



Research

High Performance Structures: Building Structures and Materials—Review

Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring

Billie F. Spencer Jr. ^{a,*}, Vedhus Hoskere ^{a,b}, Yasutaka Narazaki ^a^aDepartment of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801-2352, USA^bDepartment of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801-2302, USA

ARTICLE INFO

Article history:

Received 3 August 2018

Revised 14 October 2018

Accepted 15 November 2018

Available online 7 March 2019

Keywords:

Structural inspection and monitoring

Artificial intelligence

Computer vision

Machine learning

Optical flow

ABSTRACT

Computer vision techniques, in conjunction with acquisition through remote cameras and unmanned aerial vehicles (UAVs), offer promising non-contact solutions to civil infrastructure condition assessment. The ultimate goal of such a system is to automatically and robustly convert the image or video data into actionable information. This paper provides an overview of recent advances in computer vision techniques as they apply to the problem of civil infrastructure condition assessment. In particular, relevant research in the fields of computer vision, machine learning, and structural engineering is presented. The work reviewed is classified into two types: inspection applications and monitoring applications. The inspection applications reviewed include identifying context such as structural components, characterizing local and global visible damage, and detecting changes from a reference image. The monitoring applications discussed include static measurement of strain and displacement, as well as dynamic measurement of displacement for modal analysis. Subsequently, some of the key challenges that persist toward the goal of automated vision-based civil infrastructure and monitoring are presented. The paper concludes with ongoing work aimed at addressing some of these stated challenges.

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Much of the critical infrastructure that serves society today, including bridges, dams, highways, lifeline systems, and buildings, was erected several decades ago and is well past its design life. For example, in the United States, according to the 2017 *Infrastructure Report Card* published by the American Society of Civil Engineers, there are over 56 000 structurally deficient bridges, requiring a massive 123 billion USD for rehabilitation [1]. The economic implications of repair necessitate systematic prioritization achieved through a careful understanding of the current state of infrastructure.

Civil infrastructure condition assessment is performed by leveraging the information obtained by inspection and/or monitoring processes. Traditional techniques to assess the condition of civil infrastructure typically involve visual inspection by trained inspectors, combined with relevant decision-making criteria (e.g., ATC-20 [2], national bridge inspection standards [3]). However, such

inspection can be time-consuming, laborious, expensive, and/or dangerous (Fig. 1) [4]. Monitoring can be used to obtain quantitative understanding of the current state of the structure through measurement of physical quantities, such as accelerations, strains, and/or displacements; such approaches offer the ability to continuously observe structural integrity in real time, with the goal of enhanced safety and reliability, and reduced maintenance and inspection costs [5–8]. While these methods have been shown to produce reliable data, they typically have limited spatial resolution or require installation of dense sensor arrays. Another issue is that once installed, access to the sensors is often limited, making regular system maintenance challenging. If only occasional monitoring is required, the installation of contact sensors is difficult and time consuming. To address some of these problems, improved inspection and monitoring approaches with less intervention from humans, lower cost, and higher spatial resolution must be developed and tested to advance and realize the full benefits of automated civil infrastructure condition assessment.

Computer vision techniques have been recognized in the civil engineering field as a key component of improved inspection and monitoring. Images and videos are two major modes of data

* Corresponding author.

E-mail address: bfs@illinois.edu (B.F. Spencer Jr.).



Fig. 1. Inspectors from the US Army Corps of Engineers rappel down a Tainter gate to inspect for surface damage [4].

analyzed by computer vision techniques. Images capture visual information similar to that obtained by human inspectors. Because of this similarity, computer implementation of structural inspection that is analogous to visual inspection by human inspectors is anticipated. In addition, images can encode information from the entire field of view in a non-contact manner, potentially addressing the challenges of monitoring using contact sensors. Videos are a sequence of images in which the extra dimension of time provides important information for both inspection and monitoring applications, ranging from the assimilation of context when images are collected from multiple views, to the dynamic response of the structure when high-sampling rates are used. A significant amount of research in the civil engineering community has focused on developing and adapting computer-vision techniques for inspection and monitoring tasks. Moreover, such vision-based approaches, used in conjunction with cameras and unmanned aerial vehicles (UAVs), offer the potential for rapid and automated inspection and monitoring for civil infrastructure condition assessment.

The present paper reviews recent research on vision-based condition assessment of civil infrastructure. To put the research described in this paper in its proper technical perspective, a brief history of computer vision research is first provided in Section 2. Section 3 then reviews in detail several recent efforts dealing with inspection applications of computer vision techniques for civil infrastructure assessment. Section 4 focuses on monitoring applications, and Section 5 outlines challenges for the realization of automated structural inspection and monitoring. Section 6 discusses ongoing work by the authors toward the goal of automated inspections. Finally, Section 7 provides conclusions.

2. A brief history of computer vision research

Computer vision is an interdisciplinary scientific field concerned with the automatic extraction of useful information from image data in order to understand or represent the underlying physical world, either qualitatively or quantitatively. Computer vision methods can be used to automate tasks of the human visual cortex. Initial efforts toward applying computer vision methods began in the 1960s, and sought to extract shape information about objects using edges and primitive shapes (e.g., boxes) [9]. Computer vision methods began to consider more complex perception problems with the development of different representations of image patterns. Optical character recognition (OCR) was of major interest, as the characters and digits of any fonts needed to be recognized for the purpose of increased automation in the United States Postal Service [10], license plate recognition [11], and so

forth. Facial recognition, where the input image is evaluated in feature space obtained by applying hand-crafted or learned filters with the aim of detecting patterns representing human faces, has also been an active area of research [12,13]. Other object detection problems, such as pedestrian detection and car detection, have begun to show significant improvement in recent years (e.g., Ref. [14]), motivated by increased demand for surveillance and traffic monitoring. Computer vision techniques have also been used in sports broadcasting for applications such as ball tracking and virtual replays [15].

Recent advances in computer vision techniques have largely been fueled through end-to-end learning using artificial neural networks (ANNs) and convolutional neural networks (CNNs). In ANNs and CNNs, a complex input-output relation of data is approximated by a parametrized nonlinear function defined using units called nodes [16]. Output of each ANN node is computed by the following:

$$y_n = \sigma_n(\mathbf{w}_n^T \mathbf{x}_n + \mathbf{b}_n) \quad (1)$$

where \mathbf{x}_n is a vector of input to the node n ; y_n is a scalar output from the node; and \mathbf{w}_n and \mathbf{b}_n are vectors of weight and bias parameters, respectively. σ_n is a nonlinear activation function, such as the sigmoid function and rectifier (rectified linear unit, or ReLU [17]). Similarly, for CNNs, each node applies convolution, followed by a nonlinear activation function:

$$y_n = \sigma_n(W_n * \mathbf{x}_n + \mathbf{b}_n) \quad (2)$$

where $*$ denotes convolution and W_n is the convolution kernel. The final layer of CNNs is typically a fully connected layer (FCL) which has dense connections to the output, similar to the layers of an ANN. The CNN is particularly effective for image and video data, because recognition using CNNs is robust to translation with a limited number of parameters. By increasing the number of nodes connected with each other, an arbitrary complex parametrization of the input-output relation can be realized (e.g., multilayer perceptron with many hidden layers and/or many nodes in each layer, deep convolutional neural networks (DCNNs), etc.). The parameters of the ANNs/CNNs are optimized using a collection of input and output data (training data) (e.g., Refs. [18,19]).

These algorithms have achieved remarkable success in building perception systems for highly complex visual problems. CNNs have achieved more than 99.5% accuracy on the Modified National Institute of Standards and Technology (MNIST) handwritten digit classification problem (Fig. 2(a)) [20]. Moreover, state-of-the-art CNN architectures have achieved less than 5% top-five error (ratio of data where the true class does not mark the top five classification score) [21] on the 1000-class ImageNet classification problem (Fig. 2(b)) [22].

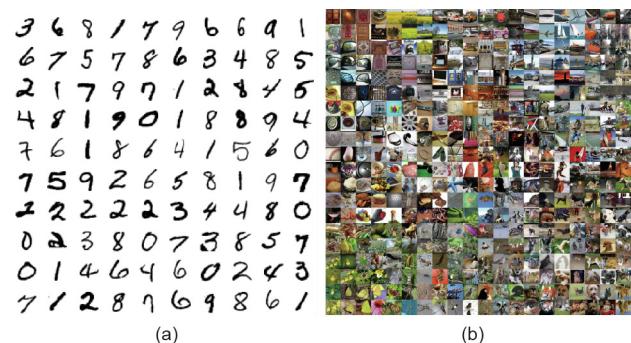


Fig. 2. Popular image classification datasets. (a) Example images from the MNIST dataset [20]; (b) ImageNet example images, visualized by t-distributed stochastic neighbor embedding (t-SNE) [22].

Use of CNNs is not limited to image classification (i.e., inferring a single label per image). A DCNN applies multiple nonlinear filters and computes a map of filter responses (\mathbf{F} in Fig. 3, which is called a “feature map”). Instead of using all filter responses simultaneously to get a per-image class (upper flow in Fig. 3), filter responses at each location in the map can be used separately to extract information about both object categories and their locations. Using the feature map, semantic segmentation algorithms assign an appropriate label to each pixel of the image [23–26]. Object detection algorithms use the feature map to detect and localize objects of interest, typically by drawing their bounding boxes [27–31]. Instance-level segmentation algorithms [32–34] further process the feature map to differentiate each instance of an object (e.g., assigning a separate label for each person in an image, instead of assigning the same label to all people in the input image). While dealing with video data, additional temporal information can also be used to conduct spatiotemporal analysis [35–37] for segmentation.

The Achilles heel of supervised learning techniques is the need for high-quality labeled data (i.e., images in which the objects are already identified) that are used for training purposes. While many software applications have been created to help ease the labeling process (e.g., Refs. [38,39]), manual labeling still remains a very cumbersome task. Weakly supervised training has been proposed to perform object detection and localization tasks without pixel-level or object-level labeling of images [40,41]; here, a CNN is trained on image-wise labels to get the object category and approximate location in the image.

Unsupervised learning techniques further reduce the need for labeled data by identifying the underlying probabilistic structure in the observed data. For example, clustering algorithms (e.g., k -means algorithm [11]) assume that the data (e.g., image patch) is generated by multiple sources (e.g., different material types), and allocate each data sample to one of the sources based on the maximum likelihood (ML) framework. For example, DeGol et al. [42] used the k -means algorithm to perform material recognition for imaged surfaces. More complex probabilistic structures can be extracted by fitting parametrized probabilistic models to the observed data (e.g., Gaussian mixture model (GMM) [16], Boltzmann machines [43,44]). In the image processing context, CNN-based architectures for unsupervised learning have been actively investigated, such as auto-encoders [45–47] and generative adversarial networks (GANs) [48–50]. These methods can automatically learn the compact representation of the input image and/or image recovery/generation process from compact representations without manually labeled ground truth. A thorough and concise review of different supervised and unsupervised learning algorithms can be found in Ref. [51].

Another set of algorithms that have spurred significant advances in computer vision and artificial intelligence (AI) across many applications are optical flow techniques. Optical flow estimates a

motion field through pixel correspondences between two image frames. Four main classes of algorithms can be used to compute optical flow, including: ① differential methods, ② region matching, ③ energy methods, and ④ phase-based techniques, for which details and references can be found in Ref. [52]. Optical flow has wide ranging applications in processing video data, from video compression [53], to video segmentation [54], motion magnification [55] and vision-based navigation of UAVs [56].

With these advances, computer vision techniques have been used to realize a wide variety of cutting-edge applications. For example, computer vision techniques are used in self-driving cars (Fig. 4) [57–59] to identify and react to potential risks encountered during driving. Accurate face recognition algorithms empower social media [60] and are also used in surveillance applications (e.g., law enforcement in airports [61]). Other successful applications include automated urban mapping [62] and enhanced medical imaging [63]. The significant improvements and successful applications of computer vision techniques in many fields provide increasing motivation for scholars to develop computer vision solutions to the civil engineering problems. Indeed, using computer vision is a natural step toward improved monitoring and inspection of civil infrastructure. With this brief history as background, the following sections describe research efforts to adapt and further develop computer vision techniques for the inspection and monitoring of civil infrastructure.

3. Inspection applications

Researchers frequently envision an automated inspection framework that consists of two main steps: ① utilizing UAVs for remote automated data acquisition; and ② data processing and inspection using computer vision techniques. Intelligent UAVs are no longer a thing of the future, and the rapid growth in the drone industry over the last few years has made UAVs a viable option for data acquisition. Indeed, UAVs are being deployed by several federal and state agencies, as well as other research organizations in the United States (e.g., Minnesota Department of Transportation [64,65], Florida Department of Transportation [66], University of Florida [67], Michigan Department of Transportation [68], South Dakota State University [69]). These efforts have primarily focused on taking photographs and videos that are used for onsite evaluation or subsequent virtual inspections by engineers. The ability to automatically and robustly convert images or video data into actionable information is still challenging. Toward this goal, the first major subsection below reviews literature in damage detection, and the second reviews structural component recognition. The third major subsection briefly reviews a demonstration that combines both these aspects: damage detection with structure-level consistency.

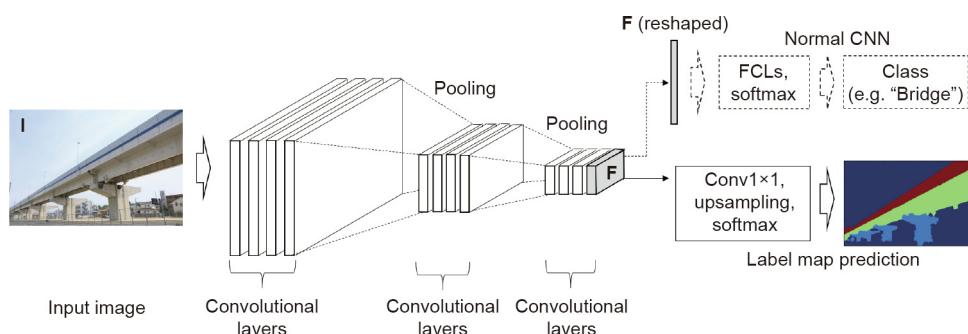


Fig. 3. Fully convolutional neural networks (FCNs) [23].



Fig. 4. A self-driving car system by Waymo [58,59].[†]

3.1. Damage detection

Automated damage detection is a crucial component of any automated or semi-automated inspection system. When characterized by the ratio of pixels representing damage to those representing the undamaged portion of the structure's surface, the presence of defects in images of a structure can be considered a relatively rare occurrence. Thus, the detection of visual defects with high precision and recall is a challenging task. This problem is further complicated by the presence of damage (DP)-like features (e.g., dark edges such as a groove can be mistaken for a crack). As described below, a great deal of research has been devoted to developing methods and techniques to reliably identify different visual defects, including concrete cracks, concrete spalling and delamination, fatigue cracks, steel corrosion, and asphalt cracks. Three different approaches for damage detection are discussed below: ① heuristic feature-extraction methods, ② deep learning-based damage detection, and ③ change detection.

3.1.1. Heuristic feature-extraction methods

Researchers have developed different heuristics methods for damage detection using image data. In principle, these methods work by applying a threshold or a machine learning classifier to the output of a hand-crafted filter for the particular damage type (DT) of interest. This section describes some of the key DTs for which heuristic feature-extraction methods have been developed.

(1) **Concrete cracks.** Much of the early work on vision-based damage detection focused on identifying concrete cracks based on heuristic filters (e.g., Refs. [70–80]). Edge detection filters were the first type of heuristics to be used (e.g., Ref. [70]). An early survey of approaches can be found in Ref. [71]. Jahanshahi and Masri [72] used morphological features, together with classifiers (neural networks and support vector machines), to identify cracks of different thicknesses. The results from this study are presented in Fig. 5 [72,81], where the first column shows the original images used in the study and the subsequent columns show the results from the application of the bottom-hat method, Canny method, and the algorithm from Ref. [72]. The same paper also puts forth a method for quantifying crack thickness by identifying the centerline of each crack and computing the distance to the edges. Nishikawa et al. [74] proposed multiple sequential image filtering for crack detection and property estimation. Other researchers have also developed methods to estimate the properties of concrete cracks. Liu et al. [79] proposed a method for automated crack assessment using adaptive image processing, in which a median filter was used to decompose a crack into its skeleton and edges. Depth and three-dimensional (3D) information was

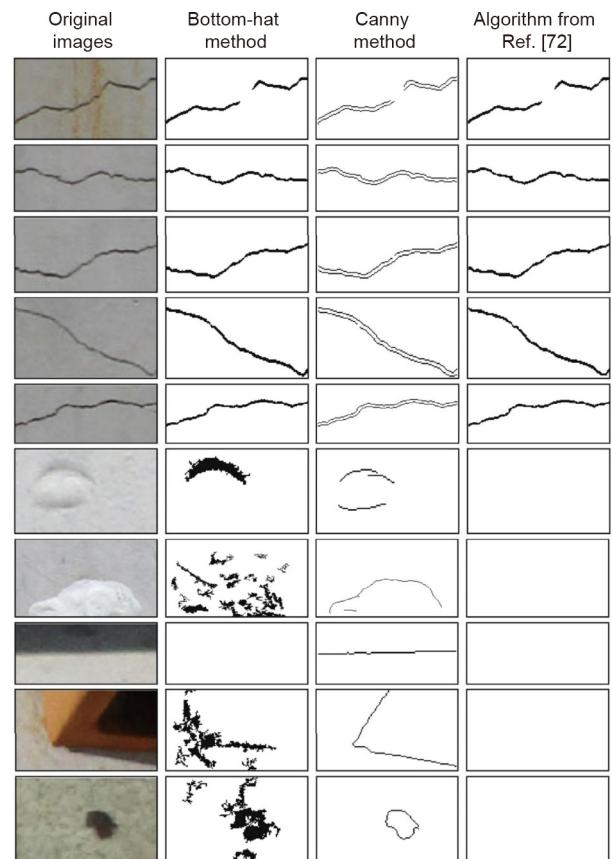


Fig. 5. A comparison of different crack detection methods conducted in Ref. [81].

also incorporated to conduct quantitative damage evaluation in Refs. [80,81]. Erkal and Hajjar [82] developed and evaluated clustering process to automatically classify defects such as cracks, corrosion, ruptures, and spalling in colorized laser scan data using surface normal based damage detection. In many of the methods discussed here, binarization is a step typically employed in crack-detection pipelines. Kim et al. [83] compared different binarization methods for crack detection. These methods have been applied to a variety of civil infrastructure, including bridges (e.g., Refs. [84,85]), tunnel linings (e.g., Ref. [76]), and post-earthquake building assessment (e.g., Ref. [86]).

(2) **Concrete spalling.** Methods have also been proposed to identify other defects in concrete, such as spalling. A novel orthogonal transformation approach combined with a bridge condition index was used by Adhikari et al. [87] to quantify degradation and subsequently map to condition ratings. The authors were able to achieve a reasonable accuracy of 85% for the detection of spalling in their dataset, but were unable to address situations in which both cracks and spalling were present. Paal et al. [88] employed a combination of segmentation, template matching, and morphological preprocessing, for both spall detection and concrete column assessment.

(3) **Fatigue cracks in steel.** Fatigue cracks are a critical problem for steel deck bridges because they can significantly shorten the lifespan of a structure. However, research on steel fatigue crack detection in civil infrastructure has been fairly limited. Yeum and Dyke [89] manually created defects on a steel beam to give the appearance of fatigue cracks (Fig. 6). They then used a combination of region localization by object detection and filtering techniques to identify the created fatigue-crack-like defects. The authors made an interesting and useful assumption that fatigue cracks generally develop around bolt holes; however, this assumption may not be

[†] All image copyrights with Waymo Inc. image used under "17 US Code § 107. Limitations on exclusive rights: fair use."

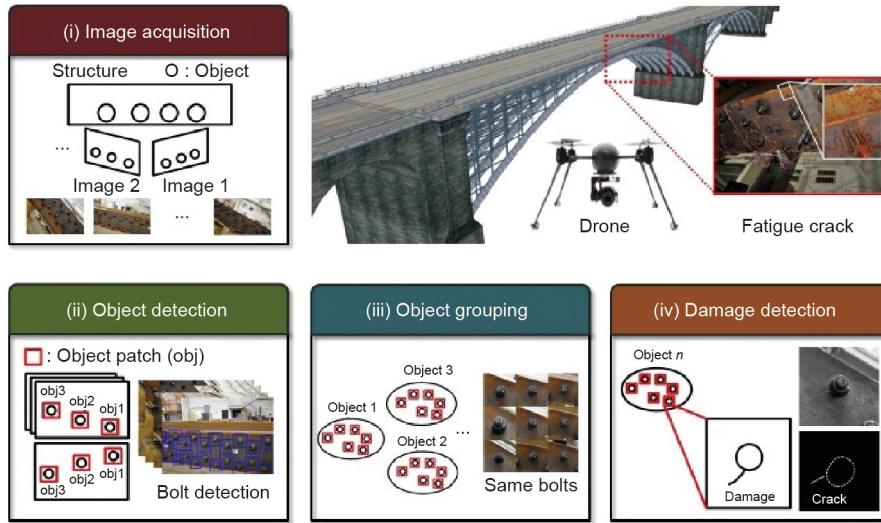


Fig. 6. Vision-based automated crack detection for bridge inspection in Ref. [89].

valid for other steel structures, for which critical members are usually welded—including, for example, navigational infrastructure such as miter gates [90]. Jahanshahi et al. [91] proposed a region growing method to segment microcracks on internal components of nuclear reactors.

(4) **Steel corrosion.** Researchers have used textural, spectral, and color information for the identification of corrosion. Ghanta et al. [92] proposed the use of wavelet features together with principal component analysis for corrosion percent estimation in images. Jahanshahi and Masri [93] parametrically evaluated the performance of wavelet-based corrosion algorithms. Methods using textural and color features have also been proposed and evaluated (e.g., Refs. [94,95]). Automated algorithms for robotic and smartphone-based maintenance systems have also been proposed for image-based corrosion detection (e.g., Refs. [96,97]). A survey of corrosion detection approaches using computer vision can be found in Ref. [98].

(5) **Asphalt defects.** Numerous techniques exist for the detection and assessment of asphalt pavement cracks and defects using heuristic feature-extraction techniques [99–105]. Hu and Zhao [101] used a local binary pattern-based (LBP) algorithm to identify pavement cracks. Salman et al. [100] proposed the use of Gabor filtering. Koch and Brilakis [99] used histogram shape-based thresholding to automatically detect potholes in pavements. In addition to RGB data (where RGB refers to 3 color channels representing red, green, and blue wavelengths of light), depth data has been used for the condition assessment of roads. For example, Chen et al. [106] reported the use of an inexpensive RGB-D sensor (Microsoft Kinect) to detect, quantify, and localize pavement defects. A detailed review of methods for asphalt defect detection can be found in Ref. [107].

For further study on identification methods for some of these defects, Koch et al. [108] provided an excellent review of computer vision defect detection techniques developed prior to 2015, classified based on the structure to which they are applied.

3.1.2. Deep learning-based damage detection

The studies and techniques described thus far can be categorized as either using machine learning techniques or relying on a combination of heuristic features together with a classifier. In essence, however, the application of such techniques in an automated structural inspection environment is limited, because these

techniques do not employ the contextual information that is available in the regions around where the defect is present, such as the nature of the material or structural components. These heuristic filtering-based techniques need to be manually or semi-automatically tuned, depending on the appearance of the target structure being monitored. Real-world situations vary extensively, and hand crafting a general algorithm that can be successful in general cases is quite difficult. The recent success of deep learning for computer vision [21,51] in a number of fields, such as general image classification [109], autonomous transportation systems [57], and medical imaging [63], has driven its application in civil infrastructure inspection and monitoring. Deep learning has greatly extended the capability and robustness of traditional vision-based damage detection for a wide variety of visual defects, ranging from cracks and spalling to corrosion. Different approaches for detection have been studied, including ① image classification, ② object detection or region-proposal methods, and ③ semantic segmentation methods. These applications are discussed below.

(1) **Image classification.** CNNs have been employed for the application of crack detection in steel decks [110], asphalt pavements [111], and concrete surfaces [112], with very high accuracy being achieved in all cases. Kim et al. [113] proposed a classification framework for identifying cracks in the presence of crack-like patterns using CNN and speeded-up robust features (SURF), and determined the pixel-wise location using image binarization. Architectures such as AlexNet have been fine-tuned for crack detection [114,115] and GoogleNet has been similarly fine-tuned for spalling [116]. Atha and Jahanshahi [117] evaluated different deep learning techniques for corrosion detection, and Chen and Jahanshahi [118] proposed the use of the naïve Bayes data fusion with a CNN for crack detection. Yeum [119] utilized CNNs for the extraction of important regions of interest in highway truss structures in order to ease the inspection process.

Xu et al. [110,120] systematically investigated the detection of steel fatigue cracks for long-span bridges using deep learning neural networks, including a restricted Boltzmann machine and a fusion CNN. The novel fusion CNN proposed was able to identify minor cracks at multiple scales, with high accuracy, and under complex backgrounds present during in-field testing. Maguire et al. [121] developed a concrete crack image dataset for machine learning applications containing 56 000 images that were classified as either having a crack or not.

Bao et al. [122] proposed the use of DCNNs as an anomaly detector to aid inspectors in filtering out anomalous data from a bridge structural health monitoring (SHM) system recording acceleration data. Dang et al. [123] used a UAV to collect close-up images of bridges, and then automatically detected structural damage using CNNs applied to image patches.

(2) **Object detection.** Object detection methods have recently been applied for damage detection [112,119]. Instead of classifying an entire image, object detection methods create a bounding box around the damaged region. Regions with CNN features (R-CNNs) have been used for spall detection in a post-disaster scenario by Yeum et al. [124], although the results (59.39% true positive accuracy) left some room for improvement. The methods discussed thus far have only been applied to a single DT. In contrast, deep learning methods have the ability to learn general representations of identifiable characteristics in images over a large number of classes. For example, DCNNs have been successful in classification problems with over 1000 classes [21]. Limited research is available on techniques for the detection of multiple DTs. Cha et al. [125] studied the use of Faster R-CNN, a region-based method proposed by Ren et al. [30] to identify multiple DTs including concrete cracks and different levels of corrosion and delamination.

(3) **Semantic segmentation.** Object-detection-based methods cannot precisely delineate the shape of the damage they are isolating, as they only aim to fit a rectangular box around the region of

interest. Another method to isolate regions of interest in an image is termed semantic segmentation. More precisely, semantic segmentation is the act of classifying each pixel in an image into one of a fixed number of classes. The result is a segmented image in which each segment is assigned a certain class. Thus, when applied to damage detection, the precise location and shape of the damage can be delineated.

Zhang et al. [126] proposed CrackNet, an efficient architecture for the semantic segmentation of pavement cracks. An object instance segmentation technique, Mask R-CNN [32], has also been recently adapted [116] for the detection of cracks, spalling, rebar exposure, and efflorescence. Although Mask R-CNN does provide pixel-level delineation of damage, it only segments the part of the image where “objects” have been found, rather than performing a full semantic segmentation of the image.

Hoskere et al. [127,128] evaluated two methods for the general localization and classification of multiple DTs: ① a multiscale pixel-wise DCNN, and ② a fully convolutional neural network (FCN). As shown in Fig. 7 [127], six different DTs were considered: concrete cracks, concrete spalling, exposed reinforcement bars, steel corrosion, steel fracture and fatigue cracks, and asphalt cracks. In Ref. [127], a new network configuration and dataset was proposed. The dataset is comprised of images from a wide variety of structures including bridges, buildings, pavements, dams, and laboratory specimens. The proposed technique consists

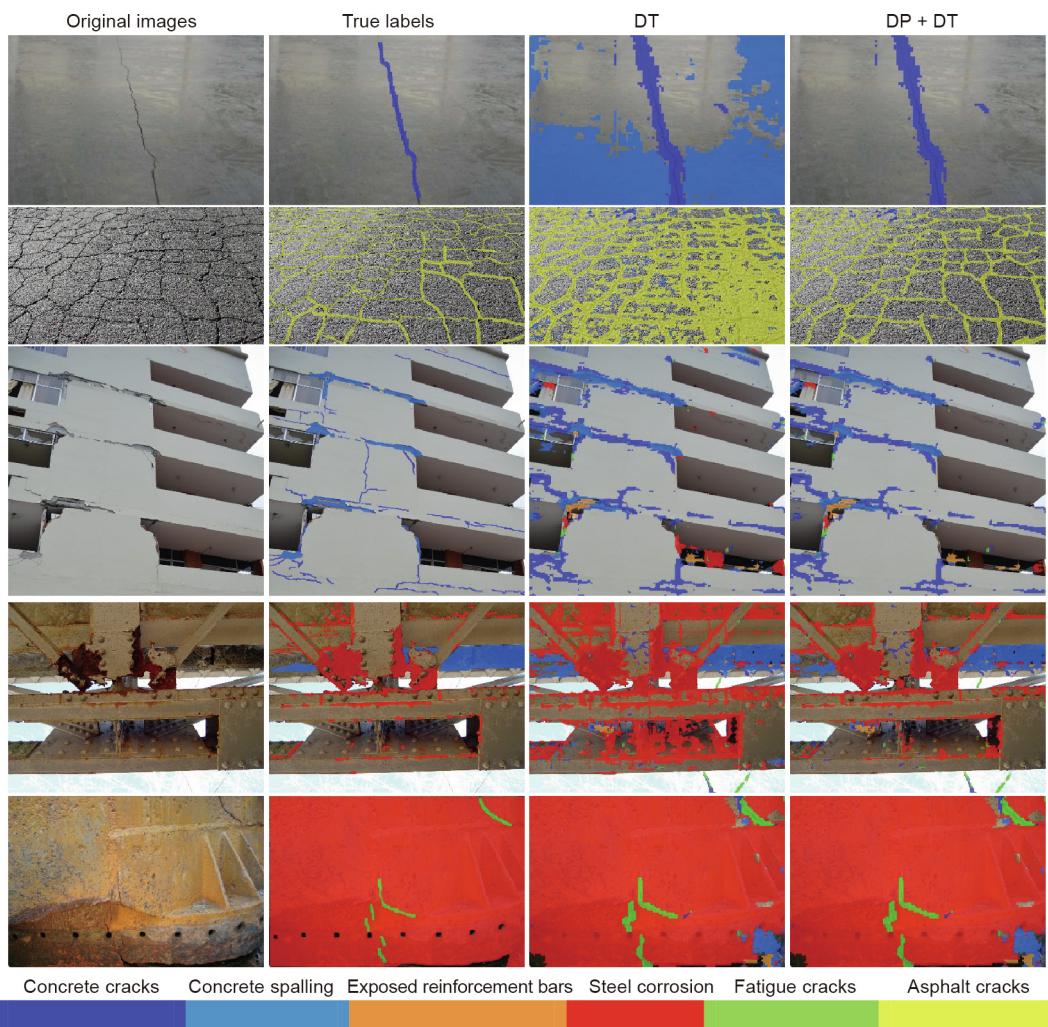


Fig. 7. Deep learning-based semantic segmentation of multiple structural DTs [127].

of a parallel configuration of two networks—that is, the DP network and DT network—to increase the efficiency of damage detection. The scale invariance of the technique is demonstrated through the variety of scales of damage present in the proposed dataset.

3.1.3. Change detection

When a structure must be regularly inspected, a baseline representation of the structure can first be established. This baseline can then be compared against using data from subsequent inspections. Any new visual damage to the structure will manifest itself as a change during a comparison with the baseline. Identifying and localizing changes can help to reduce the workload for processing data from UAV inspections. Because any damage must manifest as a change, adopting change detection approaches before conducting damage detection can help to reduce the number of false positives, as damage-like textures are likely to be present in both states. Change detection is a problem that has been studied in computer vision, with applications ranging from environmental monitoring to video surveillance. In this subsection, we examine two main approaches to change detection: ① point-cloud-based change detection and ② image-based change detection.

(1) **Point-cloud-based change detection.** Structure from motion (SFM) and multi-view stereo (MVS) [129] are vision-based techniques that can be used to generate point clouds of a structure. To conduct change detection, a baseline point cloud must first be established. These point clouds can be highly accurate, even for complex civil infrastructure such as truss bridges or dams, as described in Refs. [130,131]. Subsequent scans are then registered with the baseline cloud, with alignment typically carried out by the iterative closest point (ICP) algorithm [132]. The ICP algorithm has been implemented in open-source software such as MeshLab [133] and CloudCompare [134]. After alignment, different change detection procedures can be implemented. These techniques can be applied to both laser-scanning point clouds and point clouds generated from photogrammetry. Early studies used the Hausdorff distance from cloud to cloud (C2C) as a metric to identify changes in three dimensions [135]. Other techniques include digital elevation models of difference (DoD), cloud-to-mesh (C2M) methods, and Multiscale Model to Model Cloud Comparison (M3C2) methods. A review of these methods can be found in Ref. [136].

Combined with UAV data acquisition, change detection methods have been applied to civil infrastructure. For example, Morgenthal and Hallermann [137] identified manually inserted changes on a retaining wall using aligned orthomosaics for in-plane changes and C2C comparison using the CloudCompare package for out-of-plane changes [138]. Khaloo and Lattanzi [139] utilized the hue values of the pixels in different color spaces to aid the detection of important changes to a gravity dam. Jafari et al. [140] proposed a new method for measuring deformations using the direct point-wise distance in conjunction with statistical sampling to maximize data integrity. Another interesting application of point-cloud-based change detection is for finite-element model updating. Ghahremani et al. [141] used vision-based methods to automatically locate, identify, and quantify damage through comparative analysis of two point clouds of a laboratory structural component, and subsequently used that information to update a finite-element model of the component. Point-cloud-based change detection methods are applicable when the change in point depth is sufficient to be identified. When looking for visual changes that may not cause sufficient geometric changes, image-based change detection methods can be used.

(2) **Image-based change detection.** Change detection on images is a commonly studied problem in computer vision due to its widespread use across a number of applications [142]. One of the most prevalent use-cases for image-based change detection is seen in remotely sensed satellite imagery, where applications range from

land-cover and land-use monitoring to damage assessment and disaster monitoring. An in-depth review of change detection methods for very high-resolution satellite images can be found in Ref. [143]. Before change detection is performed, the images are preprocessed to remove environmental variations such as atmospheric and radiometric effects, followed by image registration. Similar to damage detection, heuristic and deep learning-based techniques are available, as well as point-cloud and object-detection-based techniques. Ref. [144] provides a review of these approaches.

While remotely sensed satellite imagery can provide a city-scale understanding of damage, for individual structures, the resolution and views available from such images limit the amount of useful information that can be extracted. For images from UAV or ground vehicle surveys, change detection methods can be applied as a precursor to damage detection, in order to help localize candidate pixels or regions that may represent damage. To this end, Sakurada et al. [145] proposed a method to detect 3D changes in an outdoor scene from multi-view images taken at different time instances using a probabilistically estimated scene depth. CNNs have also been used to identify changes in urban scenes [146]. Stent et al. [147] proposed the use of a CNN to identify changes on tunnel linings, and subsequently ranked the changes in importance using clustering. A schematic of the method proposed in Stent et al. [147] is shown in Fig. 8.

3.2. Structural component recognition

Structural component recognition is a process of detecting, localizing, and classifying characteristic parts of a structure, which is expected to be a key step toward the automated inspection of civil infrastructure. Information about structural components adds semantics to raw images or 3D point-cloud data. Such semantics help humans understand the current state of the structure; they also impose consistency onto error-prone data from field environments [148–150]. For example, by assigning a label “column” to point-cloud data, a collection of points becomes recognizable as a single structural component (as-built modeling). In the construction progress monitoring context, the column of the as-built model can be corresponded with the column of the 3D model developed at the design stage (the as-planned model), thus permitting the reference-based evaluation of the current state of the column. During the evaluation, points not labeled as “column” can be safely neglected, because such points are thought to originate from irrelevant objects or erroneous data. In this sense, information about structural components is one of the essential attributes of as-built models used to represent the current state of structures in an effective and consistent manner.

Structural component recognition also provides strong supportive information for the automated vision-based damage evaluation of civil structures. Similar to as-built modeling, information regarding structural components can be used to improve the consistency of the automated damage detection methods by removing damage-like patterns on objects other than the structural components of interest (e.g., a crack detected in a tree is false-positive detection). Furthermore, information about structural components is necessary for the safety evaluation of the entire structure, because damage and the structural components on which the damage appears are jointly evaluated in order to derive the safety ratings in most current structural inspection guidelines (ATC-20 [2], national bridge inspection standards [3]).

Toward fully autonomous inspection, structural component recognition is expected to be a building block of autonomous navigation and data collection algorithms for robotic platforms (e.g., UAVs). Based on the types and positions of structural components recognized by the on-board camera(s), autonomous robots are expected to plan appropriate navigation paths and data

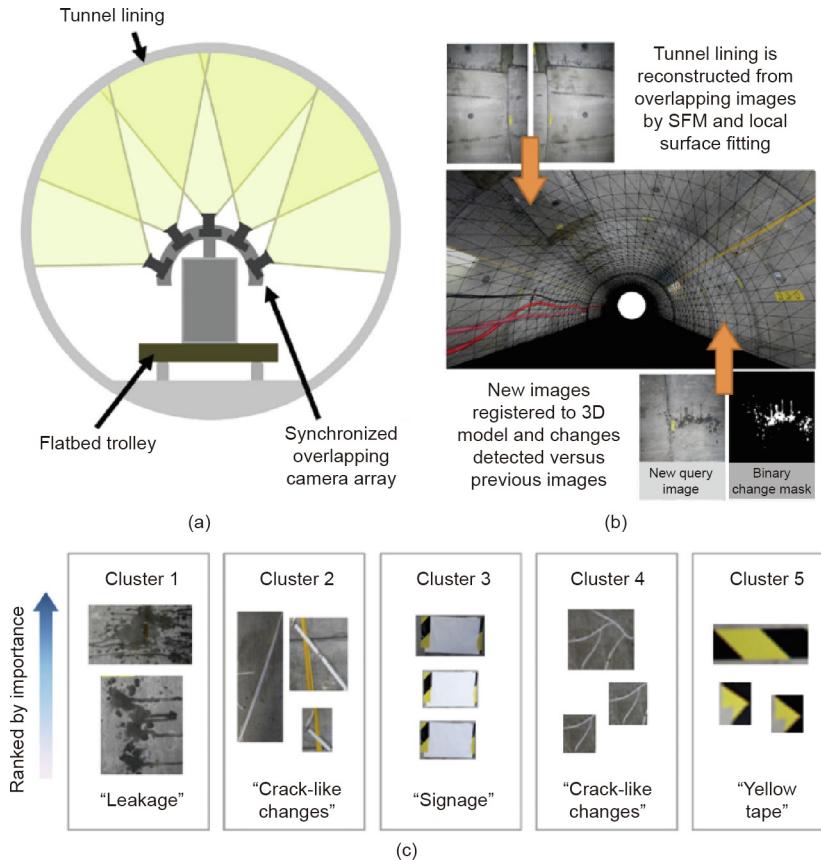


Fig. 8. Illustration of the system proposed in Ref. [147]. (a) Hardware for data capture; (b) changes detected in new images by localizing within a reconstructed reference model; (c) sample output, in which detected changes are clustered by appearance and ranked.

collection behaviors. Although full automation of structural inspection has not been achieved yet, an example of an autonomous robot based on vision-based recognition of the surrounding environment can be found in the field of agriculture (i.e., the TerraSentia [151]).

3.2.1. Heuristic-based structural component recognition using image data

Early studies use hand-crafted image filters and heuristics to extract structural components from images. For example, reinforced concrete (RC) columns in an image are recognized using line-segment pairs (Fig. 9) [152,153]. To distinguish columns from other irrelevant line-segment pairs, this method applies a threshold to select near-parallel pairs with a predetermined range of the aspect ratio. The authors of that research applied the method to 20 images in which columns were photographed as the main objects, and detected 38 out of 51 columns with seven false-positive detections. Despite the simplicity of the method, this

approach strongly depends on the values of thresholds, and tends to fail to find partially occluded or relatively distant columns. Furthermore, the method recognizes any line-segment pairs satisfying the thresholds as columns, without further contextual understanding. To improve the results and achieve fewer false-positive detections, high-level context within a scene needs to be incorporated at different scales.

3.2.2. Structural component recognition using 3D point-cloud data

Another important scenario for structural component recognition is to perform the task when dense 3D point-cloud data is available. Different segmentation and classification approaches have been proposed to perform the task of building component recognition using dense 3D point-cloud data. Xiong et al. [154] investigated an automated method for converting dense 3D point-cloud data from a room to a semantically rich 3D model represented by planar walls, floors, ceilings, and rectangular openings (a process referred to as Scan-to-Building Information Modeling (BIM)). Perez-Perez et al. [155] adopted higher dimensional features (193 dimensions for semantic features and 553 dimensions for geometric features) to perform structural and nonstructural component recognition in indoor spaces. With the rich information carried by the extracted features and post-processing performed using conditional random fields, this method was able to label both planar and complex nonplanar surfaces accurately, as shown in Fig. 10 [155]. Armeni et al. [156] proposed a method of filtering, segmenting, and classifying a dense 3D point-cloud data, and demonstrated the method by parsing an entire building to planar components.

Golparvar-Fard et al. [157] conducted a detailed comparison of image-based point clouds and laser scanning for automated performance monitoring techniques including accuracy and usability for



Fig. 9. RC column recognition results [153].



Fig. 10. Indoor semantic segmentation by Perez-Perez et al. [155] using 3D dense point-cloud data.

3D reconstruction, shape modeling, and as-built visualizations, and found that although image-based techniques were not as accurate, they provided tremendous opportunities for visualization and the extraction of rich semantic information. Golparvar-Fard et al. [158] proposed an automated method to monitor changes of 3D building elements by fusing unordered photo collections together with a building information modeling using a combination of SFM followed by voxel-based scene quantization. Recently, Lu et al. [159] proposed an approach to robustly detect four types of bridge components from point clouds of RC bridges using a top-down approach.

The effectiveness of the 3D approaches discussed in this section depends on the available data for solving the problem at hand. Compared with image data, dense 3D point-cloud data carry richer information with the extra dimension, which enables the recognition of complex-shaped structural components and/or recognition tasks that require high localization accuracies. On the other hand, in order to obtain accurate and dense 3D point-cloud data, every part of the structure under inspection needs to be photographed with sufficient resolution and overlap, which requires increased effort in data collection. Also, off-line post-processing is often required, posing challenges to the application of 3D approaches to real-time processing tasks. For such scenarios, deep learning-based structural component recognition using image data is another promising approach to perform structural component recognition tasks, as discussed in the next section.

3.2.3. Deep learning-based structural component recognition using image data

Approaches based on machine learning methods for structural component recognition have recently come under active investigation. Image classification is one of the major applications of CNNs, in which a single representative label is estimated from an input image. Yeum et al. [160] used CNNs to classify candidate image

patches of the welded joints of a highway sign truss structure to reliably recognize the region of interest. Gao and Mosalam [161] used CNNs to classify input images into appropriate structural component and damage categories. The authors then inferred rough localization of the target object based on the output of the last convolutional layer (weakly supervised learning; see Fig. 11 [161] for structural component recognition results). Object detection algorithms can also be used for structural component recognition. Liang [162] applied the Faster R-CNN algorithm to detect and localize bridge components by drawing bounding boxes around them automatically.

Semantic segmentation is another promising approach for solving structural component recognition problems [163–165]. Instead of drawing bounding boxes or inferring approximate object locations from per-image labels, semantic segmentation algorithms output label maps with the same resolutions as the input images, which is particularly effective for accurately detecting, localizing, and classifying complex-shaped structural components. To obtain high-resolution bridge-component recognition results consistent with a high-level scene structure, Narazaki et al. [164] investigated three different configurations of FCNs: ① naïve configuration, which estimates a label map directly from an input image; ② parallel configuration, which estimates a label map based on the semantic segmentation results on high-level scene classes and bridge-component classes running in parallel (Fig. 12(a)); and ③ sequential configuration, which estimates a label map based on the scene segmentation results and input image (Fig. 12(b)). Example bridge-component recognition results are shown in

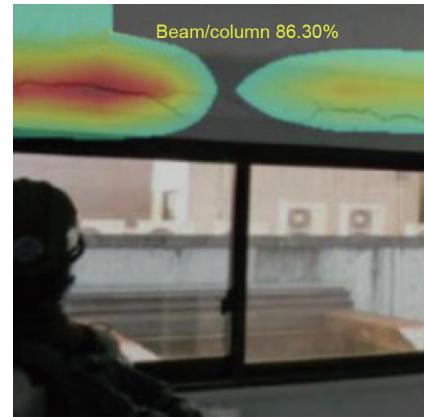


Fig. 11. Structural component recognition results using weakly supervised learning [161].

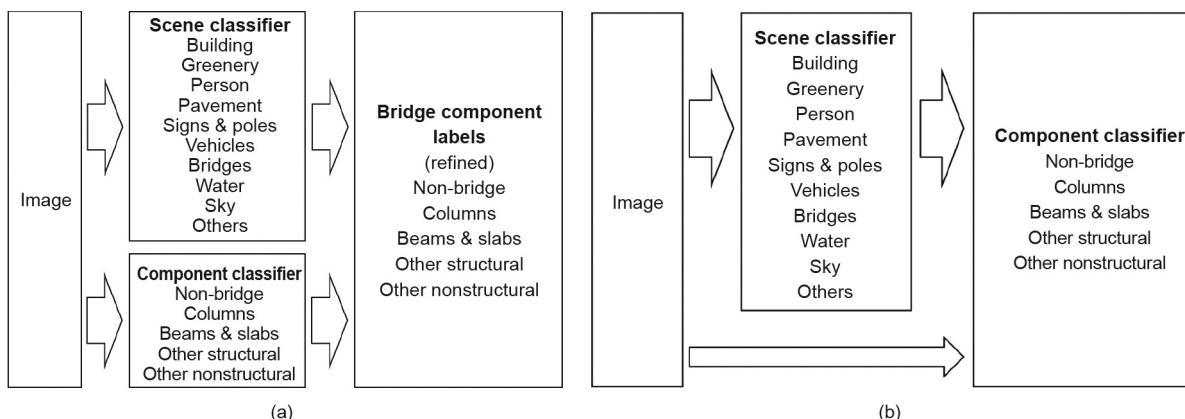


Fig. 12. Network configurations to impose scene-level consistency [164].

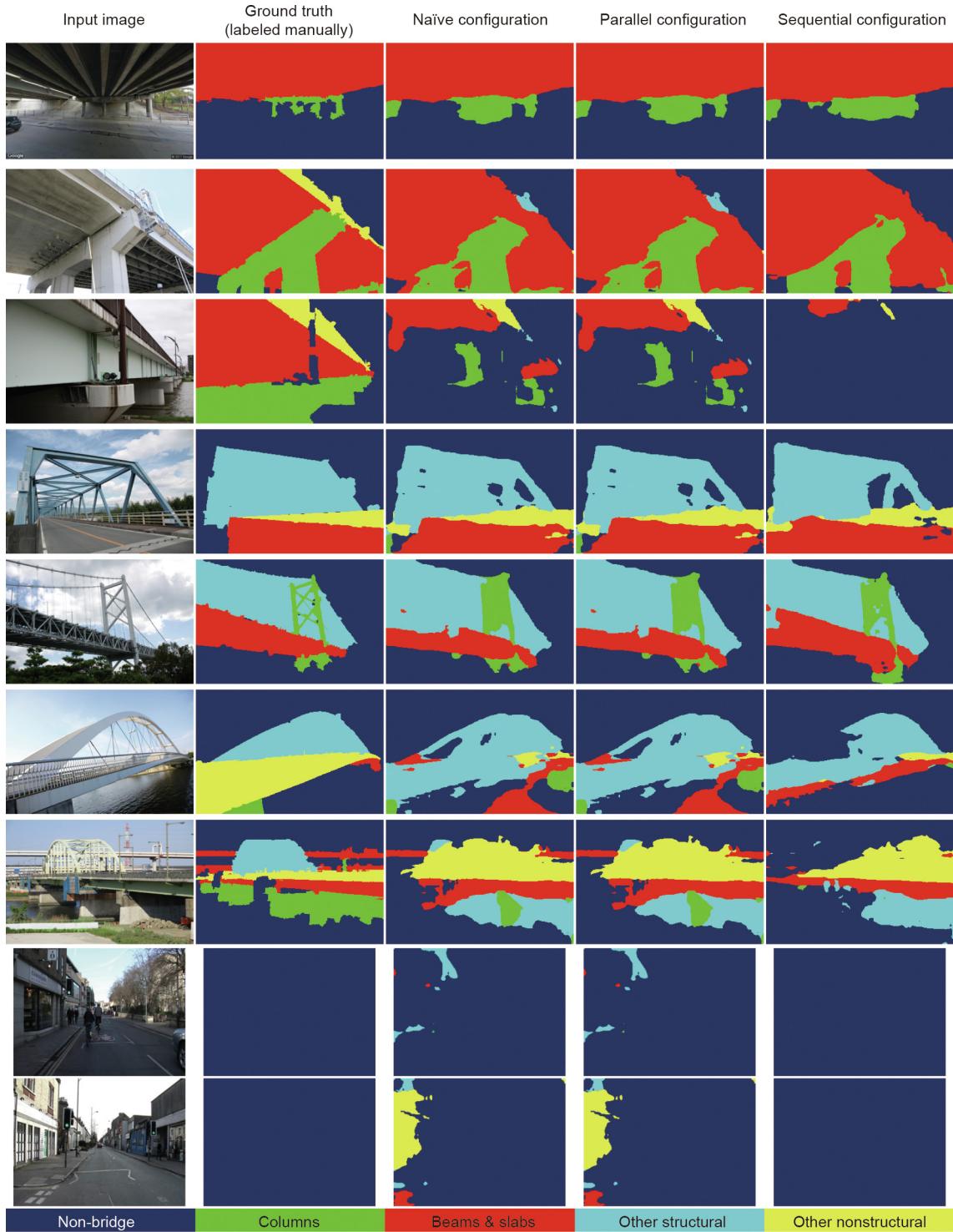


Fig. 13. Example bridge-component recognition results [164].

Fig. 13. Except for the third and seventh image, all configurations were able to identify the structural components, including far or partially occluded columns. Notable differences were observed for non-bridge images (see the last two images in Fig. 13 [164]). For the naïve and parallel configuration, false-positive detections were observed in building and pavement pixels. In contrast, FCNs with a sequential configuration did not show false positives. (Table 1 presents the rates of false-positive detections in the results for non-bridge images.) Therefore, the sequential configuration is

considered to be effective in imposing high-level scene consistency onto bridge-component recognition in order to improve the robustness of recognition for images of complex scenes.

3.3. Damage detection with structure-level consistency

To conduct automated assessments, combining information about both the structural components and the damage state of those components will be vital. German et al. [166] proposed an

Table 1

False-positive rates for the nine scene classes.

Configuration	Scene classes								
	Building	Greenery	Person	Pavement	Signs & poles	Vehicles	Water	Sky	Others
Naïve (%)	18.60	5.80	3.70	3.10	6.80	4.60	9.90	0.90	13.20
Parallel (%)	16.90	5.20	3.10	2.70	5.80	3.90	8.90	0.80	11.80
Sequential (%)	0.48	0.08	0.03	0.05	0.21	0.18	0.81	0.03	0.50

automated framework for rapid post-earthquake building evaluation. In the framework, a video was recorded from inside the damaged building and a search was conducted in each frame for the presence of columns [152] and then each column was assigned a damage index. The damage index [88] was estimated by classifying the column's failure mode into shear or flexure critical based on the location and severity of cracks, spalling, and exposed reinforcement identified by the methods proposed in Refs. [73,167]. The structural layout of the building is then manually recorded and then used to query a fragility database which provides information of the probability of being in a certain damage state.

Anil et al. [168] identified some information requirements to adequately represent visual damage information for structural walls in the aftermath of an earthquake and grouped them into five categories based on 17 damage parameters with varying damage sensitivity. This information was used to then describe a BIM-based approach in Ref. [169] to help automate the engineering

analyses introducing some heuristics to incorporate both strength analyses and visual damage assessment information. Wei et al. [170] provided a detailed review of the current status and persistent and emergent challenges related to 3D imaging techniques for construction and infrastructure management.

Hoskere et al. [128] used FCNs for the segmentation of damage and building component images to produce inspection-like semantic information. Three different networks were used: one for scene and building (SB) information, one to identify DP, and another to identify DT. The SB network had an average accuracy of 88.8% and the combined DP and DT networks together had an average accuracy of 91.1%. The proposed method was able to successfully identify the location and type of damage, together with some context about the SB; thus, this method is applicable in a much more general setting than prior work. Qualitative results for several images are shown in Fig. 14 [128], where the right column provides comments on successes and false positives and negatives.

**Fig. 14.** Qualitative results from Ref. [128].

3.4. Summary

Damage detection, change detection, and structural component recognition are key steps toward enabling automated structural inspections. While structural inspections provide valuable indicators to assess the condition of infrastructure, more quantitative measurement of structural responses is often required. Quantities such as displacement and strain can also be measured using vision-based techniques to enable structural condition assessment. The next section provides a review of civil infrastructure monitoring applications using vision-based techniques.

4. Monitoring applications

Monitoring is performed to obtain quantitative understanding of the current state of civil infrastructure through the measurement of physical quantities such as accelerations, strains, and/or displacements. Monitoring is often accomplished using wired or wireless contact sensors [171–178]. Although numerous applications can use contact sensors to gather data effectively, these sensors are often expensive to install and difficult to maintain. Vision-based techniques offer the advantages of non-contact methods, which potentially address some of the challenges of using contact sensors. As mentioned in Section 2, a key computer vision algorithm that can perform measurement tasks is optical flow, which estimates the translational motion of each pixel between two image frames [179]. Optical flow algorithms are general computer vision techniques which associate pixels in a reference image with corresponding pixels in another image of the same scene from a slightly different viewpoint by optimizing an objective function, such as sum of square difference (SSD) error, normalized cross correlation (NCC) criterion, global cost function [179], or combined local and global functionals [180,181]. A comparison of methods with different cost functions and optimization algorithms can be found in Ref. [182]. The rest of this section discusses studies of vision-based techniques for monitoring civil

infrastructure. The section is divided into two main subsections: ① static applications and ② dynamic applications.

4.1. Static applications

Measurement of static displacements and strains for civil infrastructure using vision-based techniques is often carried out using digital image correlation (DIC). According to Sutton et al. [183], DIC refers to “the class of non-contacting methods that acquire images of an object, store images in digital form, and perform image analysis to extract full-field shape, deformation, and/or motion measurements.” (p. 1) In addition to estimation of the displacement field in the image plane, DIC algorithms can include different post-processing steps to compute two-dimensional (2D) in-plane strain fields (2D DIC), out-of-plane displacement and strain fields (3D DIC), and volumetric measurement (VDIC). Highly reliable commercial DIC solutions are available today (e.g., VIC-2D™ [184] and GOM Correlate [185]). A detailed review of general DIC applications can be found in Refs. [183,186].

DIC methods have been applied for displacement and strain measurement in civil engineering. Hoult et al. [187] evaluated the performance of the 2D DIC technique using a steel specimen under uniaxial loading by comparing results with strain gage measurements (Fig. 15). The authors then proposed methods to compensate for the effect of out-of-plane deformation. The performance of the 2D DIC technique was also tested using steel and reinforced concrete beam specimens, for which theoretical values of strain and strain measurement data from strain gages were available [188]. In Ref. [189], a 3D DIC system was used as a reference to measure the static displacement of laboratory specimens. In these tests, sub-pixel accuracy for displacements was achieved, and the strain estimation agreed with the measurement by strain gages and the theoretical values.

DIC methods have also been applied to the field measurement of displacement and strain of civil structures. McCormick and Lord [190] applied the 2D DIC technique to vertical displacement measurement of a highway bridge deck statically loaded with four

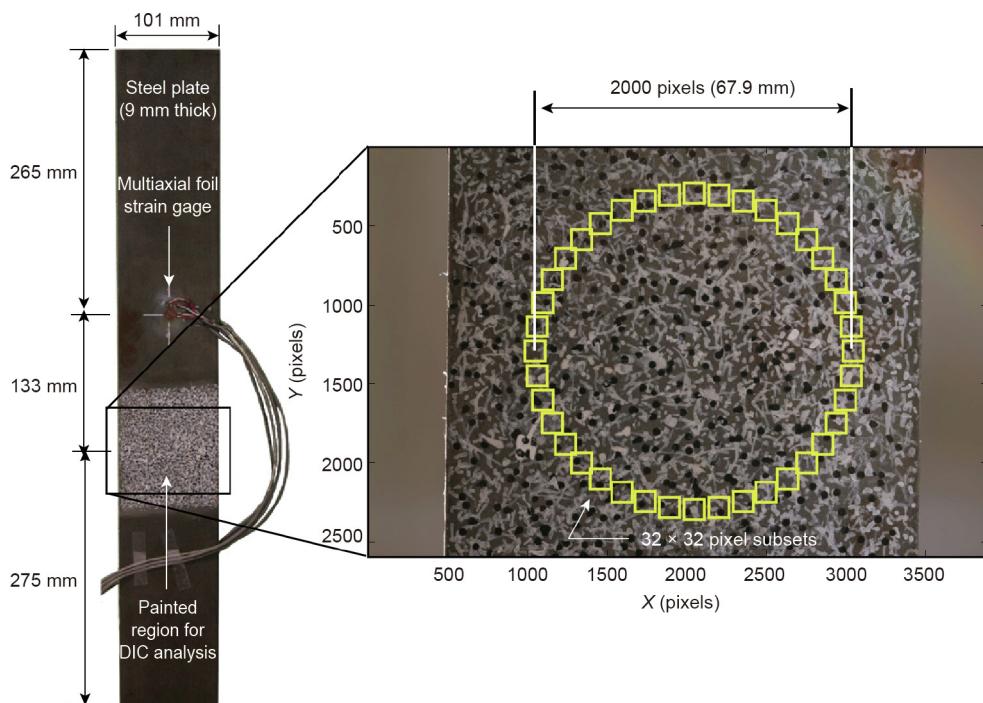


Fig. 15. A steel plate specimen used for uniaxial testing by Hoult et al. [187].

32-ton trucks. Yoneyama et al. [191] applied the 2D DIC technique to estimate the deflection of a bridge girder loaded by a 20-ton truck. The authors evaluated the accuracy of the deflection measurement with and without artificial patterns using data from displacement transducers. Yoneyama and Ueda [192] also applied the 2D DIC to bridge deflection measurement under operational loading. Helfrick et al. [193] used 3D DIC for full-field vibration measurement. Reagan [194] used a UAV carrying a stereo camera to apply 3D DIC to the long-term monitoring of bridge deformations.

Another promising civil engineering application of the DIC methods is crack mapping, in which 3D DIC is employed to extract cracked regions characterized by large strains. Mahal et al. [195] successfully extracted cracks on an RC specimen, and Ghorbani et al. [196] extended this crack-mapping method to a full-scale masonry wall specimen under cyclic loading (Fig. 16). The crack maps thus obtained are useful for analyzing laboratory test results, as well as for augmenting the information used for the structural inspection.

4.2. Dynamic applications

System identification and modal analysis are powerful tools for SHM, and can provide valuable information about the dynamic properties of structural systems. System identification and other SHM-related tasks are often accomplished using wired or wireless accelerometers due to the reliability of such sensors and their ease of installation [171–178]. Vision-based techniques offer the advantage of non-contact methods, as compared with traditional approaches. Combined with the proliferation of low-cost cameras in the market and increased computation capabilities, video-based methods have become convenient approaches for displacement measurement in structures. Several algorithms are available to accomplish displacement extraction; these work in principle by template matching or by tracking either the contours of the constant phase or intensity through time [55,197,198]. Optical flow methods have been used to measure dynamic and pseudo-static responses for several applications, including system identification, modal analysis, model updating, and direct indication of changes in serviceability based on thresholds. For further information, recent reviews of dynamic monitoring applications using computer vision techniques have been provided by Ye et al. [199] and Feng and Feng [200]. This section discusses studies that have proposed and/or evaluated different algorithms for displacement measurement with ① laboratory experiments and ② field validation.

4.2.1. Laboratory experiments

Early optical flow algorithms used for dynamic monitoring focused on natural frequency estimation [201] and displacement measurement [202,203]. Markers are commonly used to obtain streamlined and precise detection and tracking of points of interest. Min et al. [204] designed high-contrast markers to be used for displacement measurement with smartphone devices and

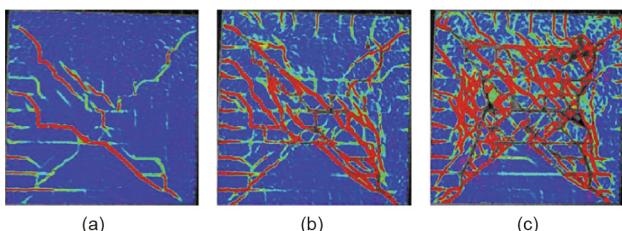


Fig. 16. Crack maps created using 3D DIC [196]. (a) First crack; (b) peak load; (c) ultimate state. Red corresponds to $+3000 \mu\text{m}\cdot\text{m}^{-1}$.

telephoto lenses, and obtained excellent results in laboratory testing (Fig. 17).

Dong et al. [205] proposed a method for the multipoint synchronous measurement of dynamic structural displacement. Celik et al. [206] evaluated different vision-based techniques to measure human loads on structures. Lee et al. [207] presented an approach for displacement measurements, which was tailored for field testing with enhanced robustness to strong sunlight. Park et al. [208] demonstrated the efficacy of the data fusion of vision-based images with accelerations to extend the dynamic range and lower signal noise. The application of vision algorithms has also been extended to the system identification of laboratory structures. Schumacher and Shariati [209] introduced the concept of virtual visual sensors that could be used for the modal analysis of a structure. Yoon et al. [210] implemented a Kanade–Lucas–Tomasi (KLT) tracker to identify a model for a laboratory-scale six-story building model (Fig. 18). Ye et al. [211] conducted multipoint displacement measurement on a scale-model arch bridge and validated the measurements using linear variable differential transformers (LVDTs). Abdelbarr et al. [212] measured 3D dynamic displacements using an inexpensive RGB-D sensor. Ye et al. [213] conducted studies on a shake table to determine the factors that influence system performance for vision-based measurements. Feng and Feng [214] implemented a template tracking approach using up-sampled cross-correlations to obtain displacements at multiple points on a vibrating structure. System identification has also been carried out on laboratory structures using vision data captured by UAVs [215]. The same authors proposed a method to measure dynamic displacements using UAVs with a stationary reference in the background.

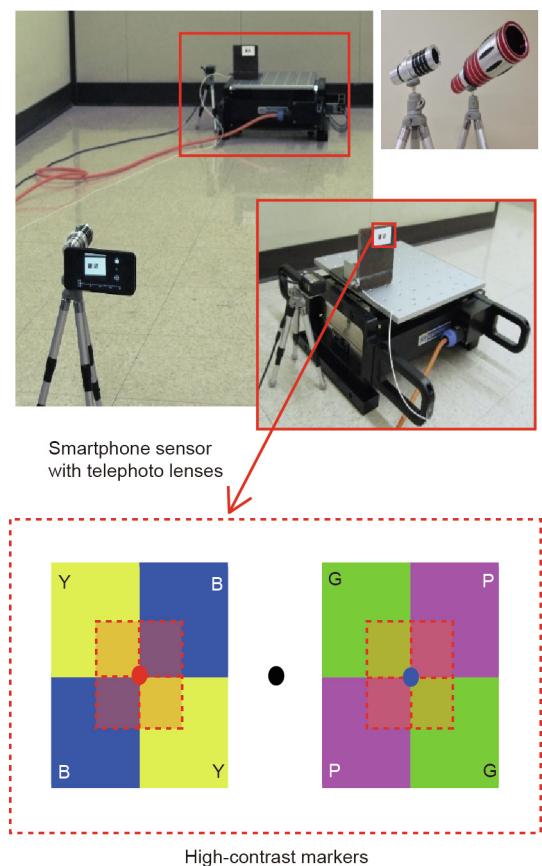


Fig. 17. A smartphone-based displacement measurement system including telephoto lenses and high-contrast markers proposed by Min et al. [204]. B: blue; G: green; P: pink; Y: yellow.

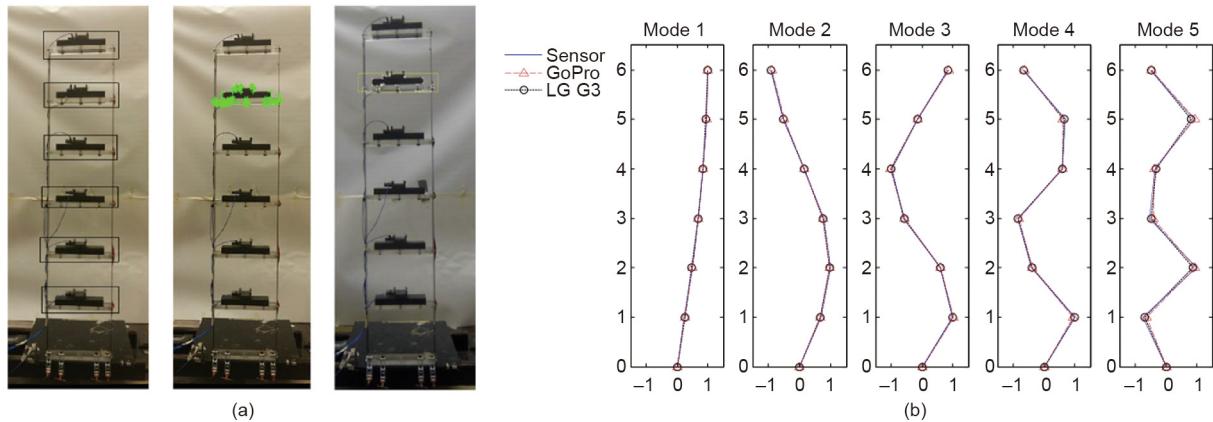


Fig. 18. The target-free approach for vision-based structural system identification using consumer-grade cameras [210]. (a) Screenshots of target tracking; (b) extracted mode shapes from different sensors. GoPro and LG G3 are the cameras used in the test.

Wadhwa et al. [55] proposed a technique called motion magnification that band-passes videos with small deformations in order to extract and amplify motion at certain frequencies. The procedure involves decomposing the video at multiple scales, applying a filter to each, and then recombining the filtered spatial bands. Subsequently, a number of papers have appeared on full-field modal analysis of structures using vision-based methods inspired by motion magnification. Chen et al. [216] successfully applied motion magnification to visualize operational deflection shapes of laboratory structures (Fig. 19). Cha et al. [217] used a phase-based approach together with the unscented Kalman filters for system identification using noisy displacement measurements. Yang et al. [218] adapted the method to include blind source separation with multiscale wavelet filters, as opposed to complex steerable filters, in order to obtain the full-field modes of a laboratory specimen. Yang et al. [219] proposed a method for the high-fidelity and realistic simulation of structural response using a high-spatial resolution modal model and video manipulation.

4.2.2. Field validation

The success of vision-based vibration measurement techniques in the laboratory has led to a number of field applications over the

past few years. The most common application has been to measure displacements on full-scale bridge structures [220–223], including the measurement of different components such as bridge decks, trusses, and hangar cables. Phase-based methods have also been applied for the displacement and frequency estimation of an antenna tower, and to obtain partial mode shapes of a truss bridge structure [224].

Some researchers have investigated the use of vision sensors to measure displacements at multiple points on a structure. Yoon [225] used a camera system to measure railroad bridge displacement during train crossings. As shown in Fig. 20 [225], the measured displacement was very close to the values predicted from a finite-element model that used the train load as input; the differences were primarily attributed to the train not having a constant speed. Mas et al. [226] developed a method for the simultaneous multipoint measurement of vibration frequencies through the analysis of high-speed video sequences, and demonstrated their algorithm on a steel pedestrian bridge.

An interesting application of load estimation was studied by Chen et al. [227], who used computer vision to identify the distribution of vehicle loads across a bridge in both space and time by automatically detecting the type of vehicle entering the bridge and combining that with information from the weigh-in-motion system at one cross section of the bridge. The developed system enables accurate load determination for the entire bridge, as compared with previous methods which can measure the vehicle loads at only one cross section of a bridge.

The use of computer vision techniques for the system identification of structures has been somewhat limited; obtaining measurements of all points on a large structure within a single video frame usually results in a pixel resolution that is insufficient to yield accurate structural displacements. Moreover, finding a good vantage point to place a camera also proves difficult in urban environments, resulting in video data that contains perspective and atmospheric distortions due to monitoring from afar with zoom lenses [216]. Finally, when using a remote camera, only points on the structure that are readily visible from the selected camera location can be monitored. Isolating modal information from video data often involves the selection of manually generated masks or regions of interest, which makes the process cumbersome.

Xu et al. [222] proposed a low-cost and non-contact vision-based system for multipoint displacement measurement based on a consumer-grade camera for video acquisition, and were able to obtain the mode shapes of a pedestrian bridge using the proposed system. Hoskere et al. [228] developed a divide-and-conquer approach to acquire the mode shapes of

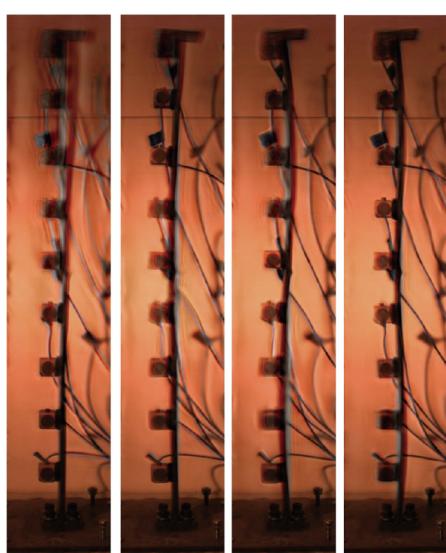


Fig. 19. Screenshots of a motion-magnified video of a vibrating cantilever beam from Ref. [216].

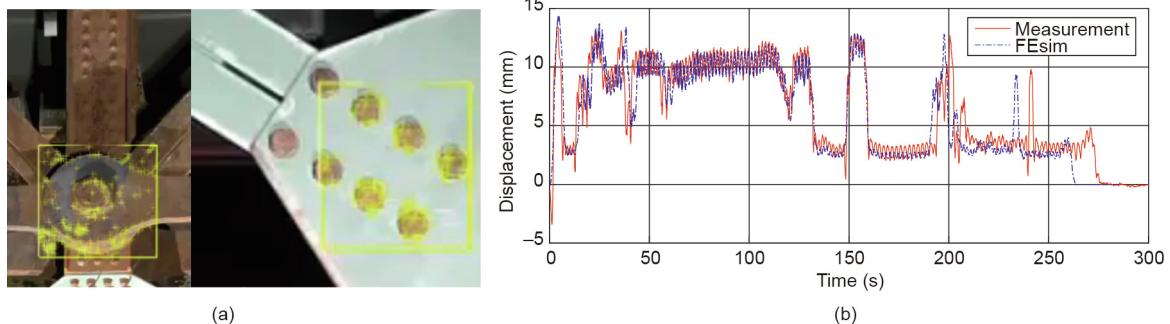


Fig. 20. Railroad bridge displacement measurement using computer vision [225]. (a) Images of the optical tracking of the railroad component; (b) comparison of vision-based displacement measurement to FE simulation estimation. FEsim: finite-element simulation.

full-scale infrastructure using UAVs. The proposed approach directly addressed a number of difficulties associated with the modal analysis of full-scale infrastructure using vision-based methods. Initial evaluation was carried out using a six-story shear-building model excited on a shaking table in a laboratory environment. Field tests of a full-scale pedestrian suspension bridge were subsequently conducted to obtain natural frequencies and mode shapes (Fig. 21) [151].

5. Challenges to the realization of vision-based automated inspection and monitoring of civil infrastructure

Although the research community has made significant progress in recent years, a number of technical hurdles must be crossed before the use of vision-based techniques for automated

SHM can be fully realized. The key difficulties broadly lie in converting the features and signals extracted by vision-based methods into more actionable information that can aid in higher level decision-making.

5.1. Automated structural inspections necessitate integral understanding of damage and context

Humans performing visual inspections have remarkable perception capabilities that are still very difficult for vision and deep learning algorithms to replicate. Trained human inspectors are able to identify regions of importance to the overall health of the structure (e.g., critical structural components, the presence of structurally significant damage, etc.). When a structure is damaged, depending on the shape, size, and location of the damage, and the type and importance of the component in which it is occurring,

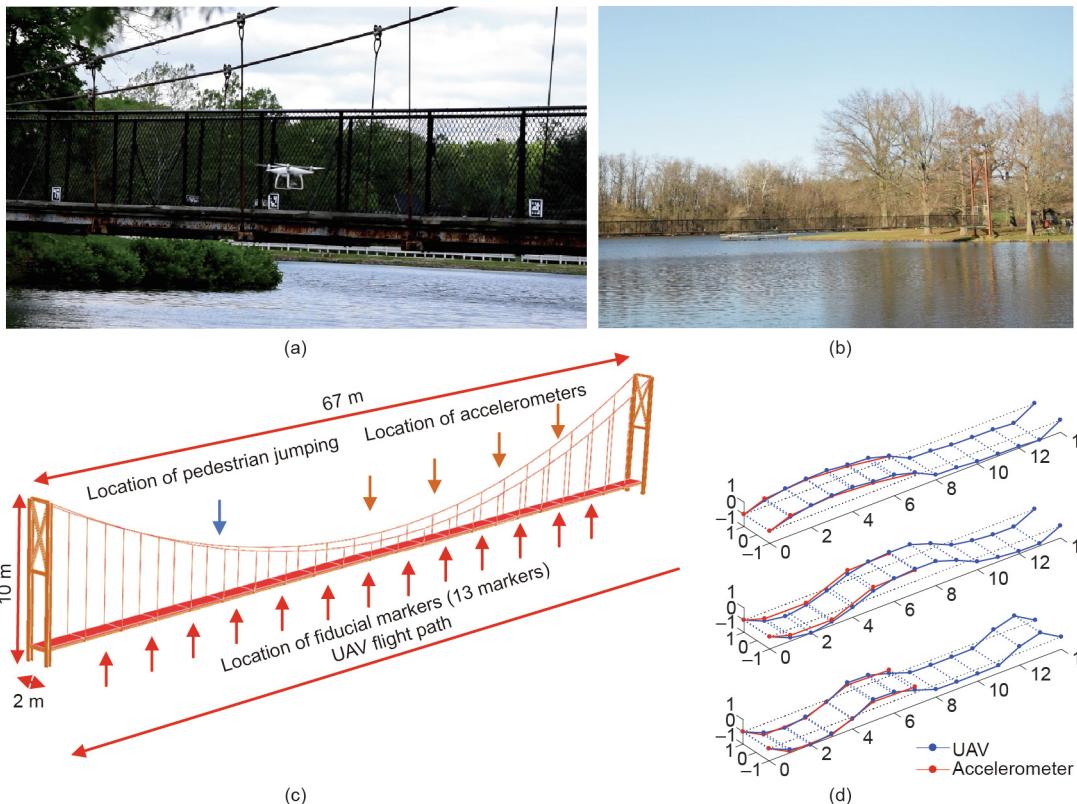


Fig. 21. (a) Image of Phantom 4 recording video of a vibrating bridge (3840×2160 pixels at 30 fps); (b) Lake of the Woods pedestrian bridge, Mahomet, IL; (c) finite-element model of the bridge; (d) extracted mode shapes [151].

a trained inspector can infer the structural importance of the damage identified. The inspector can also understand the implications the presence of multiple kinds of damage. Thus, while significant progress has been made in this field, higher accuracy damage detection and component recognition are still needed. Moreover, studies on interpreting the structural significance of identified damage and assimilating local information with global information to make structure-level assessments are almost completely absent from the literature. Addressing these challenges will be vital for the realization of fully automated vision-based inspections.

5.2. The generality of deep networks depends on the generality of data

Trained DCNN models tend to perform poorly if the features extracted from the data on which inference is conducted differ significantly from the training data. Thus, the quality of a trained deep model is directly dependent on the underlying dataset. The perception capabilities of DCNN models that have not been trained to be robust to damage-like features, such as grooves or joints, cannot be relied upon to distinguish between such textures during inference. The limited number of datasets available for the detection of structural damage is a challenge that must be overcome in order to advance the perception capabilities of DCNNs for automated inspections.

5.3. Human-like perception for inspections requires an understanding of sequential views

A single image does not always provide sufficient context for performing damage detection and component-recognition tasks. For example, damage recognition is most likely to be successful when the image is a close-up view of a component; however, component recognition from such images is very difficult. In an extreme case, the inspector may be so close to the component that concrete columns would be indistinguishable from concrete beams or concrete walls. During inspection by humans, this problem is easily resolved by first examining the entire structure, and then moving close to the structural components while keeping in mind the target structural component. To replicate this human function, the viewing sequence (e.g., using video data) must be incorporated into the process and the recognition tasks must be performed based on the current frame as well as on previous frames.

5.4. Displacements are often small and difficult to capture

For monitoring applications, recent work has been successful in demonstrating the feasibility of vision-based methods for measuring modal information, as well as displacements and strains of structures both in the laboratory and in the field. On the other hand, accurate displacement and strain measurement for *in situ* civil infrastructure is rarely straightforward. The displacement and strain ranges expected during field tests are often smaller than those in laboratory tests, because the target structures in the field are responding to operational loading. In a field environment, the accessibility of the structural components of interest is often limited. In such cases, optimal camera locations for high-quality measurement cannot be realized, and markers guiding the displacement measurement cannot be placed. For static applications, surface textures are often added artificially (e.g., speckle patterns) to help with the image-matching step in DIC methods [183], which is also difficult for structures with limited accessibility. Further research and development effort is needed both in terms of hardware and software in order to apply vision-based static displacement/strain measurement in such operational situations.

5.5. Lighting and environmental effects

Vision-based methods are highly susceptible to changes in visibility-related environmental effects such as the presence of rain, mist, or fog. While the abovementioned problems are difficult ones to circumvent, other environmental factors such as changes in lighting, shadows, and atmospheric interference are amenable to normalization, although more work is required to improve robustness.

5.6. Big data needs big data management

The realization of continuous and automated vision-based monitoring poses challenges regarding the massive amount of data produced, which becomes difficult to store and process for long-term applications. Automated real-time signal extraction will be necessary to reduce the amount of data that is stored. Methods to handle and process full-field modal information obtained from video-bandpassing techniques are also an open area of research.

6. Ongoing work toward automated inspections

Toward the goal of automated inspections, vision-based perception is still an open research problem that requires a great deal of attention. This section discusses ongoing work at the University of Illinois aimed at addressing the following challenges, which were outlined in Section 5: ① incorporating context to generate condition-aware models; ② generating synthetic, labeled data using photorealistic physics-based graphics models to address the need for more general data; and ③ leveraging video sequences for human-like recognition of structural components.

6.1. Incorporating context to generate condition-aware models

As discussed in Section 5.1, understanding the context in which damage is occurring is a key aspect of being able to make automated and high-level assessments that render detailed inspection judgements. To address this issue, Hoskere et al. [128] proposed a new procedure in which information about the type of structure, its various components, and the condition of each of the components were combined into a single model, referred to as a condition-aware model. Such models can be viewed as analogous to the as-built models used in the construction and design industry, but for the purpose of inspection and maintenance. Condition-aware models are automatically generated annotations showing the presence of visual defects on the structure. Depending on the particular inspection application being considered, the fidelity of the condition-aware model required also varies. The main advantage of building a condition-aware model, as opposed to just using images directly, is that the context of the structure and the scale of the damage are easily identifiable. Moreover, the availability of global 3D geometry information can aid the assessment process. The model serves as convenient entity to rapidly and automatically document visually identifiable defects on the structure.

The framework proposed by Hoskere et al. [128] for the generation of condition-aware models for rapid automated post-disaster inspections is shown in Fig. 22. A 3D mesh model is generated using multi-view stereo from a UAV survey of the structure. Deep learning-based condition inference is then conducted on the same set of images for the semantic segmentation of the damage and building context. The generated labels are projected onto the mesh using UV mapping (a 3D modeling process of projecting a 2D image to a 3D model), which results in a condition-aware model with averaged damage and context labels superimposed

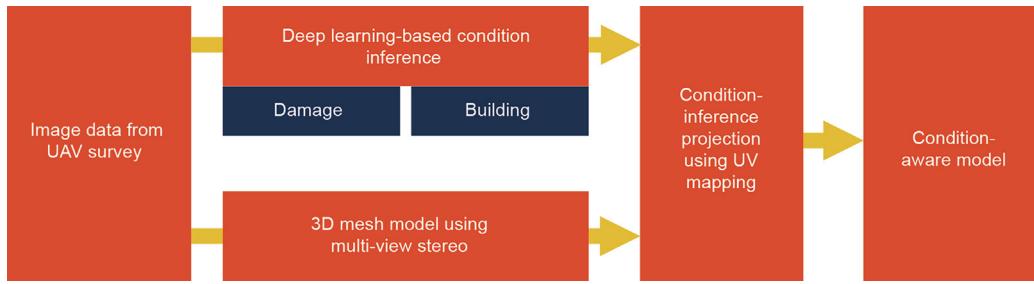


Fig. 22. A framework to generate condition-aware models for rapid post-disaster inspections.

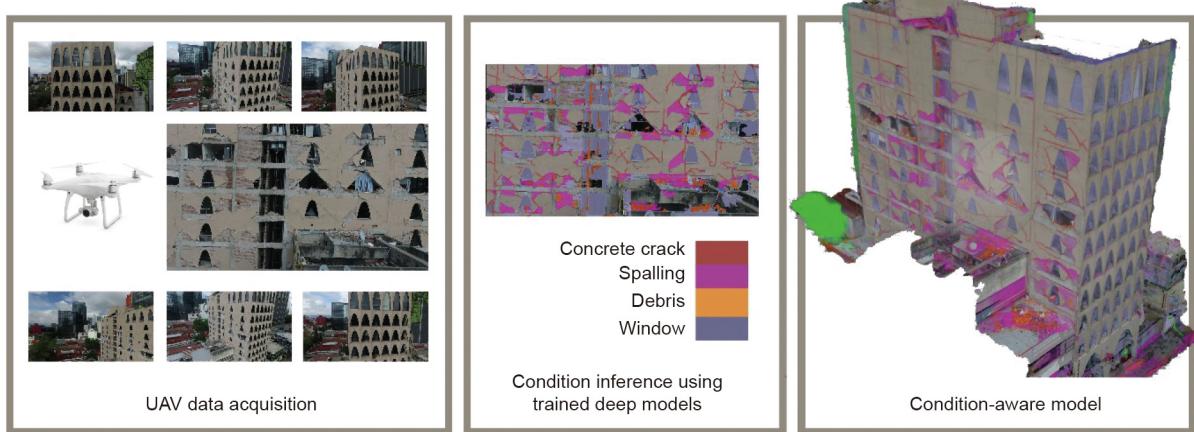


Fig. 23. A condition-aware model of a building damaged during the central Mexico earthquake in September 2017 [128].

on each cell. Fig. 23 shows the condition-aware model that was developed using this procedure for a building damaged during the central Mexico earthquake in September 2017.

6.2. Generating synthetic labeled data using photorealistic physics-based graphics models

As discussed in Section 5.2, for deep learning techniques aimed at automated inspections, the lack of vast quantities of labeled data makes it difficult to generalize training models across a wide variety of structures and environmental conditions. Each civil engineering structure is unique, which makes the identification of damage more challenging. For example, buildings could be painted in a wide range of colors (a parameter that will almost certainly have an impact in the damage detection result, especially for corrosion); thus, developing a general algorithm for damage detection is difficult without accounting for such issues. To compound the problem, high-quality data from damaged structures is also relatively difficult to obtain, as damaged structures are not a common occurrence.

The significant progress that has been made in computer graphics over the past decade has allowed the creation of photorealistic images and videos. Here, synthetic data refers to data generated from a graphics model, as opposed to a camera in the real world. Synthetic data has recently been used in computer vision to train deep neural networks for the semantic segmentation of urban scenes, and models trained on synthetic data have shown promising performance on real data [229]. The use of synthetic data offers many benefits. Two types of platforms are available to produce synthetic data: ① real-time game engines that use rasterization to render graphics images at low computational cost, at the expense of accuracy and realism; and ② renderers that use physics-based ray tracing engines for accurate simulations of light and materials in order to produce realistic images, at a high

computational cost. The generation of synthetic data can help to solve the problem of data labeling, as any data from algorithmically generated graphics models will be automatically labeled, both at the pixel and image levels. Graphics models can also provide a testbed for vision algorithms with readily repeatable conditions. Different environmental conditions, such as lighting, can be simulated, and algorithms can be studied with different camera parameters and UAV data-acquisition flight paths. Algorithms that are effective in these virtual testbeds will be more likely to work on real-world datasets.

3D modeling, simulation, and rendering tools such as Blender [230] can be used to better simulate real-world environmental effects. Combined with deformed meshes from finite-element models, these tools can be used to create graphics models of damaged structures. Understanding the damage condition of a structure requires context awareness. For example, identical cracks at different locations on the same structure could have different implications for the overall health of a structure. Similarly, cracks in bridge columns must be treated differently than cracks in a wall of a building. Hoskere et al. [231] proposed a novel framework (Fig. 24) using physics-based models of structures to create synthetic graphics images of representative damaged structures. The proposed framework has five main steps: ① the use of parameterized finite-element models to structurally model representative structures of various shapes, sizes, and materials; ② nonlinear finite-element analysis to identify structural hotspots on the generated models; ③ the application of material graphic properties for realistic rendering of the generated model; ④ procedural damage generation using hotspots from finite-element models; and ⑤ the training of deep learning models for assessment using generated synthetic data.

Physics-based graphics models can be used to generate a wide variety of damage scenarios. As the generated data will be similar

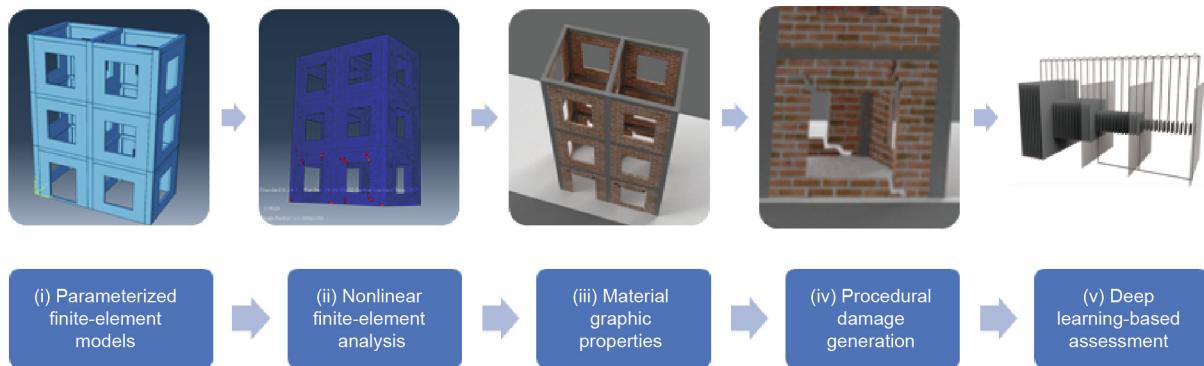


Fig. 24. Framework for physics-based graphics generation for automated assessment using deep learning [231].

to real data, the limits of deep learning methods to identify important damage and structure features can be established. These models provide high-quality labeled data at multiple levels of context, including: ① global structure properties, such as the number of stories and bays, and the structural system; ② structural and non-structural components and critical regions; and ③ the presence of different types of local and global damage, such as cracks, spalling, soft stories, and buckled columns, along with other hazards such as falling parts. This higher level of context information can enable reliable automated inspections where approaches trained on images of local damage have failed. The inherent subjectivity of onsite human inspectors can be greatly reduced through the use of objective physics-based models as the underlying basis for the training data, rather than the use of subjective hand-labeled data.

Research on the use of synthetic data for vision-based inspection applications has been limited. Hoskere et al. [232] created a physics-based graphics model of a miter gate and trained deep semantic segmentation to identify important changes on the gate in the synthetic environment. Training data for the networks was generated using physics-based graphics models, and included defects such as cracks and corrosion, while accommodating variations in lighting (Fig. 25).

Research is currently under way to transfer the applicability of successful deep learning models trained on synthetic data to work with real data as well.

6.3. Leveraging video sequences for human-like recognition of structural components

Human inspectors first survey an entire structure, and then move closer to make detailed assessments of damaged structural components. When performing a detailed inspection, they keep in mind how the damaged component fits into the global structural context; this is essential in order to understand the relevance of the damage to structural safety. However, as discussed in Section 5.3, computer vision strategies for damage detection usually operate on a frame-by-frame basis—that is, independently using single images; close-up images do not contain the necessary information about global structural context. For the human inspector, the viewing history (i.e., sequences of dependent images such as in videos) provides this context. This section discusses the incorporation of the viewing history embedded in video sequences in order to achieve more accurate structural component recognition throughout the inspection process.

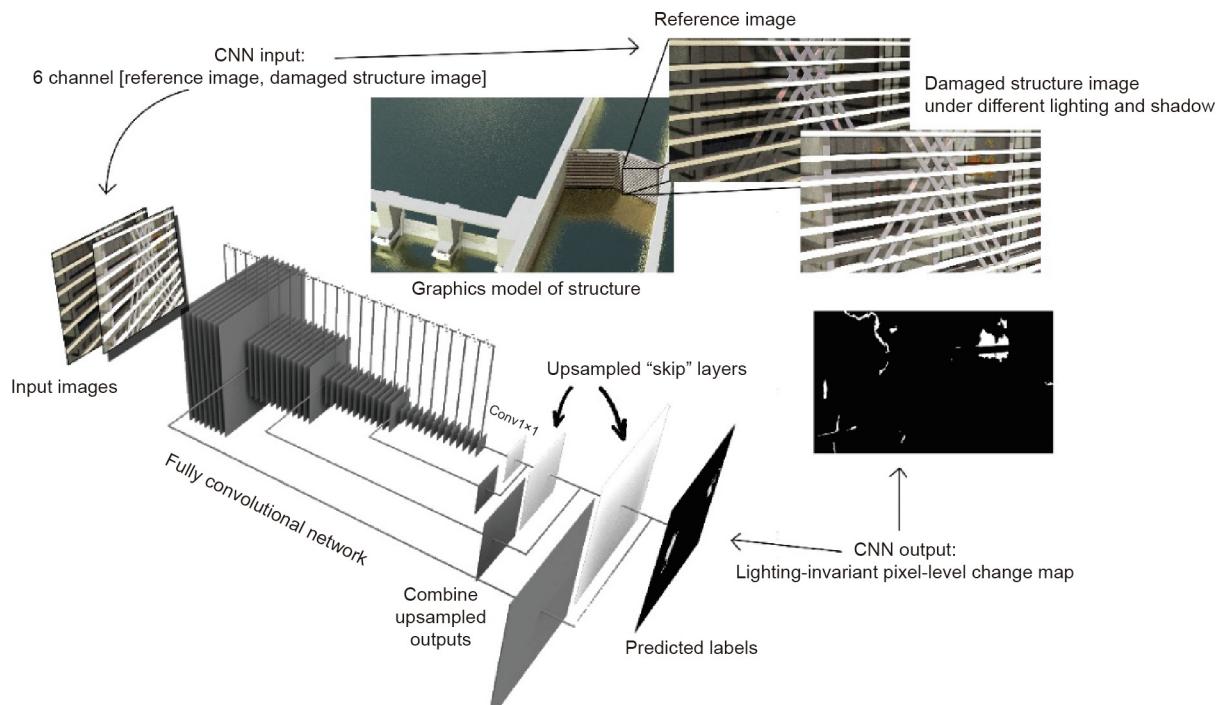


Fig. 25. Deep learning-based change detection [232].

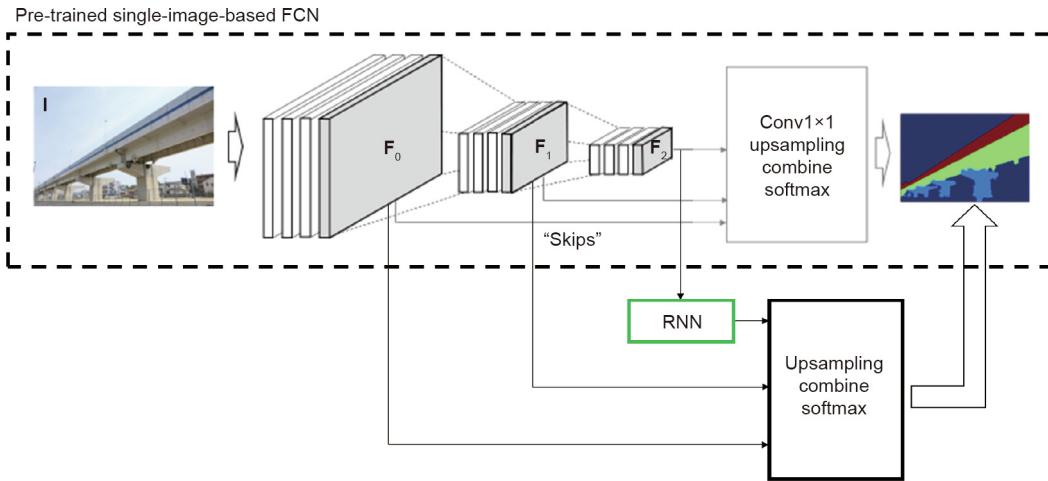


Fig. 26. Illustration of the network architecture used in Ref. [186].

Narazaki et al. [233] applied recurrent neural networks (RNNs) to implement bridge-component recognition using video data, which incorporates views of the global structure and close-up details of structural component surfaces. The network architecture used in the study is illustrated in Fig. 26. First, a deep single image-based FCN is applied to extract a map of label predictions. Next, three smaller RNN layers are added after the lowest resolution prediction layer. Finally, the output from the RNN layers and other skipped layers with higher resolution are combined to generate the final estimated label maps. The RNN units are inserted only after the lowest resolution prediction layer, because the RNN units in the study are used to memorize where the video is focused, rather than improving the level of detail of the estimated map.

Two types of RNN units were tested in the study [186]: simple RNN and convolutional long short-term memory (ConvLSTM) [234] units. For the simple RNN units, the input to the unit was augmented by the output of the unit at the previous time step, and the convolution with the ReLU activation function was applied. Alternatively, ConvLSTM units were inserted into the RNN of the architecture to model long-term patterns effectively.

One of the major challenges of training and testing RNNs for video processing is to collect video data and their ground-truth labels. In particular, manually labeling every frame of video data is impractical. Following the discussion in Section 6.2 on the benefits of synthetic data, the real-time rendering capabilities of the game engine Unity3D [235] was used in the study [186] to address the challenge. A video dataset was created from a simulation of a UAV navigating around a concrete-girder bridge. The steps used to create the dataset were similar to the steps used to create the SYNTHIA dataset [229]. However, this dataset randomly navigates in 3D space with a larger variety of headings, pitches, and altitudes. The resolution of the video was set to 240×320 , and 37 081 training images and 2000 test images were automatically generated along with the corresponding ground-truth labels. Example frames of the video are shown in Fig. 27. The depth map was also retrieved, although those data were not used for the study.

The example results in Fig. 28 show the effectiveness of the recurrent units when the FCN fails to recognize the bridge components correctly. These results indicate that ConvLSTM units combined with pre-trained FCN is an effective approach to perform automated bridge-component recognition, even if the visual cues of the global structures are temporally unavailable. The total pixel-wise accuracy for the single-image-based FCN was 65.0%. In contrast, the total pixel-wise accuracy for the simple RNN and the ConvLSTM units was 74.9% and 80.5%, respectively. Other

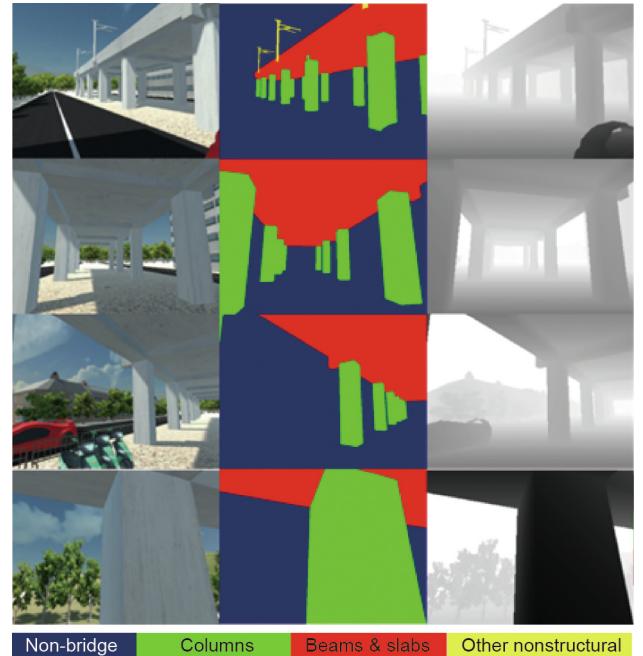


Fig. 27. Example frames of the new video dataset with ground-truth labels and ground-truth depth maps [233].

details of the dataset, training, and testing are provided in Ref. [233].

At present, this research is being leveraged to develop rapid post-earthquake inspection strategies for transportation infrastructure.

7. Summary and conclusions

This paper presented an overview of recent advances in computer vision-based civil infrastructure inspection and monitoring. Manual visual inspection is currently the main means of assessing the condition of civil infrastructure. Computer vision for civil infrastructure inspection and monitoring is a natural step forward and can be readily adopted to aid and eventually replace manual visual inspection, while offering new advantages and opportunities. However, the use of image data can be a double-edged sword; as although rich spatial, textural, and contextual information is

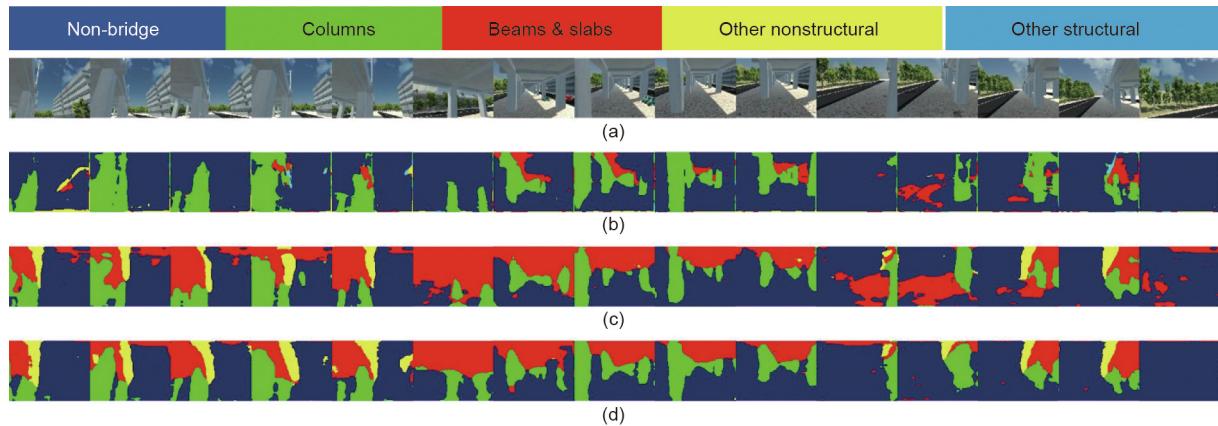


Fig. 28. Example results. (a) Input images; (b) FCN; (c) FCN–simple RNN; (d) FCN–ConvLSTM [233].

available in each image, the process of extracting actionable information from these images is challenging. The research community has been successful in demonstrating the feasibility of vision algorithms, ranging from deep learning to optical flow. The inspection applications discussed in this article were classified into the following three categories: characterizing local and global visible damage, detecting changes from a reference image, and structural component recognition. Recent progress in automated inspections has stemmed from the replacement of heuristic-based methods with data-driven detection, where deep models are built by training on large sets of data. The monitoring applications discussed spanned both static and dynamic applications. The application of full-field measurement techniques and the extension of laboratory techniques to full-scale infrastructure has provided impetus for further growth.

This paper also presented key challenges that the research community faces toward the realization of automated vision-based inspection and monitoring. These challenges broadly lie in converting the features and signals extracted by vision-based methods into actionable data that can aid decision-making at a higher level.

Finally, three areas of ongoing research aimed at enabling automated inspections were presented: the generation of condition-aware models, the development of synthetic data through graphics models, and methods for the assimilation of data from video. The rapid advances in research in computer vision-based inspection and monitoring of civil infrastructure described in this paper will enable time-efficient, cost-effective, and eventually automated civil infrastructure inspection and monitoring, heralding a coming revolution in the way that infrastructure is maintained and managed, ultimately leading to safer and more resilient cities throughout the world.

Acknowledgements

This research was supported in part by funding from the US Army Corps of Engineers under a project entitled “Cybermodeling: A Digital Surrogate Approach for Optimal Risk-Based Operations and Infrastructure” (W912HZ-17-2-0024).

Compliance with ethics guidelines

Billie F. Spencer Jr., Vedhus Hoskere, and Yasutaka Narazaki declare that they have no conflict of interest or financial conflicts to disclose.

References

- [1] ASCE's 2017 infrastructure report card: bridges [Internet]. Reston: American Society of Civil Engineers; 2017 [cited 2017 Oct 11]. Available from: <https://www.infrastructurereportcard.org/cat-item/bridges/>.
- [2] R.P. Gallagher Associates, Inc. ATC-20-1 Field manual: postearthquake safety evaluation of buildings (second edition). Redwood City: Applied Technology Council; 2005.
- [3] Federal Highway Administration, Department of Transportation. National bridge inspection standards. Fed Regist [Internet] 2004 Dec [cited 2018 Jul 30];69(239):74419–39. Available from: <https://www.govinfo.gov/content/pkg/FR-2004-12-14/pdf/04-27355.pdf>.
- [4] Corps of Engineers inspection team goes to new heights. Washington, DC: US Army Corps of Engineers; 2012.
- [5] Ou J, Li H. Structural health monitoring in mainland China: review and future trends. *Struct Health Monit* 2010;9(3):219–31.
- [6] Carden EP, Fanning P. Vibration based condition monitoring: a review. *Struct Health Monit* 2004;3(4):355–77.
- [7] Brownjohn JMW. Structural health monitoring of civil infrastructure. *Philos Trans A Math Phys Eng Sci* 1851;2007(365):589–622.
- [8] Li HN, Ren L, Jia ZG, Yi TH, Li DS. State-of-the-art in structural health monitoring of large and complex civil infrastructures. *J Civ Struct Heal Monit* 2016;6(1):3–16.
- [9] Roberts LG. Machine perception of three-dimensional solids [dissertation]. Cambridge: Massachusetts Institute of Technology; 1963.
- [10] Postal mechanization and early automation [Internet]. Washington, DC: United States Postal Service; 1956 [cited 2018 Jun 22]. Available from: https://about.usps.com/publications/pub100/pub100_042.htm.
- [11] Szeliski R. Computer vision: algorithms and applications. London: Springer; 2011.
- [12] Viola P, Jones MJ. Robust real-time face detection. *Int J Comput Vis* 2004;57(2):137–54.
- [13] Turk MA, Pentland AP. Face recognition using eigenfaces. In: Proceedings of 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 1991 Jun 3–6; Maui, HI, USA. Piscataway: IEEE; 1991. p. 586–91.
- [14] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2005 Jun 20–25; San Diego, CA, USA. Piscataway: IEEE; 2005. p. 886–93.
- [15] Pingali G, Opalach A, Jean Y. Ball tracking and virtual replays for innovative tennis broadcasts. In: Proceedings of the 15th International Conference on Pattern Recognition; 2000 Sep 3–7; Barcelona, Spain. Piscataway: IEEE; 2000. p. 152–6.
- [16] Bishop CM. Pattern recognition and machine learning. New York City: Springer-Verlag; 2006.
- [17] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: Gordon G, Dunson D, Dudík M, editors. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics; 2011 Apr 11–13; Fort Lauderdale, FL, USA; 2011. p. 315–23.
- [18] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986;323(6088):533–6.
- [19] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations; 2015 May 7–9; San Diego, CA, USA; 2015. p. 1–15.
- [20] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86(11):2278–324.
- [21] Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. *Multimed Tools Appl* 2018;77(9):10437–53.
- [22] Karpathy A. t-SNE visualization of CNN codes [Internet]. [cited 2018 Jul 27]. Available from: <https://cs.stanford.edu/people/karpathy/cnembed/>.

- [23] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition; 2015 Jun 7–12; Boston, MA, USA. Piscataway: IEEE; 2015. p. 3431–40.
- [24] Farabet C, Couprie C, Najman L, LeCun Y. Learning hierarchical features for scene labeling. *IEEE Trans Pattern Anal Mach Intell* 2013;35(8):1915–29.
- [25] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39(12):2481–95.
- [26] Couprie C, Farabet C, Najman L, LeCun Y. Indoor semantic segmentation using depth information. arXiv:1301.3572. 2013.
- [27] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer vision—ECCV 2014. Cham: Springer; 2014. p. 346–61.
- [28] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. Piscataway: IEEE; 2014. p. 580–7.
- [29] Girshick R. Fast R-CNN. In: Proceedings of 2015 IEEE International Conference on Computer Vision; 2015 Dec 7–13; Santiago, Chile. Piscataway: IEEE; 2015. p. 1440–8.
- [30] Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017;39(6):1137–49.
- [31] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 779–88.
- [32] He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In: Proceedings of 2017 IEEE International Conference on Computer Vision; 2017 Oct 22–29; Venice, Italy. Piscataway: IEEE; 2017. p. 2980–8.
- [33] De Brabandere B, Neven D, Van Gool L. Semantic instance segmentation with a discriminative loss function. arXiv:1708.02551. 2017.
- [34] Zhang Z, Schwing AG, Fidler S, Urtasun R. Monocular object instance segmentation and depth ordering with CNNs. arXiv:1505.03159. 2015.
- [35] Werbos PJ. Backpropagation through time: what it does and how to do it. Proc IEEE 1990;78(10):1550–60.
- [36] Perazzi F, Khoreva A, Benenson R, Schiele B, Sorkine-Hornung A. Learning video object segmentation from static images. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21–26; Honolulu, HI, USA. Piscataway: IEEE; 2017. p. 2663–72.
- [37] Nilsson D, Sminchisescu C. Semantic video segmentation by gated recurrent flow propagation. arXiv:1612.08871. 2016.
- [38] Labelme.csail.mit.edu [Internet]. Cambridge: MIT Computer Science and Artificial Intelligence Laboratory; [cited 2018 Aug 1]. Available from: <http://labelme.csail.mit.edu/Release3.0/>.
- [39] Hess MR, Petrovic V, Kuester F. Interactive classification of construction materials: feedback driven framework for annotation and analysis of 3D point clouds. In: Hayes J, Quimet C, Santana Quintero M, Fai S, Smith L, editors. Proceedings of ICOMOS/ISPRS International Scientific Committee on Heritage Documentation (volume XLII-2/W5); 2017 Aug 28–Sep 1; Ottawa, ON, Canada; 2017. p. 343–7.
- [40] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 2921–9.
- [41] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of 2017 IEEE International Conference on Computer Vision; 2017 Oct 22–29; Venice, Italy. Piscataway: IEEE; 2017. p. 618–26.
- [42] DeGol J, Golparvar-Fard M, Hoiem D. Geometry-informed material recognition. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 1554–62.
- [43] Hinton GE. Training products of experts by minimizing contrastive divergence. *Neural Comput* 2002;14(8):1771–800.
- [44] Salakhutdinov R, Larochelle H. Efficient learning of deep Boltzmann machines. In: Teh YW, Titterington M, editors. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics; 2010 May 13–15; Sardinia, Italy; 2010. p. 693–700.
- [45] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006;313(5786):504–7.
- [46] Pathak D, Krähenbühl P, Donahue J, Darrell T, Efros AA. Context encoders: feature learning by inpainting. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 2536–44.
- [47] Zhang R, Isola P, Efros AA. Split-brain autoencoders: unsupervised learning by cross-channel prediction. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21–26; Honolulu, HI, USA. Piscataway: IEEE; 2017. p. 645–54.
- [48] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434. 2015.
- [49] Denton E, Chintala S, Szlam A, Fergus R. Deep generative image models using a Laplacian pyramid of adversarial networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems; volume 1; 2015 Dec 7–12; Montreal, QC, Canada. Cambridge: MIT Press; 2015. p. 1486–94.
- [50] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: Proceedings of the 30th International Conference on Neural Information Processing Systems; 2016 Dec 5–10; Barcelona, Spain. New York City: Curran Associates Inc.; 2016. p. 2234–42.
- [51] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [52] Barron JL, Fleet DJ, Beauchemin SS. Performance of optical flow techniques. *Int J Comput Vis* 1994;12(1):43–77.
- [53] Richardson IEG. H.264 and MPEG-4 video compression: video coding for next-generation multimedia. Chichester: John Wiley & Sons, Ltd.; 2003.
- [54] Grundmann M, Kwatra V, Han M, Essa I. Efficient hierarchical graph-based video segmentation. In: Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2010 Jun 13–18; San Francisco, CA, USA. Piscataway: IEEE; 2010. p. 2141–8.
- [55] Wadhwa N, Rubinstein M, Durand F, Freeman WT. Phase-based video motion processing. *ACM Trans Graph* 2013;32(4):80:1–9.
- [56] Engel J, Sturm J, Cremers D. Camera-based navigation of a low-cost quadrocopter. In: Proceedings of 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2012 Oct 7–12; Vilamoura, Portugal. Piscataway: IEEE; 2012. p. 2815–21.
- [57] Bojarski M, Del Testa D, Dworakowski D, Firner B, Flepp B, Goyal P, et al. End to end learning for self-driving cars. arXiv:1604.07316v1. 2016.
- [58] Goodfellow IJ, Bulatov Y, Ibarz J, Arnoud S, Shet V. Multi-digit number recognition from street view imagery using deep convolutional neural networks. arXiv:1312.6082v4. 2014.
- [59] Templeton B, inventor; Google LLC, Waymo LLC, assignees. Methods and systems for transportation to destinations by a self-driving vehicle. United States Patent US 9665101B1. 2017 May 30.
- [60] Taigman Y, Yang M, Ranzato MA, Wolf L. DeepFace: closing the gap to human-level performance in face verification. In: Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. Piscataway: IEEE; 2014. p. 1701–8.
- [61] Airport's facial recognition technology catches 2 people attempting to enter US illegally [Internet]. New York City: FOX News Network, LLC; 2018 Sep 12 [cited 2018 Sep 25]. Available from: <http://www.foxnews.com/travel/2018/09/12/airports-facial-recognition-technology-caughts-2-people-attempting-to-enter-us-illegally.html>.
- [62] Wojna Z, Gorban AN, Lee DS, Murphy K, Yu Q, Li Y, et al. Attention-based extraction of structured information from street view imagery. In: Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition; volume 1; 2017 Nov 9–15; Kyoto, Japan. Piscataway: IEEE; 2018. p. 844–50.
- [63] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
- [64] Zink J, Lovelace B. Unmanned aerial vehicle bridge inspection demonstration project: final report. St. Paul: Minnesota Department of Transportation; 2015 Jul. Report No.: MN/RC 2015–40.
- [65] Wells J, Lovelace B. Unmanned aircraft system bridge inspection demonstration project phase II: final report. St. Paul: Minnesota Department of Transportation; 2017 Jun. Report No.: MN/RC 2017–18.
- [66] Otero LD, Gagliardo N, Dalli D, Huang WH, Cosentino P. Proof of concept for using unmanned aerial vehicles for high mast pole and bridge inspections: final report. Tallahassee: Florida Department of Transportation; 2015 Jun. Contract No.: BDV28 TWO 977–02.
- [67] Tomiczek AP, Bridge JA, Ifju PG, Whitley TJ, Tripp CS, Ortega AE, et al. Small unmanned aerial vehicle (sUAV) inspections in GPS denied area beneath bridges. In: Soules JG, editor. Structures Congress 2018: bridges, transportation structures, and nonbuilding structures; 2018 Apr 19–21; Fort Worth, TX, USA. Reston: American Society of Civil Engineers; 2018. p. 205–16.
- [68] Brooks C, Dobson RJ, Banach DM, Dean D, Oommen T, Wolf RE, et al. Evaluating the use of unmanned aerial vehicles for transportation purposes: final report. Lansing: Michigan Department of Transportation; 2015 May. Report No.: RC-1616, MTRI-MDOTUAV-FR-2014. Contract No.: 2013-067.
- [69] Duque L, Seo J, Wacker J. Synthesis of unmanned aerial vehicle applications for infrastructures. *J Perform Constr Facil* 2018;32(4):04018046.
- [70] Abdel-Qader I, Abudayeh O, Kelly ME. Analysis of edge-detection techniques for crack identification in bridges. *J Comput Civ Eng* 2003;17(4):255–63.
- [71] Jahanshahi MR, Kelly JS, Masri SF, Sukhatme GS. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Struct Infrastruct Eng* 2009;5(6):455–86.
- [72] Jahanshahi MR, Masri SF. A new methodology for non-contact accurate crack width measurement through photogrammetry for automated structural safety evaluation. *Smart Mater Struct* 2013;22(3):035019.
- [73] Zhu Z, German S, Brilakis I. Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation. *Autom Construct* 2011;20(7):874–83.
- [74] Nishikawa T, Yoshida J, Sugiyama T, Fujino Y. Concrete crack detection by multiple sequential image filtering. *Comput Civ Infrastruct Eng* 2012;27(1):29–47.
- [75] Yamaguchi T, Hashimoto S. Fast crack detection method for large-size concrete surface images using percolation-based image processing. *Mach Vis Appl* 2010;21(5):797–809.

- [76] Zhang W, Zhang Z, Qi D, Liu Y. Automatic crack detection and classification method for subway tunnel safety monitoring. *Sensors (Basel)* 2014;14(10):19307–28.
- [77] Olsen MJ, Chen Z, Hutchinson T, Kuester F. Optical techniques for multiscale damage assessment. *Geomat Nat Haz Risk* 2013;4(1):49–70.
- [78] Lee BY, Kim YY, Yi ST, Kim JK. Automated image processing technique for detecting and analysing concrete surface cracks. *Struct Infrastruct Eng* 2013;9(6):567–77.
- [79] Liu YF, Cho SJ, Spencer BF Jr, Fan JS. Automated assessment of cracks on concrete surfaces using adaptive digital image processing. *Smart Struct Syst* 2014;14(4):719–41.
- [80] Liu YF, Cho SJ, Spencer BF Jr, Fan JS. Concrete crack assessment using digital image processing and 3D scene reconstruction. *J Comput Civ Eng* 2016;30(1):04014124.
- [81] Jahanshahi MR, Masri SF, Padgett CW, Sukhatme GS. An innovative methodology for detection and quantification of cracks through incorporation of depth perception. *Mach Vis Appl* 2013;24(2):227–41.
- [82] Erkal BG, Hajjar JF. Laser-based surface damage detection and quantification using predicted surface properties. *Auton Constr* 2017;83:285–302.
- [83] Kim H, Ahn E, Cho S, Shin M, Sim SH. Comparative analysis of image binarization methods for crack identification in concrete structures. *Cement Concr Res* 2017;99:53–61.
- [84] Prasanna P, Dana KJ, Gucunski N, Basily BB, La HM, Lim RS, et al. Automated crack detection on concrete bridges. *IEEE Trans Autom Sci Eng* 2016;13(2):591–9.
- [85] Adhikari RS, Moselhi O, Bagchi A. Image-based retrieval of concrete crack properties for bridge inspection. *Auton Constr* 2014;39:180–94.
- [86] Torok MM, Golparvar-Fard M, Kochersberger KB. Image-based automated 3D crack detection for post-disaster building assessment. *J Comput Civ Eng* 2014;28(5):A4014004.
- [87] Adhikari RS, Moselhi O, Bagchi A. A study of image-based element condition index for bridge inspection. In: Proceedings of the 30th International Symposium on Automation and Robotics in Construction and Mining: building the future in automation and robotics (volume 2); 2013 Aug 11–15; Montreal, QC, Canada. International Association for Automation and Robotics in Construction; 2013. p. 868–79.
- [88] Paal SG, Jeon JS, Brilakis I, DesRoches R. Automated damage index estimation of reinforced concrete columns for post-earthquake evaluations. *J Struct Eng* 2015;141(9):04014228.
- [89] Yeum CM, Dyke SJ. Vision-based automated crack detection for bridge inspection. *Comput Civ Infrastruct Eng* 2015;30(10):759–70.
- [90] Greimann LF, Stecker JH, Kao AM, Rens KL. Inspection and rating of miter lock gates. *J Perform Constr Facil* 1991;5(4):226–38.
- [91] Jahanshahi MR, Chen FC, Joffe C, Masri SF. Vision-based quantitative assessment of microcracks on reactor internal components of nuclear power plants. *Struct Infrastruct Eng* 2017;13(8):1013–26.
- [92] Ghanta S, Karp T, Lee S. Wavelet domain detection of rust in steel bridge images. In: Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing; 2011 May 22–27; Prague, Czech Republic. Piscataway: IEEE; 2011. p. 1033–6.
- [93] Jahanshahi MR, Masri SF. Parametric performance evaluation of wavelet-based corrosion detection algorithms for condition assessment of civil infrastructure systems. *J Comput Civ Eng* 2013;27(4):345–57.
- [94] Medeiros FNS, Ramalho GLB, Bento MP, Medeiros LCL. On the evaluation of texture and color features for nondestructive corrosion detection. *EURASIP J Adv Signal Process* 2010;2010:817473.
- [95] Shen HK, Chen PH, Chang LM. Automated steel bridge coating rust defect recognition method based on color and texture feature. *Auton Constr* 2013;31:338–56.
- [96] Son H, Hwang N, Kim C, Kim C. Rapid and automated determination of rusted surface areas of a steel bridge for robotic maintenance systems. *Auton Constr* 2014;42:13–24.
- [97] Igoe D, Parisi AV. Characterization of the corrosion of iron using a smartphone camera. *Instrum Sci Technol* 2016;44(2):139–47.
- [98] Ahuja SK, Shukla MK. A survey of computer vision based corrosion detection approaches. In: Satapathy S, Joshi A, editors. Information and communication technology for intelligent systems (ICTIS 2017)—volume 2. Cham: Springer; 2018. p. 55–63.
- [99] Koch C, Brilakis I. Pothole detection in asphalt pavement images. *Adv Eng Inform* 2011;25(3):507–15.
- [100] Salman M, Mathavan S, Kamal K, Rahman M. Pavement crack detection using the Gabor filter. In: Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems; 2013 Oct 6–9; The Hague, the Netherlands. Piscataway: IEEE; 2013. p. 2039–44.
- [101] Hu Y, Zhao C. A novel LBP based methods for pavement crack detection. *J Pattern Recognit Res* 2010;5(1):140–7.
- [102] Zou Q, Cao Y, Li Q, Mao Q, Wang S. CrackTree: automatic crack detection from pavement images. *Pattern Recognit Lett* 2012;33(3):227–38.
- [103] Kirschke KR, Velinsky SA. Histogram-based approach for automated pavement-crack sensing. *J Transp Eng* 1992;118(5):700–10.
- [104] Li L, Sun L, Tan S, Ning G. An efficient way in image preprocessing for pavement crack images. In: Proceedings of the 12th COTA International Conference of Transportation Professionals; 2012 Aug 3–6; Beijing, China. Reston: American Society of Civil Engineers; 2012. p. 3095–103.
- [105] Adu-Gyamfi YO, Okine NOA, Garateguy G, Carrillo R, Arce GR. Multiresolution information mining for pavement crack image analysis. *J Comput Civ Eng* 2012;26(6):741–9.
- [106] Chen YL, Jahanshahi MR, Manjunatha P, Gan WP, Abdelbarr M, Masri SF, et al. Inexpensive multimodal sensor fusion system for autonomous data acquisition of road surface conditions. *IEEE Sens J* 2016;16(21):7731–43.
- [107] Zakeri H, Nejad FM, Fahimifar A. Image based techniques for crack detection, classification and quantification in asphalt pavement: a review. *Arch Comput Methods Eng* 2017;24(4):935–77.
- [108] Koch C, Georgieva K, Kasireddy V, Akinci B, Fieguth P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv Eng Inform* 2015;29(2):196–210. Corrigendum in: *Adv Eng Inform* 2016;30(2):208–10.
- [109] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of International Conference on Learning Representations 2015; 2015 May 7–9; San Diego, CA, USA; 2015. p. 1–14.
- [110] Xu Y, Bao Y, Chen J, Zuo W, Li H. Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. *Struct Heal Monit*. Epub 2018 Apr 2.
- [111] Zhang L, Yang F, Zhang YD, Zhu YJ. Road crack detection using deep convolutional neural network. In: Proceedings of 2016 IEEE International Conference on Image Processing; 2016 Sep 25–28; Phoenix, AZ, USA. Piscataway: IEEE; 2016. p. 3708–12.
- [112] Cha YJ, Choi W, Büyüköztürk O. Deep learning-based crack damage detection using convolutional neural networks. *Comput Civ Infrastruct Eng* 2017;32(5):361–78.
- [113] Kim H, Ahn E, Shin M, Sim SH. Crack and noncrack classification from concrete surface images using machine learning. *Struct Health Monit*. Epub 2018 Apr 23.
- [114] Kim B, Cho S. Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. *Sensors (Basel)* 2018;18(10):3452.
- [115] Kim B, Cho S. Automated and practical vision-based crack detection of concrete structures using deep learning. *Comput Civ Infrastruct Eng*. In press.
- [116] Lee Y, Kim B, Cho S. Image-based spalling detection of concrete structures incorporating deep learning. In: Proceedings of the 7th World Conference on Structural Control and Monitoring; 2018 Jul 22–25; Qingdao, China; 2018.
- [117] Atha DJ, Jahanshahi MR. Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection. *Struct Health Monit* 2018;17(5):1110–28.
- [118] Chen FC, Jahanshahi MR. NB-CNN: deep learning-based crack detection using convolutional neural network and naïve Bayes data fusion. *IEEE Trans Ind Electron* 2018;65(5):4392–400.
- [119] Yeum CM. Computer vision-based structural assessment exploiting large volumes of images [dissertation]. West Lafayette: Purdue University; 2016.
- [120] Xu Y, Li S, Zhang D, Jin Y, Zhang F, Li N, et al. Identification framework for cracks on a steel structure surface by a restricted Boltzmann machines algorithm based on consumer-grade camera images. *Struct Contr Health Monit* 2018;25(2):e2075.
- [121] Maguire M, Dorafshan S, Thomas RJ. SDNET2018: a concrete crack image dataset for machine learning applications [dataset]. Logan: Utah State University; 2018.
- [122] Bao Y, Tang Z, Li H, Zhang Y. Computer vision and deep learning-based data anomaly detection method for structural health monitoring. *Struct Heal Monit* 2019;18(2):401–21.
- [123] Dang J, Shrestha A, Haruta D, Tabata Y, Chun P, Okubo K. Site verification tests for UAV bridge inspection and damage image detection based on deep learning. In: Proceedings of the 7th World Conference on Structural Control and Monitoring; 2018 Jul 22–25; Qingdao, China; 2018.
- [124] Yeum CM, Dyke SJ, Ramirez J. Visual data classification in post-event building reconnaissance. *Eng Struct* 2018;155:16–24.
- [125] Cha YJ, Choi W, Suh G, Mahmoudkhani S, Büyüköztürk O. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Comput Civ Infrastruct Eng* 2018;33(9):731–47.
- [126] Zhang A, Wang KCP, Li B, Yang E, Dai X, Peng Y, et al. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. *Comput Civ Infrastruct Eng* 2017;32(10):805–19.
- [127] Hoskere V, Narazaki Y, Hoang TA, Spencer BF Jr. Vision-based structural inspection using multiscale deep convolutional neural networks. In: Proceedings of the 3rd Huixian International Forum on Earthquake Engineering for Young Researchers; 2017 Aug 11–12; Urbana, IL, USA; 2017.
- [128] Hoskere V, Narazaki Y, Hoang TA, Spencer BF Jr. Towards automated postearthquake inspections with deep learning-based condition-aware models. In: Proceedings of the 7th World Conference on Structural Control and Monitoring; 2018 Jul 22–25; Qingdao, China; 2018.
- [129] Seitz SM, Curless B, Diebel J, Scharstein D, Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; volume 1; 2006 Jun 17–22; New York City, NY, USA. Piscataway: IEEE; 2006. p. 519–28.
- [130] Khaloo A, Lattanzi D, Cunningham K, Dell'Andrea R, Riley M. Unmanned aerial vehicle inspection of the Placer River Trail Bridge through image-based 3D modelling. *Struct Infrastruct Eng* 2018;14(1):124–36.
- [131] Khaloo A, Lattanzi D, Jachimowicz A, Devaney C. Utilizing UAV and 3D computer vision for visual inspection of a large gravity dam. *Front Built Environ* 2018;4:31.

- [132] Besl PJ, McKay ND. A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 1992;14(2):239–56.
- [133] Meshlab.net [Internet]. [cited 2018 Jun 20]. Available from: <http://www.meshlab.net/>.
- [134] CloudCompare: 3D point cloud and mesh processing software open source project [Internet]. [cited 2018 Jun 20]. Available from: <http://www.danielgm.net/cc/>.
- [135] Girardeau-Montaut D, Roux M, Marc R, Thibault G. Change detection on points cloud data acquired with a ground laser scanner. In: Proceedings of the ISPRS Workshop—Laser Scanning 2005; 2005 Sep 12–14; Enschede, the Netherlands; 2005. p. 30–5.
- [136] Lague D, Brodin N, Leroux J. Accurate 3D comparison of complex topography with terrestrial laser scanner: application to the Rangitikei canyon (N-Z). *ISPRS J Photogramm Remote Sens* 2013;82:10–26.
- [137] Morgenthal G, Hallermann N. Quality assessment of unmanned aerial vehicle (UAV) based visual inspection of structures. *Adv Struct Eng* 2014;17(3):289–302.
- [138] Hallermann N, Morgenthal G, Rodehorst V. Unmanned aerial systems (UAS)—case studies of vision based monitoring of ageing structures. In: Proceedings of the International Symposium Non-Destructive Testing in Civil Engineering; 2015 Sep 15–17; Berlin, Germany; 2015.
- [139] Khaloo A, Lattanzi D. Automatic detection of structural deficiencies using 4D hue-assisted analysis of color point clouds. In: Pakzad S, editor. *Dynamics of civil structures*, volume 2. Cham: Springer; 2019. p. 197–205.
- [140] Jafari B, Khaloo A, Lattanzi D. Deformation tracking in 3D point clouds via statistical sampling of direct cloud-to-cloud distances. *J Nondestruct Eval* 2017;36:65.
- [141] Ghahremani K, Khaloo A, Mohamadi S, Lattanzi D. Damage detection and finite-element model updating of structural components through point cloud analysis. *J Aerosp Eng* 2018;31(5):04018068.
- [142] Radke RJ, Andra S, Al-Kofahi O, Roysam B. Image change detection algorithms: a systematic survey. *IEEE Trans Image Process* 2005;14(3):294–307.
- [143] Hussain M, Chen D, Cheng A, Wei H, Stanley D. Change detection from remotely sensed images: from pixel-based to object-based approaches. *ISPRS J Photogramm Remote Sens* 2013;80:91–106.
- [144] Tewkesbury AP, Comber AJ, Tate NJ, Lamb A, Fisher PF. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sens Environ* 2015;160:1–14.
- [145] Sakurada K, Okatani T, Deguchi K. Detecting changes in 3D structure of a scene from multi-view images captured by a vehicle-mounted camera. In: Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition; 2013 Jun 23–28; Portland, OR, USA. Piscataway: IEEE; 2013. p. 137–44.
- [146] Alcantarilla PF, Stent S, Ros G, Arroyo R, Gherardi R. Street-view change detection with deconvolutional networks. In: Proceedings of 2016 Robotics: Science and Systems Conference; 2016 Jun 18–22; Ann Arbor, MI, USA; 2016.
- [147] Stent S, Gherardi R, Stenger B, Soga K, Cipolla R. Visual change detection on tunnel linings. *Mach Vis Appl* 2016;27(3):319–30.
- [148] Kiziltas S, Akinci B, Ergen E, Tang P, Gordon C. Technological assessment and process implications of field data capture technologies for construction and facility/infrastructure management. *Electron J Inf Technol Constr* 2008;13:134–54.
- [149] Golparvar-Fard M, Peña-Mora F, Savarese S. D4AR—a 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *J Inf Technol Constr* 2009;14:129–53.
- [150] Huber D. ARIA: the Aerial Robotic Infrastructure Analyst. SPIE Newsroom [Internet] 2014 Jun 9 [cited 2019 Feb 24]. Available from: <http://spie.org/newsroom/5472-aria-the-aerial-robotic-infrastructure-analyst?SSO=1>.
- [151] Earthsense.co [Internet]. Champaign: EarthSense, Inc.; [cited 2019 Feb 25]. Available from: <https://www.earthsense.co>.
- [152] Zhu Z, Brilakis I. Concrete column recognition in images and videos. *J Comput Civ Eng* 2010;24(6):478–87.
- [153] Koch C, Paal SG, Rashidi A, Zhu Z, König M, Brilakis I. Achievements and challenges in machine vision-based inspection of large concrete structures. *Adv Struct Eng* 2014;17(3):303–18.
- [154] Xiong X, Adan A, Akinci B, Huber D. Automatic creation of semantically rich 3D building models from laser scanner data. *Autom Construct* 2013;31:325–37.
- [155] Perez-Perez Y, Golparvar-Fard M, El-Rayes K. Semantic and geometric labeling for enhanced 3D point cloud segmentation. In: Perdomo-Rivera JL, González-Quevedo A, del Puerto CL, Maldonado-Fortunet F, Molina-Bas OL, editors. *Construction Research Congress 2016: old and new*; 2016 May 31–June 2; San Juan, Puerto Rico. Reston: American Society of Civil Engineers; 2016. p. 2542–52.
- [156] Armeni I, Sener O, Zamir AR, Jiang H, Brilakis I, Fischer M, et al. 3D semantic parsing of large-scale indoor spaces. In: In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 1534–43.
- [157] Golparvar-Fard M, Bohn J, Teizer J, Savarese S, Peña-Mora F. Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques. *Autom Construct* 2011;20(8):1143–55.
- [158] Golparvar-Fard M, Peña-Mora F, Savarese S. Monitoring changes of 3D building elements from unordered photo collections. In: Proceedings of 2011 IEEE International Conference on Computer Vision Workshops; 2011 Nov 6–13; Barcelona, Spain. Piscataway: IEEE; 2011. p. 249–56.
- [159] Lu R, Brilakis I, Middleton CR. Detection of structural components in point clouds of existing RC bridges. *Comput Civ Infrastruct Eng* 2019;34(3):191–212.
- [160] Yeum CM, Choi J, Dyke SJ. Automated region-of-interest localization and classification for vision-based visual assessment of civil infrastructure. *Struct Health Monit. Epub* 2018 Mar 27.
- [161] Gao Y, Mosalam KM. Deep transfer learning for image-based structural damage recognition. *Comput Civ Infrastruct Eng* 2018;33(9):748–68.
- [162] Liang X. Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization. *Comput Civ Infrastruct Eng. Epub* 2018 December 5.
- [163] Narazaki Y, Hoskere V, Hoang TA, Spencer BF Jr. Vision-based automated bridge component recognition integrated with high-level scene understanding. In: Proceedings of the 13th International Workshop on Advanced Smart Materials and Smart Structures Technology; 2017 Jul 22–23; Tokyo, Japan; 2017.
- [164] Narazaki Y, Hoskere V, Hoang TA, Spencer BF Jr. Vision-based automated bridge component recognition with high-level scene consistency. *Comput Civ Infrastruct Eng. In Press*.
- [165] Narazaki Y, Hoskere V, Hoang TA, Spencer BF. Automated vision-based bridge component extraction using multiscale convolutional neural networks. In: Proceedings of the 3rd Huixian International Forum on Earthquake Engineering for Young Researchers; 2017 Aug 11–12; Urbana, IL, USA; 2017.
- [166] German S, Jeon JS, Zhu Z, Bearman C, Brilakis I, DesRoches R, et al. Machine vision-enhanced postearthquake inspection. *J Comput Civ Eng* 2013;27(6):622–34.
- [167] German S, Brilakis I, Desroches R. Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments. *Adv Eng Informatics* 2012;26(4):846–58.
- [168] Anil EB, Akinci B, Garrett JH, Kurc O. Information requirements for earthquake damage assessment of structural walls. *Adv Eng Inform* 2016;30(1):54–64.
- [169] Anil EB, Akinci B, Kurc O, Garrett JH. Building-information-modeling-based earthquake damage assessment for reinforced concrete walls. *J Comput Civ Eng* 2016;30(4):04015076.
- [170] Wei Y, Kasireddy V, Akinci B, Smith I, Domer B. 3D imaging in construction and infrastructure management: technological assessment and future research directions3D imaging in construction and infrastructure management: technological assessment and future research directions. In: Advanced computing strategies for engineering: proceedings of the 25th EG-ICE International Workshop 2018; 2018 Jun 10–13; Lausanne, Switzerland. Cham: Springer; 2018. p. 37–60.
- [171] Farrar CR, James III GH. System identification from ambient vibration measurements on a bridge. *J Sound Vibrat* 1997;205(1):1–18.
- [172] Lynch JP, Sundararajan A, Law KH, Kiremidjian AS, Carryer E, Sohn H. Field validation of a wireless structural monitoring system on the Alamosa Canyon Bridge. In: Liu SC, editor. *Proceedings volume 5057: Smart Structures and Materials 2003: smart systems and nondestructive evaluation for civil infrastructures*; 2003 Mar 2–6; San Diego, CA, USA; 2003.
- [173] Jin G, Sain MK, Spencer Jr BF. Frequency domain system identification for controlled civil engineering structures. *IEEE Trans Contr Syst Technol* 2005;13(6):1055–62.
- [174] Jang S, Jo H, Cho S, Mechitov K, Rice JA, Sim SH, et al. Structural health monitoring of a cable-stayed bridge using smart sensor technology: deployment and evaluation. *Smart Struct Syst* 2010;6(5–6):439–59.
- [175] Rice JA, Mechitov K, Sim SH, Nagayama T, Jang S, Kim R, et al. Flexible smart sensor framework for autonomous structural health monitoring. *Smart Struct Syst* 2010;6(5–6):423–38.
- [176] Moreu F, Kim R, Lei S, Diaz-Fañas G, Dai F, Cho S, et al. Campaign monitoring railroad bridges using wireless smart sensors networks. In: Proceedings of 2014 Engineering Mechanics Institute Conference [Internet]; 2014 Aug 5–8; Hamilton, ON, Canada, 2014 [cited 2017 Oct 11]. Available from: <http://emi2014.mcmaster.ca/index.php/emi2014/EMI2014/paper/viewPaper/139>.
- [177] Cho S, Giles RK, Spencer Jr BF. System identification of a historic swing truss bridge using a wireless sensor network employing orientation correction. *Struct Contr Health Monit* 2015;22(2):255–72.
- [178] Sim SH, Spencer Jr BF, Zhang M, Xie H. Automated decentralized modal analysis using smart sensors. *Struct Contr Health Monit* 2010;17(8):872–94.
- [179] Horn BKP, Schunck BG. Determining optical flow. *Artif Intell* 1981;17(1–3):185–203.
- [180] Bruhn A, Weickert J, Schnörr C. Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods. *Int J Comput Vis* 2005;61(3):211–31.
- [181] Liu C. Beyond pixels: exploring new representations and applications for motion analysis [dissertation]. Cambridge: Massachusetts Institute of Technology; 2009.
- [182] Baker S, Scharstein D, Lewis JP, Roth S, Black MJ, Szeliski R. A database and evaluation methodology for optical flow. *Int J Comput Vis* 2011;92(1):1–31.
- [183] Sutton MA, Orteu JJ, Schreier HW. *Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications*. New York City: Springer Science & Business Media; 2009.
- [184] The VIC-2D™ system [Internet]. Irmo: Correlated Solutions, Inc.; [cited 2019 Feb 24]. Available from: <https://www.correlatedsolutions.com/vic-2d/>.

- [185] GOM Correlate: evaluation software for 3D testing [Internet]. Braunschweig: GOM GmbH; [cited 2019 Feb 24]. Available from: <https://www.gom.com/3d-software/gom-correlate.html>.
- [186] Pan B, Qian K, Xie H, Asundi A. Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. *Meas Sci Technol* 2009;20(6):062001.
- [187] Hoult NA, Take WA, Lee C, Dutton M. Experimental accuracy of two dimensional strain measurements using digital image correlation. *Eng Struct* 2013;46:718–26.
- [188] Dutton M, Take WA, Hoult NA. Curvature monitoring of beams using digital image correlation. *J Bridge Eng* 2014;19(3):05013001.
- [189] Ellenberg A, Branco L, Krick A, Bartoli I, Kontos A. Use of unmanned aerial vehicle for quantitative infrastructure evaluation. *J Infrastruct Syst* 2015;21(3):04014054.
- [190] McCormick N, Lord J. Digital image correlation for structural measurements. *Proc Inst Civ Eng Civ Eng* 2012;165(4):185–90.
- [191] Yoneyama S, Kitagawa A, Iwasa S, Tani K, Kikuta H. Bridge deflection measurement using digital image correlation. *Exp Tech* 2007;31(1):34–40.
- [192] Yoneyama S, Ueda H. Bridge deflection measurement using digital image correlation with camera movement correction. *Mater Trans* 2012;53(2):285–90.
- [193] Helfrick MN, Niezrecki C, Avitabile P, Schmidt T. 3D digital image correlation methods for full-field vibration measurement. *Mech Syst Signal Process* 2011;25(3):917–27.
- [194] Reagan D. Unmanned aerial vehicle measurement using three-dimensional digital image correlation to perform bridge structural health monitoring [dissertation]. Lowell: University of Massachusetts Lowell; 2017.
- [195] Mahal M, Blanksvärd T, Täljsten B, Sas G. Using digital image correlation to evaluate fatigue behavior of strengthened reinforced concrete beams. *Eng Struct* 2015;105:277–88.
- [196] Ghorbani R, Matta F, Sutton MA. Full-field deformation measurement and crack mapping on confined masonry walls using digital image correlation. *Exp Mech* 2015;55(1):227–43.
- [197] Tomasi C, Kanade T. Detection and tracking of point features. Technical report. Pittsburgh: Carnegie Mellon University; 1991 Apr. Report No.: CMU-CS-91-132.
- [198] Lewis JP. Fast normalized cross-correlation. *Vis. Interface* 1995;10:120–3.
- [199] Ye XW, Dong CZ, Liu T. A review of machine vision-based structural health monitoring: methodologies and applications. *J Sens* 2016;2016:7103039.
- [200] Feng D, Feng MQ. Computer vision for SHM of civil infrastructure: from dynamic response measurement to damage detection—a review. *Eng Struct* 2018;156:105–17.
- [201] Nogueira FMA, Barbosa FS, Barra LPS. Evaluation of structural natural frequencies using image processing. In: Proceedings of 2005 International Conference on Experimental Vibration Analysis for Civil Engineering Structures; 2005 Oct 26–28; Bordeaux, France; 2005.
- [202] Chang CC, Xiao XH. An integrated visual-inertial technique for structural displacement and velocity measurement. *Smart Struct Syst* 2010;6(9):1025–39.
- [203] Fukuda Y, Feng MQ, Narita Y, Kaneko S, Tanaka T. Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm. *IEEE Sens J* 2013;13(12):4725–32.
- [204] Min JH, Gelo NJ, Jo H. Real-time image processing for non-contact monitoring of dynamic displacements using smartphone technologies. In: Lynch JP, editor. Proceedings volume 9803: sensors and smart structures technologies for civil, mechanical, and aerospace systems 2016; 2016 Mar 20–24; Las Vegas, NV, USA; 2016. p. 98031B.
- [205] Dong CZ, Ye XW, Jin T. Identification of structural dynamic characteristics based on machine vision technology. *Measurement* 2018;126:405–16.
- [206] Celik O, Dong CZ, Catbas FN. Measurement of human loads using computer vision. In: Pakzad S, editor. Dynamics of civil structures, volume 2: proceedings of the 36th IMAC, a conference and exposition on structural dynamics 2018; 2018 Feb 12–15; Orlando, FL, USA. Cham: Springer; 2019. p. 191–5.
- [207] Lee J, Lee KC, Cho S, Sim SH. Computer vision-based structural displacement measurement robust to light-induced image degradation for in-service bridges. *Sensors (Basel)* 2017;17(10):2317.
- [208] Park JW, Moon DS, Yoon H, Gomez F, Spencer Jr BF, Kim JR. Visual-inertial displacement sensing using data fusion of vision-based displacement with acceleration. *Struct Contr Health Monit* 2018;25(3):e2122.
- [209] Schumacher T, Shariati A. Monitoring of structures and mechanical systems using virtual visual sensors for video analysis: fundamental concept and proof of feasibility. *Sensors (Basel)* 2013;13(12):16551–64.
- [210] Yoon H, Elanwar H, Choi H, Golparvar-Fard M, Spencer Jr BF. Target-free approach for vision-based structural system identification using consumer-grade cameras. *Struct Contr Health Monit* 2016;23(12):1405–16.
- [211] Ye XW, Yi TH, Dong CZ, Liu T, Bai H. Multi-point displacement monitoring of bridges using a vision-based approach. *Wind Struct* 2015;20(2):315–26.
- [212] Abdelbarr M, Chen YL, Jahanshahi MR, Masri SF, Shen WM, Qidwai UA. 3D dynamic displacement-field measurement for structural health monitoring using inexpensive RGB-D based sensor. *Smart Mater Struct* 2017;26(12):125016.
- [213] Ye XW, Yi TH, Dong CZ, Liu T. Vision-based structural displacement measurement: system performance evaluation and influence factor analysis. *Measurement* 2016;88:372–84.
- [214] Feng D, Feng MQ. Vision-based multipoint displacement measurement for structural health monitoring. *Struct Contr Health Monit* 2016;23(5):876–90.
- [215] Yoon H, Hoskere V, Park JW, Spencer BF Jr. Cross-correlation-based structural system identification using unmanned aerial vehicles. *Sensors (Basel)* 2017;17(9):2075.
- [216] Chen JG, Wadhwa N, Cha YJ, Durand F, Freeman WT, Buyukozturk O. Modal identification of simple structures with high-speed video using motion magnification. *J Sound Vibrat* 2015;345:58–71.
- [217] Cha YJ, Chen JG, Büyükoztürk O. Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters. *Eng Struct* 2017;132:300–13.
- [218] Yang Y, Dorn C, Mancini T, Talken Z, Kenyon G, Farrar C, et al. Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. *Mech Syst Signal Process* 2017;85:567–90.
- [219] Yang Y, Dorn C, Mancini T, Talken Z, Kenyon G, Farrar C, et al. Spatiotemporal video-domain high-fidelity simulation and realistic visualization of full-field dynamic responses of structures by a combination of high-spatial-resolution modal model and video motion manipulations. *Struct Contr Health Monit* 2018;25(8):e2193.
- [220] Zaurin R, Khuc T, Catbas FN. Hybrid sensor-camera monitoring for damage detection: case study of a real bridge. *J Bridge Eng* 2016;21(6):05016002.
- [221] Pan B, Tian L, Song X. Real-time, non-contact and targetless measurement of vertical deflection of bridges using off-axis digital image correlation. *NDT E Int* 2016;79:73–80.
- [222] Xu Y, Brownjohn J, Kong D. A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Struct Contr Health Monit* 2018;25(5):e2155.
- [223] Kim SW, Kim NS. Dynamic characteristics of suspension bridge hanger cables using digital image processing. *NDT E Int* 2013;59:25–33.
- [224] Chen JG. Video camera-based vibration measurement of infrastructure [dissertation]. Cambridge: Massachusetts Institute of Technology; 2016.
- [225] Yoon H. Enabling smart city resilience : post-disaster response and structural health monitoring [dissertation]. Urbana: University of Illinois at Urbana-Champaign; 2016.
- [226] Mas D, Ferrer B, Acevedo P, Espinosa J. Methods and algorithms for video-based multi-point frequency measuring and mapping. *Measurement* 2016;85:164–74.
- [227] Chen Z, Li H, Bao Y, Li N, Jin Y. Identification of spatio-temporal distribution of vehicle loads on long-span bridges using computer vision technology. *Struct Contr Health Monit* 2016;23(3):517–34.
- [228] Hoskere V, Park JW, Yoon H, Spencer BF Jr. Vision-based modal survey of civil infrastructure using unmanned aerial vehicles. *J Struct Eng*. In press.
- [229] Ros G, Sellart L, Materzynska J, Vazquez D, Lopez AM. The SYNTHIA dataset: a large collection of synthetic images for semantic segmentation of urban scenes. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway: IEEE; 2016. p. 3234–43.
- [230] Blender.org [Internet]. Amsterdam: Stichting Blender Foundation; [cited 2018 Aug 1]. Available from: <https://www.blender.org/>.
- [231] Hoskere V, Narazaki Y, Spencer BF Jr, Smith MD. Deep learning-based damage detection of miter gates using synthetic imagery from computer graphics. In: Proceedings of the 12th International Workshop on Structural Health Monitoring; 2019 Sep 10–12; Stanford, CA, USA; 2019. In press.
- [232] Hoskere V, Narazaki Y, Spencer BF Jr. Learning to detect important visual changes for structural inspections using physics-based graphics models. In: 9th International Conference on Structural Health Monitoring of Intelligent Infrastructure; 2019 Aug 4–7; Louis, MO, USA; 2019. In press.
- [233] Narazaki Y, Hoskere V, Hoang TA, Spencer BF Jr. Automated bridge component recognition using video data. In: Proceedings of the 7th World Conference on Structural Control and Monitoring; 2018 Jul 22–25; Qingdao, China; 2018.
- [234] Shi X, Chen Z, Wang H, Yeung DY, Wong W, Woo W. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems: volume 1; 2015 Dec 7–12; Montreal, QC, Canada. Cambridge: MIT Press; 2015. p. 802–10.
- [235] Unity3d.com [Internet]. San Francisco: Unity Technologies; c2019 [cited 2019 Feb 24]. Available from: <https://unity3d.com/>.