Council on Tall Buildings and Urban Habitat Korea

International Journal of
High-Rise Buildings
www.ctbuh-korea.org/ijhrb/index.php

International Journal of High-Rise Buildings
December 2023, Vol 12, No 4, 321-333
https://doi.org/10.21022/IJHRB.2023.12.4.321

Computer Vision-based Structural Health Monitoring: A Review

Jun Su Park, Joohyun An, and Hyo Seon Park†

Department of Architecture and Architectural Engineering, Yonsei University, Seoul 03722, Republic of Korea

Abstract Structural health monitoring is a technology or research field that extends the service life of structures and contributes to the prevention of disaster accidents by continuously evaluating the safety, stability, and serviceability of structures as well as allowing timely and proper maintenance. However, the contact-type sensors used for it require considerable time, cost, and labor for installation and maintenance. As an alternative, computer vision has attracted attention recently. Computer vision has the potential to make quality, deformation, and damage monitoring for structures contactless and automated. In this study, research cases in which computer vision was utilized for structural health monitoring are introduced, and its effects and limitations are summarized. Therefore, the applicability and future research directions of computer vision-based structural health monitoring are discussed.

Keywords Structural health monitoring, Computer vision

†Corresponding author:
Hyo Seon Park
E-mail:
hspark@yonsei.ac.kr

1. Introduction

Structural health monitoring (SHM) (Azhar, et al., 2023, Mishra, et al., 2022, Sabato, et al., 2023) is a technology or research field that extends the service life of structures and contributes to the prevention of disaster accidents by continuously evaluating the safety, stability, and serviceability of structures, thereby allowing the required maintenance in a timely manner (Azhar, et al., 2023, Choi, et al., 2023, Kwan Oh, et al., 2023, Park, et al., 2018, Sabato, et al., 2023, Yun, et al., 2021). Traditionally, SHM has depended on contact-type sensors, which have limitations because considerable time, cost, and labor are required for installation and maintenance (Lee, et al., 2010, Oh, et al., 2022). Efforts have been made to remove complex wire connections by constructing a wireless sensor network (Azhar, et al., 2023, Kim, et al., 2019, Lee, et al., 2010, Yu, et al., 2023); however, they showed limitations because data loss frequently occurs during sensing and transmission (Azhar, et al., 2023, Kim, et al., 2019, Oh, et al., 2020, Oh, et al., 2017). Currently, the trend towards taller and larger structures is increasing interest in SHM, but ultimately, the stability and durability of SHM are still not considered reliable.

The introduction of computer vision (CV) (Prince, 2012) is one of the attempts to overcome the aforementioned limitations of SHM. CV is a non-contact type and can be used in monitoring the condition of structures using visual data, which includes images and videos (Dong, et al., 2020, Guo, et al., 2024, Khuc, et al., 2018, Oh, et al., 2015, Park, et al., 2017, Park, et al., 2016, Ye, et al., 2023). It is particularly favorable for monitoring structures that are difficult to access. During construction, a considerable amount of visual data is collected, which is being accelerated with the development of cutting-edge systems, such as drones, ground robots, surveillance cameras, and cellphones (Paneru, et al., 2021). On average, 400,000 images are collected from a site (Han, et al., 2017). Traditionally, collected visual data have been used for report preparation, work status identification, and security/surveillance, but they can be directly utilized for SHM if the significant information extracted from them using CV is analyzed. They can be used to examine the dimensions of building components, inspect the spacing between rebars, and detect defects in concrete members. They can be used to measure the deformed shape of major structural members and their safety, and modal analysis can be conducted through dynamic displacement measurement. CV has the potential to significantly contribute to SHM automation (Paneru, et al., 2021). Therefore, research has been actively conducted to introduce CV to the construction industry (Xu, et al., 2021).

In particular, conventional CV before the 2010s when deep learning began to develop was effective only in simplified environments, and its performance could not be guaranteed in complex and chaotic real-world environments, such as construction sites (Xu, et al., 2021). This is expressed as “non-robust”. This is because it was practically impossible to mathematically model real-world environments that are changeable with many variables as conventional CV models image processing and pattern recognition processes as explicit mathematical representations. Therefore, there

were few cases in which conventional CV was successfully introduced to actual construction sites until the 2000s. However, since the 2010s, CV has been rapidly developed owing to advances in machine learning technologies, such as deep learning. Currently, CV is practically applied to complex problems in a robust manner beyond the laboratory scale, as well as to various engineering fields.

In this study, research cases in which CV was utilized for SHM are introduced, and its effects and limitations are summarized. Therefore, the applicability of CV-based SHM and future research directions are discussed. This paper is organized as follows. In section 1, the need for SHM and CV is briefly introduced. In section 2, the definition, development, and major research performances of CV are briefly introduced. In section 3, research cases of CV-based SHM are introduced in four categories. In the final conclusion, the performances and limitations of previous studies are examined and future research directions are discussed.

2. Computer Vision

CV is a technology or research field enabling computers to perform specific tasks, by extracting significant information from visual data and analyzing them (Paneru, et al., 2021, Xu, et al., 2021). Significant information refers to patterns, distinctive features, and spatial arrangements that are useful in designing the mapping relationship between visual data and the target output (Paneru, et al., 2021). CV is divided into image processing and pattern recognition. Image processing refers to technologies to process and analyze images, such as image quality improvement, transformation, filtering, contrast adjustment, and three-dimensional (3D) reconstruction. Pattern recognition is a technology to recognize significant patterns, which are helpful in performing object detection and face recognition, from visual data. CV integrates image processing and pattern recognition, and provides vision to computers.

Representative visual data that are managed with CV include images, videos, and 3D point clouds. Images are two-dimensional (2D) data represented by still photographs, and they consist of square grids referred to as pixels. Each pixel contains such information as color, transparency, and location. A video is a set of continuous images. It visually expresses information that is changing over time by displaying images sequentially and rapidly. The 3D point clouds are data that represent the surface of an object using a set of points with 3D coordinate information. They are suitable for expressing the 3D geometry and dimensions of an object.

The main tasks of CV include image classification, object detection, semantic segmentation, object tracking, image translation, and 3D reconstruction. Image classification is a task to identify the types of objects or scenes included in images, and object detection is a task to estimate the types and locations of objects included in images. Semantic segmentation is a task to identify objects to which each pixel of an image belongs. Object tracking is a task to track the locations of objects over time by continuously identifying the locations of specific objects for each image that constitutes a video. Image translation is a task to convert images of a certain style to images of a different style. Examples of this task include the conversion of a captured object cross-section into a drawing and the conversion of a photograph taken at night into that of during the day. The 3D reconstruction is a task to estimate the 3D geometry of the target object from 2D images. It includes the task of reconstructing 3D geometry information of the captured object by restoring the depth information lost by projecting the object in real space into 2D images using geometric relationships between the images taken from different angles.

CV can be classified as a field of artificial intelligence (AI) in that computers aim to recognize, understand, and judge in a similar way to humans (Xu, et al., 2021). Research on CV began in earnest with Lawrence Roberts' doctoral dissertation in 1963 (Roberts, 1963, Shapiro, 2020). Active research and development have been conducted over approximately 60 years. Roberts (1963) announced a methodology to estimate the original 3D geometry of simple block-type objects using a single 2D image. It became an early methodology to process and interpret visual data using computers, and later became the basis for research on CV. Sobel, et al. (1968) and Canny (1986) developed an edge detection algorithm, which became the most basic tool of CV to identify and recognize the structures of the objects included in images. Duda, et al. (1972) developed the hough transformation for detecting simple geometric shapes, such as straight lines and circles, from images. It has been established as the basic tool of CV. Harris, et al. (1988) and Jianbo, et al. (1994) developed a corner detection algorithm, which has become a major tool for motion tracking, image matching, and 3D reconstruction. Lucas, et al. (1981) and Tomasi, et al. (1991) developed a feature-tracking algorithm that can detect feature points from a sequential image dataset and track their locations. It became the basic tool of CV for video analysis and real-time location change monitoring. The histogram of oriented gradients (HOG) by Dalal, et al. (2005) detected patterns related to the geometry and directions of objects, thereby contributing to object recognition, such as pedestrian detection. Feature detection and description algorithms, which are represented by the scale-invariant feature transform (SIFT) of Lowe (1999), speeded up robust features (SURF) of Bay, et al. (2006), oriented fast and rotated brief (ORB) of Rublee, et al. (2011), and KAZE of Alcantarilla, et al. (2012), detected distinctive feature points from images and described their characteristics. These significantly contributed to the development of image matching technology that finds and matches the same points among the images captured from different angles. These feature detection and description

algorithms are currently being used as the main tools of CV for object recognition, object tracking, and 3D reconstruction. Stereo vision, which was established over the long term through many researchers, and structure from motion (SfM), which was established by Ullman (1979) and Longuet-Higgins, et al. (1980), have become the main methodologies of CV to reconstruct 3D representations from 2D images. Despite these monumental studies and performances, conventional CV before the 2010s managed toy examples under simplified environmental conditions with a focus on examining its applicability. This is because it was practically impossible to mathematically model actual environmental conditions that are changeable with many variables as the image processing and pattern recognition processes of conventional CV perform modeling as explicit mathematical representations. Consequently, conventional CV could not implement sufficiently robust generalization performance for real-world environments with complexity and uncertainty. Therefore, for CV during this period, simplified and limited environments were established, resulting in industrial applications being feasible only in limited situations.

However, after the 2010s, CV has undergone rapid development due to advancements in deep learning. Deep convolutional neural network (CNN) models (LeCun, et al., 2015), which are represented by AlexNet of Krizhevsky, et al. (2012), visual geometry group network (VGG) of Simonyan, et al. (2014), GoogLeNet of Szegedy, et al. (2015), and residual network (ResNet) of He, et al. (2016), showed remarkable performance improvements in the image classification task. The faster region-based convolutional neural network (Faster R-CNN) of Ren, et al. (2015) and You Only Look Once (YOLO) of Redmon, et al. (2016) achieved significant performance improvements for the task of detecting objects in real-time from visual data. The siamese network of Bertinetto, et al. (2016) improved the accuracy and speed of object tracking. Pixel-2-pixel (Pix2Pix) of Isola, et al. (2017) and the cycle-consistent generative adversarial network (CycleGAN) of Zhu, et al. (2017) achieved significant advances in the image-to-image translation task. The development and introduction of deep learning technology improved the robustness of CV for real-world environments, which could not be achieved through sophisticated mathematical models, to a significant level. Consequently, present CV can make practical contributions in addressing the issue beyond toy examples in simple environments (Paneru, et al., 2021). Cases where CV was successfully applied and utilized have been reported in various engineering fields (Xu, et al., 2021).

3. Computer Vision in Structural Health Monitoring

The introduction of CV to SHM can contribute to automation (Paneru, et al., 2021), provide a non-contact monitoring solution (Spencer, et al., 2019), and exhibit potential effects, such as measurement accuracy and precision improvement, sensor maintenance cost reduction, and safety accident risk prevention. Research to apply CV to SHM has been conducted for several decades. In particular, cases of applying CV to actual construction sites have increased over the last decade (Xu, et al., 2021). In general, CV-based SHM collects visual data using equipment, such as cameras and terrestrial laser scanning (TLS), and obtains such information as the geometry, dimensions, and displacements of structures through feature extraction and analysis that use deep learning models and CV algorithms. The obtained information is used to evaluate the safety and stability of the structures or identify their dynamic characteristics. In this chapter, CV application cases are introduced in four categories.

3.1. Dimensional Quality Inspection

Dimensional quality inspection is the task of inspecting the dimensions, layout, and quality of construction components. It is an important task for preventing safety accidents or construction delays caused by quality problems in advance (Park, et al., 2023). Traditionally, it has been manually performed using measurement tools, such as rulers and tapes. However, these methods require considerable time and labor, and errors may occur owing to the subjective opinions of the inspector. The introduction of CV to dimensional quality inspection has the potential to contribute to a reduction in measurement time and labor, objective inspection, and an improvement in accuracy and precision owing to automation. The introduction of CV to dimensional quality inspection has been considered a fairly difficult task. This is because sophisticated object detection and segmentation technology was required as all parts of the target object become the targets of the task compared to other tasks in which displacements are measured or tracked for specific measurement points. This task was difficult for conventional CV. Its difficulty was particularly higher owing to the complex geometry, diversity, and randomness of building components as well as the complexity and diversity of construction sites, which are the background of images. Consequently, research cases on dimensional quality inspection for construction components using CV before deep learning are limited. Therefore, in this subsection, dimensional quality inspection cases where advanced CV technology with deep learning technology was used are introduced.

Kardovskyi, et al. (2021) proposed an artificial intelligence quality inspection model (AI-QIM) that can inspect the rebars placed at construction sites (Figure 1). To ensure the safety and durability of structures, rebars must be installed with high dimensional quality and without omission according to the design drawings. AI-QIM performs the segmentation of rebars included in images using Mask R-CNN (He, et al., 2017), a deep learning model, and estimates the quantity, spacing, diameter, and

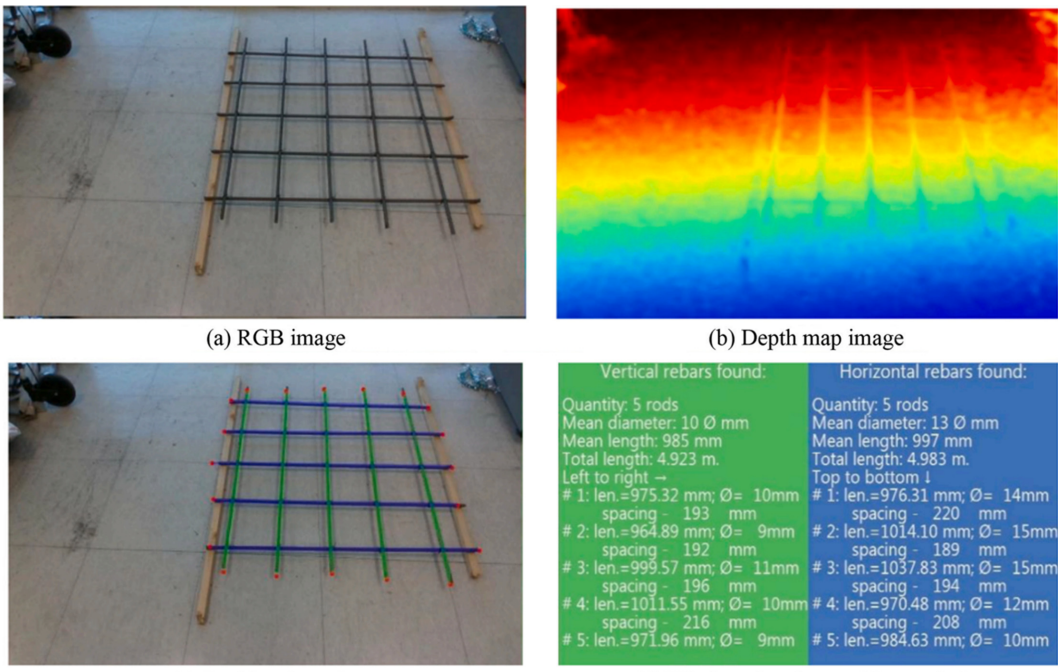


Figure 1. Artificial intelligence quality inspection model (AI-QIM). (Reprinted from Kardovskyi and Moon: Artificial intelligence quality inspection of steel bars installation by integrating mask R-CNN and stereo vision. Automation in construction. 2021, 130, 103850 with permission by Elsevier).

length of rebars through stereo vision. It achieved a measurement accuracy of ±6 mm or less at a distance of approximately 2 m. The research by Kardovskyi, et al. (2021) is significant because it showed the possibility of automating rebar inspection at construction sites by introducing a deep learning model and CV. However, it was verified at the laboratory scale as shown in Figure 1 and its robust performance was not verified in the actual construction site environment. Furthermore, it has limitations in terms of applicability because only the case of placing a layer of rebars is addressed. For practical application in actual construction sites, it is crucial to develop and validate a model for inspection scenarios that involve two or more layers of rebars.

Mirzaei, et al. (2023) proposed a model that performs the segmentation of each member from the 3D point cloud

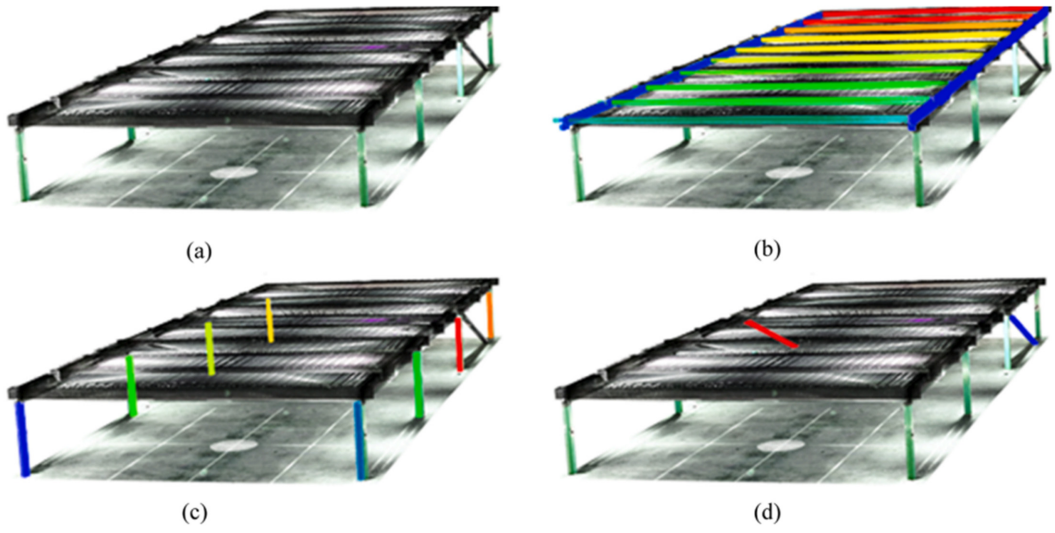


Figure 2. Geometric segmentation results for dimensional inspection of each member (a) raw point cloud, (b) extraction of potential beam points, (c) extraction of potential column points, and (d) extraction of potential bracing points (Reprinted from Mirzaei, et al.: End-to-end point cloud-based segmentation of building members for automating dimensional quality control. Advanced engineering informatics. 2023, 55, 101878 with permission by Elsevier)

data measured with TLS using PointNet (Qi, et al., 2017), a deep learning model, and measures the widths, heights, and lengths of the members (Figure 2). In this process, the bounding box method proposed in the study was utilized. Mirzaei, et al. (2023) achieved a measurement accuracy of ±19 mm or less in member dimension measurement. This is estimated to be a measurement accuracy result for a measurement distance of approximately 27 m considering that the performance results were derived for components in a building with a size of 12.3 m (width) × 23.6 m (length) × 2.5 m (height) (Park, et al., 2023). The research by Mirzaei, et al. (2023) is significant in that it automatically performed the segmentation of each member from 3D point cloud measurement data and measured their dimensions. However, structures in real-world construction sites rarely consist of only frames, devoid of non-structural elements, as assumed in this study (Figure 2). When the model is applied to the actual construction site environment, non-structures in the building can be measured together. Failure to separate them may cause errors during dimensional measurement. In addition, when occlusion occurs during measurement owing to various objects, the dimensional measurement results can be significantly distorted for parts where point clouds cannot be measured. Therefore, for the model of Mirzaei, et al. (2023) to be applied to actual construction sites, further research is required on measures to secure segmentation robustness and consider the unmeasured parts owing to occlusion.

Park, et al. (2023) proposed an automated model that captures stereo images and reconstructs them into cross-sectional drawing images to inspect the cross-sectional dimensions of building components (Figure 3). The proposed model uses telephoto lenses to capture high-resolution images. Camera calibration that is difficult to apply owing to the use of telephoto lenses was omitted. Instead, a new method that optimizes camera poses using relationships between stereo images was proposed and applied. In this process, image matching that is applied to stereo vision was also omitted. Because image matching involves many errors, its omission contributed to an improvement in reconstruction accuracy. Faster R-CNN and Pix2Pix, which are deep learning models, were applied to detect the target object from stereo images and translate the images into sectional drawings. An artificially generated synthetic image dataset was used for the training of the deep learning models. This presented a direction to mitigate the problem of securing training data, which is insufficient in the construction industry. When the proposed model was used and dimensional measurement was performed using the restored cross-sectional drawings for the curtain wall profile, a type of non-structural construction component, a high measurement accuracy of approximately 1.0 mm or less was achieved from an object distance of 10 m. The research by Park, et al. (2023) contributed to increasing the automation level for the

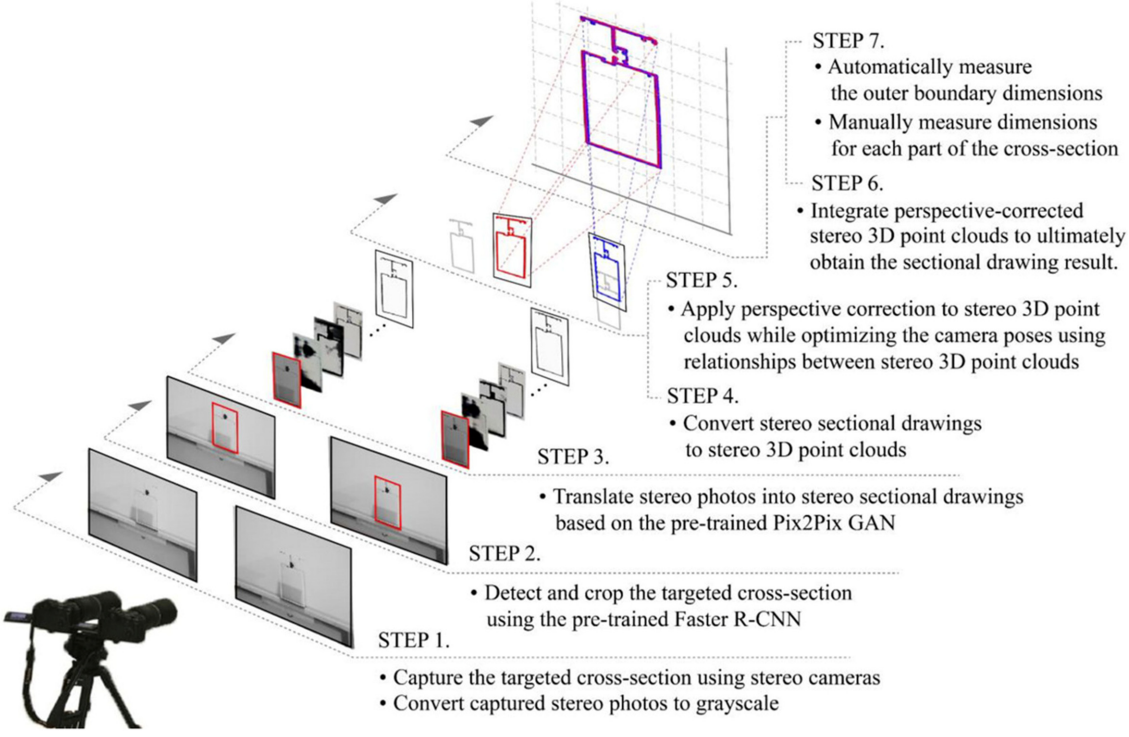


Figure 3. Overall procedure of automated cross-sectional reconstruction model. (Reprinted from Park and Park: Automated reconstruction model of a cross-sectional drawing from stereo photographs based on deep learning. Computer-aided civil and infrastructure engineering. 2023, with permission by Wiley)

dimensional quality inspection task. It is significant in that it presented a direction to reduce the cost required for training dataset construction by showing the possibility of using a synthetic dataset. However, the research normally operates only for 2D plane objects (e.g. member cross-section), and has limited applicability because the distance between stereo cameras and the target construction component must be roughly measured. While it implemented the automation of the entire process of translating stereo-captured images into cross-sectional drawings, it could not fully implement automation as the process of locating and measuring the targeted parts within the translated cross-sectional drawings remained manual.

3.2. Defects Inspection

Defects inspection is required to maintain structures in safe condition, and depends on visual and manual inspections by humans. This requires considerable time, and may involve errors as it reflects the subjective opinions of the inspector (Paterson, et al., 1997). The introduction of CV to defects inspection is expected to improve such limitations. It is possible to reduce the time, labor, and cost required for defects inspection as well as the safety accident risk by automatically detecting and classifying the defects included in images and performing segmentation of defects using deep learning models and CV.

The research by Paterson, et al. (1997) is an early study on defects inspection. Paterson, et al. (1997) captured images of the exterior of a building using a single camera and CURIO, a building inspection robot developed by the City University, London, and performed image processing to inspect the presence/absence of defects, such as cracks, on the building surface. Here, the location of the detected defects is estimated by performing image processing for the images of the robot installed on the building surface and matching the result with the drawings of the building façade. Figure 4(a) shows the inspection setup of the research while Figure 4(b) presents the photograph of CURIO. The research by Paterson, et al. (1997) is significant in that it shows the applicability of CV to the problem of finding defects and their location on the building exterior and its automation possibility. The research is applicable only to buildings with visually simple facades, and it cannot secure robustness for changes in actual environmental elements, including light reflection and shadows. In addition, it cannot be automated as it requires human intervention in the homography transformation process.

Sinha, et al. (2006) developed a crack segmentation model that uses a statistical filter for the automatic segmentation of cracks on the concrete surface. The model consists of two stages. In the first stage, crack candidates are determined using two crack detectors. In the second stage, cracks are selected from the candidates through the post-treatment processes of cleaning and linking. The model showed higher crack segmentation performance than canny edge detection, otsu's method (Otsu, 1979). Figure 5 shows the results of detecting cracks on the concrete surface using the model. Because the model depends on the statistical filter and does not have robust performance for environmental changes, such as brightness contrast, sharpness, and lighting conditions; however, its application to various types of structures or environments is difficult. Crack detection in this period before the introduction of deep learning had common limitations in that it was not robust owing to dependency on conventional CV, and it was difficult to achieve generalization performance, resulting in difficulty in application to actual construction sites.

Mirzazade, et al. (2023) proposed a model that can reconstruct a 3D model for the inside of a tunnel using photogrammetry, perform segmentation of crack defects using a deep learning model, and specify their location on the reconstructed 3D model. For the segmentation of cracks, InceptionV3 (Szegedy, et al., 2015) and U-Net (Ronneberger, et al., 2015), which are CNN models, were used. Mirzazade, et al. (2023) compared and evaluated the performance of U-Net and SegNet (Badrinarayanan, et al., 2017) in performing crack segmentation. Consequently,

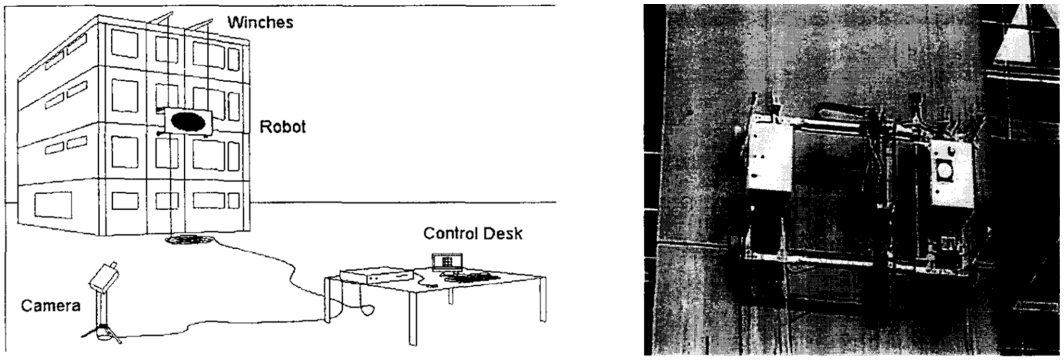


Figure 4. Building defects inspection system: (a) System setup, (b) Inspection robot, CURIO. (Reprinted from Paterson, et al.: Building inspection: Can computer vision help?. Automation in construction. 1997, 7(1), 13-20 with permission by Elsevier)

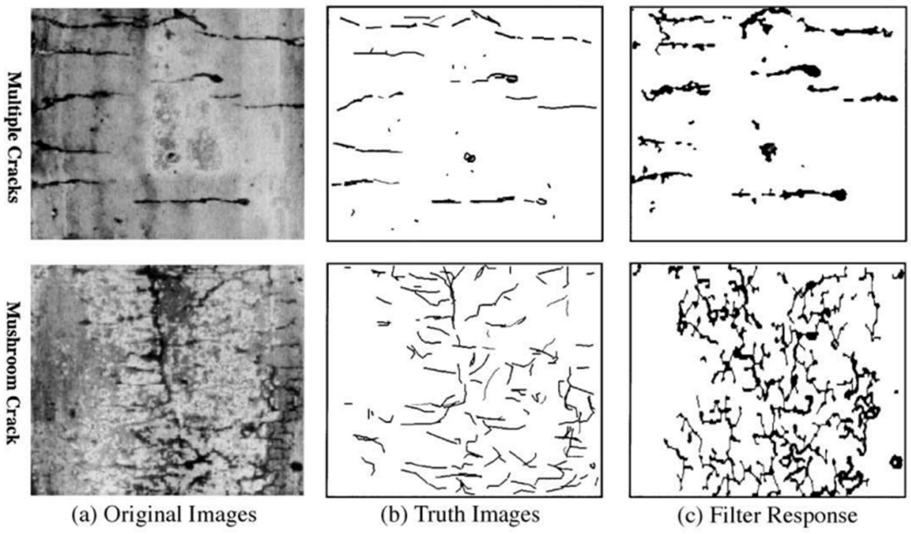


Figure 5. Automated crack detection results. (Reprinted from Sinha, et al.: Automated detection of cracks in buried concrete pipe images. Automation in construction. 2006, 15(1), 58-72 with permission by Elsevier)

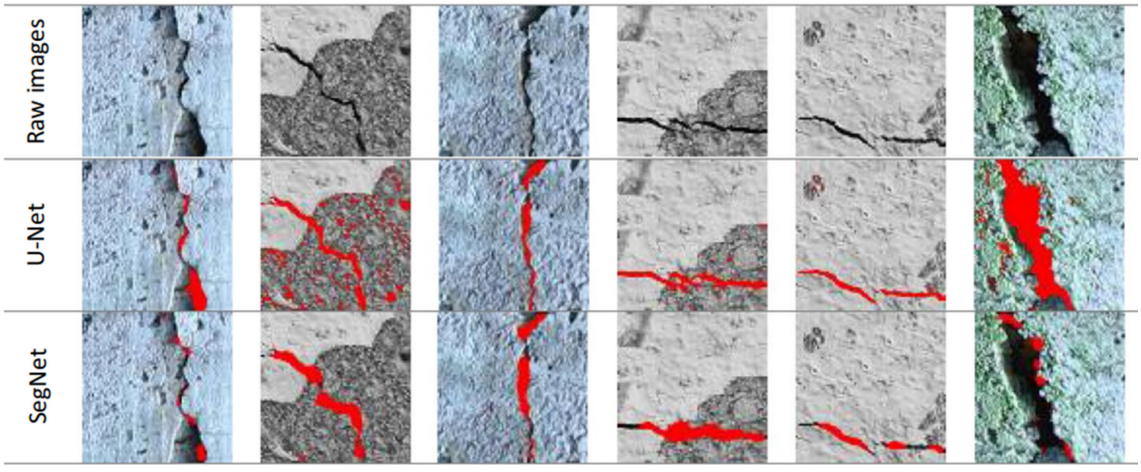


Figure 6. Crack segmentation results. (Reprinted from Mirzazade, et al.: Semi-autonomous inspection for concrete structures using digital models and a hybrid approach based on deep learning and photogrammetry. Journal of civil structural health monitoring. 2023, 1-20 used under creative commons CC-BY license)

they insisted that U-Net had higher performance (Figure 6). Because the purpose of the proposed model included the dimensional measurement of cracks, the process of translating the crack segmentation results to ortho views from a position perpendicular to the surface was included. The dimensional accuracy of the 3D model reconstructed using the proposed model was less than a centimeter from a distance of 2 m. The proposed model contributed to crack detection and segmentation automation, but it is not robust for actual environmental changes, such as shadows, dust, snow, and water droplets. In particular, its micro-crack detection performance is lowered in the presence of such environmental changes. In addition, the proposed model could not achieve full automation as it required human intervention.

3.3. Static Displacement Measurement

Static displacement is the important structural response for structural safety and serviceability evaluation. It can be used to examine whether the structure is constructed as designed and safe. It can be used to examine whether structural performance is lowered and whether reinforcement is required for old structures. The traditional static displacement measurement method uses contact-type sensors, which require considerable time and labor. The introduction of CV enables contactless and remote measurement of static displacement through the use of cameras and TLS.

As an early study on static displacement measurement, Yoneyama, et al. (2007) measured the deflection distribution of a bridge structure using digital image correlation (DIC). This study is significant in that it evaluated the applicability of the DIC technology whose applicability had been

evaluated for toy examples mainly in the laboratory environment by applying it to the actual bridge structure under the load of a heavy truck. This study shows the possibility of replacing the existing displacement measurement method that uses contact-type sensors with DIC; however, it cannot secure sufficient robustness to be continuously applied to actual environments owing to its sensitivity to changes in external environments, such as lighting and weather. It has limitations, such as severe deviations in deflection measurements, and the possibility of limiting the measurable bridge span depending on the camera's field of view because the entire bridge should be placed on one screen.

Park, et al. (2007) proposed a deflection distribution measurement model for steel beam structures using TLS (Figure 7(a)). TLS calculates 3D coordinates of points that constitute the surface of the target structure using the round-trip time of the laser pulse and measures them in the form of 3D point clouds. Contrary to the DIC-based measurement model of Yoneyama, et al. (2007) introduced above, the model of Park, et al. (2007) is robust against the influence of illumination or shadows and enables measurement even in the dark evening. The deflection measurements obtained using TLS exhibited small differences and deviations from the deflection measurements of LVDT, which were assumed to be true values. This study is significant as a pioneering study that showed the applicability of TLS to static displacement measurement along with a measurement error of less than 1.0 mm. However, TLS is costly compared to cameras and it cannot be used for dynamic measurement because it takes considerable time to capture a scene. It also requires considerable manual work that involves human intervention and the application of an appropriate numerical method to obtain the results as shown in Figure 7(b) from the measured 3D point clouds. This is labor-intensive and time-consuming, and requires advanced knowledge in related fields. Moreover, TLS has a significantly shorter measurable distance than cameras that use telephoto lenses.

Yin, et al. (2023) proposed a vision monitoring system for multipoint deflection measurement on large bridges (Figure 8(a)). This system consists of a camera network system, markers, an infrared lighting device, and a data collection/transmission device. The lighting device enables 24-hour continuous monitoring even at night without light. The applicability of the system was evaluated for a bridge with a length of approximately 1 km (Figure 8(b) and 8(c)). The deflection measurement precision was found to be at a submillimeter level. Compared to the research by Yoneyama, et al. (2007) and Park, et al. (2007) introduced above, markers are used for measurements and the installation and maintenance of the camera station and data transmission device are complex. However, the system is significant in that it achieved a high level of measurement precision even for a long-span structure with a length of approximately 1 km. The study could not sufficiently secure robustness against changes in environmental conditions.

3.4. Dynamic Displacement Measurement

Dynamic displacement measurement is required in modal analysis to understand the structural response to dynamic loads, such as earthquakes, wind, and traffic loads. CV enables remote and contactless real-time measurement of dynamic displacement using high-speed cameras.

The research by Olaszek (1999) is an early study on dynamic displacement measurement. Olaszek (1999) proposed a method for calculating dynamic characteristics by measuring the dynamic displacement over time for a specific measurement point set on a bridge structure using a camera equipped with a telephoto lens (Figure 9(a)). The measurement point was set on the bridge structure by attaching a marker with a black cross on the white background. The measurement point is recognized and tracked using CV. The study is significant in that it proposed a noncontact-type dynamic displacement measurement model and evaluated its applicability to an actual structure (Figure 9(b)). However, its applicability is practically limited because the marker must be installed in

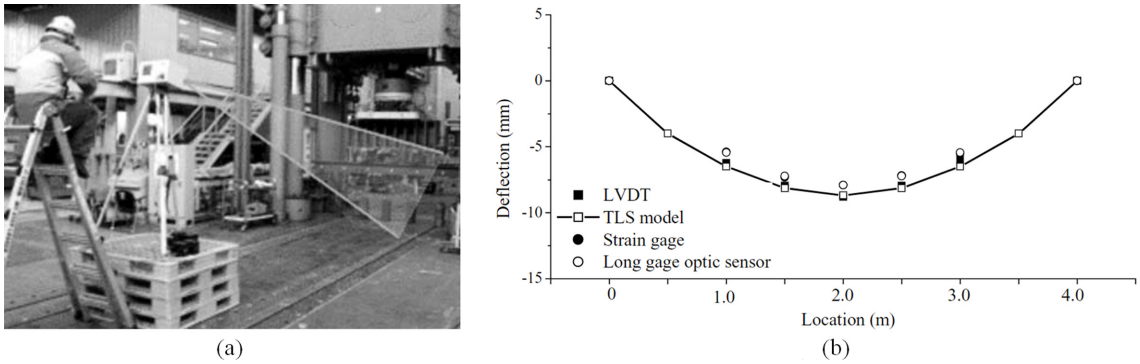


Figure 7. Steel beam deflection measurement using TLS: (a) Experimental setup, (b) Experimental result. (Reprinted from Park, et al.: New approach for health monitoring of structures: Terrestrial laser scanning. Computer-aided civil and infrastructure engineering. 2007, 22(1), 19-30 with permission by Wiley)

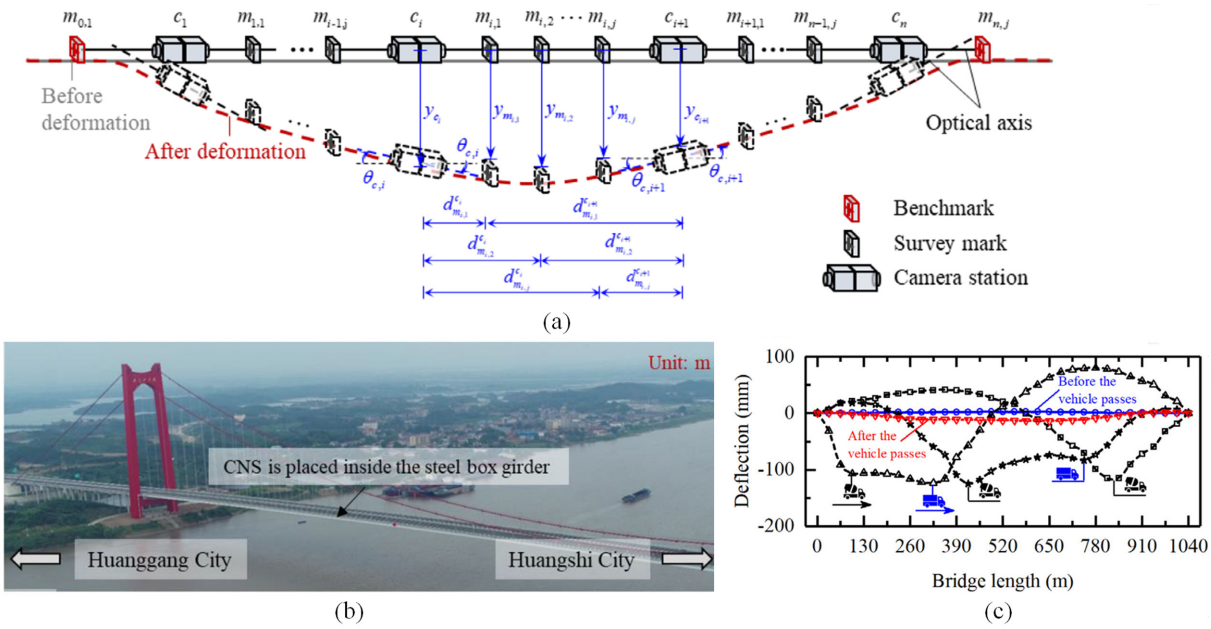


Figure 8. Long-span bridge deflection measurement: (a) Camera network system, (b) Experimental target bridge, (c) Experimental results when applying vehicle loads. (Reprinted from Yin, et al.: A vision monitoring system for multipoint deflection of large-span bridge based on camera networking. Computer-aided civil and infrastructure engineering. 2023, 38(13), 1879-1891 with permission by Wiley).

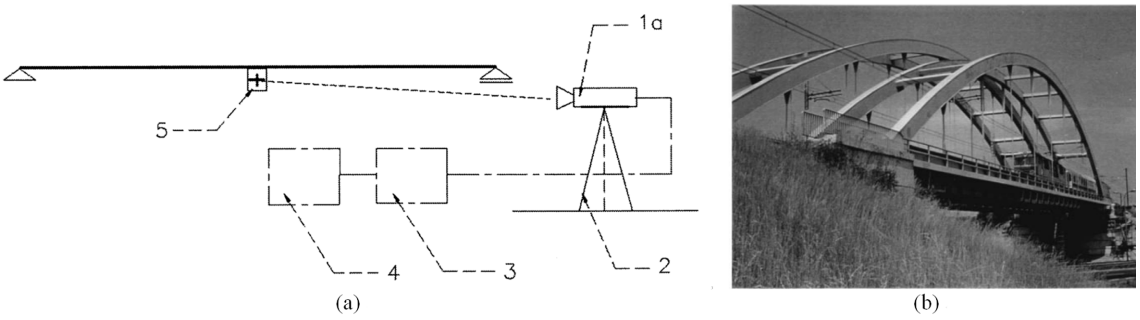


Figure 9. Bridge dynamic displacement measurement system: (a) System setup: 1a, camera and one telephotolense; 2, stand; 3, computer and image digitizer; 4, monitor; 5, measuring pattern; 6, reference pattern, (b) Experimental target bridge structure. (Reprinted from Olaszek: Investigation of the dynamic characteristic of bridge structures using a computer vision method. Measurement. 1999, 25(3), 227-236 with permission by Elsevier).

advance, and there are limitations in targeting structures with high natural frequency or evaluating high-order dynamic characteristics because images are captured at a low sampling rate of approximately 25 fps. This study achieved a measurement accuracy of approximately ±1.0 mm at a distance of 100 m using a telephoto lens, and made a significant contribution by enabling real-time measurement for parts of the structure where direct measurement is difficult owing to difficulty in access.

Park, et al. (2015) proposed a dynamic displacement measurement model using the motion capture system (MCS) (Figure 10(a)). MCS acquires 3D coordinate information for each marker by converting the 2D coordinates of markers measured using two or more cameras into 3D coordinates using triangulation. As with the model of Olaszek (1999), there is a restriction that markers must be attached to points to be measured. However, as can be observed from Figure 10(b), each marker exhibits a high measurement precision of less than ±0.1 mm for each of the x, y, and z axes. In addition, measurements are possible at a high sampling rate of approximately 100 Hz. This model is vulnerable to sunlight interference as it uses the reflection of infrared rays on markers, contrary to the model of Olaszek (1999). This causes improper measurements in outdoor environments or when sunlight passes through windows. This indicates that the model can be utilized only for the monitoring of indoor building components.

Ma, et al. (2022) proposed a model for long-term measurement of the dynamic displacement of a building structure by applying the Kalman filter to CV and accelerometer measurements Figure 11. The proposed model

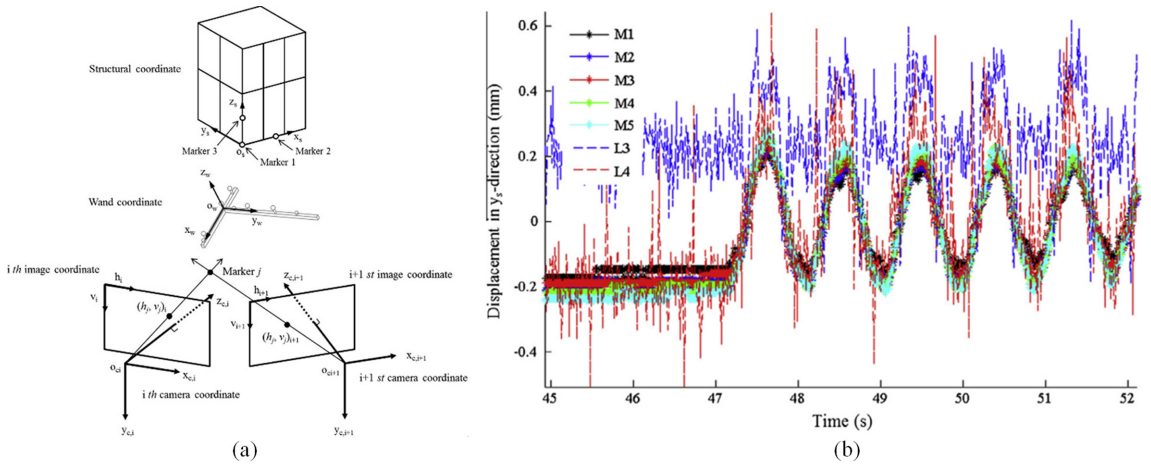


Figure 10. Dynamic displacement measurement using MCS: (a) Measurement system, (b) Measurement result. (Reprinted from Park, et al.:3D displacement measurement model for health monitoring of structures using a motion capture system. Measurement. 2015, 59, 352-362 with permission by Elsevier)

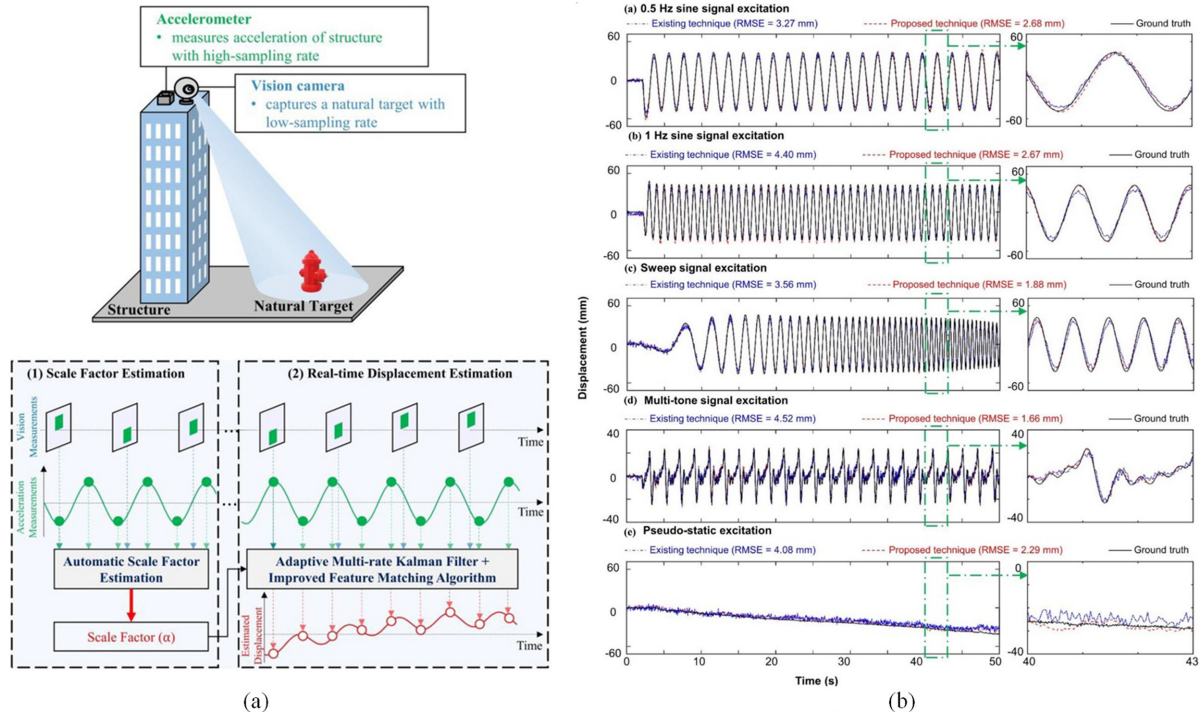


Figure 11. Dynamic displacement measurement using kalman filter and computer vision: (a) Measurement system, (b) Measurement result. (Reprinted from Ma. et al.:Real-time structural displacement estimation by fusing asynchronous acceleration and computer vision measurements. Computer-aided civil and infrastructure engineering. 2022, 37(6), 688-703 with permission by Wiley)

obtains the pixel displacement from the natural target images captured with a camera using the improved feature-matching algorithm. The scale factor to convert this into the physical quantity for the actual distance is estimated through the relationship between accelerometer measurements and CV. The Kalman filter is applied to improve the accuracy of dynamic displacement measurement. The proposed method was tested for the shaking table, four-story building model, and pedestrian steel box girder bridge, resulting in an absolute error in displacement measurement within a maximum of 3 mm. A measurement method at this level is significant because precise measurement of the displacement of a building is a difficult problem in SHM research. However, it has limitations, such as the need to manually set the region of interest (ROI) at the beginning, the possibility of considering only one degree of freedom during displacement measurement, insufficient robustness against changes in

various external environmental factors, and the impossibility of monitoring at night.

4. Conclusions

In this study, CV-based SHM was classified into four major areas. Research cases in each area were introduced, and their effects and limitations were summarized. The introduction of CV and deep learning has demonstrated potential for contributing to the automation of SHM, enabling the avoidance of conventional contact-type sensors, and enhancing both accuracy and precision. Recent related studies show the significant performance of CV-based SHM in areas, such as dimensional quality inspection, defect inspection, static displacement measurement, and dynamic displacement measurement. However, owing to the diversity and complexity of construction sites securing robustness for applicability to construction sites is challenging. Most studies have limited applicability to actual sites because they are focused on experimental verification under limited conditions. Therefore, future research on CV-based SHM must focus on achieving robust performance and full automation for actual construction sites. Implementation of robust operation and full automation even in complex and dynamic real-world construction environments remains the final goal of CV-based SHM even in the deep learning era.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government Ministry of Science, ICT & Future Planning (MSIP) (No. 2021R1A2C3008989 and 2018R1A5A1025 137).

References

Alcantarilla, P. F., Bartoli, A. and Davison, A. J. (2012). "Kaze Features." Computer Vision - ECCV 2012, Berlin, Heidelberg, 2012//, 214-227.

Azhar, A. S., Kudus, S. A., Jamadin, A., Mustaffa, N. K. and Sugiura, K. (2023). "Recent Vibration-Based Structural Health Monitoring on Steel Bridges: Systematic Literature Review." *Ain Shams Engineering Journal*, 102501, https://doi.org/10.1016/j.asej.2023.102501.

Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017). "Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.

Bay, H., Tuytelaars, T. and Van Gool, L. (2006). "Surf: Speeded up Robust Features." Computer Vision – ECCV 2006, Berlin, Heidelberg, 2006//, 404-417.

Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A. and Torr, P. H. (2016). "Fully-Convolutional Siamese Networks for Object Tracking." Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14, 850-865.

Canny, J. (1986). "A Computational Approach to Edge Detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 679-698, 10.1109/TPAMI.1986.4767851.

Choi, J., Kang, D. S., Lee, M. G., Bae, S. G., Hong, T., Lee, D.-E. and Park, H. S. (2023). "Vibration Safety Evaluation Model and Sensor Network-Based Monitoring System for Coke Drums in Operation." *Journal of Asian Architecture and Building Engineering*, 22(3), 1399-1412, 10.1080/13467581.2022.2085719.

Dalal, N. and Triggs, B. (2005). "Histograms of Oriented Gradients for Human Detection." 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 20-25 June 2005, 886-893 vol. 881, 10.1109/CVPR.2005.177.

Dong, C.-Z. and Catbas, F. N. (2020). "A Review of Computer Vision–Based Structural Health Monitoring at Local and Global Levels." *Structural Health Monitoring*, 20(2), 692-743, 10.1177/1475921720935585.

Duda, R. O. and Hart, P. E. (1972). "Use of the Hough Transformation to Detect Lines and Curves in Pictures." Commun. ACM, 15(1), 11-15, 10.1145/361237.361242.

Guo, J., Liu, P., Xiao, B., Deng, L. and Wang, Q. (2024). "Surface Defect Detection of Civil Structures Using Images: Review from Data Perspective." *Automation in Construction*, 158(105186, https://doi.org/10.1016/j.autcon.2023.105186.

Han, K. K. and Golparvar-Fard, M. (2017). "Potential of Big Visual Data and Building Information Modeling for Construction Performance Analytics: An Exploratory Study." *Automation in Construction*, 73(184-198, https://doi.org/10.1016/j.autcon.2016.11.004.

Harris, C. and Stephens, M. (1988). "A Combined Corner and Edge Detector." *Alvey vision conference*, 10-5244.

He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017). "Mask R-Cnn." *Proceedings of the IEEE international conference on computer vision*, 2961-2969.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). "Deep Residual Learning for Image Recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A. (2017). "Image-to-Image Translation with Conditional Adversarial Networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125-1134.

Jianbo, S. and Tomasi (1994). "Good Features to Track." *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 21-23 June 1994, 593-600, 10.1109/CVPR.1994.323794.

Kardovskyi, Y. and Moon, S. (2021). "Artificial Intelligence Quality Inspection of Steel Bars Installation by Integrating Mask R-Cnn and Stereo Vision." *Automation in Construction*, 130(103850.

Khuc, T. and Catbas, F. N. (2018). "Structural Identification Using Computer Vision-Based Bridge Health Monitoring." *Journal of Structural Engineering*, 144(2), 04017202, 10.1061/(ASCE)ST.1943-541X.0001925.

Kim, Y., Park, J. S., Oh, B. K., Cho, T., Kim, J. M., Kim, S.

H. and Park, H. S. (2019). "Practical Wireless Safety Monitoring System of Long-Span Girders Subjected to Construction Loading a Building under Construction." *Measurement*, 146(524-536, 10.1016/j.measurement.2019.05.110.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). "Imagenet Classification with Deep Convolutional Neural Networks." *Advances in neural information processing systems*, 25.

Kwan Oh, B., Seon Park, H. and Glisic, B. (2023). "Time-Dependent Structural Response Estimation Method for Concrete Structures Using Time Information and Convolutional Neural Networks." Engineering Structures, 275(115193, https://doi.org/10.1016/j.engstruct.2022.115193.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). "Deep Learning." *nature*, 521(7553), 436-444.

Lee, H. M., Kim, J. M., Sho, K. and Park, H. S. (2010). "A Wireless Vibrating Wire Sensor Node for Continuous Structural Health Monitoring." *Smart Materials and Structures*, 19(5), 055004, 10.1088/0964-1726/19/5/055004.

Longuet-Higgins, H. C. and Prazdny, K. (1980). "The Interpretation of a Moving Retinal Image." *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 208(1173), 385-397.

Lowe, D. G. (1999). "Object Recognition from Local Scale-Invariant Features." *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 20-27 Sept. 1999, 1150-1157 vol.1152, 10.1109/ICCV.1999.790410.

Lucas, B., D and Kanade, T. (1981). "An Iterative Image Registration Technique with an Application to Stereo Vision." IJCAI'81: 7th international joint conference on Artificial intelligence, Vancouver, Canada, 1981-08-24, 674-679.

Ma, Z., Choi, J. and Sohn, H. (2022). "Real-Time Structural Displacement Estimation by Fusing Asynchronous Acceleration and Computer Vision Measurements." *Computer-Aided Civil and Infrastructure Engineering*, 37(6), 688-703, https://doi.org/10.1111/mice.12767.

Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Mahdiyar, A. and Gonzalez, V. (2023). "End-to-End Point Cloud-Based Segmentation of Building Members for Automating Dimensional Quality Control." *Advanced Engineering Informatics*, 55(101878.

Mirzazade, A., Popescu, C., Gonzalez-Libreros, J., Blanksvärd, T., Täljsten, B. and Sas, G. (2023). "Semi-Autonomous Inspection for Concrete Structures Using Digital Models and a Hybrid Approach Based on Deep Learning and Photogrammetry." *Journal of Civil Structural Health Monitoring*, 10.1007/s13349-023-00680-x.

Mishra, M., Lourenço, P. B. and Ramana, G. V. (2022). "Structural Health Monitoring of Civil Engineering Structures by Using the Internet of Things: A Review." *Journal of Building Engineering*, 48(103954, https://doi.org/10.1016/j.jobe.2021.103954.

Oh, B. K., Glisic, B., Kim, Y. and Park, H. S. (2020). "Convolutional Neural Network–Based Data Recovery Method for Structural Health Monitoring." Structural Health Monitoring, 19(6), 1821-1838, 10.1177/1475921719897571.

Oh, B. K., Hwang, J. W., Kim, Y., Cho, T. and Park, H. S. (2015). "Vision-Based System Identification Technique for Building Structures Using a Motion Capture System." *Journal of Sound and Vibration*, 356(72-85, https://doi.org/10.1016/j.jsv.2015.07.011.

Oh, B. K., Kim, K. J., Kim, Y., Park, H. S. and Adeli, H. (2017). "Evolutionary Learning Based Sustainable Strain Sensing Model for Structural Health Monitoring of High-Rise Buildings." *Applied Soft Computing*, 58(576-585, https://doi.org/10.1016/j.asoc.2017.05.029.

Oh, B. K. and Park, H. S. (2022). "Urban Safety Network for Long-Term Structural Health Monitoring of Buildings Using Convolutional Neural Network." Automation in Construction, 137(104225.

Olaszek, P. (1999). "Investigation of the Dynamic Characteristic of Bridge Structures Using a Computer Vision Method." *Measurement*, 25(3), 227-236, https://doi.org/10.1016/S0263-2241(99)00006-8.

Otsu, N. (1979). "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems*, Man, and Cybernetics, 9(1), 62-66, 10.1109/TSMC.1979.4310076.

Paneru, S. and Jeelani, I. (2021). "Computer Vision Applications in Construction: Current State, Opportunities & Challenges." Automation in Construction, 132(103940.

Park, H. S., Lee, H. M., Adeli, H. and Lee, I. (2007). "A New Approach for Health Monitoring of Structures: Terrestrial Laser Scanning." *Computer-Aided Civil and Infrastructure Engineering*, 22(1), 19-30, https://doi.org/10.1111/j.1467-8667.2006.00466.x.

Park, H. S. and Oh, B. K. (2018). "Real-Time Structural Health Monitoring of a Supertall Building under Construction Based on Visual Modal Identification Strategy." *Automation in Construction*, 85(273-289.

Park, H. S., Park, J. S. and Oh, B. K. (2017). "Vision-Based Stress Estimation Model for Steel Frame Structures with Rigid Links." *Measurement Science and Technology*, 28(7), 075104, 10.1088/1361-6501/aa6f50.

Park, J. S., Oh, B. K., Choi, S. W. and Park, H. S. (2016). "Sensor-Free Stress Estimation Model for Steel Beam Structures Using a Motion Capture System." *IEEE Sensors Journal*, 16(8), 2701-2713, 10.1109/JSEN.2016.2519033.

Park, J. S. and Park, H. S. (2023). "Automated Reconstruction Model of a Cross-Sectional Drawing from Stereo Photographs Based on Deep Learning." *Computer-Aided Civil and Infrastructure Engineering*, n/a(n/a), https://doi.org/10.1111/mice.13083.

Park, S. W., Park, H. S., Kim, J. H. and Adeli, H. (2015). "3d Displacement Measurement Model for Health Monitoring of Structures Using a Motion Capture System." Measurement, 59(352-362, https://doi.org/10.1016/j.measurement.2014.09.063.

Paterson, A. M., Dowling, G. R. and Chamberlain, D. A. (1997). "Building Inspection: Can Computer Vision Help?" *Automation in Construction*, 7(1), 13-20, https://doi.org/10.1016/S0926-5805(97)00031-9.

Prince, S. J. (2012). Computer Vision: Models, Learning, and Inference, Cambridge University Press.

Qi, C. R., Su, H., Mo, K. and Guibas, L. J. (2017). "Pointnet: Deep Learning on Point Sets for 3d Classification and Segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652-660.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016).

"You Only Look Once: Unified, Real-Time Object Detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779-788.

Ren, S., He, K., Girshick, R. and Sun, J. (2015). "Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks." *Advances in neural information processing systems*, 28(

Roberts, L. G. (1963). Machine Perception of Three-Dimensional Solids, Massachusetts Institute of Technology.

Ronneberger, O., Fischer, P. and Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation." Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 234-241.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011). "Orb: An Efficient Alternative to Sift or Surf." 2011 *International Conference on Computer Vision*, 6-13 Nov. 2011, 2564-2571, 10.1109/ICCV.2011.6126544.

Sabato, A., Dabetwar, S., Kulkarni, N. N. and Fortino, G. (2023). "Noncontact Sensing Techniques for Ai-Aided Structural Health Monitoring: A Systematic Review." *IEEE Sensors Journal*, 23(5), 4672-4684, 10.1109/JSEN.2023.3240092.

Shapiro, L. G. (2020). "Computer Vision: The Last 50 Years." *International Journal of Parallel, Emergent and Distributed Systems*, 35(2), 112-117, 10.1080/17445760.2018.1469018.

Simonyan, K. and Zisserman, A. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv preprint arXiv:1409.1556.

Sinha, S. K. and Fieguth, P. W. (2006). "Automated Detection of Cracks in Buried Concrete Pipe Images." *Automation in Construction*, 15(1), 58-72, https://doi.org/10.1016/j.autcon.2005.02.006.

Sobel, I. and Feldman, G. (1968). "A 3×3 Isotropic Gradient Operator for Image Processing." a talk at the Stanford Artificial Project in, 271-272.

Spencer, B. F., Hoskere, V. and Narazaki, Y. (2019). "Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring." *Engineering*, 5(2), 199-222, https://doi.org/10.1016/j.eng.2018.11.030.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015). "Going Deeper with Convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.

Tomasi, C. and Kanade, T. (1991). "Detection and Tracking of Point Features." Carnegie Mellon University Technical Report CMU-CS-91-132.

Ullman, S. (1979). "The Interpretation of Structure from Motion." *Proceedings of the Royal Society of London*. Series B. Biological Sciences, 203(1153), 405-426.

Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M. and Wang, X. (2021). "Computer Vision Techniques in Construction: A Critical Review." *Archives of Computational Methods in Engineering*, 28(5), 3383-3397, 10.1007/s11831-020-09504-3.

Ye, W., Ren, J., Zhang, A. A. and Lu, C. (2023). "Automatic Pixel-Level Crack Detection with Multi-Scale Feature Fusion for Slab Tracks." *Computer-Aided Civil and Infrastructure Engineering*, 38(18), 2648-2665, https://doi.org/10.1111/mice.12984.

Yin, Y., Yu, Q., Hu, B., Zhang, Y., Chen, W., Liu, X. and Ding, X. (2023). "A Vision Monitoring System for Multi-point Deflection of Large-Span Bridge Based on Camera Networking." *Computer-Aided Civil and Infrastructure Engineering*, 38(13), 1879-1891, https://doi.org/10.1111/mice.13050.

Yoneyama, S., Kitagawa, A., Iwata, S., Tani, K. and Kikuta, H. (2007). "Bridge Deflection Measurement Using Digital Image Correlation." *Experimental Techniques*, 31(1), 34-40, 10.1111/j.1747-1567.2006.00132.x.

Yu, X., Fu, Y., Li, J., Mao, J., Hoang, T. and Wang, H. (2023). "Recent Advances in Wireless Sensor Networks for Structural Health Monitoring of Civil Infrastructure." *Journal of Infrastructure Intelligence and Resilience*, 100066, https://doi.org/10.1016/j.iintel.2023.100066.

Yun, D. Y., Kim, D., Kim, M., Bae, S. G., Choi, J. W., Shim, H. B., Hong, T., Lee, D.-E. and Park, H. S. (2021). "Field Measurements for Identification of Modal Parameters for High-Rise Buildings under Construction or in Use." *Automation in Construction*, 121(103446, https://doi.org/10.1016/j.autcon.2020.103446.

Zhu, J.-Y., Park, T., Isola, P. and Efros, A. A. (2017). "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." *Proceedings of the IEEE international conference on computer vision*, 2223-2232.