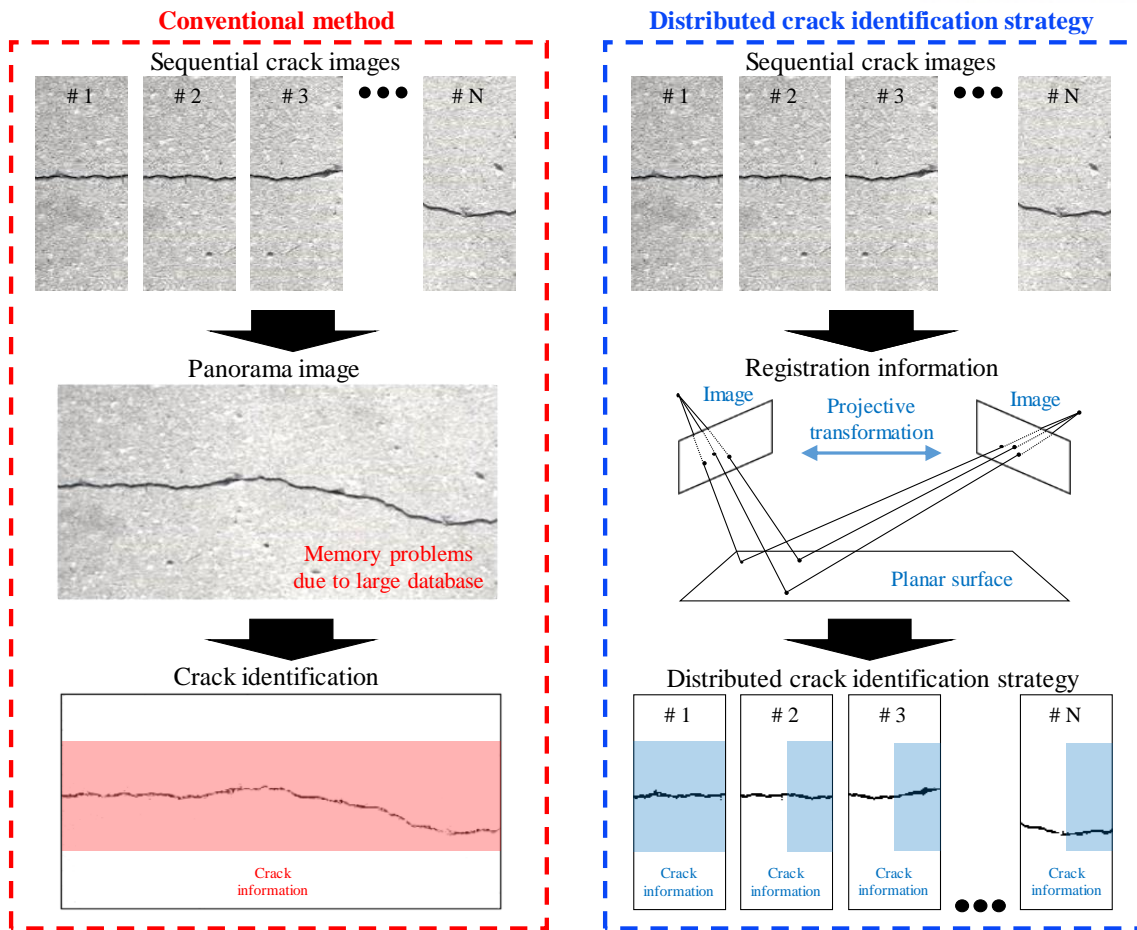
Fig. 5.7 Distributed crack identification strategy based on registration information.

5.3 Experimental Validation

5.3.1 Skewed Angle

To evaluate the crack evaluation performance of the proposed RGB-D camera system, a field test is conducted on a concrete wall of a gymnasium building in the UNIST campus (Fig. 5.8). Five color targets are utilized, from which 50 points between the edges in each target are used to calculate the average width, as shown in Fig. 5.9. An optical microscope is employed to provide the reference widths at the specified locations where the color targets are attached. Table 5.3 lists the specifications of the optical microscope.

Fig. 5.8 Experimental validation of the proposed approach applied to a concrete wall.

Digital camera

Optical microscope

Fig. 5.9 Comparison of average crack width.

Table 5.3 Specifications of the optical microscope.

Company	Model	Specification
DIGIBIRD	MSP2000	RGB resolution: 1600×1200 pixels Weight: 124 g

The proposed camera system is used to acquire a high-resolution crack image and its corresponding depth information with respect to various angles of view. Based on the vertical line starting from the concrete surface (0°), the camera is located approximately 2.5 m away and moved to the left side with an interval of 20°, as shown in Fig. 5.10. Consequently, the

71

same crack region with five color targets is captured from four different angles of view (i.e., 0, 20, 40, and 60°).   Fig. 5.11 and Fig. 5.12 show the RGB and depth data obtained from the RGB-D camera, respectively.   As the RGB-D camera captures low-resolution crack images because of the relatively short focal length, it is inappropriate for identifying cracks from a distance, as shown in Fig. 5.11.   In the case of depth information, shown in Fig. 5.12, the measurement noise is significant on the mirror part, because the projected IR patterns are dispersed in such materials.   However, the skewness of the camera with respect to the concrete surface is reflected with respect to the angle of view.   In contrast to the RGB-D camera, the digital camera with a relatively longer focal length lens captures a high-resolution crack image, as shown in Fig. 5.13.
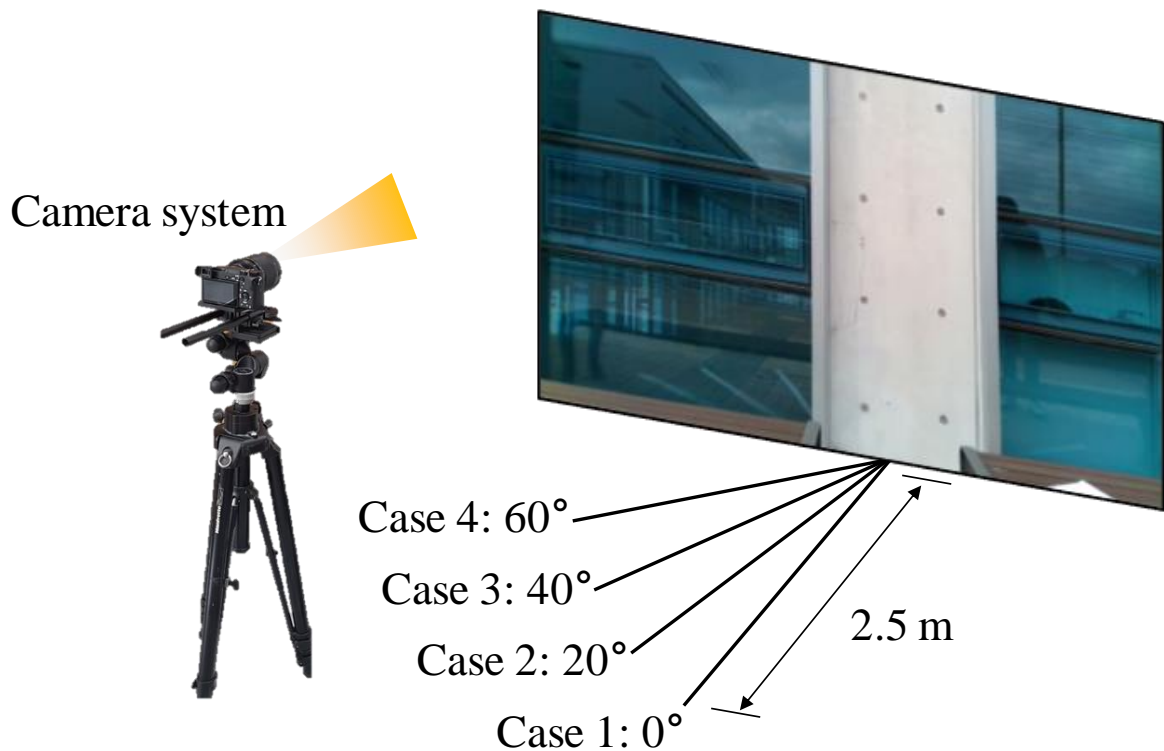
Fig. 5.10 Experimental cases with four different angles of view.

(a)

(b)

(c)

(d)
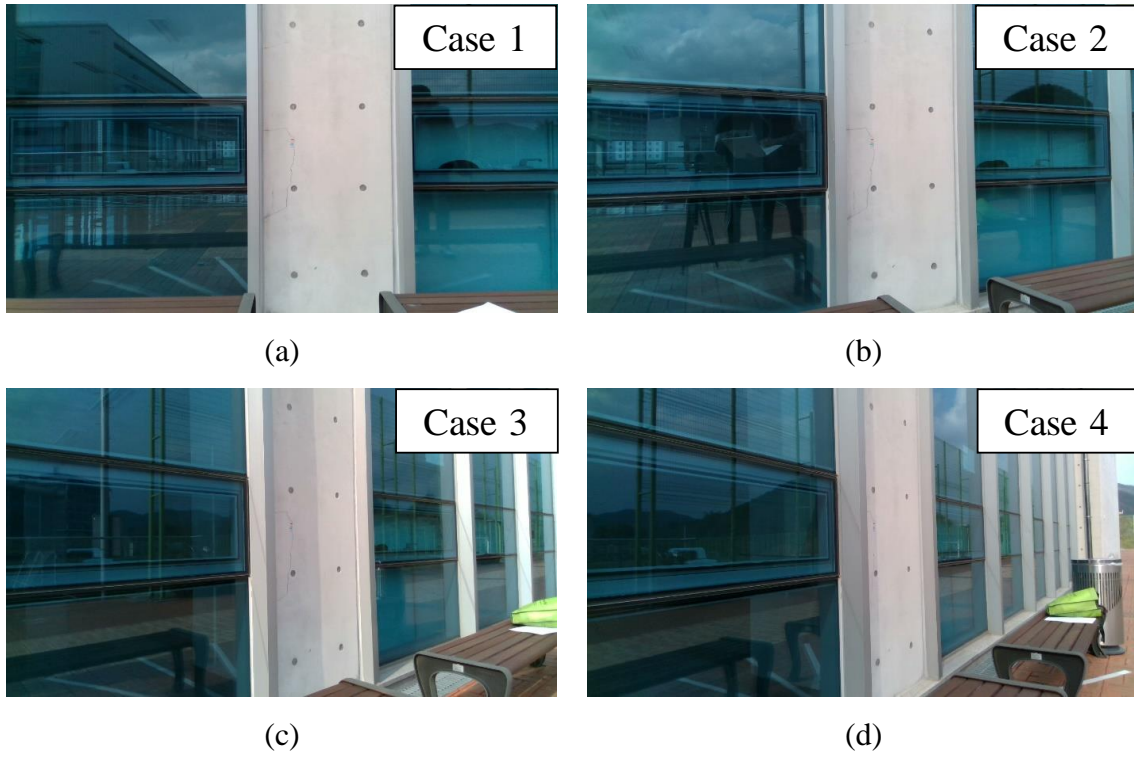
Fig. 5.11 RGB images captured using an RGB-D camera: (a) Case 1 with 0°, (b) Case 2 with 20°, (c) Case 3 with 40°, and (d) Case 4 with 60°.

(a)

(b)

(c)

(d)
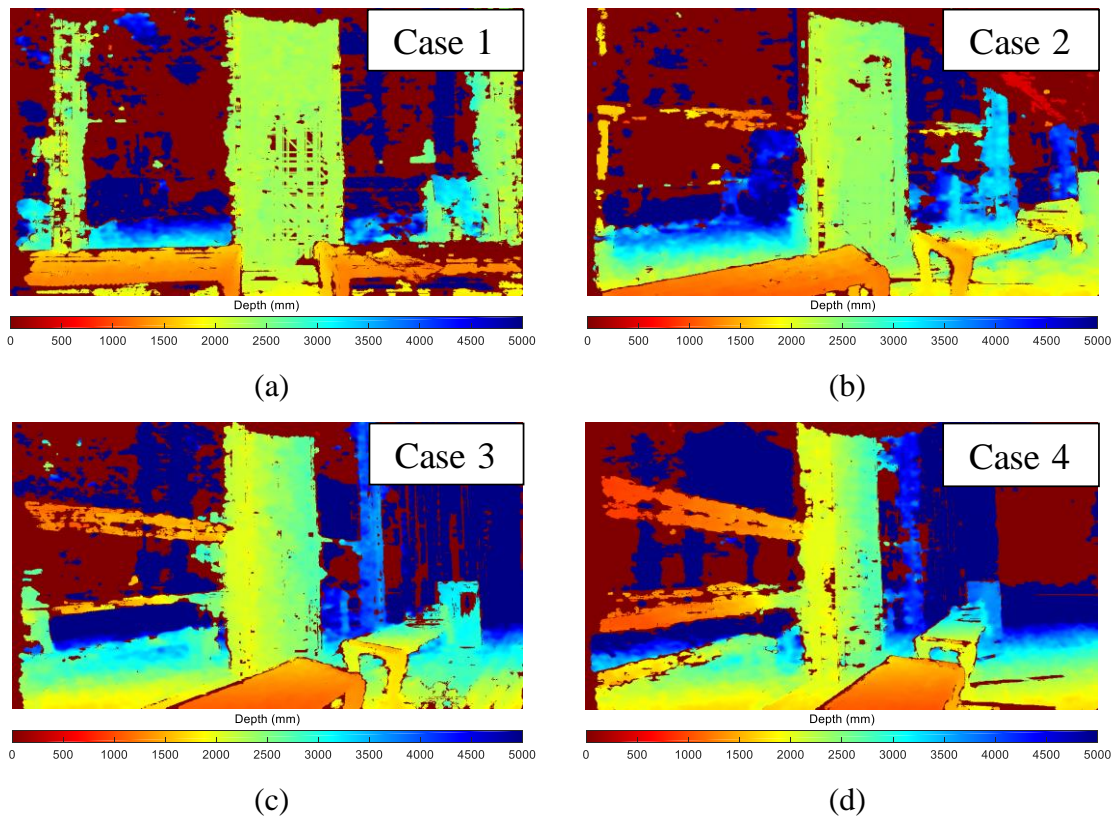
Fig. 5.12 Depth information obtained using an RGB-D camera: (a) Case 1 with 0°, (b) Case 2 with 20°, (c) Case 3 with 40°, and (d) Case 4 with 60°.

(a)   (b)

(c)   (d)
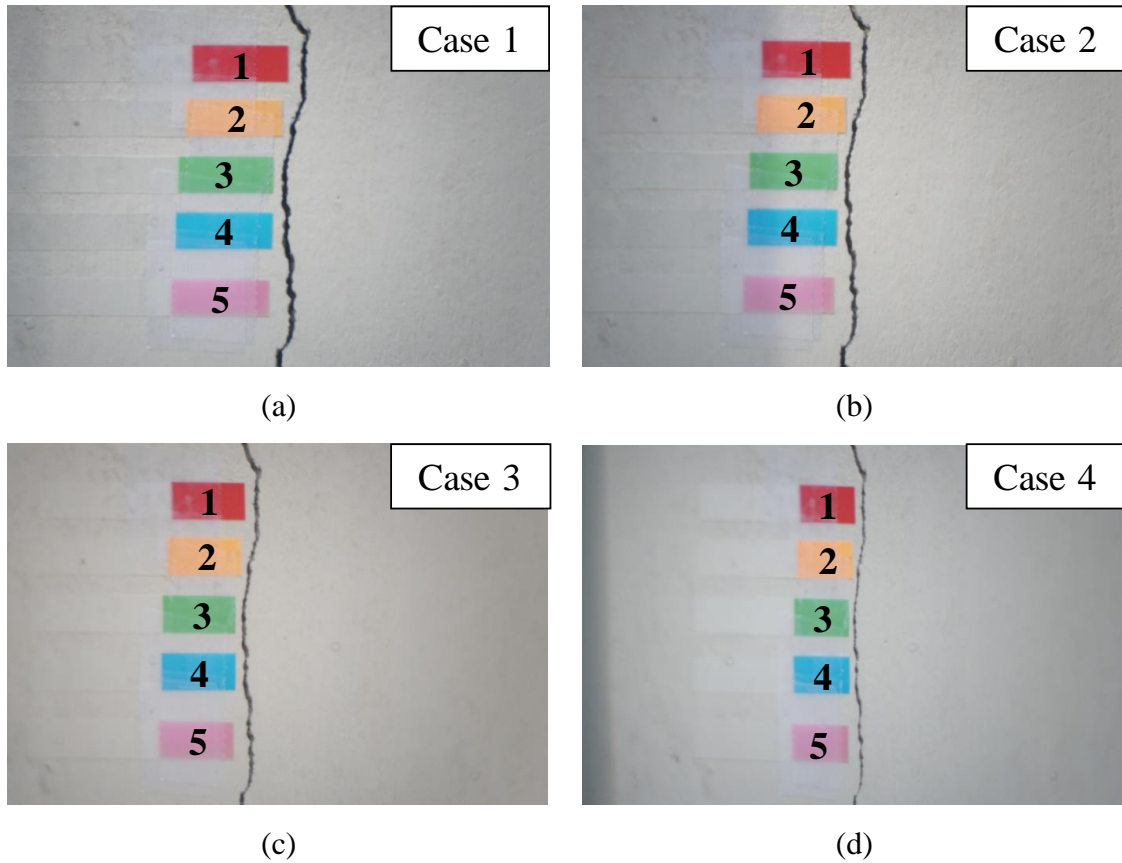
Fig. 5.13 RGB images captured using a digital camera: (a) Case 1 with 0°, (b) Case 2 with 20°, (c) Case 3 with 40°, and (d) Case 4 with 60°.

The proposed sensor fusion strategy is applied to the high-resolution crack image and the corresponding depth information obtained from the camera system to calculate the crack widths. The point cloud within the concrete surface is first generated from the RGB-D camera, in which RANSAC is used to determine the approximate plane with a maximum allowable distance of 1 mm. Subsequently, the high-resolution crack image captured using the digital camera is processed by the proposed frameworks of crack detection and quantification in the previous chapters. Based on the coordinate transformation between the two cameras, the crack pixels

of the digital camera are transferred onto the approximate plane of the RGB-D camera to estimate the depth value of each crack pixel.   The 3D information is further mapped onto the camera coordinate of the RGB-D camera, from which two edge pixels nearest to the skeleton pixel are utilized to measure the crack width.

The crack widths calculated using the proposed sensor fusion approach are compared with the reference widths to validate the measurement accuracy with respect to the angle of view. Here, the conventional method utilized only the distance between the camera and the concrete surface for calibrating the pixels, which is additionally employed for verifying the performance of the proposed approach.   Table 5.4 lists the results of the crack widths determined using the optical microscope, proposed approach, and conventional method for the four cases with different angles.   The proposed sensor fusion methodology accurately measures the crack widths regardless of the angle of view.   In contrast, the measurement accuracy of the conventional method decreases significantly with the increase in the angle.   This is because, as the distance between the camera and the surface is used to calibrate the pixels in the conventional method, the skewness of the camera with respect to the concrete surface could not be reflected.   Thus, the crack widths can be accurately measured only when the camera is conveniently aligned with the target surface (Case 1).   In Cases 2, 3, and 4, the crack widths are gradually underestimated with the increase in the camera angles, because of the lack of depth information in each pixel.   Fig. 5.14 shows the changes in the measured widths at each location with respect to the four different angles.   In contrast to the conventional method, the crack widths measured using the proposed approach are reasonably close to the corresponding reference widths determined using the optical microscope.   The average difference between the proposed methodology and the optical microscopes is approximately 0.05 mm in all the cases.

Table 5.4 Crack measurement results.

Crack	Reference width (mm)	Case 1		Case 2		Case 3		Case 4	
		Prop[a]	Conv[b]	Prop[a]	Conv[b]	Prop[a]	Conv[b]	Prop[a]	Conv[b]
1	0.94	0.88	0.88	0.82	0.74	0.96	0.62	1.00	0.38
2	0.88	0.87	0.87	0.80	0.72	0.92	0.60	0.89	0.34
3	0.84	0.89	0.88	0.80	0.73	0.90	0.59	0.81	0.31
4	0.89	0.94	0.93	0.83	0.76	0.91	0.59	0.81	0.31
5	1.1	1.07	1.07	0.96	0.88	1.11	0.72	1.12	0.42

[a]Proposed approach.
[b]Conventional approach.

(e)
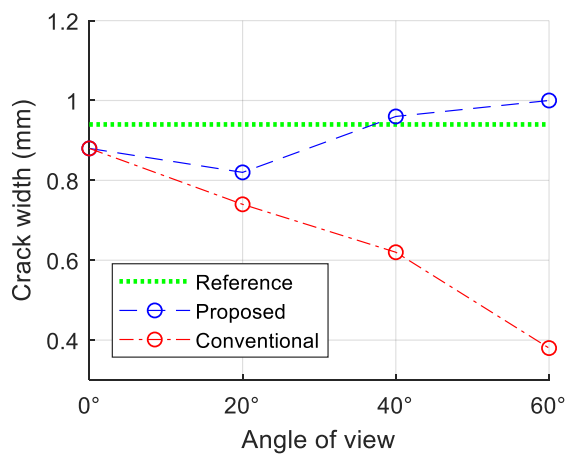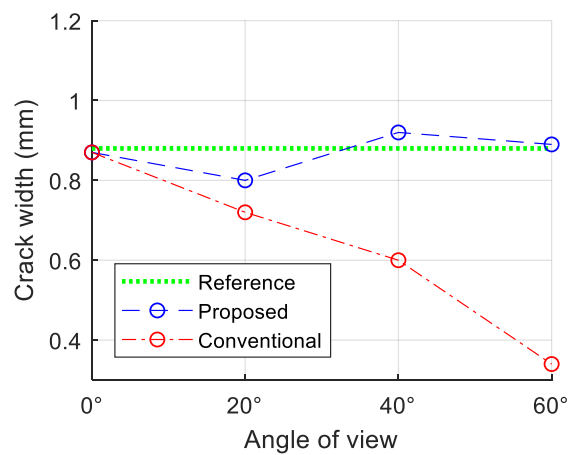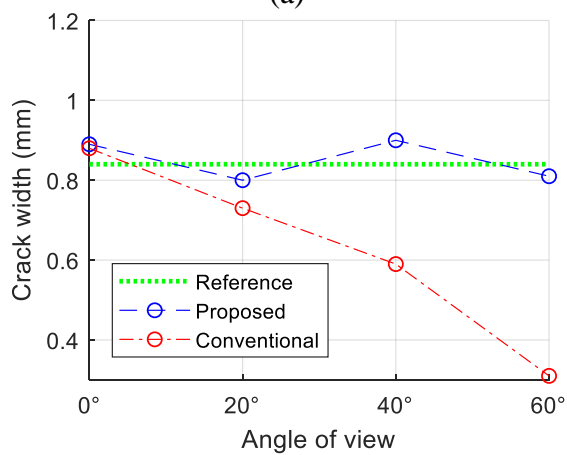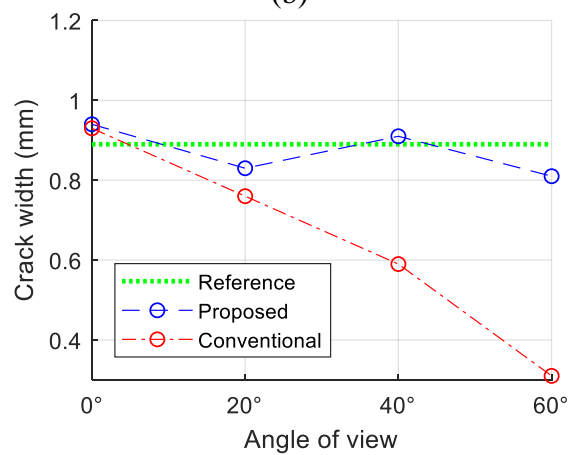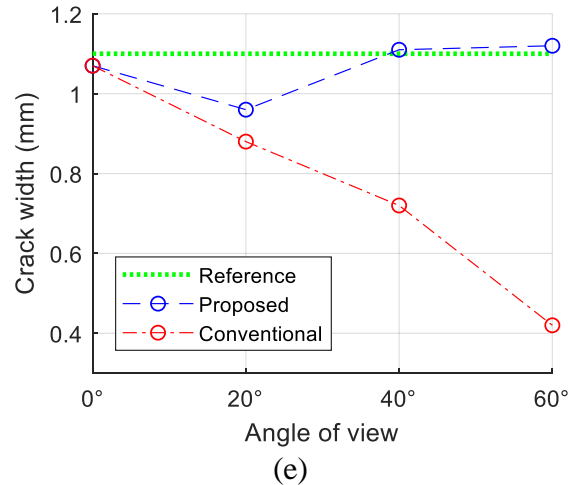
Fig. 5.14 Changes in measured crack widths with respect to four different angles of view: (a) location 1, (b) location 2, (c) location 3, (d) location 4, and (e) location 5.

5.3.2   Full Crack Measurement

The proposed crack evaluation strategy is applied to the monitoring of the self-healing process in concrete to validate the overall performance.   Here, three cases of concrete are designed with the following mixtures: (1) ordinary Portland cement (OPC), (2) capsules including the solid type of cement powder (CS), and (3) supplementary cementitious materials (SCM), such as ground granulated blast-furnace slag (GGBFS), fine and coarse clinker, and gypsum.   Note that the self-healing maters of CS and SCM are supported by other organizers.   All the concrete specimens are fabricated with 500-mm length, 250-mm width, and 120-mm depth (see Fig. 5.15), which have two reinforcements at the bottom of the specimen to prevent fractures.

A vertical crack is generated at the center of each specimen to induce the self-healing process.   After 28 days of curing, a universal testing machine (UTM, Instron HDX-1500) is utilized to perform a three-point bending test to generate the vertical crack at the center, as shown in Fig. 5.16a.   Here, two Pi-shape displacement transducers (Tokyo Measuring Instruments Laboratory, PI-5-100) are installed on the concrete surface to control the crack

mouth opening displacement (CMOD).   When the CMOD ranges from 1 mm to 1.2 mm, the

maximum crack widths are approximately 0.3 mm after stopping the load.   Thus, all the

specimens can be controlled to have a similar width with the loading rate of 0.2 mm/min for a

pair comparison.   Note that all the specimens are in the water to promote self-healing process

nearby crack (see Fig. 5.16b).
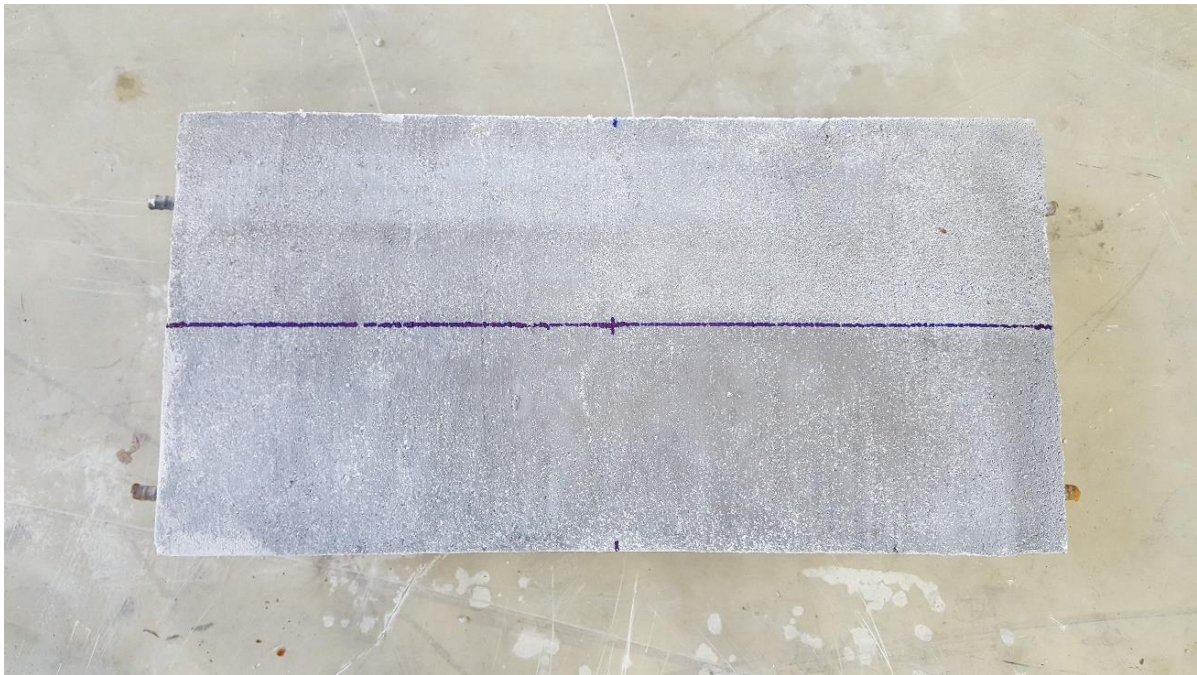
Fig. 5.15 Test specimen for monitoring self-healing performances.
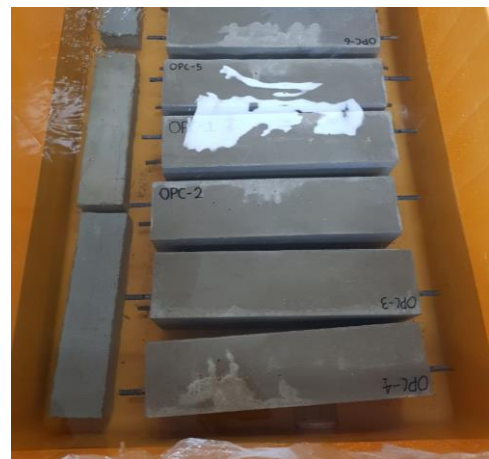
| (a) | (b) |

Fig. 5.16 Test procedures: (a) three-point bending test and (b) self-healing in the water.

The proposed camera slider-based system is used to capture the sequential crack images for crack evaluation. The digital camera, which is mounted on the automatic slider, provides a set of close-up crack images with a high-resolution, as shown in Fig. 5.17. The working distance between the camera and the concrete surface in the experiment is less than 100 mm, which results in around 0.01 mm/pixel. Thus, the crack widths less than 0.3 mm and the corresponding healing can be estimated reasonably. In this experiment, the concrete specimens are assessed to quantify crack opening areas on 7, 14, 28, and 63 days after the self-healing process.
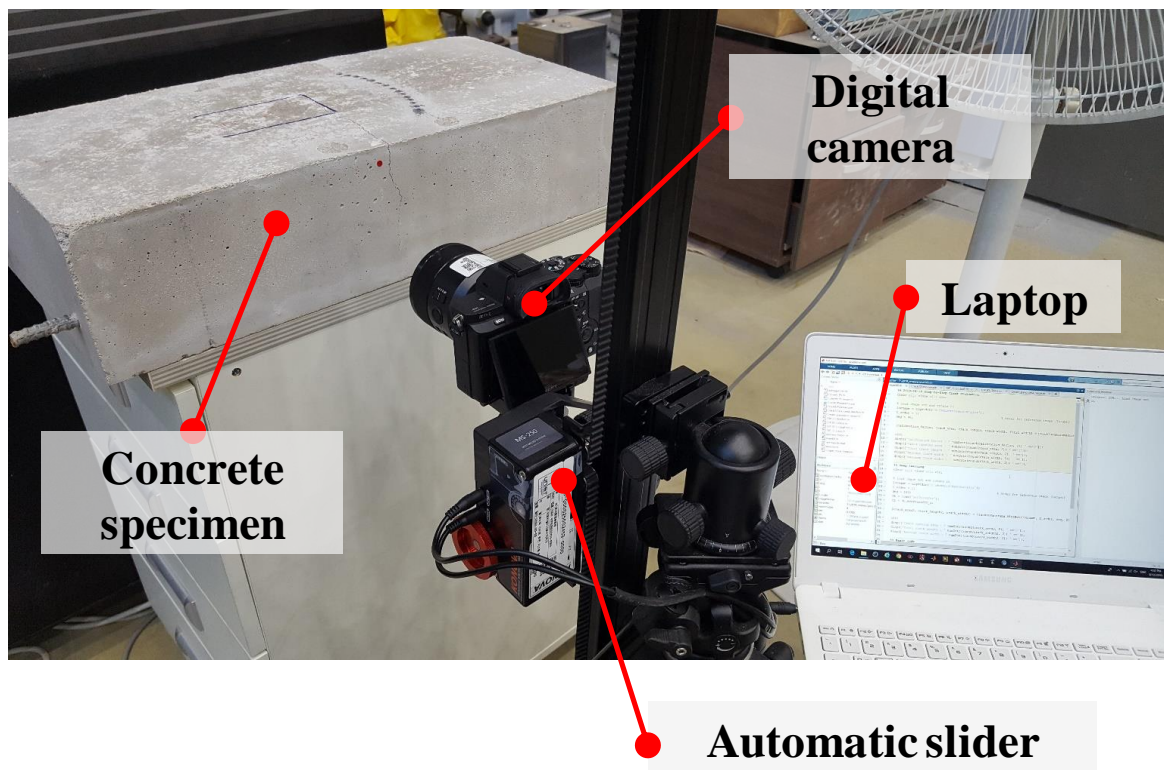
Fig. 5.17 Experimental setup.

The self-healing performances are investigated from the sequential crack images taken

from the proposed camera system. A color target, which has a diameter of 5 mm, is attached to the top of the concrete surface for calculating the pixel resolution from the first of the sequential crack images (see Fig. 5.18). The proposed frameworks for crack detection and quantification in the previous chapters are subsequently employed for identifying the locations of crack pixels in each image, as shown in Fig. 5.19. The crack information in the non-overlapping area is calculated from each image, as indicated by the green box, in which the total crack opening area is finally obtained. While the direct use of a full panorama image results in memory problems because of a combination of high-resolution images, the proposed approach can efficiently perform efficient crack identification from each image. Thus, high-resolution crack images can be directly processed for crack evaluation without resizing.

Fig. 5.20 shows the self-healing performances of the three cases of concrete (i.e., OPC, CS, and SCM) in terms of the obtained crack information. Here, the maximum crack width and crack opening area are selected to evaluate the self-healing performance. As shown in Fig. 5.20b, d, and f, most of the maximum crack widths are set to around 0.3 mm at the beginning. Fig. 5.20a, c, and e, which are the crack opening area, shows a similar trend with the maximum crack width. When the number of days for the self-healing in the water is increased, the crack opening areas are decreased, particularly 7 days after the self-healing process. Note that the self-healing performances are slightly different from each other even with the same mixture, because concrete is a heterogeneous material. The average of the crack opening areas is calculated in each case (see Table 5.5), from which the self-healing ratio is computed based on the reduction of crack opening areas, as shown in Table 5.6. The crack areas are significantly decreased on 7 days. Afterwards, the self-healing process is still observed on 56 days. Among them, the self-healing performance of CS outperforms the other mixtures, resulting an average of around 80 % of the crack opening areas is filled.

Original image   Binary image
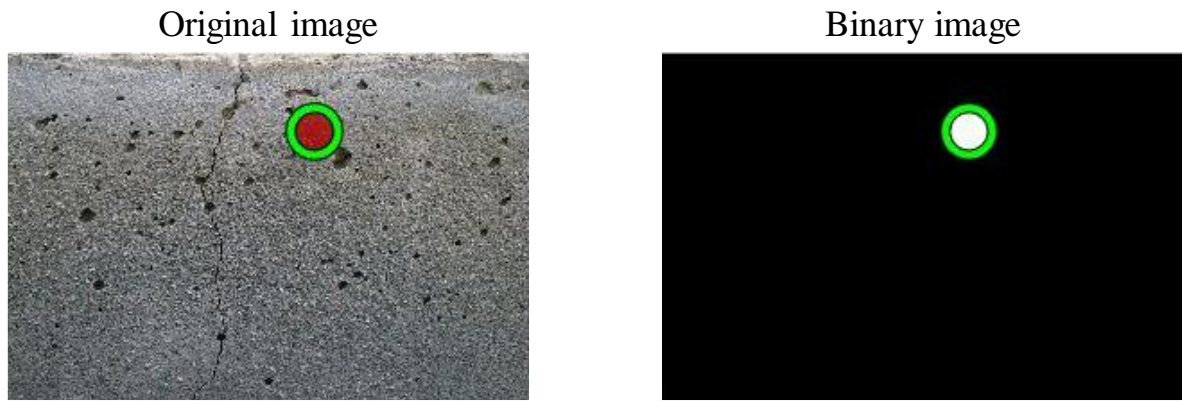
Fig. 5.18 Calculation of pixel resolution.

Sequential crack images
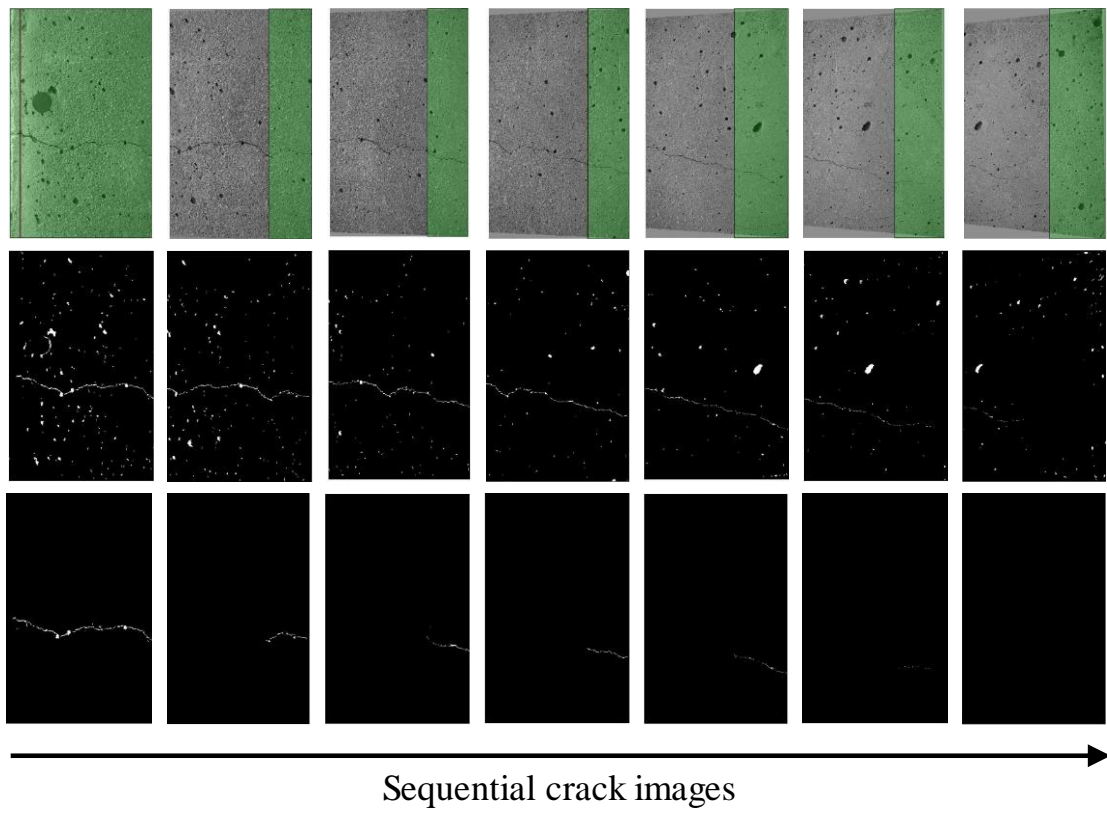
Fig. 5.19 Distributed crack identification strategy for crack measurements.

Table 5.5 Change of average crack opening areas.

Self-healing concrete	Average crack opening area				
	0 day	7 day	14 day	28 day	56 day
OPC	15.4820	7.9897	5.3548	4.5458	3.8168
CS	12.3084	6.3576	5.3622	3.3387	2.5159
SCM	11.8081	7.2927	5.2726	4.5526	3.1789

Table 5.6 Self-healing ratio based on reduction of average crack opening areas.

Self-healing concrete	Average self-healing ratio				
	0 day	7 day	14 day	28 day	56 day
OPC	-	0.4839	0.6541	0.7064	0.7535
CS	-	0.4835	0.5643	0.7287	0.7956
SCM	-	0.3824	0.5535	0.6145	0.7308

(a)   (b)
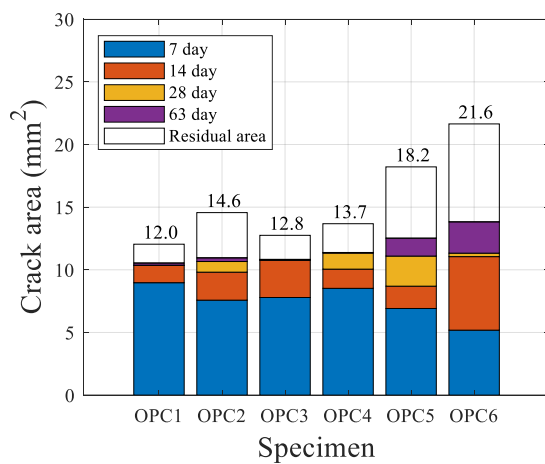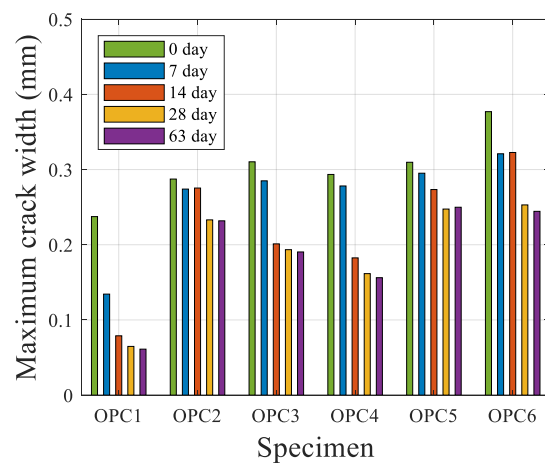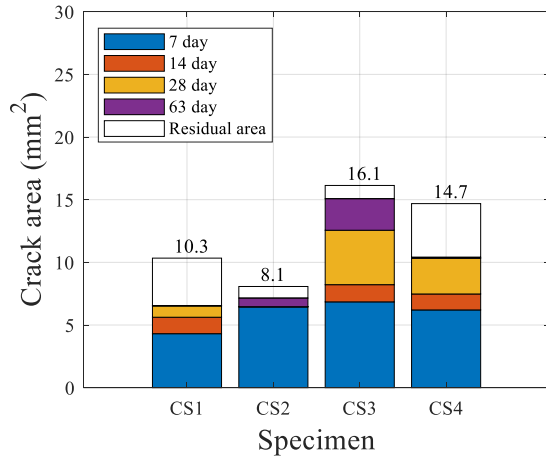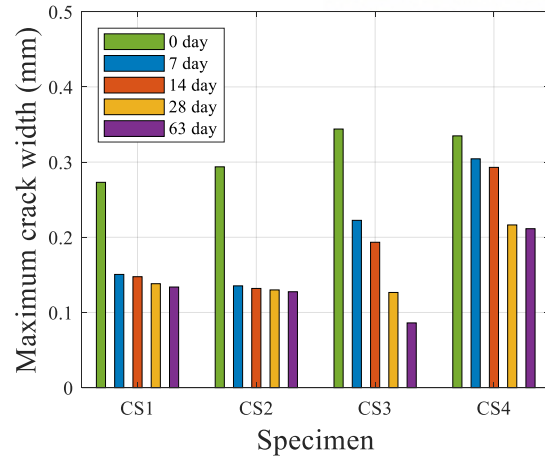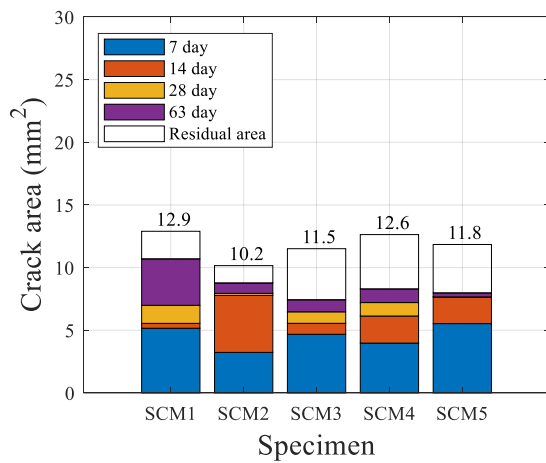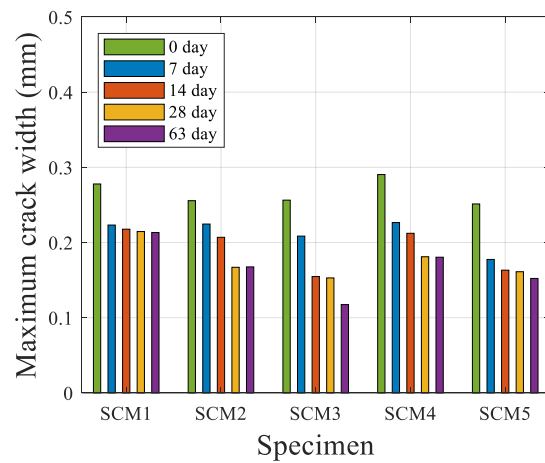
Fig. 5.20 Self-healing performances in terms of the obtained crack information: (a) maximum crack width in the case of OPC, (b) crack opening area in the case of OPC, (c) maximum crack width in the cases of CS, (d) crack opening area in the cases of CS, (e) maximum crack width in the case of SCM, and (f) crack opening area in the case of SCM.

The self-healing performances in each concrete are observed from the original crack images.   Fig. 5.21, 22, and 23 show the positive samples in OPC, CS, and SCM, respectively, in which the self-healing process can be clearly tracked over time.   In contrast, Fig. 5.24, 25, and 26 represent the negative samples in OPC, CS, and SCM, respectively.   From the results, while the crack widths less than 0.2 mm are effectively healed in the most cases, the other shows only a partial success.   Note that hydration products (e.g., calcium silicate hydrate and calcium hydroxide), resulting from the reaction between cement and water, are the main factors for the self-healing process.

(a) 0 day   (b) 7 day   (c) 14 day   (d) 28 day   (e) 63 day
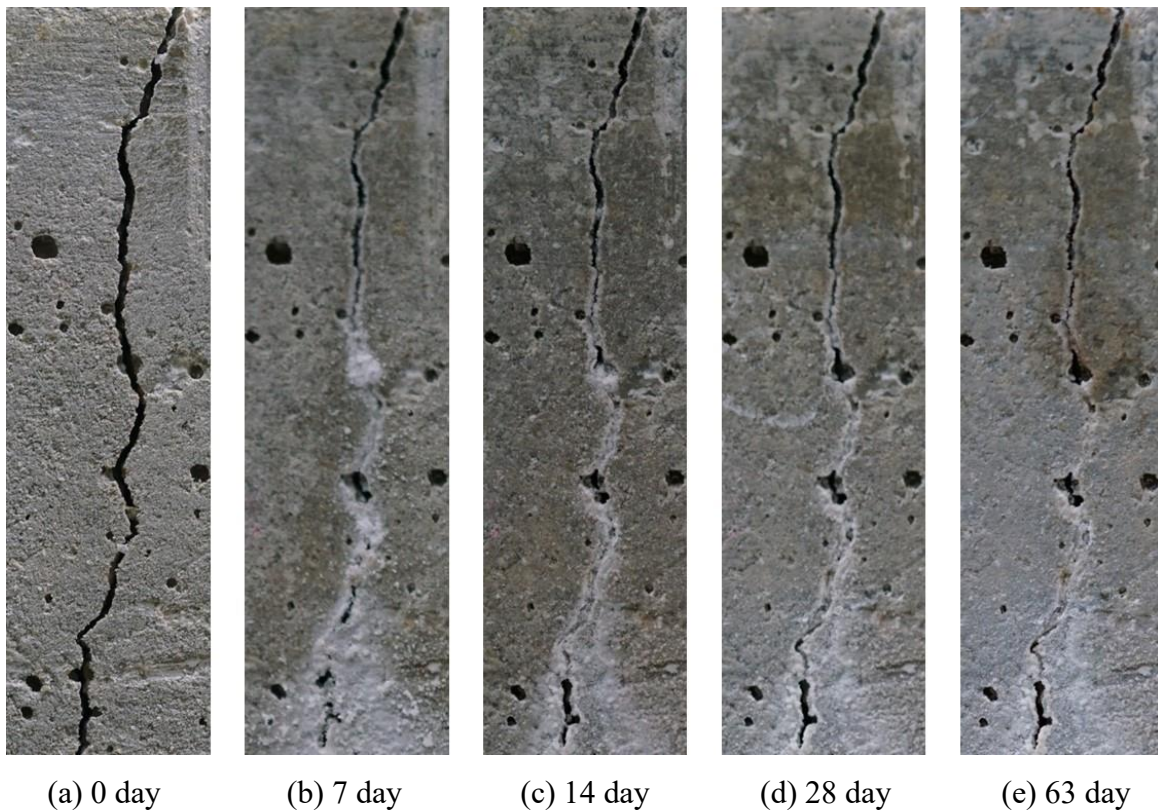
Fig. 5.21 Positive sample for OPC.
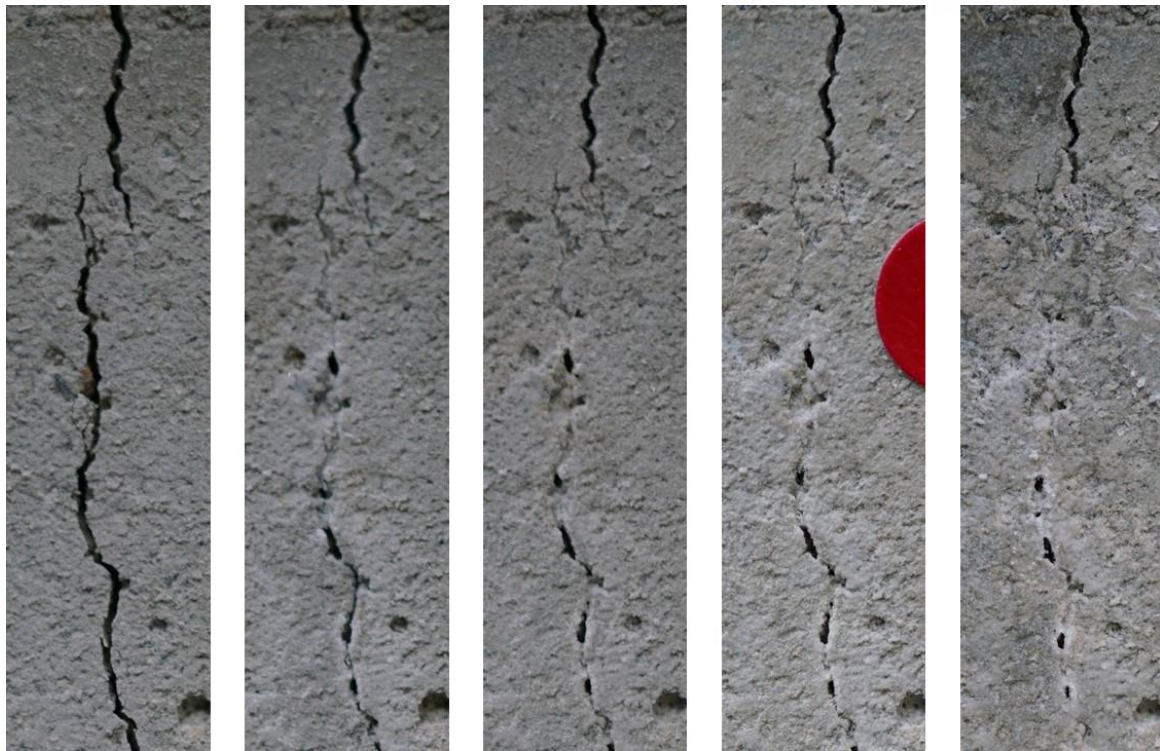
(a) 0 day	(b) 7 day	(c) 14 day	(d) 28 day	(e) 63 day

Fig. 5.22 Positive sample for CS.
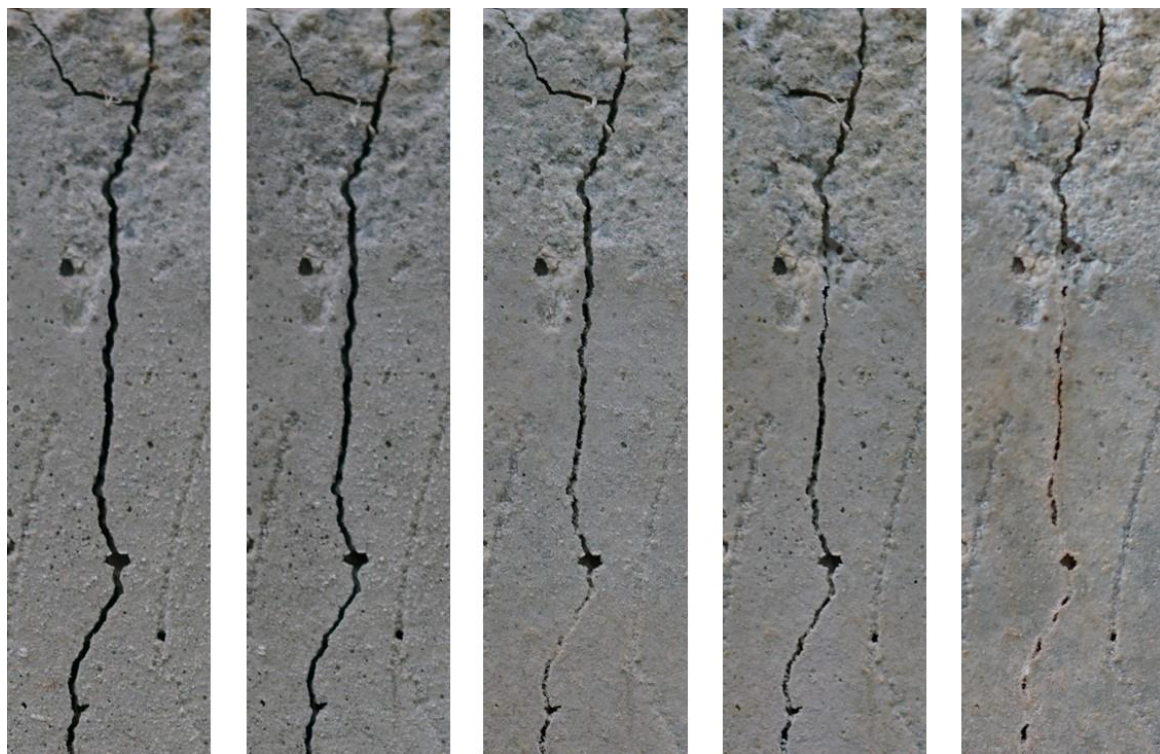
(a) 0 day	(b) 7 day	(c) 14 day	(d) 28 day	(e) 63 day

Fig. 5.23 Positive sample for SCM.

| (a) 0 day | (b) 7 day | (c) 14 day | (d) 28 day | (e) 63 day |
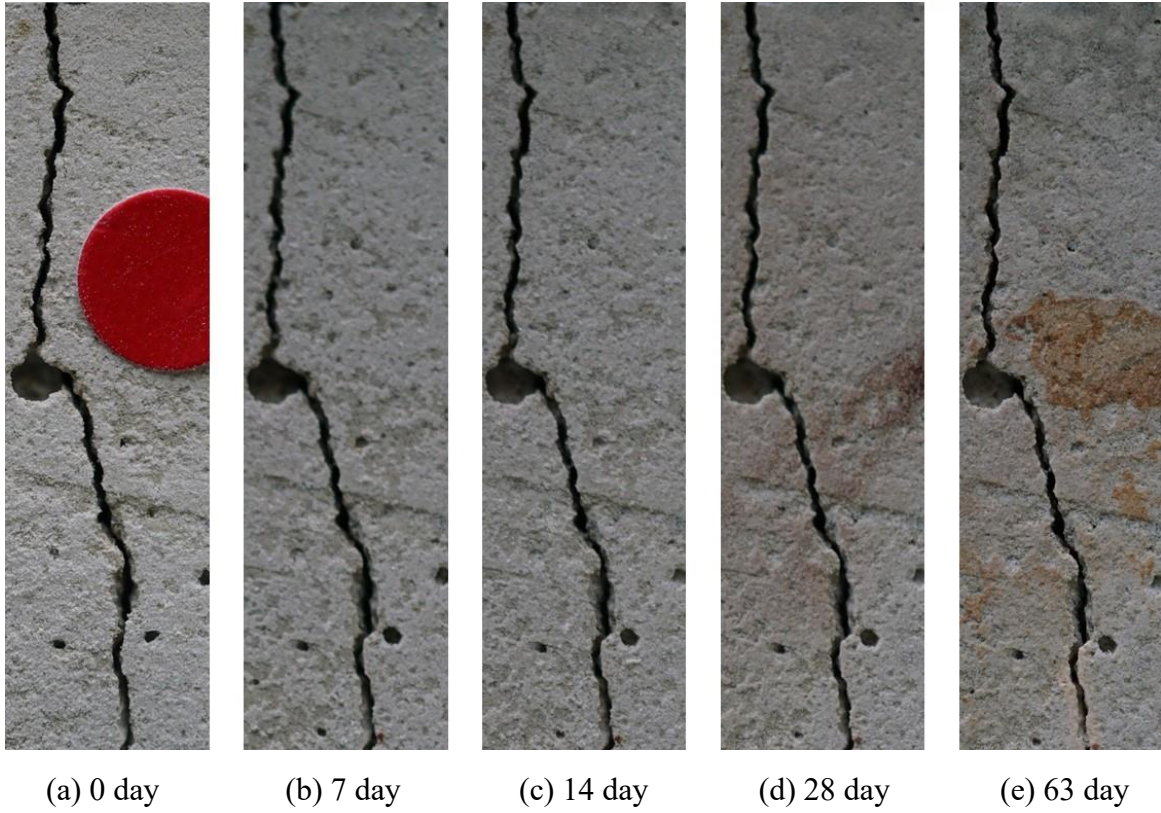
Fig. 5.24 Negative sample for OPC.

| (a) 0 day | (b) 7 day | (c) 14 day | (d) 28 day | (e) 63 day |
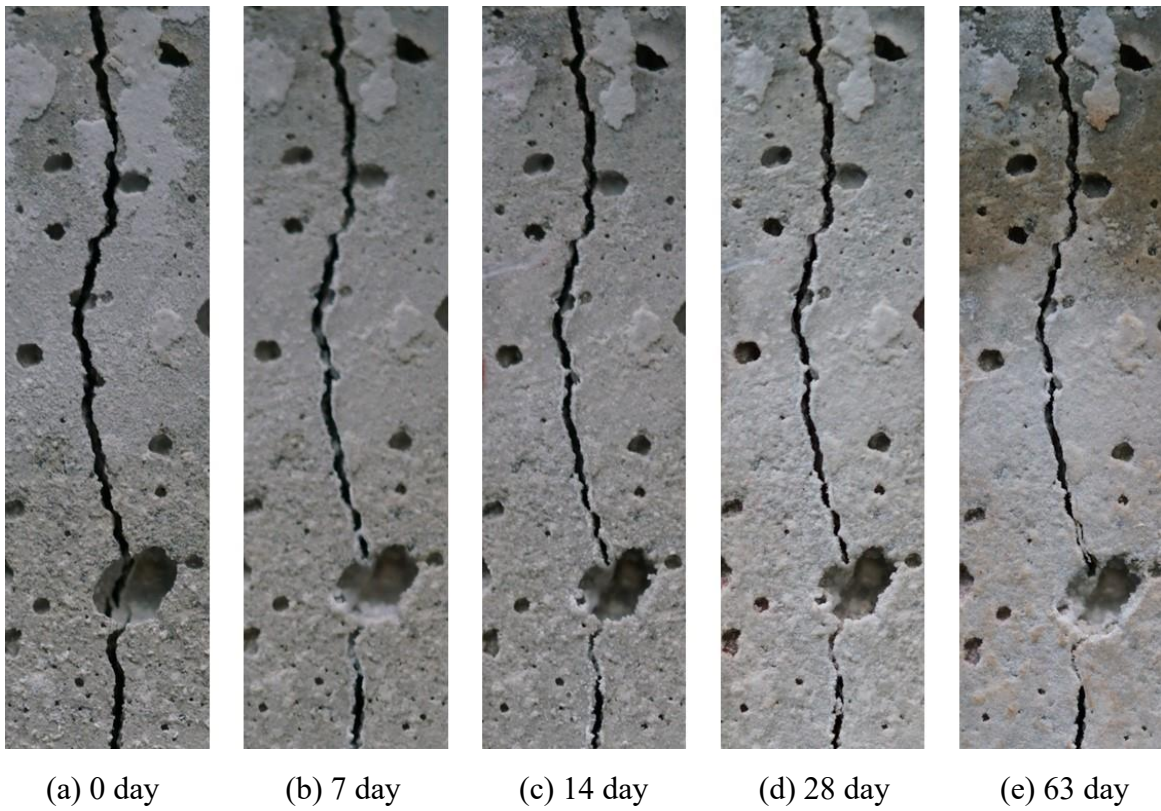
Fig. 5.25 Negative sample for CS.

| (a) 0 day | (b) 7 day | (c) 14 day | (d) 28 day | (e) 63 day |
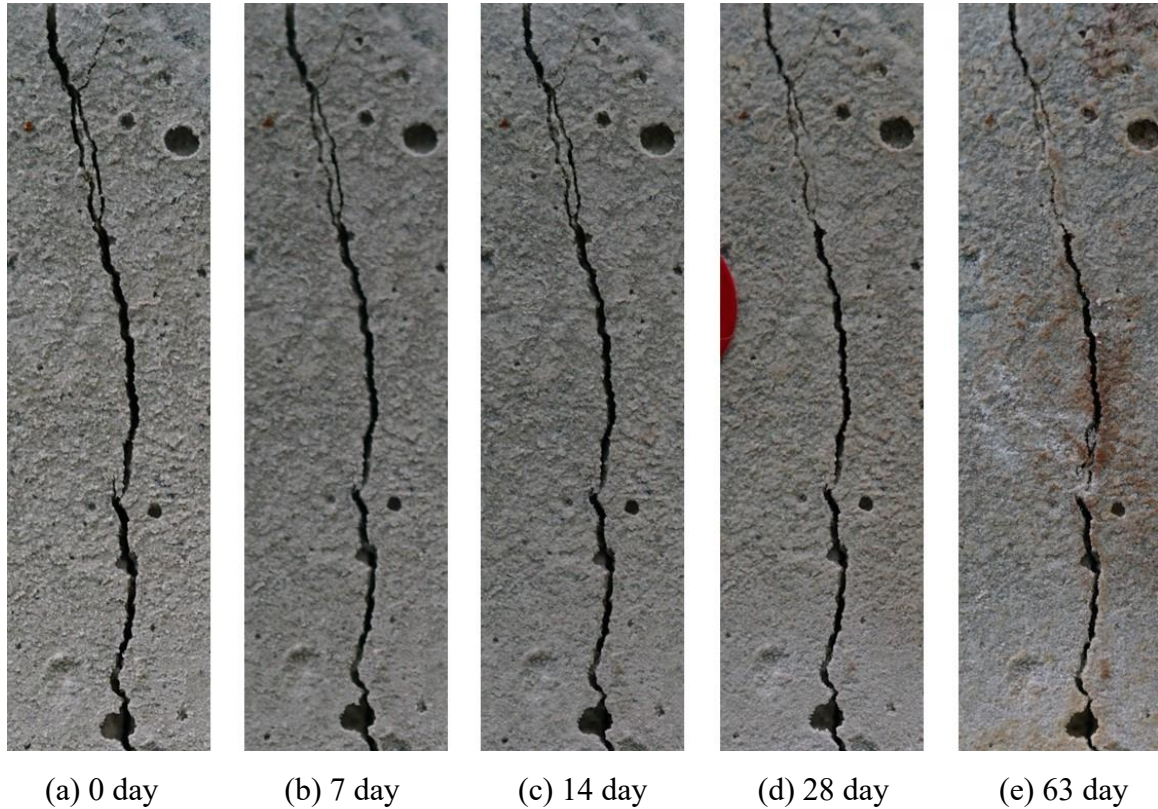
Fig. 5.26 Negative sample for SCM.

5.4 Summary

This research presented a couple of camera systems to properly handle the practical issues on civil infrastructure for accurate crack evaluation. Firstly, the RGB-D camera-based system was designed to provide technical improvements for accurate crack evaluation: a robust depth estimation considering the skewed angle problem and a high measurement resolution when images are captured from a distance. In the proposed approach, the RGB and depth information obtained using an RGB-D camera were utilized in RANSAC to construct a plane model corresponding to the concrete surface with minimal noise in the depth data. Subsequently, the crack image captured using the high-resolution digital camera with a relatively long focal length lens was processed by the proposed crack detection and quantification in the previous chapters. Based on a coordinate transformation between the digital and RGB-D cameras, the extracted crack pixels were transferred onto an approximate

plane to estimate the depth value in each pixel.   The 3D information was further mapped onto the camera coordinate of the RGB-D camera, in which the edge pixels nearest to the skeleton pixel were used to calculate the crack width.   To evaluate the performance of the proposed approach, a field test was conducted on a concrete wall of a gymnasium building considering the skewness of the camera with respect to the concrete surface.   The experimental result was compared with those obtained using an optical microscope and a conventional approach with only the high-resolution image.   In the experimental validation, the proposed approach accurately measured the crack widths regardless of the angle of view (i.e., 0, 20, 40, and 60°) from a distance of 2.5 m.   The average difference was approximately 0.05 mm for the cracks in all the cases, ranging from 0.84 to 1.1 mm.   In the conventional method, the crack widths were significantly underestimated at higher camera angles, because the skewness of the camera with respect to the concrete surface was not reflected.   With the proposed sensor fusion approach, cracks on concrete surfaces can be accurately measured regardless of the angle of view.

Secondly, the camera slider-based system was presented for efficient crack evaluation on the sequential crack images.   The proposed camera system consisted of the automatic slider and digital camera with a macro lens, providing a high-resolution set of close-up crack images from the concrete surface.   In the proposed approach, the registration information between the sequential crack images was computed based on the local features, from which the crack identification was efficiently conducted in each image instead of a full panorama image.   To evaluate the performance of the proposed approach, the monitoring of self-healing concrete was performed.   Here, the self-healing performances in three types of concrete (i.e., OPC, CS, and SCM) were quantified based on the reduction of crack opening areas investigated by the proposed approach.   In the experimental results, the crack opening areas were effectively decreased by the self-healing process in the most cases, particularly 7 days.   Among them, the

use of CS outperformed the other mixtures in terms of the self-healing ratio calculated by the reduction of average crack opening areas.   With the proposed distributed crack identification strategy, accurate crack evaluation on concrete surfaces can be efficiently performed.

CHAPTER 6   CONCLUSIONS AND FUTURE STUDIES

6.1   Conclusions

The proposed research focused on accurate crack evaluation for civil infrastructure using computer vision and deep learning.   Digital image processing was considered as a powerful alternative to manual visual inspection, particularly for measuring surface cracks.   Among the diverse crack identification methods, image binarization was seen to have a strong potential to effectively extract crack pixels on digital images.   However, the crack assessment inevitably depended on the predetermined binarization technique and the associated parameters.   In addition, although previous studies on the use of computer vision and deep learning had shown enormous potential for automated crack detection, crack-like noncracks on realistic surface images critically limits its full automation.   Furthermore, a tailored camera system and the corresponding strategy were necessary to properly address a couple of the practical issues in real-world applications, such as the skewed angle problem and the absence of efficient crack measurement for the sequential images.   In this study, the following approaches were presented for robust and accurate crack evaluation for the practical maintenance of civil engineering structures: (1) a deep learning-based crack detection with the concept of CCRs, (2) a hybrid image processing for crack quantification, (3) RGB-D camera-based and camera slider-based systems for civil infrastructure.

Firstly, a deep learning-based framework was presented for automated crack detection on realistic surface images that may contain actual cracks and crack-like noncracks.   The crack candidates, which can be a crack or a crack-like noncrack, were initially determined by the binary information and further utilized for both the training and testing stages.   Here, the concept of CCRs implemented on the deep learning network, optimizing the utilization of cracks and crack-like noncracks.   The obtained classification model was applied to new

images in which most of actual cracks and crack-like noncracks were successfully categorized.

Secondly, a hybrid image processing was developed for accurate crack quantification on surface images.   The UAV-based prototype designed in this study was capable of image and distance sensing as well as wireless communication, which makes it possible to control sensing and data transmission while the UAV was in the air.   The obtained information was subsequently processed by the proposed hybrid image processing to identify crack width accurately while minimizing the loss of crack length.   Here, both the crack width and length were accurately measured by a combination of two sets of optimal binarization parameters

Lastly, two camera systems and the associated strategies were proposed for accurate crack evaluation on civil infrastructure.   The proposed RGB-D camera-based system was able to accurately measure crack information regardless of the angle of view.   In this approach, the RGB-D camera generated a point cloud to construct an approximate plane corresponding to the concrete surface with minimal noise in the depth data.   The obtained high-resolution image taken by a digital camera was processed to localize the locations of the crack pixels; subsequently, the corresponding 3D information on the predetermined plane was utilized to calculate the crack widths.   Furthermore, the proposed camera slider-based system in conjunction with the distributed crack identification strategy were able to perform efficient crack measurements.   The proposed camera system consisted of the automatic slider and digital camera with a macro lens, providing a high-resolution set of close-up crack images from the concrete surface.   The sequential crack images were further utilized to calculate the registration information based on the local features, from which crack measurements were efficiently performed in each image instead of a full panorama image.

The proposed research, combining computer vision and deep learning, would allow accurate crack evaluation of civil infrastructure, providing a proper maintenance strategy for structural soundness in practice.

6.2   Future Studies

In the future work, two kinds of approaches will be presented: (1) an improved technique of the RGB-D camera-based system for accurate crack evaluation on curved surface and (2) an integrated system implemented on the UAV that also combines the RGB-D camera-based sensor fusion and the distributed crack identification strategy.   Although the proposed RGB-D camera-based system can accurately measure crack information regardless of the camera angle with respect to the target surface, crack evaluation on curved surface is intrinsically difficult owing to the approximation of the concrete surface to a plane.   Because concrete cracks can be generated on curved surface of civil infrastructure, this issue needs to be properly addressed.   Thus, an improved technique that can estimate various shapes of the target surface will be developed to perform accurate crack evaluation even with curved surface of the structure.   Furthermore, an integrated system will be designed based on the proposed research. Although the proposed crack detection and quantification are successfully applied to each of the RGB-D camera-based and camera slider-based systems for accurate crack evaluation, an integrated system has not been presented.   In other words, a robust depth estimation of the RGB-D camera system and an efficient distributed strategy of the camera slider-based system can promote crack evaluation on civil infrastructure.   Particularly, a prototype of the UAV will be designed with containing four essential components: a single-board computer, a high-resolution camera, an RGB-D camera, and a WiFi module, providing the sequential crack images and the corresponding depth information in proximity to surface of civil engineering structures.   The obtained information will be further utilized to perform accurate crack evaluation using the distributed crack identification strategy, regardless of the camera angle as well as the shape of structural surface.   The completion of the proposed approach would provide technical improvements for accurate crack evaluation of civil infrastructure, supporting a systematic maintenance.

REFERENCES

1. S.P. Karthick, S. Muralidharan, V. Saraswathy, S.J. Kwon. Effect of different alkali salt additions on concrete durability property. *J Struct Integr Maint* 2016; 1(1): 35–42.

2. R.J. Frosch. Another look at cracking and crack control in reinforced concrete. *ACI Struct J* 1999; 96(3): 437–442.

3. S.J. Kwon, U.J. Na, S.S. Park, S.H. Jung. Service life prediction of concrete wharves with early-aged crack: Probabilistic approach for chloride diffusion. *Struct Saf* 2009; 31(1): 75–83.

4. S. Liu, Z.B. Bundur, J. Zhu, R.D. Ferron. Evaluation of self-healing of internal cracks in biomimetic mortar using coda wave interferometry. *Cem Concr Res* 2016; 83: 70–78.

5. D.G. Aggelis, T. Shiotani, D. Polyzos. Characterization of surface cracks depth and repair evaluation using Rayleigh waves. *Cem Concr Compos* 2009; 31: 77–83.

6. S.W. Shin, J. Zhu, J. Min, J.S. Popovics. Crack depth estimation in concrete using energy transmission of surface waves. *ACI Mater J* 2008; 105(5): 510–516.

7. H. Kim, X. Liu, E. Ahn, M. Shin, S.W. Shin, S.-H. Sim. Performance assessment method for crack repair in concrete using PZT-based electromechanical impedance technique. *NDT&E Int* 2019; 104: 90–97.

8. A.R. Vosoughi. A developed hybrid method for crack identification beams. *Smart Struct Syst* 2015; 16(3): 401–414.

9. S. Moradi, P.J. Moghadam. Crack identification in post-buckled beam-type structures. *Smart Struct Syst* 2015; 15(5): 1233–1252.

10. S.H. Man, C.C. Chang, M. Hussan, A. Bermak. Design and calibration of a wireless laser-based optical sensor for crack propagation monitoring. *Smart Struct Syst* 2015; 15(6): 1543–1567.

11. J.D. Helm. Digital image correlation for specimens with multiple growing cracks. *Exp Mech* 2008; 48(6): 753–762.

12. J.D. Carroll, W. Abuzaid, J. Lambros, H. Sehitoglu. High resolution digital image correlation measurements of strain accumulation in fatigue crack growth. *Int J Fatigue* 2013; 57 140–150.

13. P. Dare, H. Hanley, C. Fraser, B. Riedel, W. Niemeier. An operational application of automatic feature extraction: The measurement of cracks in concrete structures. *Photogramm Rec* 2002; 17(99): 453–464.

14. T.C. Hutchinson, Z. Chen. Improved image analysis for evaluating concrete damage. *J Comput Civil Eng* 2006; 20(3): 210–216.

15. I. Abdel-Qader, O. Abudayyeh, M.E. Kelly. Analysis of edge-detection techniques for crack identification in bridges. *J Comput Civil Eng* 2003; 17(4): 255–263.

16. J. Bernsen. Dynamic thresholding of grey-level images. *In: Proceedings of International Conference on Pattern Recognition*, 1986; 1251–1255.

17. W. Niblack. An introduction to digital image processing. Strandberg Publishing Company, Birkeroed, 1986.

18. J. Sauvola, M. Pietikainen. Adaptive document image binarization. *Pattern Recognit* 2000; 33(2): 225–236.

19. C. Wolf, J.M. Jolion. Extraction and recognition of artificial text in multimedia documents. *Pattern Anal Appl* 2004; 6(4): 309–326.

20. K. Khurshid, I. Siddiqi, C. Faure, N. Vincent. Comparison of Niblack inspired Binarization methods for ancient documents. *In: Proceedings of SPIE-IS&T Electronic Imaging*, 2009; 72470U–72470U-9.

21. Y.F. Liu, S. Cho, B.F. Spencer Jr, J.S. Fan. Automated assessment of cracks on concrete surfaces using adaptive digital image processing. *Smart Struct Syst* 2014; 14(4): 719–741.

22. Y.F. Liu, S. Cho, B.F. Spencer Jr, J.S. Fan. Concrete crack assessment using digital image processing and 3D scene reconstruction. *J Comput Civil Eng* 2014; 30(1): 04014124-1–04014124-19.

23. N. Tanaka, K. Uematsu. A crack detection method in road surface images using morphology. *In: Proceedings of Machine Vision Applications*, 1998; 154–157.

24. S. Lyer, S.K. Sinha. Segmentation of pipe images for crack detection in buried sewers. *Comput-Aided Civil Infrastruct Eng* 2006; 21(6): 395–410.

25. S.K. Sinha, P.W. Fieguth. Segmentation of buried concrete pipe images. *Autom Constr* 2006; 15(1): 47–57.

26. M.R. Jahanshahi, J.S. Kelly, S.F. Masri, G.S. Sukhatme. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Struct Infrastruct Eng* 2009; 5(6): 455–486.

27. Y.-J. Cha, W. Choi, O. Büyüköztürk. Deep learning-based crack damage detection using convolutional neural networks. *Comput-Aided Civil Infrastruct Eng* 2017; 32(5): 361–378.

28. K. Gopalakrishnan, S.K. Khaitan, A. Choudhary, A. Agrawal. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr Build Mater* 2017; 157: 322–330.

29. Z. Tong, J. Gao, Z. Han, Z. Wang. Recognition of asphalt pavement crack length using deep convolutional neural networks. *Road Mater Pavement Des* 2017; 13: 1–16.

30. A. Zhang, K.C.P. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J.Q. Li, C. Chen. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. *Comput-Aided Civil Infrastruct Eng* 2017; 32(10): 805–819.

31. Y. LeCun, B. Boser, J.S. Denker, D. Henderson. Backpropagation applied to handwritten zip code recognition. *Neural comput* 1989; 1(4): 541–551.

32. R.S. Adhikari, O. Moselhi, A. Bagchi. A study of image-based element condition index for bridge inspection. *In Proceedings of International Symposium on Automation and Robotics in Construction*, 2013; 1–12

33. S.G. Paal, J.-S. Jeon, I. Brilakis, R. DesRoches. Automated damage index estimation of reinforced concrete columns for post-earthquake evaluations. *J Struct Eng* 2014; 141(9): 04014228.

34. C.M. Yeum, S.J. Dyke. Vision-based automated crack detection for bridge inspection. *Comput-Aided Civil Infrastruct Eng* 2015; 30(10): 759–770.

35. M.R. Jahanshahi, S.F. Masri. Parametric performance evaluation of wavelet-based corrosion detection algorithms for condition assessment of civil infrastructure systems. *J Comput Civil Eng* 2012; 27(4): 345–357.

36. H.-K. Shen, P.-H. Chen, L.-M. Chang. Automated steel bridge coating rust defect recognition method based on color and texture feature. *Autom Constr* 2013; 31: 338–356.

37. C. Koch, I. Brilakis. Pothole detection in asphalt pavement images. *Adv Eng Inform* 2011; 25(3): 507–515.

38. Q. Zou, Y. Cao, Q. Li, Q. Mao, S. Wang. CrackTree: Automatic crack detection from pavement images. *Pattern Recognit Lett* 2012; 33(3): 227–238.

39. Y. Gao, K.M. Mosalam. Deep transfer learning for image-based structural damage recognition. *Comput-Aided Civil Infrastruct Eng* 2018; 33(9): 748–768.

40. X. Liang. Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization. *Comput-Aided Civil Infrastruct Eng* 2019; 34(5): 415–430.

41. Y. Narazaki, V. Hoskere, T.A. Hoang, Y. Fujino, A. Sakurai, B.F. Spencer Jr. Vision-based automated bridge component recognition with high-level scene consistency. *Comput-Aided Civil Infrastruct Eng* 2019. (In press)

42. N.A. Hoult, W.A. Take, C. Lee, M. Dutton. Experimental accuracy of two dimensional strain measurements using digital image correlation. *Eng Struct* 2013; 46: 718–726.

43. M. Dutton, W.A. Take, N.A. Hoult. Curvature monitoring of beams using digital image correlation. *J Bridge Eng* 2013; 19(3): 05013001.

44. R. Zaurin, T. Khuc, F.N. Catbas. Hybrid sensor-camera monitoring for damage detection: case study of a real bridge. *J Bridge Eng* 2016; 21(6): 05016002.

45. Y. Xu, J. Brownjohn, D. Kong. A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Struct Control Health Monit* 2018; 25(5): e2155.

46. S. Jang, H. Jo, S. Cho, K. Mechitov, J.A. Rice, S.-H. Sim, H.-J. Jung, C.-B. Yun, B.F. Spencer Jr, G. Agha. Structural health monitoring of a cable-stayed bridge using smart sensor technology: deployment and evaluation. *Smart Struct Syst* 2010; 6(5-6): 439–459.

47. J.A. Rice, K. Mechitov, S.-H. Sim, T. Nagayama, S. Jang, R. Kim, B.F. Spencer Jr, G. Agha, Y. Fujino. Flexible smart sensor framework for autonomous structural health monitoring. *Smart Struct Syst* 2010; 6(5-6): 423–438.

48. B.F. Spencer Jr, V. Hoskere, Y. Narazaki. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering* 2019; 5(2): 199–222.

49. L. Breiman. Random forests. *Mach Learn* 2001; 45(1): 5–32.

50. C. Cortes, V. Vapnik. Support-vector networks. *Mach Learn* 1995; 20(3): 273–297.

51. G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray. Visual categorization with bags of keypoints. *In: Proceedings of European Conference on Computer Vision*, 2004; 1–16.

52. W. Zhang, Z. Zhang, D. Qi, Y. Liu. Automatic crack detection and classification method for subway tunnel safety monitoring. *Sensors* 2014; 14(10): 19307–19328.

53. P. Prasanna, K.J. Dana, N. Gucunski, B.B. Basily, H.M. La, R.S. Lim, H. Parvardeh. Automated crack detection on concrete bridges. *IEEE Trans Autom Sci Eng* 2016; 13(2): 591–599.

54. Y. Shi, L. Cui, Z. Qi, F. Meng, Z. Chen. Automatic road crack detection using random structured forests. *IEEE Trans Intell Transp Syst* 2016; 17(12): 3434–3445.

55. G. Li, X. Zhao, K. Du, F. Ru, Y. Zhang. Recognition and evaluation of bridge cracks with modified active contour model and greedy search-based support vector machine. *Autom Constr* 2017; 78: 51–61.

56. T. Lindeberg. Feature detection with automatic scale selection. *Int J Comput Vis* 1998; 30(2): 79–116.

57. D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 2004, 60(2): 91–110.

58. H. Bay, A. Ess, T. Tuytelaars, L.V. Gool. Speeded-up robust features (SURF). *Comput Vis Image Underst* 2008; 110(3): 346–359.

59. L. Juan, O. Gwun O. A comparison of sift, pca-sift and surf. *Int J Image Process* 2009; 3(4): 143–152.

60. O. Duda, P.E. Hart, D.G. Stork. Pattern classification, John Wiley & Sons, Ltd, 2000.

61. A. Krizhevsky, I. Sutskever, G.E. Hinton. ImageNet classification with deep convolutional neural networks. *In: Proceedings of Advances in Neural Information Processing Systems*, 2012; 1–9.

62. K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. *In: Proceedings of International Conference on Learning Representations*, 2015; 1–14.

63. M.D. Zeiler, R. Fergus. Visualizing and understanding convolutional networks. *In: Proceedings of European Conference on Computer Vision*, 2014; 818–833.

64. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. Going deeper with convolutions. *In: Proceedings of Computer Vision and Pattern Recognition*, 2015; 1–9.

65. K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. *In: Proceedings of Computer Vision and Pattern Recognition*, 2016; 770–778.

66. L. Zheng, Y. Yang, Q. Tian. SIFT meets CNN: A decade survey of instance retrieval. *IEEE Trans Pattern Anal Mach Intell* 2018; 40(5): 1224–1244.

67. B. Kim, S. Cho. Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. *Sensors* 2018; 18(10): 3452.

68. S. Dorafshan, R.J. Thomas, M. Maguire. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Constr Build Mater* 2018; 186: 1031–1045.

69. Y. Xu, Y. Bao, J. Chen, W. Zuo, H. Li. Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. *Struct Health Monit* 2019; 18(3): 653–674.

70. F.C. Chen, M.R. Jahanshahi. NB-CNN: deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion. *IEEE Trans Ind Electron* 2017; 65(5): 4392–4400.

71. K. Jang, N. Kim, Y.-K. An. Deep learning–based autonomous concrete crack evaluation through hybrid image scanning. *Struct Health Monit* 2019; 18(5-6): 1722–1737.

72. R. Girshick, J. Donahue, T. Darrell, J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *In: Proceedings of Computer Vision and Pattern Recognition*, 2014; 580–587.

73. K. He, X. Zhang, S. Ren, J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Tran Pattern Anal Mach Intell* 2015; 37(9): 1904–1916.

74. R. Girshick. Fast R-CNN. *In: Proceedings of International Conference on Computer Vision*, 2015; 1440–1448.

75. S. Ren, K. He, R. Girshick, J. Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *In: Proceedings of Advances in Neural Information Processing Systems*, 2015; 1–9.

76. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg. SSD: Single shot multibox detector. *In: Proceedings of European Conference on Computer Vision*, 2016; 21–37.

77. J. Dai, Y. Li, K. He, J. Sun. R-FCN: Object detection via region-based fully convolutional networks. *In: Proceedings of Advances in Neural Information Processing Systems*, 2016; 379–387.

78. J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You only look once: Unified, real-time object detection. *In: Proceedings of Computer Vision and Pattern Recognition*, 2016; 779–788.

79. Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, O. Büyüköztürk. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Comput-Aided Civil Infrastruct Eng* 2018; 33(9): 731–747.

80. X. Xue, Y. Li. A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. *Comput-Aided Civil Infrastruct Eng* 2018; 33(8): 638–654.

81. H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata. Road damage detection and classification using deep neural networks with smartphone images. *Comput-Aided Civil Infrastruct Eng* 2018; 33(12): 1127–1141.

82. Y. Li, Z. Han, H. Xu, L. Liu, X. Li, K. Zhang. YOLOv3-Lite: A lightweight crack detection network for aircraft structure based on depthwise separable convolutions. *Appl Sci-Basel* 2019; 9(18): 3781.

83. J.C.P. Cheng, M. Wang. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Autom Constr* 2018; 95: 155–171.

84. X. Yin, Y. Chen, A. Bouferguene, H. Zaman, M. AI-Hussein, L. Kurach. A deep learning-based framework for an automated defect detection system for sewer pipes. *Autom Constr* 2020; 109: 102967.

85. S. Jiang, J. Zhange. Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system. *Comput-Aided Civil Infrastruct Eng* 2019. (In press)

86. J. Long, E. Shelhamer, T. Darrell. Fully convolutional networks for semantic segmentation. *In: Proceedings of Computer Vision and Pattern Recognition*, 2015; 3431–3440.

87. O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional networks for biomedical image segmentation. *In: Proceedings of Medical Image Computing and Computer-Assisted Intervention*, 2015; 234-241.

88. V. Badrinarayanan, A. Kendall, R. Cipolla. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Tran Pattern Anal Mach Intell* 2017; 39(12): 2481–2495.

89. H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia. Pyramid scene parsing network. *In: Proceedings of Computer Vision and Pattern Recognition*, 2017; 2881–2890.

90. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Tran Pattern Anal Mach Intell* 2017; 40(4): 834–848.

91. H.-W. Huang, Q.-T. Li, D.-M. Zhang. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunn Undergr Space Technol* 2018; 77: 166–176.

92. S. Li, X. Zhao, G. Zhou. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. *Comput-Aided Civil Infrastruct Eng* 2019; 34(7): 616–634.

93. X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, X. Yang. Automatic pixel-level crack detection and measurement using fully convolutional network. *Comput-Aided Civil Infrastruct Eng* 2018; 33(12): 1090–1109.

94. C.V. Dung, L.D. Anh. Autonomous concrete crack detection using deep fully convolutional neural network. *Autom Constr* 2019; 99: 52–58.

95. M.M. Islam, J.-M. Kim. Vision-based autonomous crack detection of concrete structures using a fully convolutional encoder–decoder network. *Sensors* 2019; 19(19): 4251.

96. Z. Liu, Y. Cao, Y. Wang, W. Wang. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Autom Constr* 2019; 104: 129–139.

97. Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, S. Wang. Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE Trans Image Process* 2018; 28(3): 1498–1512.

98. X. Zhang, D. Rajan, B. Story. Concrete crack detection using context-aware deep semantic segmentation network. *Comput-Aided Civil Infrastruct Eng* 2019; 34(11): 951–971.

99. B. Kim, S. Cho. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Struct Control Health Monit* 2019; 26(8): e2381.

100. J. Zhang, C. Lu, J. Wang, L. Wang, X.-G. Yue. Concrete cracks detection based on FCN with dilated convolution. *Appl Sci-Basel* 2019; 9(13): 2686.

101. S. Bang, S. Park, H. Kim, H. Kim. Encoder–decoder network for pixel-level road crack detection in black-box images. *Comput-Aided Civil Infrastruct Eng*. 2019; 34(8): 713–727.

102. Q. Mei, M. Gül, M.R. Azim. Densely connected deep neural network considering connectivity of pixels for automatic crack detection. *Autom Constr* 2020; 110: 103018.

103. Y. Ren, J. Huang, Z. Hong, W. Lu, J. Yin, L. Zou, X. Shen. Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Constr Build Mater* 2020; 234: 117367.

104. H. Kim, E. Ahn, M. Shin, S.-H. Sim. Crack and noncrack classification from concrete surface images using machine learning. *Struct Health Monit* 2019; 18(3): 725–738.

105. H. Kim, E. Ahn, S. Cho, M. Shin, S.-H. Sim. Comparative analysis of image binarization methods for crack identification in concrete structures. *Cem Concr Res* 2017; 99: 53–61.

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

107. J.Y. Bouguet, Camera calibration toolbox for Matlab. Available online: http://www.vision.caltech.edu/bouguetj/calib_doc/ (accessed on 11 December 2019).

108. I. Colomina, P. Molina. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS-J. Photogramm Remote Sens* 2014; 92: 79–97.

109. F. Nex, F. Remondino. UAV for 3D mapping applications: a review. *Appl Geomat* 2014; 6: 1–15.

110. I.S. Campos, E.R. Nascimento, G.M. Freitas, L. Chaimowicz. A height estimation approach for terrain following flights from monocular vision. *Sensors* 2016; 16: 2071.

111. S. Chen, D.F. Laefer, E. Mangina. State of technology review of civilian UAVs. *Recent Pat Eng* 2016; 10: 160–174.

112. L.F. Gonzalez, G.A. Montes, E. Puig, S. Johnson, K. Mengersen, K.J. Gaston. Unmanned Aerial Vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors* 2016; 16: 97.

113. C. Sampedro, H. Bavle, J.L. Sanchez-Lopez, R.A.S. Fernández, A. Rodríguez-Ramos, M. Molina, P. Campoy. A flexible and dynamic mission planning architecture for UAV swarm coordination. *In: Proceedings of International Conference on Unmanned Aircraft Systems*, 2016; 355–363.

114. A.R. Vetrella, G. Fasano, D. Accardo, A. Moccia. Differential GNSS and vision-based tracking to improve navigation performance in cooperative multi-UAV systems. *Sensors* 2016; 16: 2164.

115. L.I.N. Zongjian. UAV for mapping—low altitude photogrammetric survey. *Int Arch Photogramm Remote Sens Spatial Inform Sci* 2008; 37: 1183–1186.

116. F. Remondino, L. Barazzetti, F. Nex, M. Scaioni, D. Sarazzi. UAV photogrammetry for mapping and 3d modeling–current status and future perspectives. *Int Arch Photogramm Remote Sens Spatial Inform Sci* 2011; 38: 25–31.

117. S. Siebert, J. Teizer. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Autom Constr* 2014; 41: 1–14.

118. J. Byrne, E. O'Keeffe, D. Lennon, D.F. Laefer. 3D Reconstructions using unstabilized video footage from an unmanned aerial vehicle. *J Imaging* 2017; 3: 15.

119. S.E. Chen, C. Rice, C. Boyle, E. Hauser. Small-format aerial photography for highway-bridge monitoring. *J Perform Constr Facil* 2011; 25: 105–112.

120. C. Zhang, A. Elaksher. An unmanned aerial vehicle-based imaging system for 3D measurement of unpaved road surface distresses1. *Comput-Aided Civil Infrastruct Eng* 2012; 27: 118–129.

121. L. Díaz-Vilariño, H. González-Jorge, J. Martínez-Sánchez, M. Bueno, P. Arias. Determining the limits of unmanned aerial photogrammetry for the evaluation of road runoff. *Measurement* 2016; 85: 132–141.

122. S. Srinivasan, H. Latchman, J. Shea, T. Wong, J. McNair. Airborne traffic surveillance systems: video surveillance of highway traffic. *In: Proceedings of Video Surveillance and Sensor Networks,* 2004; 131–135.

123. A.A. Puri. A survey of unmanned aerial vehicles (UAV) for traffic surveillance. Technical report, Department of computer science and engineering, University of South Florida, USA, 2005.

124. F. Heintz, P. Rudol, P. Doherty. From images to traffic behavior-a uav tracking and monitoring application. *In: Proceedings of International Conference on Information Fusion*, 2007; 1–8.

125. C. Eschmann, C.-M. Kuo, C.-H. Kuo, C. Boller. High-resolution multisensor infrastructure inspection with unmanned aircraft systems. *Int Arch Photogramm Remote Sens Spatial Inform Sci* 2013; 1: 125–129.

126. S.-S. Choi, E.-K. Kim. Building crack inspection using small UAV. *In: Proceedings of International Conference on Advanced Communication Technology*, 2015; 235–238.

127. F.C. Pereira, C.E. Pereira. Embedded image processing systems for automatic recognition of cracks using UAVs. *IFAC-PapersOnLine* 2015; 48: 16–21.

128. S. Sankarasrinivasan, E. Balasubramanian, K. Karthik, U. Chandrasekar, R. Gupta. Health monitoring of civil structures with integrated UAV and image processing system. *Procedia Comput Sci* 2015; 54: 508–515.

129. A. Ellenberg, A. Kontsos, F. Moon, I. Bartoli. Bridge related damage quantification using unmanned aerial vehicle imagery. *Struct Control Health Monit* 2016; 23: 1168–1179.

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

131. ACI 224R-90. Control of Cracking in Concrete Structures; America Concrete Institute: Farmington Hills, MI, USA, 1990.