IMAGE-BASED HYBRID STRUCTURAL HEALTH MONITORING THROUGH

ARTIFICIAL INTELLIGENCE

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Xin Bai

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Civil, Construction and Environmental Engineering

June 2022

Fargo, North Dakota

North Dakota State University
Graduate School

Title

IMAGE-BASED HYBRID STRUCTURAL HEALTH MONITORING
THROUGH ARTIFICIAL INTELLIGENCE

By

Xin Bai

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota

State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Yang, Mijia

Chair

Tangpong, Annie

Kong, Jun

Yellavajjala, Ravi

Sun, XIN

Approved:

06/27/2022	Xufeng Chu
Date	Department Chair

ABSTRACT

Bridges are widely used in human life. Understanding structural performance, assessing structural conditions, and providing in-time decision are crucial components in structural health monitoring (SHM), to avoid catastrophic events and improve public safety. However, traditional SHM needs traffic closure, extensive sensor deployment, and in-contact measurements. The main purpose of this thesis is to develop a vision-based sensor of high accuracy that does not need artificial targets. When the vibration of the UAV itself is removed, the UAV is a convenient method to record video of the vibrations. Based on the recorded images and vibration data, a new deep learning method is developed and used to analyze vibrations of the structure and detect damage locations and conditions automatically.

In the thesis, a non-contact vision sensor system for monitoring structural displacements with an advanced Zernike subpixel edge detection technique is first suggested. A new method to filter the effect of camera motions through background templates is proposed in the study. Several experiments on the MTS machine were performed with different frequencies and amplitudes to verify the method. The results show that filtering of vibrations of the camera significantly improves the displacement monitoring accuracy from 53.0% to 97.0%. Three translations and three rotations of the unmanned aerial vehicle (UAV) were derived through the suggested fast Normalized Cross Correlation (NCC) based template matching method, and their effect on the monitored structural displacement is analyzed. To verify the concept, a series of lab and field experiments were performed. Excellent precision and consistency were obtained for the UAV monitored displacement, the MTS piston motion, and the fixed camera derived displacement.

Further in the thesis, a novel deep learning-based structural health monitoring method was developed, which could detect damages using both defects and vibration data. Two ABAQUS

models on a beam and an ABAQUS model on a truss were conducted to test if the proposed CNN model could detect damages successfully. Seven transfer learning methods were compared on detecting crack images. From the outputs of the deep learning models, it is apparent that the AlexNet CNN model with defect images shows higher accuracy in estimating damage status.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xvi

1. INTRODUCTION

1.1. Background

Engineering infrastructure consists part of human life's necessities, such as roadways and bridges [1, 2]. Their safety and continuous operations are critical, as aging problems and damage accumulations often lead to serious public safety issues and large economic loss.

In the past 20 years, image-based displacement sensors become increasingly attractive because of their advantages of low cost and efficiency. The unmanned aerial vehicle (UAV) can be used conveniently to detect both static characteristics such as cracks and dynamic behaviors such as natural frequencies and vibrations under different loads using image-based sensors. When UAV is used to detect dynamic responses of structures, the most important thing is filtering out vibrations of the UAV because the displacement of structures is always very small compared to movements of the UAV due to its flying motions and the environment. Therefore, a UAV based damage detection method with the capacity to filter UVA's self-vibration is needed. Deep learning is then used to classify the damage condition based on the filtered vibrations and the images captured by the UAV.

1.2. Problem Statement

Most existing structural health monitoring (SHM) techniques are based on measured acceleration data. Such practice, however, is highly expensive to operate, mainly due to the cumbersome, time-consuming, expensive installation of sensors, and their data acquisition systems. As an emerging noncontact method, the vision-based displacement sensor systems have been more attractive to many researchers. However, most existing vision-based methods need physical access to the structure to set up artificial targets that have a higher contrast pattern and are easy to track. In this thesis, a new vision-based displacement monitoring method is suggested,

in which three translations and three rotations of the UAV were derived through the suggested Normalized Cross Correlation (NCC) based template matching computer vision method and their effect on the monitored structural displacement is filtered out. Moreover, the vibration data captured by the UAV always accompanies the target images. To detect damages using global vibration data and local image information, it is important to develop a deep learning system of high accuracy to utilize both vibration data and images.

1.3. Literature Review

1.3.1. Vibration-based Structural Health Monitoring

Conventionally, contact-type displacement sensors such as linear variable differential transformers (LVDTs) and laser-based displacement sensors [1] are employed for measuring structural displacements. However, such sensors measure relative displacements and need a stationary point to be set as a reference on the measured structure. Usually, for a large-scale civil engineering structure, such as a bridge spanning a wide river, a stationary reference is difficult to set up. Moreover, these sensors need to be wired to a data acquisition board with power supply and the package box and wiring are cumbersome and costly for monitoring a large structure. On the other hand, accelerometers, which do not require a stationary reference, can be used to measure displacement through double integration of acceleration data, however, they are usually accompanied by numerical integration errors. Global positioning systems (GPS)) [2] offer potential advantages due to its non-contact nature, but high cost and low accuracy of these systems prevent their wide applications in monitoring civil engineering structures.

1.3.2. Noncontact Vision-Based Sensor System for Displacement Measurement

In the past 20 years, image-based displacement sensors become increasingly attractive because of their advantages of low cost and efficiency. To adopt such a computer vision method

2

in structural health monitoring, template matching is typically used to trace displacements of structures. The template matching method could find the position of a template in the target image, and sometimes the bolt group or other connections in the target structure could be used as templates. Olaszek [3] first used the template matching method to obtain the dynamic characteristics of bridges. Wahbeh et al. [4] proposed a novel vision-based approach to detect the absolute displacement time history at selected locations of civil infrastructure systems using a high-resolution camera and a laser sensor as comparison. An experiment in the lab was conducted to successfully verify the method. Lee et al. [5] developed a real-time vision-based system for the structural displacement measurement of bridges and made the displacement monitoring right at the site. Park et al. [6] proposed a novel vision-based displacement measurement technique using a partition approach to detect horizontal displacement of a high-rise building structure. Ho et al. [7] presented an advanced multipoint vision-based system to detect dynamic displacements of civil infrastructures. Several shaking table tests were performed to verify the effectiveness of the proposed system and the displacements obtained agreed well with those from a conventional sensor. Sładek [8] detected the in-plane displacement using template matching on a beam. Wu et al. [9] developed a vision system that uses digital image processing and computer vision technologies to monitor the 2D plane vibrations of a reduced-scale frame mounted on a shake table. Henke et al. [10] took the LED light as the visual target and measured the deformation of building structures by use of the digital image processing technique. On the other front, some researchers advanced the original template matching method using normalization and made it more efficient in displacement monitoring. Di et al. [11] described the mapping of a modified template matching algorithm based on NCC. Wei et al. [12] proposed a fast pattern matching algorithm using NCC by combining adaptive multilevel partition with the winner update scheme to make the

3

original NCC more effective. Lei et al. [13] conducted a fast NCC-based template matching method which could reduce the computation time and increase efficiency significantly using a modified local search algorithm when motions of the object are small and could be negligible. If there are several suitable templates in one image, template matching could be used to obtain vibrations of these points at the same time. Multiple templates could be obtained in one region of interest (ROI) to detect displacements of multi-locations at the same time. Lin et al. [14] proposed a monitoring system using videogrammetry and terrestrial laser scanning to obtain the dynamic behavior of the as-built membrane roof structures at multiple locations. Jurjo et al. [15] presented a digital image processing technique to obtain displacements on thin members with several templates. However, if there are no suitable nature targets on the measured points and it is difficult and dangerous to attach artificial targets on the surface of the structure, such as bridges over a big river or a canyon, there is a need for a template-free vision-based displacement monitoring method, which will advance this field tremendously. So, in this thesis, an edge detection-based method was researched and used for structural condition assessment.

Besides the template matching method, other computer vision methods such as edge detection and digital image correlation (DIC) could detect displacement with high accuracy. Pan et al. [16] enhanced DIC for surface deformation measurement at the macroscopic to nano scale. Chan et al. [17] proposed a charge-coupled device (CCD) camera-based method to measure the vertical displacement of bridges. Kohut et al. [18] proposed a vision-based method to calculate the displacement field of the analyzed structure using the so-called digital image correlation coefficient. Cigada et al. [19] obtained the displacement responses using three vision methods (Pattern Matching, Edge Detection, and Digital Image Correlation) and two types of cameras to acquire the motions of multiple targets fixed on a railway bridge during the passage of a train. Ri

et al. [20] proposed a nontarget vision-based method for cable force estimation using handheld shooting through smartphone cameras. The edge of the cable is selected as a target and was obtained by the edge detection method. Vibrations of the cable were obtained based on the changes in its locations in the ROI of video image sequences, captured by a smartphone camera.

1.3.3. UAV-based Sensor System for Displacement Measurement

UAV is an attractive tool to perform structural health monitoring in detecting both static characteristics such as cracks and dynamic ones such as natural frequencies and vibrations under different loads. It is easy to fly UAVs at a close position to the structures even if structures are in an inaccessible area such as a bridge over a wide river or a great canyon. When the UAV is used to detect dynamic characteristics of structures, the most important thing is filtering out vibrations of the UAV because the displacement of structures is always very small compared to movements of the UAV due to its flying motion and environment. Due to motions of the UAV, very large effect on the results of vibrations of the structure could be generated. Yoon et al. [21] proposed an image-based method to detect relative displacements between stories using UAV. A fixed camera was used to compare the results with those from UAVs. The method could get rid of the three translations of the UAV without calculating the camera motion. This method could only be used to detect structures with several stories when the rotations of the camera could be negligible. The rotations of the UAV caused by wind would cause different translations of different stories, thus the real displacement of the structure could not be obtained using this method. Yoon et al. [22] also used a UAV to detect the vibration of a structure in the lab. To obtain the vibrations of the UAV, a checking barcode was used to be the background point. In the real environment, it might be difficult to attach a manmade checking barcode at the background of the target structure. In summary, the main challenge of the vision-based displacement sensor is to filter out the motions

of the camera (3 translations and 3 rotations). Some researchers used relative displacement or a manmade checking board to eliminate motions of the camera. However, the apparent distortion due to the angle between structure plane and the camera was not considered.

1.3.4. Deep Learning-based Damaged Detection Using Vibration Data and Defects' Images

Structural damages, such as corrosions and cracks, always represents degraded structural health conditions. In the last several decades, visual inspection by trained workers is the main approach to detect structural damages, which is time-consuming and costly. Because of the significant advantages of computer-vision-based sensors, researchers are increasingly attracted to adopt them. Civil structures, particularly concrete structures, are likely to suffer cracks due to changing load conditions and other reasons. Cracks in concrete are always expressed as lines or curve lines with different orientations and intensities. Compared to the background color of concrete, these lines are darker and connected. Through image processing methods like the edge detection, crack detection could be conducted using properly prescribed thresholds. The image binarization method and the sequential image processing method are the two main methods to detect defects in structures, especially for cracks. Liu [23] proposed an image binarization method to detect cracks and had satisfactory accuracy. Ebrahimkhanlou [24] researched the sequential image processing method that was carried out to detect cracks in images of concrete surfaces successfully. The vision-based crack detection methods for cracks in concrete structures were mostly based on edge detection algorithms through some filters, such as Roberts, Prewitt, Sobel, and Gaussian [25]. Usually, these methods are time-consuming and could only detect the existence of cracks in the image. These methods could not evaluate the severity of cracks directly.

Some researchers intended to use deep learning to classify and evaluate structural defects like cracks and corrosion. Rawat and Wang [26] provided the background of the CNN method for

classifying the images. Chen and Jahanshahi [27] proposed a CNN-based crack detection system and used a Naïve Bayes data fusion scheme to extract cracks from the video frames captured for the structure. Kim et al. [28] proposed a faster CNN with image binarization to obtain the location of the crack in the pixel level precision. Li et al. [29] trained and tested a deep-learning model using 6000 tunnel crack images and compared different crack mechanism networks. Their model significantly outperformed basic U-net, fully convolutional networks (FCN), SegNet, and multi-scale fusion crack detection (MFCD) for detecting cracks in tunnel through noisy images. Soloviev et al. [30], Li et al. [31], Tong et al. [32], and Fan et al. [33] proposed deep CNN models to detect and find the cracks and their lengths and sizes on pavement surfaces.

Structure assessment could also use dynamic characteristics such as the natural frequency, the damping ratio, and the mode shape. Moreover, damage locations could be calculated using these dynamic characteristics, because, in theory, any structural damage would cause the change in the mass and stiffness distributions of structures, lead to the change in natural frequency, damping ratio, and mode shape [34-36]. Vibration-based structural damage detection methods with system identification could build the correlations between the vibration features and the damage information [37-39]. However, natural frequencies are not sensitive to minor damages and could be easily contaminated by environmental factors. This issue will be resolved using the deep learning method proposed in this thesis because the developed deep learning method could connect the raw vibration data with the damage information directly through data fusion.

Lin, Nie, and Ma [40] proposed a CNN-based deep learning method, containing six convolutional layers and three maximum pooling layers. This CNN-based system was then trained using the raw vibration data of a finite element beam and the damage was detected successfully

with high accuracy (94.57%). Some researchers used the CNN-based method and found the damages caused by losing bolts of a steel truss with high curacy [41].

Some researchers have proposed deep learning-based defects detection using images as input data and some other researchers have validated the performance of deep learning-based damage detection using vibration data. However, there is no research to evaluate a civil structure comprehensively by using both defect's images and vibration data in a deep learning-based model. In this thesis, a novel deep-learning model was developed to evaluate civil structures comprehensively through infusion of static images and vibration signals.

1.4. Organization of Dissertation

The whole thesis will include 7 chapters. A literature review will be conducted in Chapter 1, followed by Chapter 2 that presents a non-contact vision sensor system for monitoring structural displacements with an advanced Zernike subpixel edge detection technique. Experiments on MTS machine is performed to verify the proposed displacement sensor. Chapter 3 presents a field testing on a street sign to obtain its vibrations and a lab experiment on a steel plate to detect the damage inside. Chapter 4 presents a new method to filter the effect of cameras' motions through background templates when the vibration of cameras could not be ignored. Moreover, the effect of template sizes and brightness on the accuracy of monitored displacement are analyzed and compared. Several experiments on the MTS machine are performed with different frequencies and amplitudes to verify the method. Chapter 5 presents a method to filter out the effect of vibrations of the UAV. A fast NCC-based template matching method is created, which could accelerate the original NCC-based template matching method significantly. Experiments on the MTS and a field test on a shaker are performed to verify the proposed method. Chapter 6 provides a novel deep learning-based structural health monitoring method. In this section, two ABAQUS models on a

beam with different segmentations and an ABAQUS model on a truss are performed and vibration data is collected. Then these vibration data is combined with structural defects' images. This database is used to train the proposed deep learning models with high accuracy. Finally, Chapter 7 summarizes the primary conclusions and address future research directions.

1.5. Reference

1.   Nassif, H.H., Gindy, M., and Davis, J. (2005) Comparison of laser doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. NDT E Int., 38, 213–218.

2.   Stephen, G., Brownjohn, J., and Taylor. (1993) Measurements of static and dynamic displacement from visual monitoring of the Humber Bridge. Eng. Struct., 15(3) 197-208.

3.   Olaszek, P. (1999) Investigation of the dynamic characteristic of bridge structures using a computer vision method. Measurement. 25, 227–236.

4.   Wahbeh, A.M., Caffrey, J.P., and Masri, S.F. (2003) A vision-based approach for the direct measurement of displacements in vibrating systems. Smart Mater. Struct. 12, 785.

5.   Lee, J. J., and Shinozuka, M. (2006) A vision-based system for remote sensing of bridge displacement. Ndt & E International, 39(5) 425-431.

6.   Park, J. W., Lee, J. J., Jung, H. J., and Myung, H. (2010) Vision-based displacement measurement method for high-rise building structures using partitioning approach. Ndt & E International, 43(7) 642-647.

7.   Ho, H. N., Lee, J. H., Park, Y. S., and Lee, J. J. (2012) A synchronized multipoint vision-based system for displacement measurement of civil infrastructures. The Scientific World Journal.

8. Sładek, J., Ostrowska, K., Kohut, P., Holak, K., G ¸aska, A., and Uhl, T. (2013) Development of a vision-based deflection measurement system and its accuracy assessment. Measurement. 46, 1237–1249.

9. Wu, L.J., Casciati, F., and Casciati, S. (2014) Dynamic testing of a laboratory model via vision-based sensing. Eng. Struct. 60, 113–125.

10. Henke, K., Pawlowski, R., and Schregle, P. (2015) Use of digital image processing in the monitoring of deformations in building structures. J. Civ. Struct. 5(2) 141-152.

11. Di Stefano, L., Mattoccia, S., and Tombari, F. (2005) Speeding-up NCC-based template matching using parallel multimedia instructions. In Seventh International Workshop on Computer Architecture for Machine Perception (CAMP'05) (pp. 193-197). IEEE.

12. Wei, S. D., and Lai, S. H. (2008) Fast template matching based on normalized cross correlation with adaptive multilevel winner update. IEEE Transactions on Image Processing, 17(11) 2227-2235.

13. Lei, Xiujun, Yi Jin, and Jie Guo. (2015) Vibration extraction based on fast NCC algorithm and high-speed camera. Applied optics 54.27 8198-8206.

14. Lin, S. Y., Mills, J. P., and Gosling, P. D. (2008) Videogrammetric monitoring of as-built membrane roof structures. The Photogrammetric Record, 23(122) 128-147.

15. Jurjo, D. L. B. R., Magluta, C., Roitman, N., and Gonçalves, P. B. (2015) Analysis of the structural behavior of a membrane using digital image processing. MECH SYST SIGNAL PR. 54 394-404.

16. Pan, B., Qian, K., Xie, H. and Asundi, A. (2009) Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. Meas Sci Technol, 20(6) 062001.

17. Chan, T.H.T., Ashebo, D.B., and Tam, H.Y. (2009) Vertical displacement measurements for bridges using optical fiber sensors and CCD cameras—A Preliminary Study. Struct. Health Monit. 8, 243–249.

18. Kohut, P., Holak, K., Uhl, T., Ortyl, Ł., Owerko, T., Kuras, P., and Kocierz, R. (2013) Monitoring of a civil structure's state based on noncontact measurements. Struct. Health Monit. 12, 411–429.

19. Busca, G., Cigada, A., Mazzoleni, P., and Zappa, E. (2014) Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system. Exp. Mech. 54, 255–271.

20. Ri, K., and Wang, N. (2017) Experimental Verification of Vision-based Cable Force Evaluation using a Smartphone. Struct Health Monit.

21. Yoon, H., Hoskere, V., Park, W. (2017). Cross-correlation-based structural system identification using unmanned aerial vehicles. Sensors 17(9) 2075.

22. Yoon, H., Shin, J., and Spencer, Jr. (2018). Structural displacement measurement using an unmanned aerial system. COMPUT-AIDED CIV INF. 33(3)183-192.

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

24. Ebrahimkhanlou, A., Farhidzadeh, A., and Salamone, S. (2016) Multifractal analysis of crack patterns in reinforced concrete shear walls. Struct. Health Monit. 15, 81–92.

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

26. Rawat, W., and Wang, Z. (2017) Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput. 29, 2352–2449.

27. Chen, F., and Jahanshahi, M.R. (2018) Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion. IEEE Trans. Ind. Electron. 65, 4392–4400.

28. Dorafshan, S., Maguire, M., Hoffer, N.V., and Coopmans, C. (2017) Fatigue Crack Detection Using Unmanned Aerial Systems in Under-Bridge Inspection. Ida. Transp. Dep. 2, 1–120.

29. Li, G.; Ma, B., He, S., Ren, X., and Liu, Q. (2020) Automatic Tunnel Crack Detection Based on U-Net and a Convolutional Neural Network with Alternately Updated Clique. Sensors. 20, 717.

30. Soloviev, A., Sobol, B., and Vasiliev, P. (2019) Identification of Defects in Pavement Images Using Deep Convolutional Neural Networks. Adv. Mater. 4, 615–626.

31. Li, B, Wang, K.C., Zhang, A., Yang, E., and Wang, G. (2020) Automatic classification of pavement crack using deep convolutional neural network. Int. J. Pavement Eng. 21, 457–463.

32. Tong, Z., Gao, J., Han, Z., and Wang, Z. (2018) Recognition of asphalt pavement crack length using deep convolutional neural networks. Road Mater. Pavement Des. 19, 1334–1349.

33. Fan, Z., Li, C., Chen, Y., Mascio, P.D., Chen, X., Zhu, G., and Loprencipe, G. (2020) Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement. Coatings. 10, 152.

34. Adeli, H., and Jiang, X. (2008). Intelligent infrastructure: Neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures (pp. 305–332). Boca Raton, FL: CRC Press.

35. Cawley, P., and Adams, R. D. (1979). The location of defects in structures from measurements of natural frequencies. J Strain Anal Eng Des, 14(2), 49–57.

36. Pandey, A. K., Biswas, M., and Samman, M. M. (1991). Damage detection from changes in curvature mode shapes. J SOUND VIB, 145(2), 321–332.

37. Chang, K. C., and Kim, C. W. (2016). Modal-parameter identification and vibration-based damage detection of a damaged steel truss bridge. Eng. Struct., 122, 156–173.

38. Reynders, E., Wursten, G., and Roeck, G. (2014). Output-only fault detection in structural engineering based on kernel PCA. BIL2014 workshop on data-driven modeling methods and applications, Leuven, Belgium, 14–15 July.

39. Yan, A. M., Kerschen, G., Boe, P., and Golinval, J. C. (2005). Structural damage diagnosis under varying environmental conditions. Part I: A linear analysis. MECH SYST SIGNAL PR, 19(4), 847–864.

40. Lin, Y. Z., Nie, Z. H., and Ma, H. (2017). Structural damage detection with automatic feature extraction through deep learning. COMPUT-AIDED CIV INF, 32(12), 1025–1046.

41. Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., and Inman, D. J. (2018). Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks. J SOUND VIB, 424, 158–172.

2. DEVELOPMENT OF VISION-BASED SENSOR FOR STRUCTURAL DISPLACEMENT USING FIXED CAMERA

2.1. Introduction

Bridges are exposed to many external loads, such as traffic, wind, flooding, and earthquakes. The monitoring of bridges' conditions becomes increasingly important due to safety concerns. There are some direct measurement methods such as the use of linear variable differential transformer (LVDT) and laser-based displacement sensors [1]. However, many LVDTs are required to measure the displacement across a bridge, which can be both costly and time-consuming. Another method to obtain displacements is the non-contact method, such as using GPS [2], but the precision of GPS displacement measurements is not satisfactory.

Nowadays, vision-based displacement sensors are more and more popular [3-9]. Moreover, the template matching is the most used method to calculate the displacement and the deformation of structures [10-13]. Olaszek [14] first used the template matching method to obtain the dynamic characteristics of bridges. Poudel et al. [15] proposed an algorithm to determine the edge location with sub-pixel precision through information from neighboring pixels and used these edge locations to trace structural deformations. A non-contact vision sensor system for monitoring structural displacements with an advanced Zernike subpixel edge detection technique is suggested in this chapter. Edge detection can detect features of objects effectively, without using templates. Subpixel techniques provide more accurate and cost-effective results when compared to integer pixel methods. Built on these two techniques, a new version sensor method is developed to detect the vibrations of structures in this chapter. Satisfactory agreements were found between the displacements measured by the vision sensor system and those recorded by the Multipurpose Testing System (MTS).

2.2. Proposed Vision Sensor System: Hardware and Basic Principle

2.2.1. Setup of the Vision Sensor System

Experiments are performed to verify the vison sensing mechanism. Figure 1 shows the setup of the experiment in the lab. An iPhone6s with 30 fps data acquisition rate and 1920×1080-pixel resolution was fixed on a tripod which is set at a reference point and used to capture vibrations of the MTS machine. The MTS system performs a 4-Hz vibration, while the displacement of its piston is measured using the deployed iPhone6s. The iPhone was placed 1m from the MTS machine. Sinusoidal waves were initiated through the MTS system with controlled frequencies (3 Hz, 2 Hz, 1 Hz, and 0.5 Hz) and amplitudes (2.5 mm, 5 mm, 10 mm, and 10 mm). To simulate the vibration of the camera, continuous vibrations were triggered manually.

Figure 1. Setup of the experiment on MTS. (a) Front view; (b) Side View.

2.2.2. Basic Principle for Displacement Measurement

An edge-detection method to monitor displacement through images captured by a single camera is illustrated in Figure 2. Once the original images (Figure 2(a)) are obtained through cameras, they are converted to gray-scale images (Figure 2(b)). The Edge detection using Canny edge detector, which is defined on the integer pixel level, is then performed (Figure 2(c)). The white lines shown in Figure 2(c) indicate edges found while the background is set to black through the greyscale thresholding process. Once the edges of a structure are found, a self-developed algorithm in MATLAB is executed to calculate the coordinates of the edges at measured points. Through this algorithm, displacements at any location of the target structure could be measured. To reduce the runtime, the ROI is used to calculate displacements at certain locations of the structure. Most of the time, the displacements of a structure are small, so an image of the entire structure is not required.

The proposed displacement measurement method is like a group of laser displacement sensors or LVDTs. The cameras are the reference points because they do not move, similar to fixtures of laser sensors or LVDTs. Every point on the edges of measured structures could be selected as there is a laser sensor at each point of the edges. It is also convenient to change measured points since they are captured simultaneously.
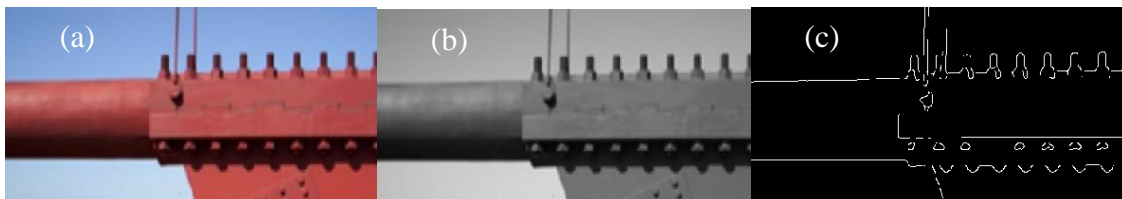
Figure 2. Processed image: (a) Original image; (b) Grayscale image; (c) Edges in the image.

When the camera is not very close to the object or the camera does not have a high resolution, the accuracy obtained is usually not adequate. There are two approaches to enhance it.

16

The one is to use a higher resolution camera. The other way is to use the sub-pixel technology to improve the accuracy of the image analysis. Some researchers have incorporated the subpixel technique to the conventional template matching methodology, mostly using the interpolation. In this chapter, the edge detection method is combined with the subpixel technique to achieve better accuracy. After the displacement at an integer-pixel level is obtained, the subpixel-technique based on Zernike moment method is used to obtain subpixel level displacements and achieve more accurate results. At the same time, to convert the displacement in pixel to the real physical distance, a relationship between the pixel and physical coordinates needs to be established.

2.2.3. Principles of Zernike Moment-based Subpixel Edge Detection

Photo images are very sensitive to noises such as changes in brightness and the vibration of cameras. Zernike moment is an integral operator that filters these effects and helps improving displacement measurement accuracy [16,17]. The Zernike moment has a property of rotation invariance, which can be written as:

$$Z'_{n,m} = Z_{n,m} \, exp(-j\phi) \tag{1}$$

Where $Z_{n,m}$ is the original n by m Zernike moment matrix; $Z'_{n,m}$ is the transferred Zernike moment matrix; j is the imagery identifier; and ϕ is the rotation angle.
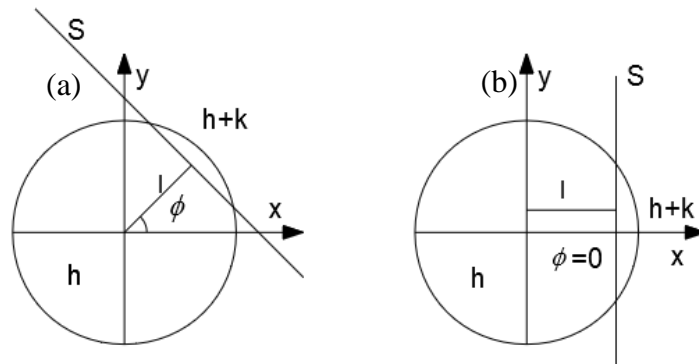
Figure 3. Model of edge: (a) Original edge; and (b) Edge after rotation.

Figure 3 contains a model of the edge. Figure 3(b) is obtained when Figure 3(a) is rotated clockwise about the origin by ϕ. In the figure, S is the edge, h and $h+k$ are the values of grayscale of two sides about S, l is the perpendicular length from the origin to S, and ϕ is the angle between l and x-axis. In Equation (1) and Figure 3, $Z'_{n,m}$ is the Zernike moment after the rotation. The exact coordinates of the edges will require k, l, and ϕ.

Each Zernike moment element is defined in Equation (2).

$$Z_{n,m} = \frac{n+1}{\pi} \iint_{x^2+y^2\leq 1} \bar{V}_{n,m} f(x,y) dx dy \tag{2}$$

$Z_{n,m}$ is the Zernike moment of $f(x,y)$ at rank n. $\bar{V}_{n,m}$ is the conjugate function of $V_{n,m}$, which is the integral core function.

Discrete form of Zernike moment can be written in Equation (3).

$$Z_{n,m} = \frac{n+1}{\pi(N-1)^2} \sum_{i=1}^{N}\sum_{j=1}^{N} f(x_i,y_j) V_{n,m}(x_i,y_j) \tag{3}$$

Where N is the number of the integration points, x_i, y_j are the coordinates of the integration points, and $f(x_i, y_j)$ is the gray scale of the pixel.

Table 1. Zernike orthogonal complex polynomials ($\bar{V}_{n,m}$) (Abdallah et al. 1997).

m/n	0	1
0	1	/
1	/	$x - yj$
2	$2x^2 + 2y^2 - 1$	/
3	/	$(3x^3 + 3xy^2 - 2x)$ $+ (3y^3 + 3x^y - 2y)j$
4	$6x^4 + 6y^4 + 12x^2y^2 - 6x^2 - 6y^2 + 1$	/

$Z_{0,0}$, $Z_{1,1}$, and $Z_{2,0}$ are used to calculate the edges at subpixel level. Their integral core functions are $\bar{V}_{0,0} = 1$, $\bar{V}_{1,1} = x - yj$, and $\bar{V}_{2,0} = 2x^2 + 2y^2 - 1$, respectively (Table 1).

The edge after rotation is symmetric about x axis, so

$$\iint_{x^2+y^2\leq1} yf'(x,y)dxdy = 0 \tag{4}$$

Where, $f'(x,y)$ is the image function after rotation.

$$Z'_{1,1} = Z_{1,1}\, exp(-j\phi)$$

$$= R\,e(Z'_{1,}1) + jI\,m(Z'_{1,}1) \tag{5}$$

Where, Re is the real part of $Z'_{1,1}$, Im is its imaginary part.

$$\phi = tan^{-1}(\frac{Im(Z_{1,1}')}{Re(Z_{1,1}')}) \tag{6}$$

In which, $Z'_{1,1}$ and $Z'_{2,0}$ are calculated using Equation (7) and (8).

$$Z'_{1,1} = \iint_{x^2+y^2\leq1} f'(x,y)(x-yj)dxdy = \iint_{x^2+y^2\leq1} f'(x,y)xdxdy = \frac{2k(1-l^2)^{\frac{3}{2}}}{3} \tag{7}$$

$$Z'_{2,0} = \iint_{x^2+y^2\leq1} f'(x,y)(2x^2+2y^2-1)dxdy = \frac{2kl(1-l^2)^{\frac{3}{2}}}{3} \tag{8}$$

l and k are calculated from Equation (9a) and Equation (9b).

$$l = \frac{Z'_{2,0}}{Z'_{1,1}} = \frac{Z'_{2,0}}{Z_{1,1}\, exp(-j\phi)} \tag{9a}$$

$$k = \frac{3Z'_{1,1}}{2(1-l^2)^{\frac{3}{2}}} = \frac{3Z_{1,1}\, exp(-j\phi)}{2(1-l^2)^{\frac{3}{2}}} \tag{9b}$$

When the values of *l, k, h* and ϕ are obtained, coordinates at subpixel level (a_{sc} and b_{sc}) are calculated using Equation (10), where *a* and *b* are integer coordinates of the edge.

$$\begin{bmatrix} a_{sc} \\ b_{sc} \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} + l \begin{bmatrix} cos(\phi) \\ sin(\phi) \end{bmatrix} \tag{10}$$

An example test is conducted to verify the accuracy of the subpixel method using MATLAB (Figure 4). First, a circle with the center at (50, 75) and a diameter of 50 units is drawn using the circle equation with color of red. The color of the outside region is black, and the inside region is white. The line for the circle is the true edge between the black and white regions. The edges are calculated using both the integer-pixel edge detection and the subpixel edge detection, and they are compared with the real edge in red (Figure 4(b)). The improved accuracy of the subpixel method over the integer-pixel method is clearly seen when the detected edge is compared with the true edge.
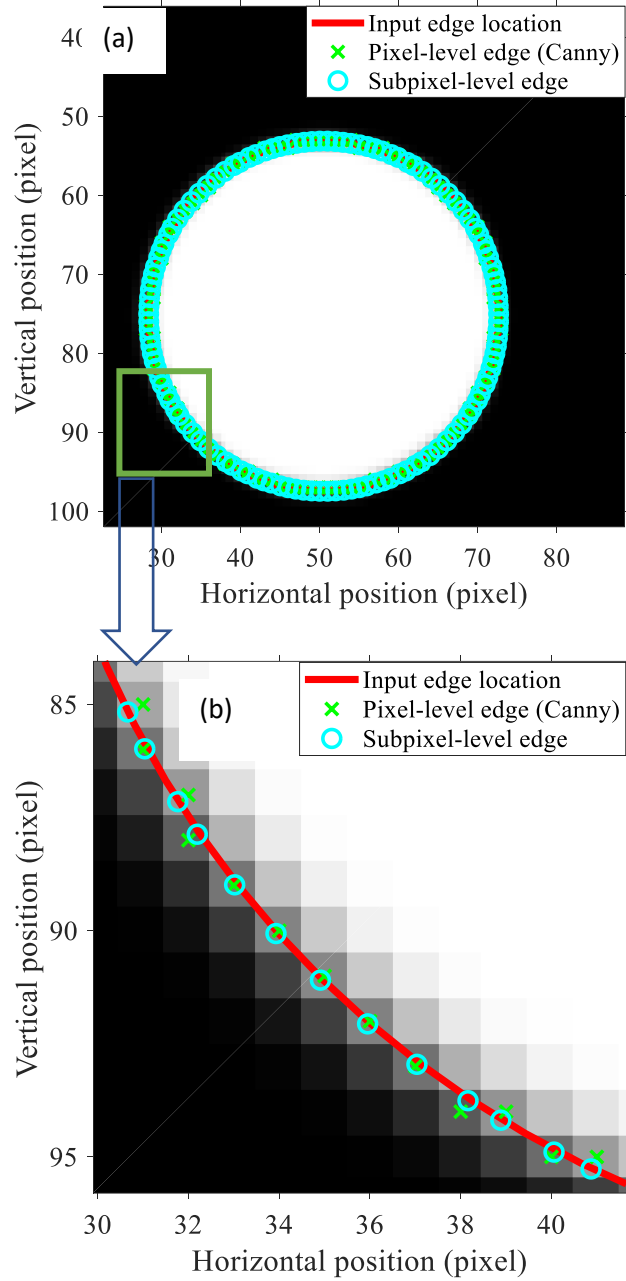
Figure 4. Performance of subpixel method and integer-pixel method. (a) The overall image, (b) Zoom-in details of (a).

2.2.4. Scaling Factor Determination

Two ways to calculate the scale factor are shown blow:

$$SF_1 = \frac{L_{known}}{O_{known}} \tag{11}$$

21

$$SF_2 = \frac{D}{F} d_{pixel} \tag{12}$$

L_{known} is the known physical length on the structure surface. O_{known} is the pixel length on the image plane. D is the distance between the camera and the structure. F is the focal length of the camera. And d_{pixel} is the pixel length such as μm/pixel.

The Zernike moment-based subpixel method could improve accuracy of edge points by about 90%. In a physical test setting, the resolution of the camera, distance from the camera to the structure, and the focal length of the camera together determines the resolution of the monitored displacement in physical unit. By applying the Zernike moment-based subpixel method, the final resolution in subpixel precision could be improved 10 times compared to the resolution in pixel level.

2.3. Effect of Threshold on Edge Detection

The edge detection is used to obtain vibrations of structures. Edge detection is more flexible and timesaving than the template matching method since the template matching method needs to select a suitable template before testing. Suitable templates may not be everywhere in images, but edges present extensively in images, especially for bridges and other structures. Figure 5(a) shows a grayscale image at one measured point of the MTS machine. Figure 5(b), (c), (d), (e), (f) show the edges obtained by Canny Edge Detector on the measured points of the MTS at different thresholds of 0.05, 0.1, 0.5, 0.6, and 0.8. Images are very sensitive to noises such as brightness changes and vibrations of cameras. Since the Zernike moment is an integral operator which is not sensitive to noise, the Zernike moment method is selected in this chapter. Different thresholds were used to filter out fake edges on structures. As shown in Figure 5(b) and Figure 5(c), some fake edges at very dark background were detected, which could prove that structural edges will be detected too using the Canny edge detector. However, structural edges would be lost when the

threshold is high. Right thresholds are necessary to remain structural edges and filter out fake edges at background. The threshold could be selected according to gradients of grayscale values on the two sides of an edge. The proposed Zernike moment edge-finding method could trace any edge on the structure conveniently.
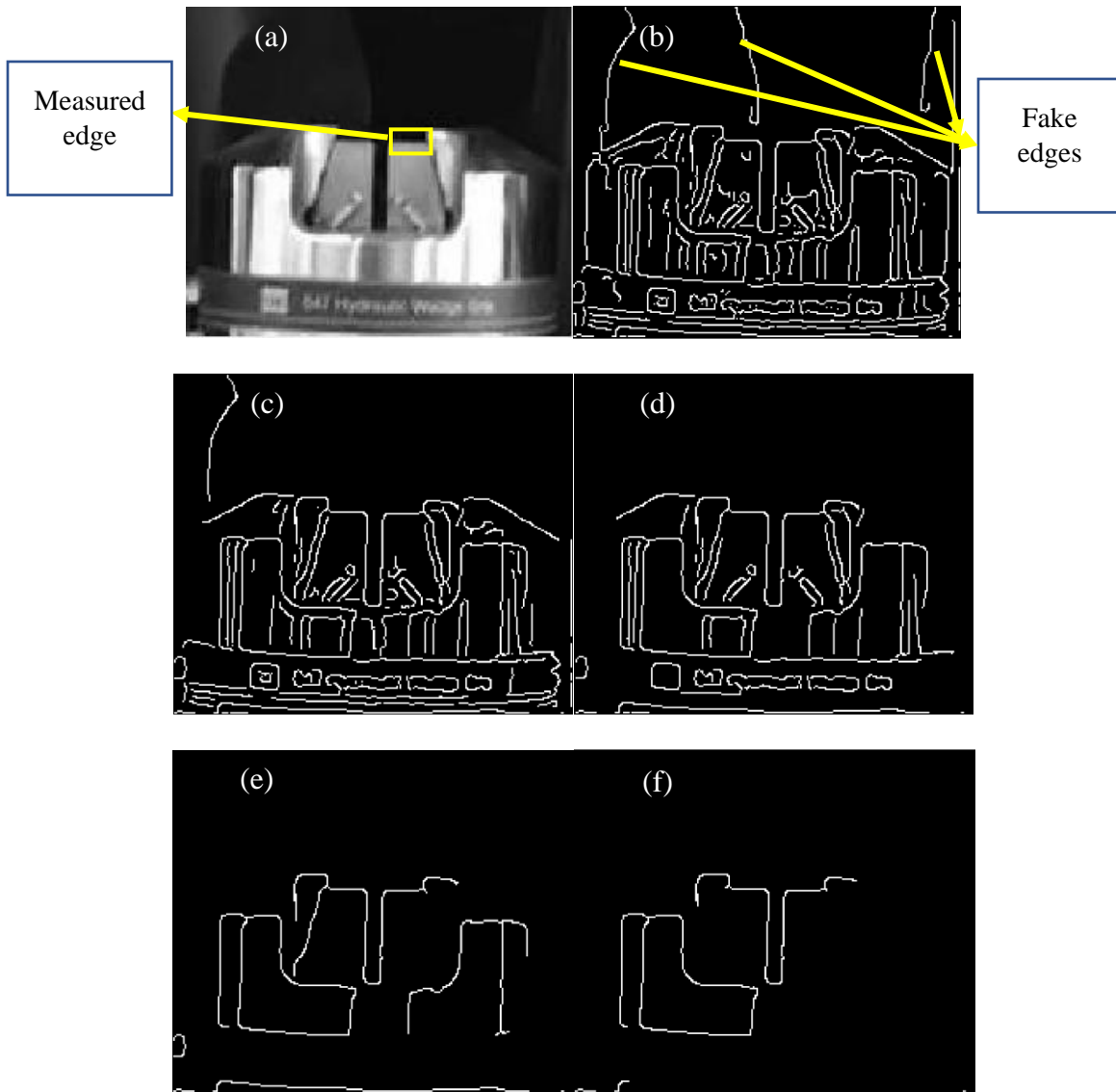
Figure 5. (a) Grayscale image at the measured point; (b) Edges found at the measured point with threshold at 0.05; (c) Edges found at the measured point with threshold at 0.1; (d) Edges found at the measured point with threshold at 0.5; (e) Edges found at the measured point with threshold at 0.6; (f) Edges found at the measured point with threshold at 0.8.

2.4. Performance Evaluation

Three tests were conducted in the lab to verify the proposed method. The tests were aimed at measuring the one-dimensional vibrations of the piston of the MTS machine. The displacement output by MTS, whose error is less than 0.5% with respect to the given input command, is taken as reference of the measurement methods. An accelerometer was also used to detect the vibrations, which were compared to the displacement data obtained through image processing and output of the MTS. The acceleration data from the accelerometer will be integrated to obtain the velocity and displacement data using the detrended double integration method through a filter for frequencies higher than 60 Hz.

2.4.1. MTS Test Setup

Figure 6 shows the experimental set-up in the lab. A XiaoYi's commercial camera with 120 fps acquisition rate and 1920×1080-pixel resolution was placed 0.5 m from the MTS machine. A MEMS accelerometer was glued on the piston and connected with a laptop which collected and stored the data from the MEMS accelerometer. The piston was mounted on the MTS machine, whose vibrations were computer controlled. Sinusoidal wave-type motions were adopted with controlled frequencies (2 Hz, 2.5 Hz, and 3 Hz) and amplitudes (4 mm, 2 mm, and 1.5 mm, respectively).
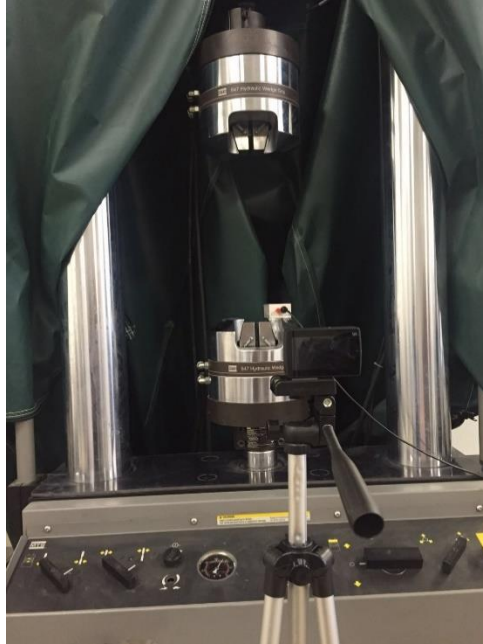
Figure 6. Set-up for lab experiment.

2.4.2. Results With Subpixel Resolution

Figure 7(a) presents the results when a sinusoidal wave with an amplitude of 4 mm and a frequency of 2 Hz was applied. Comparison of vibrations between the subpixel image processing, the integer pixel image processing, the MTS, and the accelerometer measurements are shown in Figure 7(b). Very good agreement between the subpixel image measurement and the displacement output by MTS is shown. The accuracy of the subpixel image processing method is better than that of the accelerometer method in monitoring displacement. For the integer image processing method, 56% of points per cycle are at the same value for the displacement shown. One possible reason is that the integer image processing method could only detect displacements at integer pixel values. Especially, when the piston moves slowly, more points at the same value of displacements will be obtained.
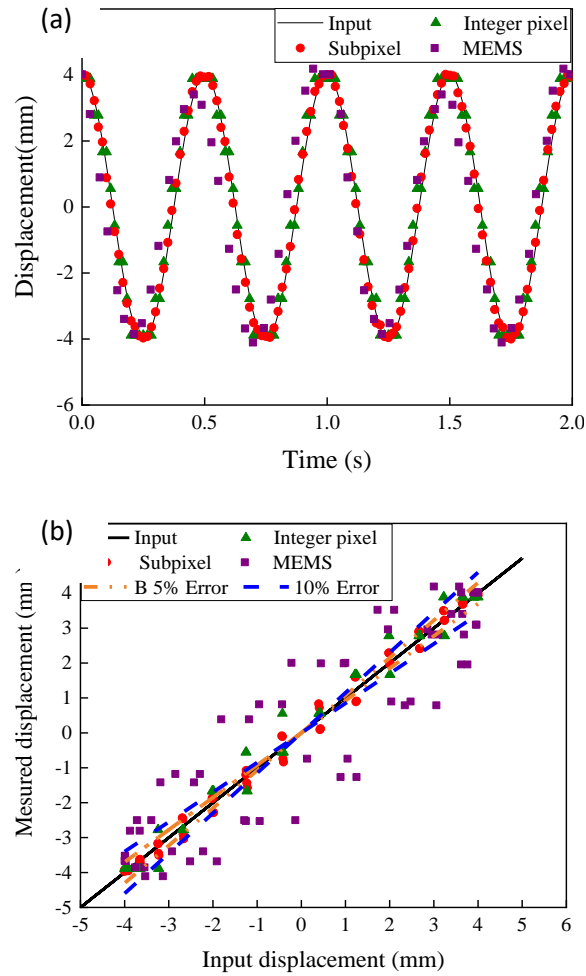
Figure 7. (a) Comparison of captured displacements with the input vibration with a frequency of 2 Hz and amplitude of 4 mm, and (b) Error analysis.

The results of vibration measurements for the case with an amplitude of 2 mm and frequency of 2.5 Hz are shown in Figure 8(a). Moreover, comparison of vibration measurements between the subpixel image processing method, the MTS input, the integer pixel image processing method, and the accelerometer measurements are shown in Figure 8(b). From Figure 8(b), higher accuracy is achieved for the subpixel image processing method.
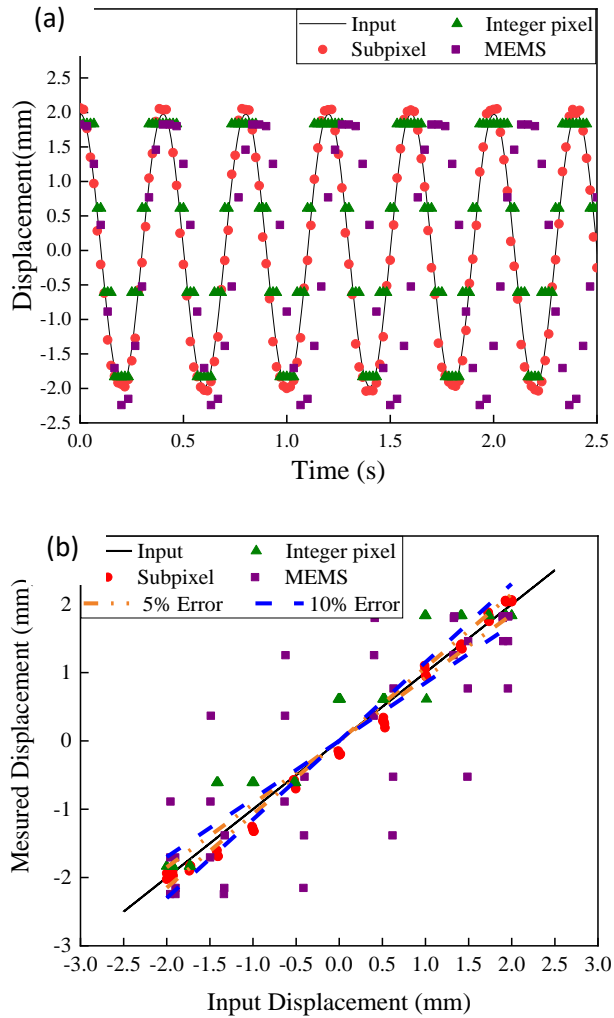
Figure 8. (a) Comparison of captured displacement with the input vibration at a frequency of 2.5 Hz and amplitude of 2 mm, and (b) Error analysis.

Figure 9(a) shows the results of vibrations measured by the subpixel image processing method, integer pixel image processing method, the MEMS accelerometer, when a sinusoidal wave with amplitude of 1.5 mm and frequency of 3 Hz was applied. Comparisons of errors between the subpixel image processing and MTS, the integer pixel image processing and MTS, and the accelerometer measurements and the MTS are shown in Figure 8(9).
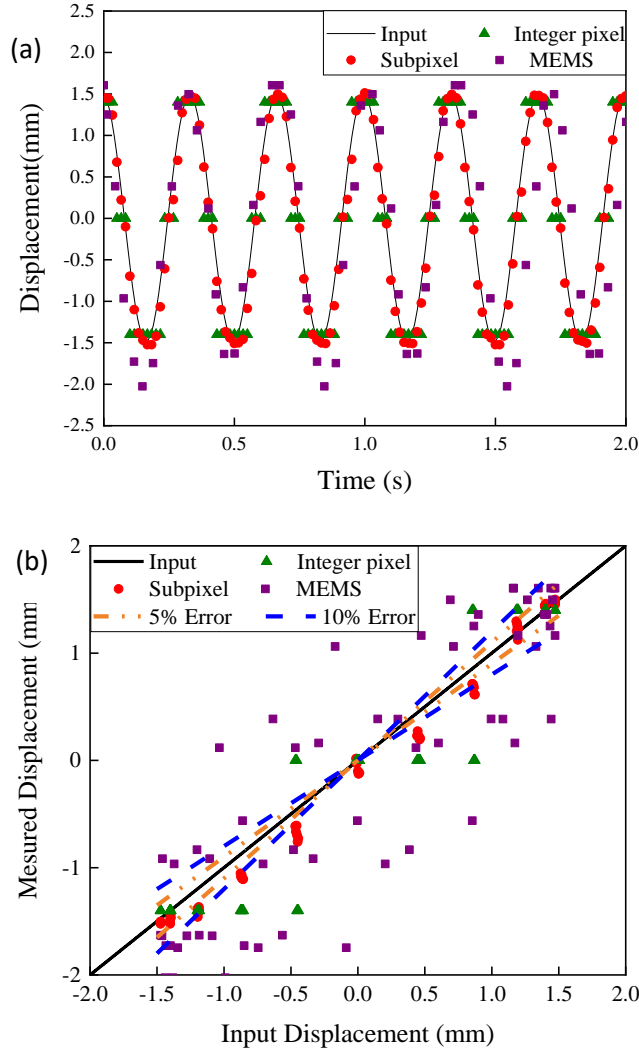
Figure 9. (a) Comparison of captured displacement data with input sinusoidal function with frequency of 3 Hz and amplitude of 1.5 mm and (b) Error analysis.

According to Figures 7(b), 8(b), and 9(b), the maximum error of subpixel method is less than 5%. Some measured displacements from the integer pixel method are larger than 10%. However, for MEMS accelerometer, the maximum error is typically larger than 20%.

2.4.3. Error Quantification

As shown in Figure 10, when the frequency is higher, larger errors are observed because fewer points per cycle are obtained to describe the vibration. Higher resolution images could help but are also very expensive, so the subpixel image processing method is adopted instead.

Moreover, the proposed subpixel image processing method does not need any target on the object and can detect displacements at any location on the structure. In comparison to the template matching method, the edge detection method is timesaving and avoids issues from events such as excessive deformation and failure of targets.
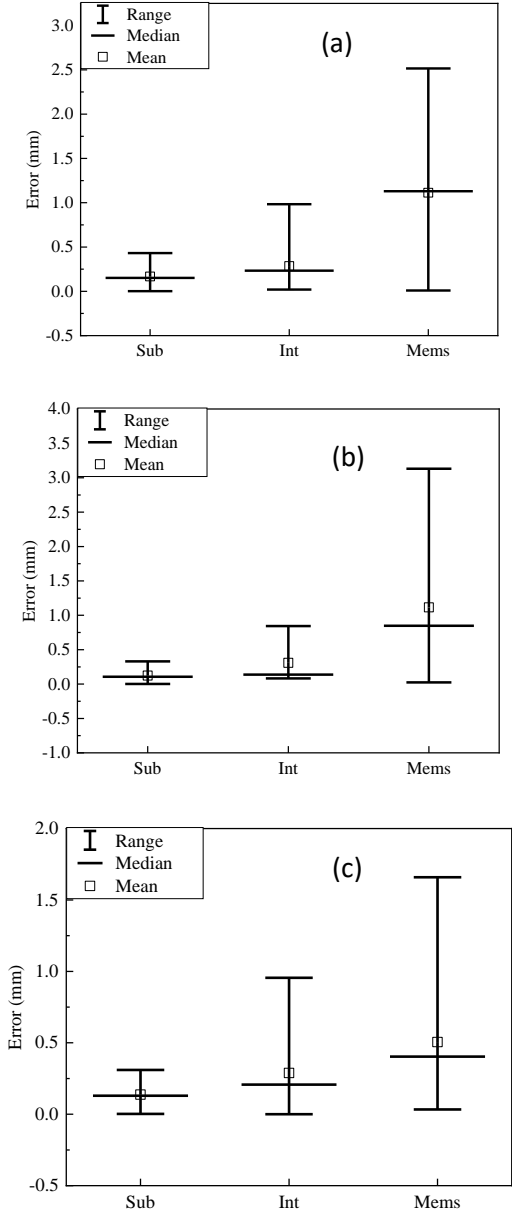
Figure 10. Error analysis of experiments: (a) 4 mm at 2 Hz; (b) 2 mm at 2.5 Hz; (c) 1.5 mm at 3 Hz. (Sub – subpixel method, Int – integer pixel method, Mems – MEMS accelerometer method)

2.5. Conclusion

In this chapter, a novel vision system was developed for noncontact displacement measurement of structures using a subpixel Zernike edge detection technique. Comprehensive experiments using an MTS machine were carried out to verify the accuracy of the proposed method. The following conclusions are reached:

- In MTS tests, satisfactory agreements were observed between the displacement measured by the vision system and those from the MTS output.

- Subpixel based Zernike matrix method is an innovative edge detection that could monitor structural displacements accurately at any location on the detected edges.

- The proposed edge-detection-based displacement sensor is a novel non-contract sensor to detect displacement and vibration successfully.

2.6. Reference

1.   Nassif, H.H., Gindy, M., and Davis, J. (2005) Comparison of laser doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. NDT E Int. 38, 213–218.

2.   Stephen, G., Brownjohn, J., and Taylor, C. (1993) Measurements of static and dynamic displacement from visual monitoring of the Humber Bridge. Eng. Struct., 15(3) 197-208.

3.   Abdallah, S.M., Nebot, E.M., and Rye, D.C. (1998). "Object recognition and orientation via Zernike moments." In Asian Conference on Computer Vision (pp. 386-393). Springer, Berlin, Heidelberg.

4.   Busca, G., Cigada, A., Mazzoleni, P., and Zappa, E. (2014). "Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system." Exp. Mech. 54, 255–271.

5. Chan, T. H. T., Ashebo, D. B., and Tam, H. Y. (2009). "Vertical displacement measurements for bridges using optical fiber sensors and CCD cameras—A Preliminary Study," Struct Health Monit, vol. 8, no. 3, pp. 243–249.

6. Chen, C.-C., Wu, W.-H., Tseng, H.-Z., Chen, C.-H., and Lai, G. (2015). "Application of digital photogrammetry techniques in identifying the mode shape ratios of stay cables with multiple camcorders," Measurement, vol. 75, pp. 134–146.

7. Feng, D., Feng, M.Q., Ozer, E., and Fukuda, Y. (2015). "A vision-based sensor for noncontact structural displacement measurement," Sensors, vol. 15, no. 7, pp. 16557–16575.

8. Fukuda, Y., Feng, M.Q., and Shinozuka, M. (2010). "Cost-Effective vision-based system for monitoring dynamic response of civil engineering structures." Struct. Control Health Monit. 17, 918–936.

9. Debella-Gilo, M., and Kääb, A. (2011). "Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation." Remote Sens. Environ. 115 (1), 130–142.

10. Fioriti, V., Roselli, I., Tatì, A., Romano, R., and De Canio, G. (2018). "Motion Magnification Analysis for Structural Monitoring of Ancient Constructions". Measurement. 129, 375-380.

11. Gentile, C., and Bernardini, G. (2009). "An interferometric radar for non-contact measurement of deflections on civil engineering structures: Laboratory and full-scale tests." Struct. Infrastruct. Eng. 6, 521–534.

12. Kohut, P., Holak, K., Uhl, T., Ortyl, Ł., Owerko, T., Kuras, P., and Kocierz, R. (2013). "Monitoring of a civil structure's state based on noncontact measurements." Struct. Health Monit. 12, 411–429.

13. Mao, J.X., Wang, H., and Spencer, B.F. (2020). "Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders." Struct. Health Monit, 1-18.

14. Olaszek, P. (1999) "Investigation of the dynamic characteristic of bridge structures using a computer vision method." Measurement; 25(3):227-36.

15. Poudel, U.P., Fu, G., and Ye, J. (2005). "Structural damage detection using digital video imaging technique and wavelet transformation". J SOUND VIB, vol. 286, no. 4-5, pp. 869–895.

16. Wang, M.H. (2004). "The Application of Invariants Theory in Pattern Recognition". Master's Thesis, Yanshan University, Qinhuangdao, China.

17. Teague, M.R. (1980) "Image analysis via the general theory of moments". JOSA 70, 920–930.

3. PERFORMANCE EVALUATION OF THE VISION-BASED DISPLACEMENT SENSOR

3.1. Introduction

To adopt computer vision methods in structural health monitoring, the template matching is typically used to trace deformation of structures [1-5]. Olaszek [6] first used the template matching method to obtain the dynamic characteristics of bridges. Sładek et al. [7] detected the in-plane displacement using template matching on a beam. Wu et al. [8] developed a vision system that uses digital image processing and computer vision technologies to monitor the 2D plane vibrations of a reduced-scale frame mounted on a shake table.

Without template targets, computer vision methods can still be used to trace structural deformations [9,10]. Poudel et al. (2005) proposed an algorithm to determine the edge location with sub-pixel precision through information from neighboring pixels and used these edge locations to trace structural deformations. Wahbeh et al. (2003) followed a similar approach to obtain direct measurements of displacement. In this chapter, a field testing on a street sign was performed to verify the proposed image-based displacement sensor. The vibration of the street sign under wind was obtained and three nature frequencies were calculated using the vibration data. For the test on a steel plate, vibration curves and mode shapes of the damaged steel plate and the healthy one was obtained. The damage location on the steel plate was then obtained by comparing these two mode shapes. Finally, a field test was conducted to see how different template sizes and contrast levels affect the accuracy.

3.2. Field Tests and Performance Evaluation on a Street Sign

3.2.1. Test Setup and Test Scenarios

A field experiment was carried out using a street sign at the North Dakota State University (NDSU) campus to verify the proposed subpixel image edge detection-based displacement sensor. As shown in Figure 11, the street sign can be considered as a cantilever column fixed on the ground. An accelerometer was attached on the street sign 40 cm above the ground to compare with the subpixel image processing method. The rate of the data acquisition of MEMS is 100Hz. A moving averaging method is used to filter high frequency noises over 40 Hz to minimize the error. A camera was fixed on a tripod and set up 85 cm from the street sign. A pulse was applied to the tip of the street sign to initiate the vibration.
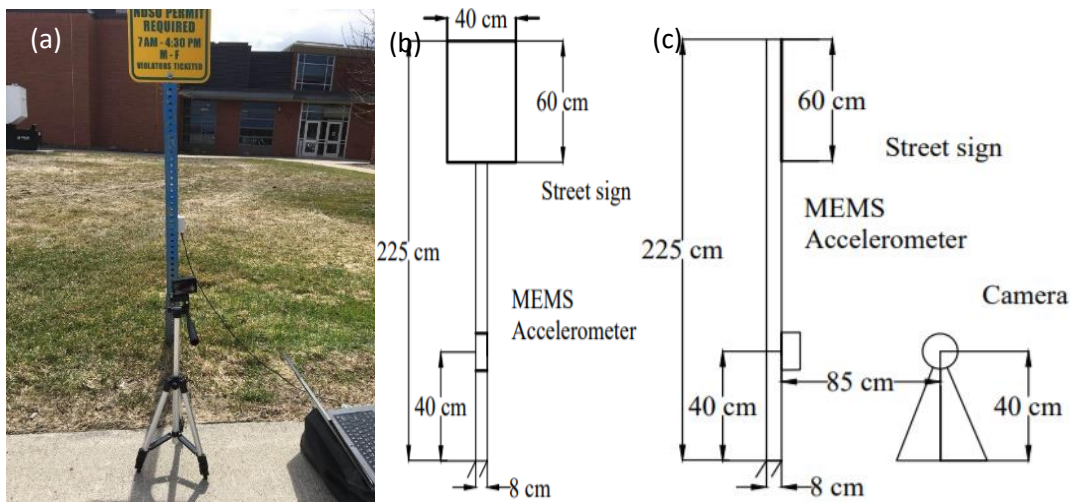
Figure 11. Setup of street sign experiment (a) Field photo; (b) Front view with dimensions; (c) Side view with dimensions.

3.2.2. Displacement Time Histories of the Street Sign

Figure 12 presents the vibration data from the image processing method. As shown in Figure 12, the magnitude of vibrations of the street sign was less than 2 mm. The MEMS accelerometer failed to detect the vibration because the magnitude of vibrations of the street sign is too small, and the street sign vibrated too fast.
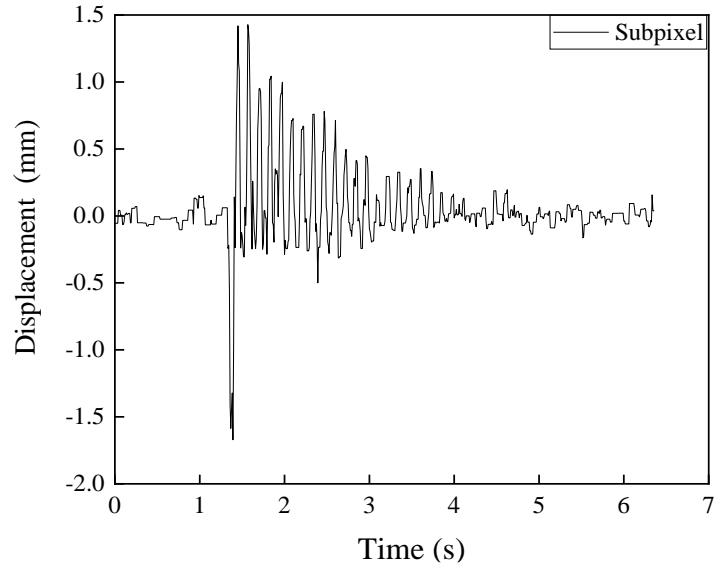
Figure 12. Vibration of street sign using subpixel image processing.

3.2.3. Analysis of Frequency Characteristics

The displacement time history and several natural frequencies were detected through the subpixel image processing method, although the amplitude of the street sign's vibration is less than 2 mm. Figure 13 shows the modal frequencies obtained through the subpixel image processing method. Through the proposed subpixel image processing, these frequencies are found at 7.85 Hz, 15.86 Hz, and 31.73 Hz respectively. The proposed algorithm can be used at any location on the street sign since it has many edges and infinite points on each edge. The edge detection technique could avoid noises such as brightness changing, so the displacement in bad weather can also be detected. Subpixel methods could enhance the edge detection method and detect small displacements with high accuracy.
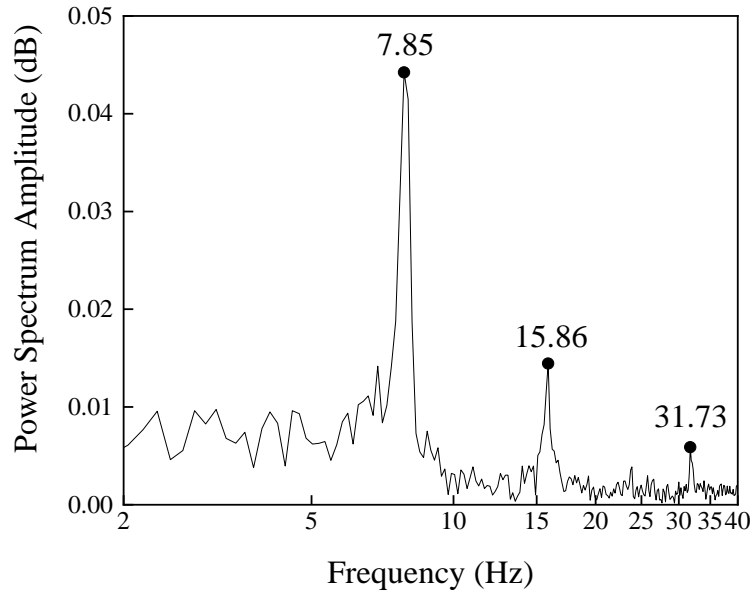
Figure 13. Modal frequencies of street sign at measured location.

3.3. Test on a Cantilever Beam and Damage Detection Using Its Mode Shapes

3.3.1. Test Setup and Test Scenarios

The main purpose of the proposed method is to evaluate damages on structures based on changes in their natural frequencies. To do so, a laboratory experiment was conducted to track the changes in the natural frequencies of a steel beam. The natural frequency of the original steel beam with no damage was first measured using the proposed method. Then a rectangular incision was made on the steel beam and the natural frequency was measured again. The incision is made at 1 foot (0.305 m) away from the fixed end with a depth of 0.5 in. (12.7 mm) and a width of 1.0 in. (25.4 mm). Finally, the location of damage was determined based on the differences between the natural frequencies of the original and damaged steel beam.

To initiate vibration of the steel beam, an arbitrary manual tapping was applied. Figure 14 shows that the camera was fixed on a tripod and set up 1 foot (0.304 m) away from the steel beam. A MEMS accelerometer connected with a laptop was used to measure the displacement of the steel beam too.

36
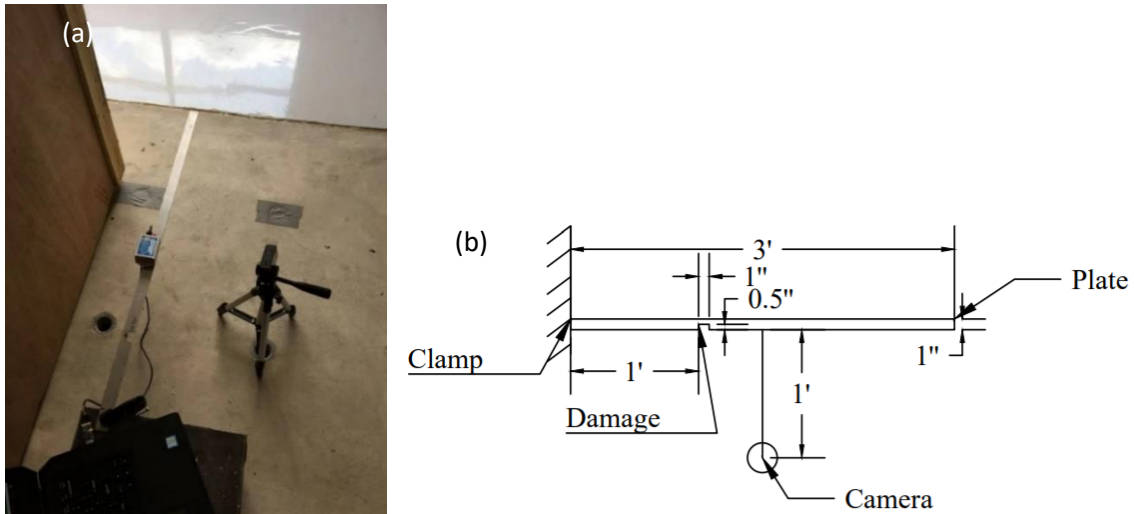
Figure 14. (a) Setup of the steel beam vibration test, (b) Schematic diagram of the experiment setup (1' = 12'' = 0.3048 m).

3.3.2. Analysis of Frequency Characteristics

Figure 15 shows the natural frequencies of the undamaged steel beam through the image processing method and MEMS measurements, while Figure 16 shows the natural frequencies of the damaged steel beam through the proposed method and the MEMS measurements. There is a gap between peaks obtained by MEMS and subpixel image method, whose reason is that MEMS has errors to obtain the vibrations due to the times of integration. The data in Figure 15 and Figure 16 was zoomed out to show the comparison of the first natural frequency, which will have the similar Power Spectrum Amplitude curve as shown in Figure 11 if the x axis scale is set to 0-40 Hz.
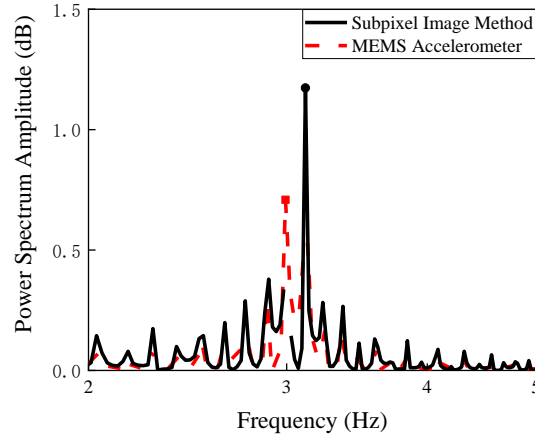
Figure 15. Natural frequency of the undamaged steel beam through the MEMS accelerometer and subpixel image method.
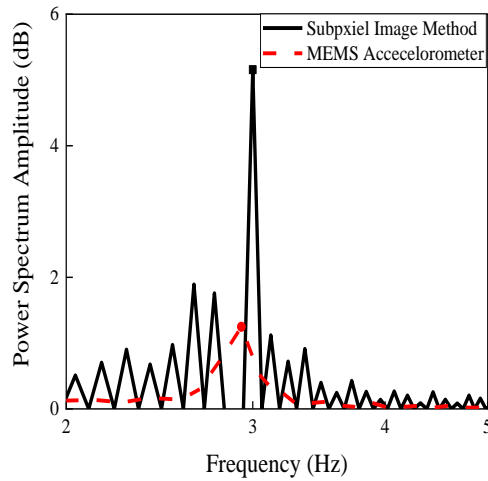
Figure 16. Natural frequency of the damaged steel beam through the MEMS accelerometer and subpixel image method.

Through the two tests, the reduction of stiffness of the steel beam due to incision was successfully found using the proposed method. The natural frequency was reduced from 3.10 Hz to 3.00 Hz after the beam was damaged. Thus, the proposed vision sensor can detect the reduction of stiffness of a structure.

3.3.3. Mode Shape and Damage Detection on the Beam

The location of the damage can also be identified if mode shapes of the structure could be found. Time domain decomposition (TDD) is a proven method used to extract mode shapes of a

38

structure and identify structural damages, even though other methods such as neural networks shown in Mao et al. (2020) could be also adopted for this purpose. Figure 17(a) shows the undamaged and damaged mode shapes. The two mode shapes are similar with small differences observed in the middle region. Normalized position is the ratio of the distance between the points on the bar and the left end (x) and the total length of the steel plate (L). The ratio of the displacement (d) and the maximum displacement (D) is defined as the normalized displacement. Figure 17(b) shows that there are some differences between the undamaged mode shape and damaged mode shapes. As shown in Figure 17(b), the peak difference happened at a normalized location of 0.3 indicates the location of the damage. Through the analysis of mode shape difference, the damage was detected successfully. Since the mode shape and the damage detection method adopted here uses 11 points (10 segments), which makes the highest precision for the normalized damage location is 0.1.
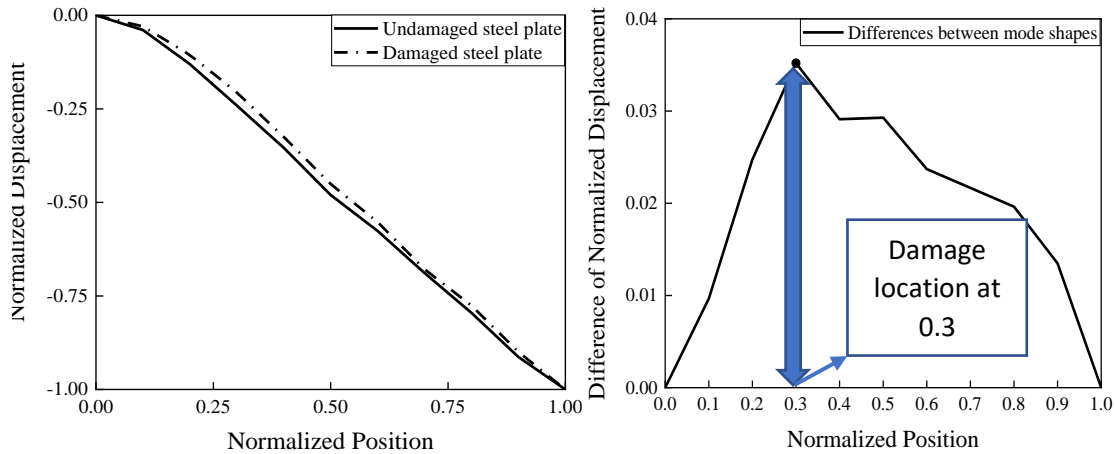
Figure 17. Mode shape analysis. (a) Mode shapes of the undamaged and the damaged steel beam, (b) Differences in the normalized displacement between the undamaged mode shape and damaged mode shape.

3.4. Template Matching Technique: NCC

When camera movement is unavoidable such as due to wind or self-vibrations of UAVs, displacement calculation through edge detection only will not be adequate. Filtering the noise from

environment factors such as wind or self-vibration is necessary and needs to be conducted through

a background template using NCC. When applying the NCC method, a template on the background

of the images is used to calculate the correlation between the template and the regions in the target

image. Then positions of the template in every image are obtained by searching the highest

correlation between the template and the target region in the target image using the NCC method.

$$NCC(u,v) = \frac{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}][t(x-u,y-v) - \bar{t}]}{\{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}]^2[t(x-u,y-v) - \bar{t}]^2\}^{0.5}} \tag{13}$$

Where f is the pixel value of the image; $\bar{f}_{u,v}$ is the mean of the pixel values of the image; t is the

pixel value of the template; $\bar{t}$ is the mean of the pixel values of the template; u and v are the

positions in the image.

Figure 18 shows NCC results of one example. X-direction and Y-direction represent the

searching region of the template in the target image. X-direction runs through the pixel horizontal

location of the image, while Y-direction runs through the pixel vertical location of the image. The

peak point indicates the position of the template in the image. In this process, every NCC between

each position in the target image and the template are calculated using Eq. (13). The coordinates

of the largest NCC could be found and indicate the position of the template in the target image.
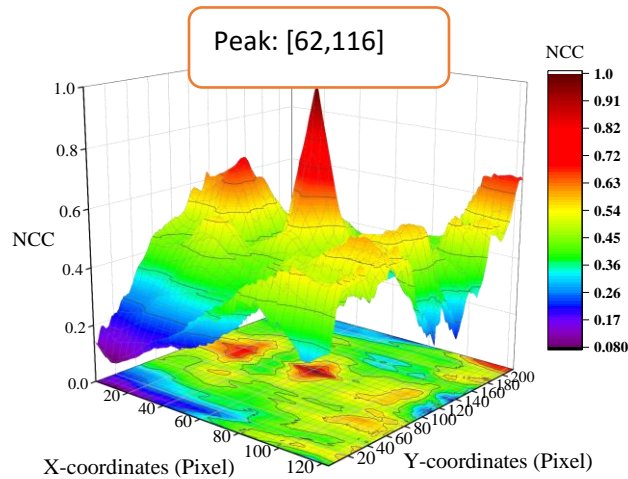
Figure 18. The matching result of NCC.

40

3.5. Subpixel Through 3-D Fitting

Once positions of the template in images are obtained through the above method, the resolution is in the pixel level. To improve the precision, 3D parabolic curve fitting could be used to obtain the sub-pixel level accuracy.

As shown in Eq. (14), the pixel with the maximum value of NCC correlation coefficients and its adjacent eight pixels in the image are fitted to construct a 3D parabolic equation.

$$p(x, y) = ax^2 + by^2 + cxy + dx + ey + f \tag{14}$$

p0 is the maximum NCC value; $p1$, …, $p8$ are its eight adjacent pixel NCC values; $x_0, y_0, x_1, …y_n$ are the coordinators of the pixels in the image. Eq. (15) is used to solve the 3D problem.

$$AX = B \tag{15}$$

$$A = \begin{bmatrix} x_0^2 & y_0^2 & x_0y_0 & x_0 & y_0 & 1 \\ …. & …. & . … & . … & … . & … . \\ x_8^2 & y_8^2 & x_8y_8 & x_8 & y_8 & 1 \end{bmatrix} \tag{16}$$

$$X = [a\ b\ c\ d\ e\ f]^T \tag{17}$$

$$B = [p0\ p1\ p2\ p3\ p4\ p5\ p6\ p7\ p8]^T \tag{18}$$

So, X could be calculated using Eq. (19):

$$X = (A^T A)^{-1} A^T B \tag{19}$$

Thus, the NCC coefficient surface is obtained. The zero gradient of the surface equation with respect to x and y is used to obtain the x and y coordinates of the peak. The overall computational process is free of iteration and efficient.

3.6. Effect of Template Size and Contrast Level

Figure 19 (a) shows four different templates used in the experiment. Figure 19 (b) shows the setup of the experiment, and the camera is placed at 1.5 m away from the shaker in a windy day. There are four templates which were traced to calculate the vibration of the camera considering the effect of environmental factors such as sunlight intensity and wind. Templates used in this thesis are different, including the car on the parking lot, and the tree near the larger building. They have different distances from the camera, different sizes, and different pixel values. The average of NCC values for each different template is used to describe how the templates are matched in the target images. The higher the average of NCC values for each template, the better the matching is, and the more accurate motion of the camera could be obtained.
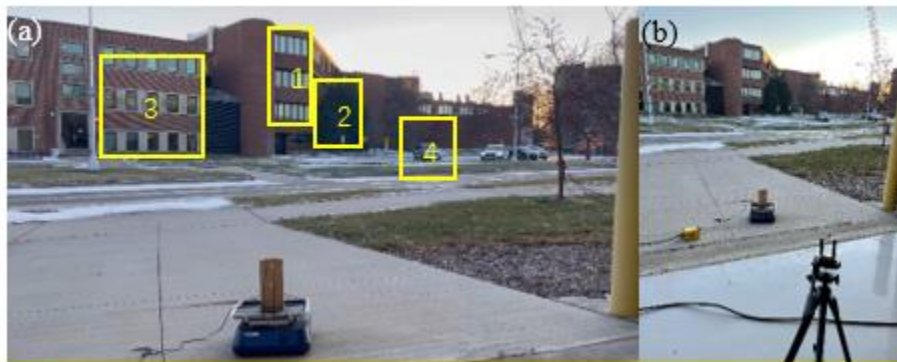
Figure 19. (a) Camera view with templates selected for the field test of the suggested method, (b) Setup of the experiment.

Table 2 shows the average of the NCC values for different templates. Two buildings provide similar values of the average of the NCC values. Car and Tree supply lower values, because the size of them and the contrast of the pixel value are smaller. It is obvious accuracy is higher when the average of the NCC calculated in the process of template matching is higher.

Table 2. Pixel size and average of the NCC values for each template.

	(1) Building 1	(2) Tree	(3) Building 2	(4) Car
Pixel Size	215x589	310x410	639x570	260x200
Average of NCC values	0.997	0.960	0.989	0.954
NCC between obtained vibration with the input	0.998	0.980	0.992	0.975

Figure 20(a) shows the monitored displacement of the shaker using four different templates and the proposed vision-based displacement finding method. Displacements obtained using the 4 different templates are shown in Figure 20(b). When the vibration of the camera is small, the monitored displacements using the four templates agreed with each other very well. When the motions of the camera increase, the monitored displacements using the four templates have different level of errors.
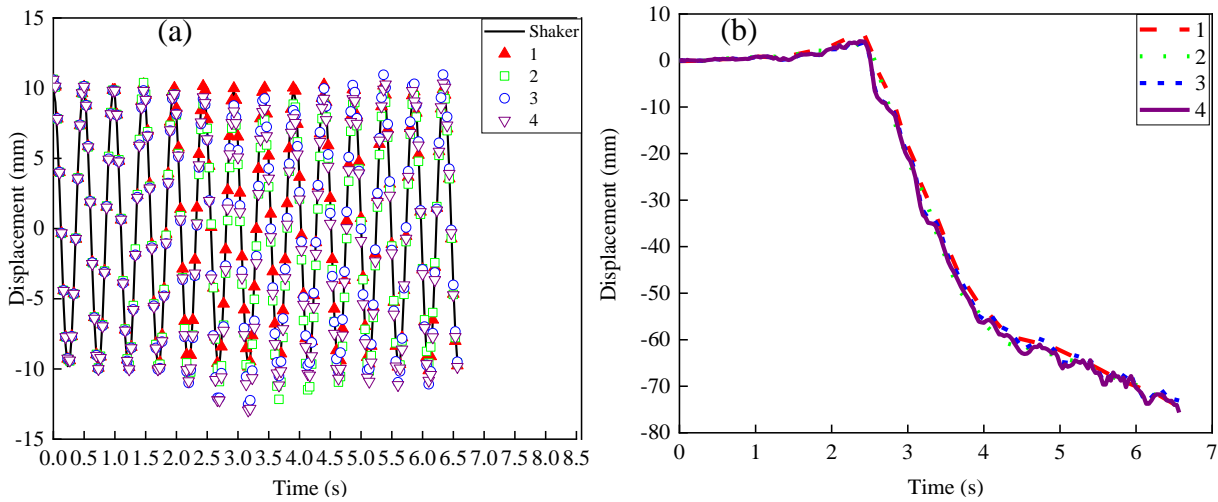
Figure 20. (a) Monitored Vibration of the shaker using four different templates and the input of the shaker, (b) Vibration of camera obtained through different templates using the suggested method.

3.7. Conclusions

In this chapter, a novel vision system was developed for noncontact displacement measurement of structures using a subpixel Zernike edge detection technique. Comprehensive experiments, including vibration tests on street sign and steel beams, were carried out to verify the accuracy of the proposed method.

If movement of camera is unavoidable, NCC-based template matching could be used to solve this issue. Different templates would have different impacts on obtaining displacement of structures. Typically, larger templates contain more feature points and provide higher accuracy in monitored displacement when NCC methods are used. One needs to select templates carefully and avoid any shadows that could come in and out of the templates, because the shadows would cause errors in the process of template matching. If there are two templates with similar size, the template with more feature points would give better accuracy in monitored displacements. The following conclusions are reached:

- From the street sign test, vibrations of the street sign after an excitation pulse were detected successfully. Using Fast Fourier Transform, several natural frequencies of the street sign were founded successfully.

- The subpixel-based Zernike matrix method is an innovative edge detection method that could monitor structural displacements accurately at any location on the detected edges.

- Through the analysis of mode shapes obtained by the proposed image processing method, the damage location on the steel beam was accurately detected.

- Four different templates were selected to obtain vibrations of the camera and showed different accuracies. To achieve higher accuracy, the template is better to be fully

contained in the image, needs to have high contrast, and is not flat on the grayscale values. In other words, templates should have enough feature points to make them stand out in the target image.

By tracking existing natural edges of structures, the vision sensor method developed in this thesis provides flexibility to change measured locations on structures. The availability of such a remote sensor will facilitate cost-effective monitoring of civil engineering structures, with consideration of camera movement due to environmental or operational factors.

3.8. Reference

1. Lee, J. J., and Shinozuka, M. (2006) "A vision-based system for remote sensing of bridge displacement". Ndt & E International, 39(5) 425-431.

2. Park, J. W., Lee, J. J., Jung, H. J., and Myung, H. (2010) "Vision-based displacement measurement method for high-rise building structures using partitioning approach". Ndt & E International, 43(7) 642-647.

3. Ho, H. N., Lee, J. H., Park, Y. S., and Lee, J. J. (2012) "A synchronized multipoint vision-based system for displacement measurement of civil infrastructures". Sci. World J

4. Sładek, J., Ostrowska, K., Kohut, P., Holak, K., and Uhl, T. (2013) "Development of a vision-based deflection measurement system and its accuracy assessment". Measurement 46, 1237–1249.

5. Wu, L.J., Casciati, F., and Casciati, S. (2014) "Dynamic testing of a laboratory model via vision-based sensing". Eng. Struct. 60, 113–125.

6. Olaszek, P. (1999). "Investigation of the dynamic characteristic of bridge structures using a computer vision method." Measurement; 25(3):227-36.

7. Sładek, J., Ostrowska, K., Kohut, P., Holak, K., Gąska, A., and Uhl, T. (2013). "Development of a vision-based deflection measurement system and its accuracy assessment." Measurement 46, 1237–1249.

8. Wu, L.J., Casciati, F., and Casciati, S. (2014). "Dynamic testing of a laboratory model via vision-based sensing." Eng. Struct. 60, 113–125.

9. Debella-Gilo, M., and Kääb, A. (2011). "Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation." Remote Sens. Environ. 115 (1), 130–142.

10. Chen, C.-C., Wu, W.-H., Tseng, H.-Z., Chen, C.-H., and Lai, G. (2015). "Application of digital photogrammetry techniques in identifying the mode shape ratios of stay cables with multiple camcorders," Measurement, vol. 75, pp. 134–146.

11. Poudel, U.P., Fu, G., and Ye, J. (2005). "Structural damage detection using digital video imaging technique and wavelet transformation". J SOUND VIB, vol. 286, no. 4-5, pp. 869–895.

12. Wahbeh, A.M., Caffrey, J.P. and Masri, S.F. (2003). "A vision-based approach for the direct measurement of displacements in vibrating systems." Smart Mater Struct, 12(5), p.785.

4. VISION-BASED DISPLACEMENT MEASUREMENT CONSIDERING TRANSLATIONAL AND ROTATIONAL MOTIONS OF CAMERA

4.1. Introduction

With massive bridges constructed recently, health monitoring of these infrastructures becomes more and more important, to ensure that the structural behavior and the security level are appropriate throughout the bridge's usage life. Railway bridges are one of them, which is particularly related to the dynamic displacement of the structure due to the passage of trains.

To filter out the effect of camera movement due to environmental or operational factors, template matching is typically adopted. The area-based method calculates the correlation between the template image and the target image and obtains the location of template image in the target image [1-7].

Nowadays, UAVs are widely used to obtain the videos of structures' vibrations. The most important thing to be considered is that the effect of the UAV vibration needs to be eliminated. Yoon [8] used an UAS to measure relative vibration between stories of a 6-story structure in the lab. The nature frequencies are obtained successfully but the actual displacement of the story and motions of the UAS could not be obtained. Yoon [9] also proposed a method which could obtain vibrations of the UAS verified by experiments in lab and field. The main drawback of this method is that it needs a manmade checking board on the background to obtain vibrations of the UAS.

A new method to filter the effect of cameras' motions through background templates is proposed in this chapter. All six motions of the camera (three translations and three rotations) were considered and the proposed method filtered the effect of these motions successfully. The results show that filtering out vibrations of the camera significantly improves the displacement monitoring accuracy from 53.0% to 97.0%. Adopting scale factors for every frame and different positions in

each frame also helps improving accuracy of the monitored displacement from 97.0% to 99.0%.

Through the developed filtering method, frequencies of target motions have been successfully identified.

4.2. Effect of Camera Motions on Displacement Monitoring

4.2.1. Displacement of Target and Movement of Camera

When the motions of cameras are not negligibly small, it could affect the accuracy of the monitored displacement significantly. For example, a bridge is monitored on a windy day and the motion of the camera is large due to wind. It is vital to find a method to filter out vibrations of the camera.

Figure 21 shows the process of the proposed system. Images are captured using a high-speed camera with good resolutions. Then fixed points on the background of the structure are selected by users and the NCC is used to obtain the positions of these fixed points in every image. To obtain the subpixel resolution, the 3-D fitting is applied based on the results of NCC. With the obtained camera motions, the images are converted to the cases without distortion through filtering the obtained camera motions. An edge is then selected as the measure point on the structure. And the proposed Zernike moment matrix edge-finding method is used to obtain positions of this edge in every image. To convert coordinates at pixels to physical coordinates, the scale factor in every image is calculated and applied. Finally, the displacement history of the measured points is obtained.
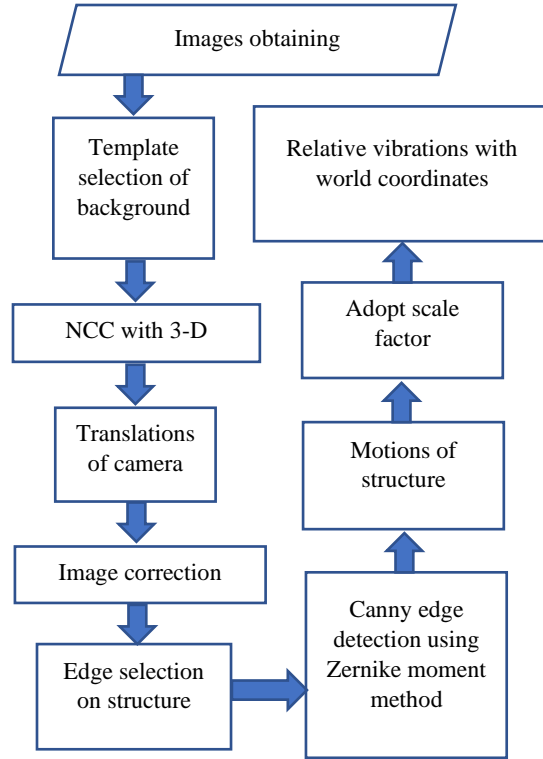
Figure 21. Flowchart of the proposed displacement monitoring method.

4.2.2. Determination of Scale Factors

Template matching methods will be used to filter the rigid motion of the camera when the motion is perpendicular to the image plane as shown in section 2.5. The motion of the camera, when moving toward the structure, would cause the distance change between the camera and the structure. And the distance change between the camera and the structure would subsequently cause changes of scale factors of every image. Two ways to calculate the scale factor are shown blow:

$$SF_1 = \frac{L_{known}}{O_{known}} \tag{20}$$

$$SF_2 = \frac{D}{F} d_{pixel} \tag{21}$$

L_{known} is the known physical length on the structure surface. O_{known} is the pixel length on the image plane. D is the distance between the camera and the structure. F is the focal length of the camera. And d_{pixel} is the pixel length such as μm/pixel.

4.2.3. Getting Rid of the Translations and Rotations of the Camera

When the camera rotates, the positions of background and measure points would change in image [10]. Since there is an angle between the camera and the structure, a keystone effect with an apparent distortion (Figure 22) will appear if the camera rotates toward the target (around x or z axis). For example, it would cause a square in the original image to look like a trapezoid in the distorted image, the shape of an architectural keystone.
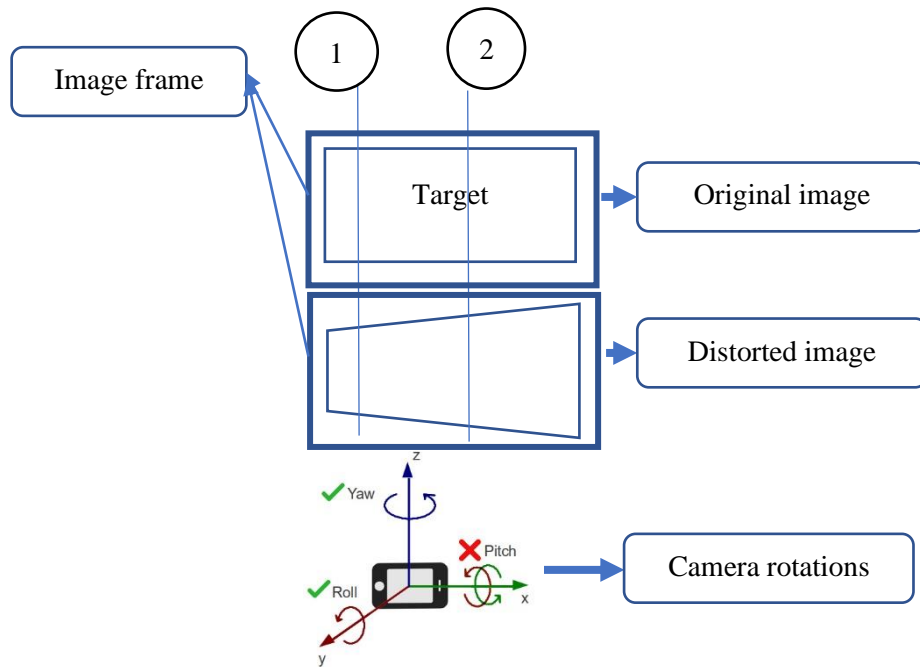
Figure 22. Keystone effect due to rotation of the camera.

As shown in Figure 22, the shape of the original image is a rectangle, and the shape of the distorted image is a trapezoid. At position 1 and position 2, there are two different scale factors, which would cause error on the measured vibrations. The most frequently used technique to solve this issue is interpolation. But in the process of interpolation, it cannot convert the distorted image to the original image perfectly, which could cause errors. The method suggested in this chapter could be called the "adaptive scale method", which uses the adaptive scale factors for the measured features in every image. For example, if the measure point is at location 1 in Figure 22, scale

50

factors would be different from each other for the whole series of images. If every scale factor at the measure point for every image could be calculated, the error due to the rotation of the camera could be easily solved. As shown in Eq. (8), to calculate *SF1*, only two parameters need to be obtained. It is easy to find a known distance on the surface of a structure and the relative distance of the pixels on the image.
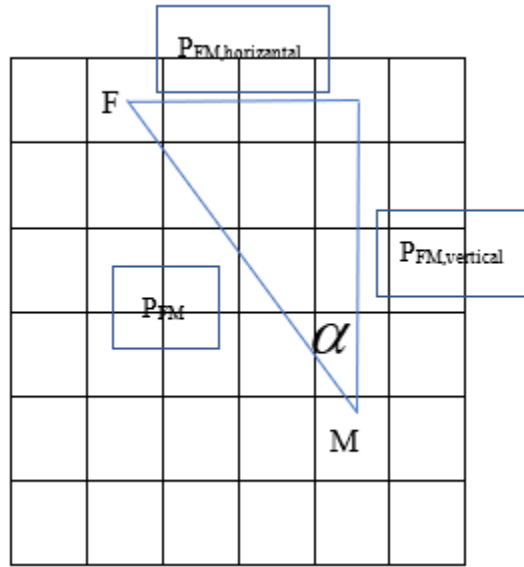
Figure 23. Calculations of average scale factors.

Figure 23 shows a 6x6 pixels image. *F* is a fixed point and *M* is a measurement point and they are approximately in the same plane. $D_{FM,vertical}$ is the vertical physical distance between *F* and *M*. $P_{FM,vertical}$ is the vertical pixel distance between *F* and *M*. Because of distortion due to rotation of the camera, the scale factors between *F* and *M* are different. So $D_{FM,vertical}$ could be calculated using the equation below:

$$D_{FM,verticle} = \sum_{FM,vertical} SF_i \times 1 \tag{22}$$

Because the distortion causes linear changes of scale factors between *F* and *M*, there is a linear change of scale factors too. So $D_{FM,vertical}$ could be calculated blow:

$$D_{FM} = SF_{average} \times P_{FM} \tag{23}$$

$$D_{FM,vertical} = SF_{average} \times P_{FM} \times Cos\,\alpha \tag{24}$$

$$D_{FM,vertical} = SF_{average} \times P_{FM,vertical} \tag{25}$$

$$SF_{average} = \frac{SF_F + SF_M}{2} \tag{26}$$

Similar procedures could apply to the horizontal physical distance and horizontal displacement. When the fixed point and the measure point could not be assumed in the same plane. The motion of the camera and the motion of the structure are calculated using their own average of scale factors and coordinates. For the rotation of the camera that is parallel to the structure plane, it would only cause the rotation of the image. The Speeded-Up Robust Features (SURF) algorithm could be used to find blob features, and M-estimator Sample Consensus (MSAC) algorithm was used to find the translations according to the feature points. Then changes of angle and scale on images due to arbitrary motion of cameras could be obtained.

4.2.4. Speeded-Up Robust Features (SURF)

SURF was first published in 2006 and has three main parts: interested point detection, local neighborhood description, and matching. SURF uses square-shaped filters as an approximation of Gaussian smoothing and adopts a blob detector based on Hessian matrix to find points of interest. The determinant of the Hessian matrix is used as a measure of local change around the point and points are chosen where this determinant is the maximum.

Given a point $P = (x, y)$ in an image I, the Hessian matrix $H(x, \sigma)$ in x at scale σ could be defined as follow:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{27}$$

$L_{xx}(x, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ at point x in image I, and similarly for $L_{xy}(x, \sigma)$ and $L_{yy}(x, \sigma)$.
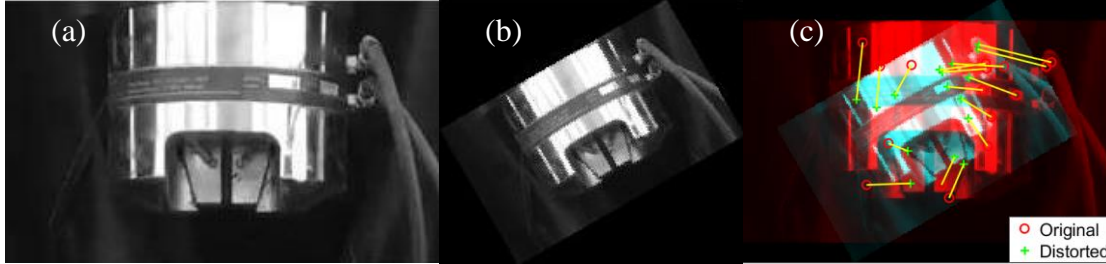
Figure 24. (a) Original image; (b) Scaled and rotated image; (c) Matching results considering scale factor and angle of rotation.

Figure 24 shows the results of feature matching on the MTS between the images before tilting and after tilting and scaling. The original image was first rotated by 30 degrees and scaled to 0.7. Then M-estimator Sample Consensus (MSAC) algorithm was used to calculate the angle and scale based on the results of the feature matching. The outputs are 29.89 degrees for angle and 0.69 for scale, which verifies the accuracy of the suggested method.

4.3. Performance Evaluation Through a Shaking Table Test in Field

4.3.1. Shaking Table Test Setup

Experiments were performed to verify the proposed concept. Figure 25 shows the setup of the experiment in the lab. An iPhone6s with 30 fps acquisition rate and 1920×1080-pixel resolution was fixed on a tripod and used to capture vibrations of the MTS machine. The MTS system performs a 4-Hz vibration, while the displacement of its piston is measured using the deployed iPhone6s. The iPhone was placed 1 m away from the MTS machine. Sinusoidal waves were initiated through the MTS system with controlled frequencies (3 Hz, 2 Hz, 1 Hz, and 0.5 Hz) and amplitudes (2.5 mm, 5 mm, 10 mm, and 10 mm), respectively. To simulate vibrations of the camera, continuous vibrations were triggered manually including translations and rotations in all three directions.

53

Figure 25. Setup of the experiment on MTS. (a) Front view, (b) Side View.

4.3.2. Subpixel Resolution Performance

Figure 26(a) shows the displacement monitoring results measured by the suggested subpixel image processing method for the cases of considering noise correction, without noise correction, and considering noise correction and adaptive scale factor adjustment, when a sinusoidal motion with an amplitude of 5.0 mm and frequency of 2 Hz was applied to the MTS system. NCCs calculated for the cases of considering noise correction and scale factor adjustment, considering noise correction only, and without noise correction and scale factor adjustment are 0.99, 0.97, and 0.53, respectively. The adopted scale factor adjustment improved the monitoring accuracy marginally, however the suggested method improved 45% accuracy of the monitored displacement when considering vibration correction of the camera even without the scale factor adjustment. Figure 26(c) shows the rotation angle at the position of the template and the measure point. The rotation angle varied from -2.5 degrees to 0.5 degrees, which brought tremendous effect

54

to the monitored displacement. The scale factors calculated at the position of the template and the measure point are shown in Figure 26(d). The scale factor ranged from 0.99 to 1.03 and did not change much, because the out-of-plane motion of the camera is not significant enough.

Figure 26(b) shows fast Fourier transform (FFT) results of the monitored displacement for the case of considering noise correction and without noise correction. The FFT results of the monitored displacement for the case of without noise correction have three peaks, where two smaller peaks are frequencies due to noises and the highest peak is the frequency of the MTS motion at 2 Hz. When the data was processed through the suggested method considering noise correction, only one peak at 2 Hz was obtained in the FFT results.
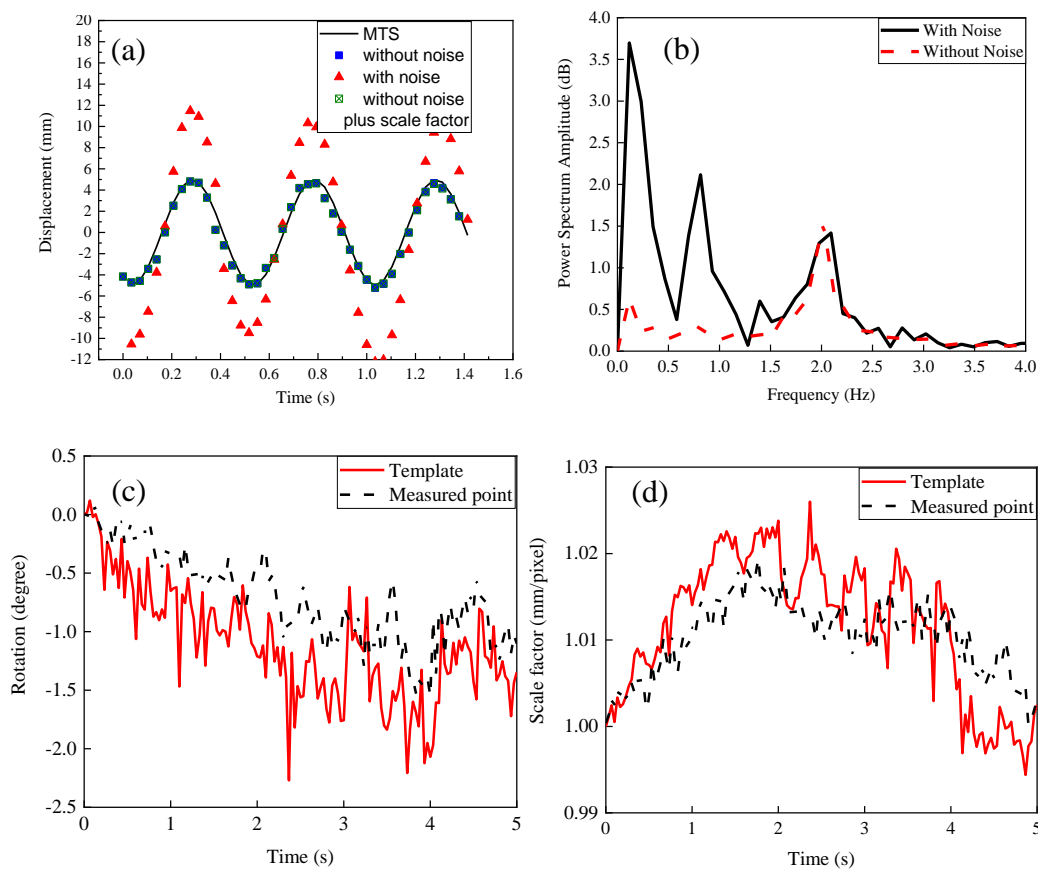
Figure 26. Comparison of captured displacement and the input sinusoidal function with a frequency of 3 Hz and amplitude of 1.5 mm. (a) Monitored Displacement, (b) FFT result of the monitored displacement, (c) Angle of camera rotations, (d) Scale factors calculated.

Figure 27(a) shows four templates in the camera view of the MTS system. Figure 27(b) shows four template ROIs (Range of Interest) and their matching results. The displacement captured using the four different templates are shown and compared with the outputs from the MTS system in Figure 27(c). Figure 27(e) shows NCCs between the captured displacements using the four different templates and the input of MTS. The proposed edge-coordinate-finding method was used to obtain the displacements of the measure point. The small-MTS-template and the large-MTS-template have similar NCCs when the displacements captured through the two templates were compared. The cases using the No.3 and No.4 templates have smaller NCCs, when compared to the cases using template No. 1 and 2. The case using No.4 template has the lowest NCC, because there is a steel column at the top of this template that changes the template when the piston vibrates. The case using No. 1 template has slightly better precision than the case using No. 2 template, because No. 1 template contains more feature points. But larger templates would need more time to operate. Larger NCCs of the template has higher precision.

(b) Template   ROI   Matching result

(c)
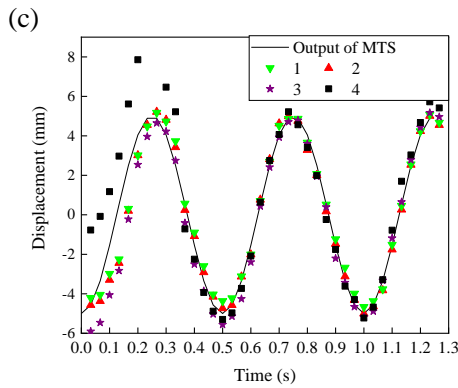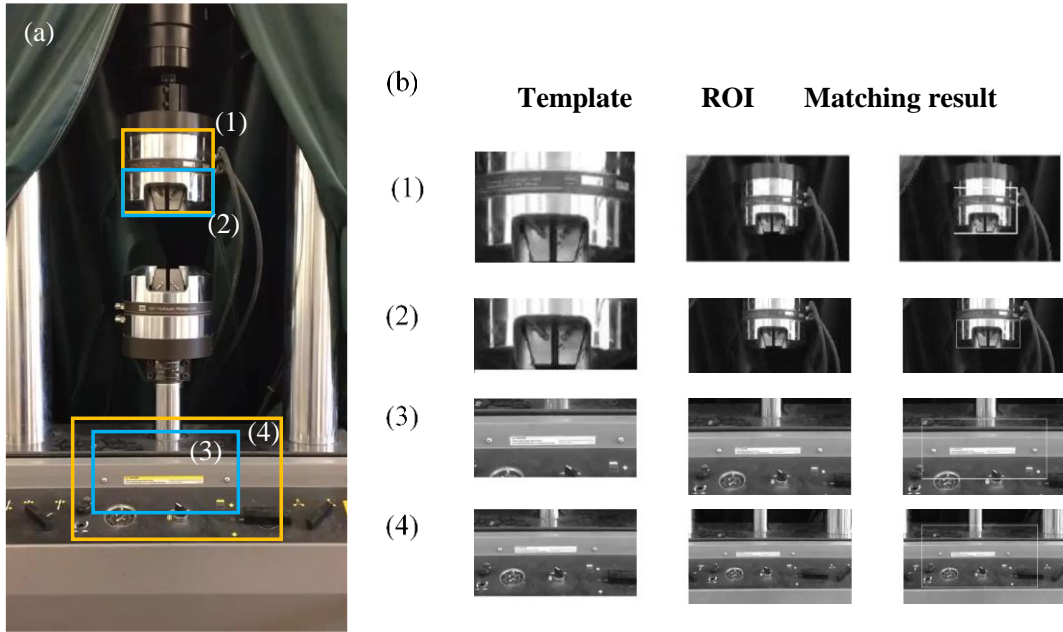
(d)

Template	1	2	3	4
Average of NCC	0.97	0.97	0.96	0.96

(e)

Template	1	2	3	4
Precision between captured vibrations with input of MTS	0.94	0.93	0.91	0.79

Figure 27. Effect of template on the accuracy of displacement monitored. (a) Camera views, (b) Four different templates, (c) Comparison of captured displacement using the four different templates, (d) NCCs calculated for the four different templates used, (e) NCCs between captured vibration using four different templates with input of MTS.

Figure 28(a) shows the results of displacement measured by the subpixel image processing method for the case of considering noise correction, without noise correction, and considering noise correction and scale factor adjustment, when a sinusoidal wave motion with an amplitude of 10.0 mm and frequency of 1 Hz was applied to the MTS system. NCCs for the case of considering noise correction and scale factor adjustment, the case of considering noise correction only, and the case without noise correction are 0.99, 0.99 and 0.56. Similarly, adopting scale factor adjustment did not much improve the monitored displacement accuracy, because out-of-plane motions of the camera was small comparing to the distance between the MTS and the camera. However, vibrations of the camera affected the monitored displacement very much. The displacement obtained with the correction of camera translations and rotations improved its accuracy by 44%. Figure 28(c) shows the rotation angles at the position of the template and the measure point. Figure 28(d) shows the scale factors at the position of the template and the measure point.

Figure28(b) shows FFT results for the cases of considering noise correction and without noise correction. The FFT results for the case with noise have four peaks, where 3 smaller peaks are frequencies due to noises and the highest peak is the frequency of MTS at 1 Hz. After noises were filtered through the suggested method, only one peak at 1 Hz was obtained in the FFT results of monitored displacement.
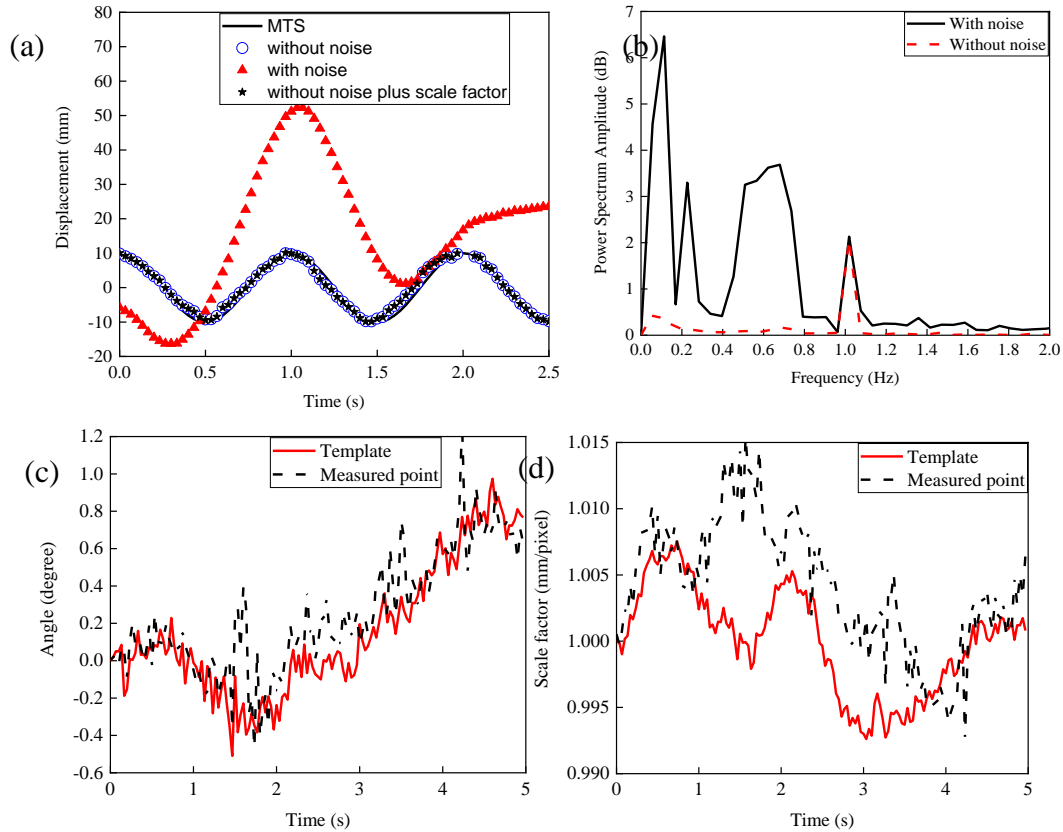
Figure 28. Comparison of captured displacement and the input sinusoidal function with a frequency of 1 Hz and amplitude of 10 mm. (a) Monitored Displacement, (b) FFT result of the monitored displacement, (c) Angle of camera rotations, (d) Scale factors calculated.

Figure 29(a) shows the four templates in the camera view of the MTS system. Figure 29(b) shows the four template ROIs and the matching results. The displacements captured using the four different templates are shown and compared with the outputs from the MTS in Figure 29(c). Figure 29(d) shows the NCCs using the four different templates. Figure 29(e) shows NCCs between the captured vibration using the four different templates and the input of MTS. The four different templates were used to obtain the vibrations of the camera and the proposed edge-coordinate-finding method was used to obtain the displacements of the measure point. The captured displacements of the measure point using the four different templates are compared with the input of the MTS, respectively. The small-MTS-template (No.2) and the large-MTS-template (No.1) have similar NCCs during the whole testing process, while No.3 and No.4 templates have smaller

NCCs because there is a shadow caused by the curtain on the MTS system. Another reason is these two templates have low contrast and flat grayscale values; thus, these two templates do not have enough feature points to make them highlighted. Larger NCCs of the template have higher precision. No.1 template has better precision than No.2 template because No.1 template contains more feature points. But larger templates would need more time to process.

Template	1	2	3	4
Average of NCC	0.977	0.969	0.636	0.588

Template	1	2	3	4
NCCs between captured vibrations with input of MTS	0.98	0.97	0.70	0.65
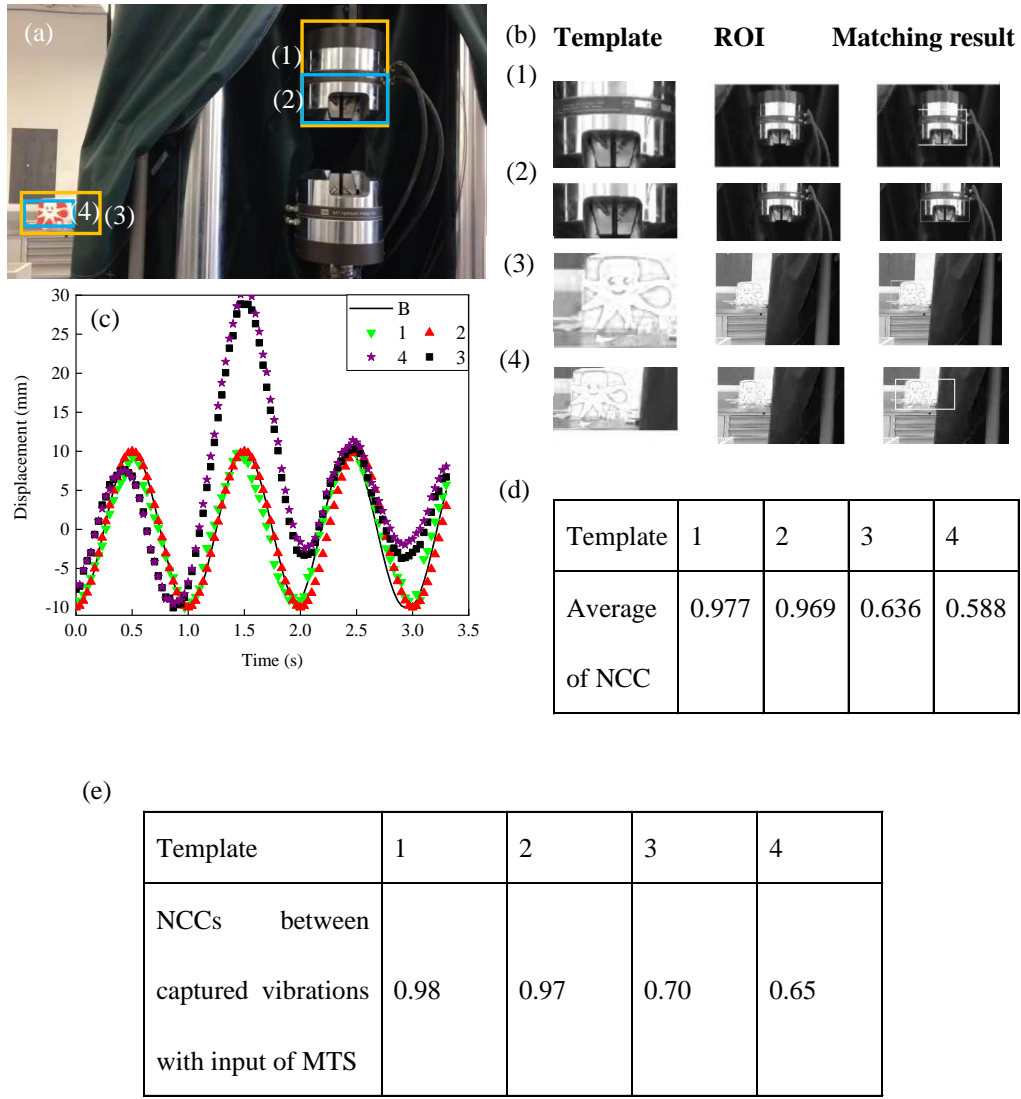
Figure 29. Effect of template on the accuracy of displacement monitored. (a) Camera views, (b) Four different templates, (c) Comparison of captured displacement using the four different templates, (d) NCCs calculated for the four different templates used, (e) NCCs between captured vibration using four different templates with input of MTS.

Above experiments tested the suggested method on displacements with the large-amplitude and regular noises. Cameras on UAVs could be affected by small-amplitude noises such as wind or the pass-by traffic. Figure 30(a) shows the displacement obtained by the proposed image processing method considering small-amplitude noises on the camera. The motion of the MTS is set to move with 10 mm of the amplitude and 0.5 Hz of the frequency. The amplitude of the noise brought in by movement of camera is less than 2 mm and irregular (Figure 30(a)). Only small difference exists for the obtained displacement considering the noise or not considering noise induced by motions of the camera and both displacements give the similar FFT distribution as shown in Figure 30(b). NCCs for the case with noise correction and the case without noise correction are 0.99 and 0.98, respectively.
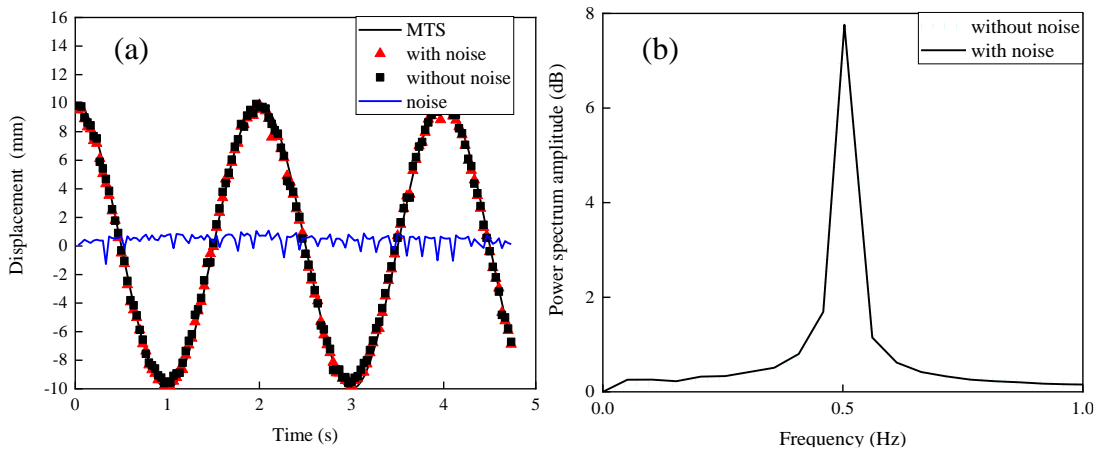
Figure 30. (a) Comparison of captured displacement data with input sinusoidal function with frequency of 0.5 Hz and amplitude of 10 mm, (b) FFT results.

Figure 31(a) shows the displacement obtained through the proposed image processing method and the small-amplitude noise induced by motions of the camera. The input motion of MTS is at 2.5 mm of amplitude and 3 Hz of frequency. The amplitude of the noise is less than 1.5 mm and irregular. Only small difference exists for the monitored displacement for the case with noise correction and the case without noise correction and both capture the motion of MTS very

well. NCCs for the case with noise correction and the case without noise correction are 0.97 and 0.92, respectively. Filtering the motion effect of the camera improved the NCC value nearly 5%. As shown in Figure 31(b), frequencies obtained using the captured displacement without noise correction and the captured displacement with noise correction are the same, both at 3 Hz.
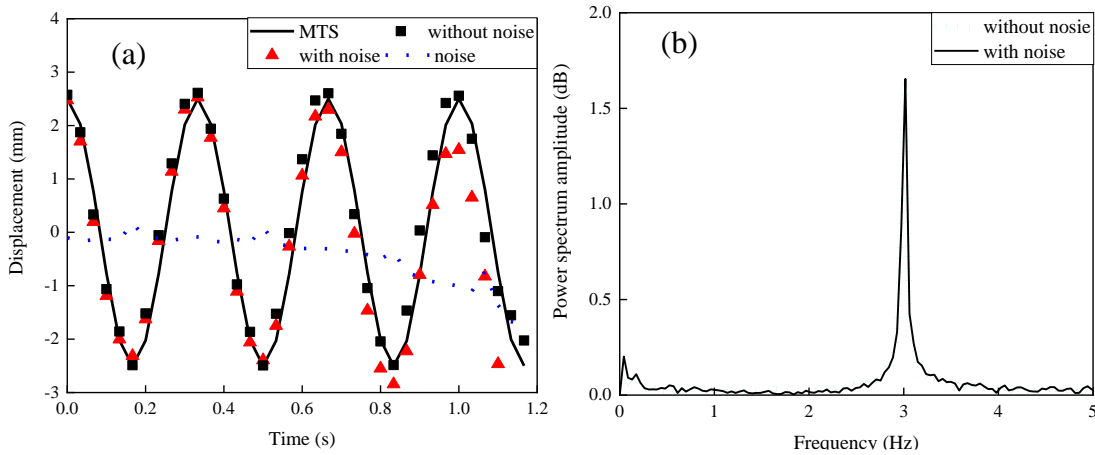
Figure 31. Comparison of captured displacement data between different processing methods for an input motion with frequency of 1 Hz and amplitude of 2.5 mm. (a) Displacement and noise, (b) FFT results.

4.4. Conclusions

The proposed vision-sensor system obtained vibrations of measure points on the structure successfully. The accuracy of monitored displacement was improved significantly by filtering vibrations of the camera through applying the subpixel curve fitting technique on NCC and the proposed Zernike moment edge-coordinate-finding method. Excellent agreements were obtained between displacements calculated by the proposed image-based displacement sensor and the outputs of MTS. Adaptive scale factor adjustment for each image improved the monitored displacement accuracy but only slightly, because the out-of-plane motion of the camera is small.

In summary, the suggested image-based sensor system has many advantages compared to existing image sensor systems in literature, such as

- The proposed image-based sensor system could filter out vibrations of the camera, which improves the accuracy of monitored displacements significantly.

- All six motions of the camera (three translations and three rotations) were considered in this chapter. Two translations which are parallel to the target image were eliminated by tracing the positions of the background point. The translation which is perpendicular to the image plane was eliminated by updating scale factors of each image. The two rotations which are perpendicular to the image plane would cause the changes of the angle between the camera and the image plane. These angles would cause the keystone effect which was eliminated by updating the average scale factor between the background point and the measure point. The scale factor and the rotation angle parallel to the image plane at the background point and the measure point were calculated using the Speeded-Up Robust Features method.

4.5. Reference

1. Chan, T. H., Ashebo, D. B., Tam, H., Yu, Y., Chan, T., Lee, P., and Perez Gracia, E. (2009). "Vertical displacement measurements for bridges using optical fiber sensors and CCD cameras—a preliminary study." Struct Health Monit, 8(3), 243-249.

2. Cheng, C., and Kawaguchi, K. (2015). "A preliminary study on the response of steel structures using surveillance camera image with vision-based method during the Great East Japan Earthquake." Measurement, 62, 142-148.

3. Cigada, A., Mazzoleni, P., and Zappa, E. (2014). "Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system." Experimental Mechanics, 54(2), 255-271.

4. Feng, M., Fukuda, Y., and Feng, D. "Nontarget vision sensor for remote measurement of bridge dynamic response". J. Bridg. Eng (2015). "Asce." Bridg. Eng, 20, 04015023.

5.   Henke, K., Pawlowski, R., Schregle, P., and Winter, S. (2015). "Use of digital image processing in the monitoring of deformations in building structures." J. Civ. Struct. Health Monit, 5(2), 141-152.

6.   Ni, Y., Ye, X., and Ko, J. (2010). "Monitoring-based fatigue reliability assessment of steel bridges: analytical model and application." J Struct Eng, 136(12), 1563-1573.

7.   Pan, B., Xie, H.-m., and Xu, B. (2006). "Performance of sub-pixel registration algorithms in digital image correlation." Meas Sci Technol, 17(6), 1615-1621.

8.   Yoon, H., Hoskere, V., Park, J.-W., and Spencer, B. (2017). "Cross-correlation-based structural system identification using unmanned aerial vehicles." Sensors, 17(9), 2075.

9.   Yoon, H., Shin, J., and Spencer Jr. (2018). "Structural displacement measurement using an unmanned aerial system." COMPUT-AIDED CIV INF, 33(3), 183-192.

5. UAV-BASED DISPLACEMENT MEASUREMENT CONSIDERING TRANSLATIONAL AND ROTATIONAL MOTIONS OF THE UAV

5.1. Introduction

In the past 20 years, image-based displacement sensors become increasingly attractive because of their advantages of low-cost and efficiency. To adopt such a computer vision method in structural health monitoring, template matching is typically used to trace displacements of structures [1-5]. Olaszek [6] first used the template matching method to obtain the dynamic characteristics of bridges. Besides the template matching method, other computer vision methods such as edge detection and digital image correlation (DIC) could detect displacement with high accuracy. Ri et al. [7] proposed a nontarget vision-based method for cable force estimation using handheld shooting through smartphone cameras. The edge of cable is selected as a target and was obtained by the edge detection method. Vibrations of the cable were obtained based on the changes in its locations in ROI of video image sequences, captured by a smartphone camera. Pan et al. [8] enhanced DIC for surface deformation measurement at the macroscopic to micro- or even nanoscale.

The UAV is an increasingly attractive tool to perform structural health monitoring in detecting both static characters such as cracks and dynamic characters such as natural frequencies and vibrations under different loads. It is easy to fly UAVs at a close position to the structures even if structures are in an inaccessible area such as a bridge over a wide river or a great canyon. Yoon et al. [9] proposed an image-based method to detect relative displacements between stories using UAV. A fixed camera was used to compare the results with these from UAVs. The method could get rid of the three translations of the UAV without calculating the camera motion. This method only could be used to detect structures with several stories when the rotations of the camera could

be negligible. The rotations of the UAV caused by wind would cause different translations of different stories, thus the real displacement of the structure could not be obtained using this method. Yoon et al. [10] used a UAV to detect the vibration of a structure in the lab. To obtain the vibrations of the UAV, a checking barcode was used to be the background point. In real environment, it is always difficult to attach a manmade checking barcode at the background of the target structure.

Both the term motion and vibration are used in dynamics. Motion is a general term to describe the movement of an object from one place to another while vibration typically refers to a periodic motion. If the rate of the data acquisition system is high enough, both vibrations with different frequencies and/or motions of the object could be obtained.

In this chapter, three translations and three rotations of the unmanned aerial vehicle were derived through the suggested Normalized Cross Correlation based template matching computer vision method and their effect on the monitored structural displacement was analyzed. The NCC-based template matching was used to obtain the vibration of UAVs by tracing far-field background points at the recorded images. The images captured at any general orientations were rotated back to their original positions which could detect errors in the monitored displacement induced by the rotation angles. A fast NCC-based template matching method was proposed in this chapter, which could accelerate the original NCC-based template matching method significantly and reaches the same level accuracy as that of the original NCC-based template matching method. To verify the concept, a series of experiments were performed on a Multipurpose Testing System machine with amplitudes of 10 mm, 5 mm, and 2.5 mm and frequencies of 2 Hz, 1 Hz and 0.5 Hz, respectively, at the same time a UAV was used to record the motions of the MTS piston. The derived displacements through the UAV and the fixed camera were compared with the true motions of the

MTS machine. Excellent precision and consistence were obtained for the UAV monitored displacement, the MTS piston motion, and the fixed camera derived displacement. In comparison to the case of the fixed camera, scales and rotation angles of the UAV are critical and largely affect the accuracy of monitored displacements. A field experiment was conducted to test the proposed method in a windy environment. Excellent agreement was found between the monitored displacement obtained by the UAV and that derived from the fixed camera.

5.2. Displacement Monitoring Through UAV

5.2.1. Hardware of the Vision Sensor System

Table 3. Main devices used in the experiment.

	DAJIANG Mavic Air 2	Record vibrations
	iPhone 6s	Record vibrations
	Multipurpose Testing System (MTS)	Provide vibrations with input amplitudes and frequencies
	Shaker	Provide vibrations with the fixed amplitude and controlled frequencies

5.2.2. Displacement Measurement Through UAV

Figure 32 shows the process of the proposed image-based displacement sensor system which could obtain the vibrations of a structure using UAVs. Images are captured using a UAV with 4k resolution and modest frame rate of 60 fps. Then rotations of each frame are calculated using the Speeded-Up Robust Features (SURF) and M-estimator Sample Consensus (MSAC) algorithm if the range of target is not too large and the keystone effect could be neglected. Each frame is then rotated to the original position based on the captured angles. Fixed points on the background of the structure are selected by users at this time and the skipped fast NCC is used to obtain the positions of these fixed points in each target image. To obtain the subpixel resolution, 3-D fitting is applied based on the results of NCC. An edge is then selected as the measured point on the structure. And the proposed edge-finding method through the subpixel Zernike moment method in Chapter 2 is used to obtain positions of this edge in every image. To convert coordinates at pixels to physical coordinates, the average of scale factors is calculated and applied to obtain world displacement of the UAV and measured point. Finally, the displacements at the measured points are obtained using the derived coordinate time history.
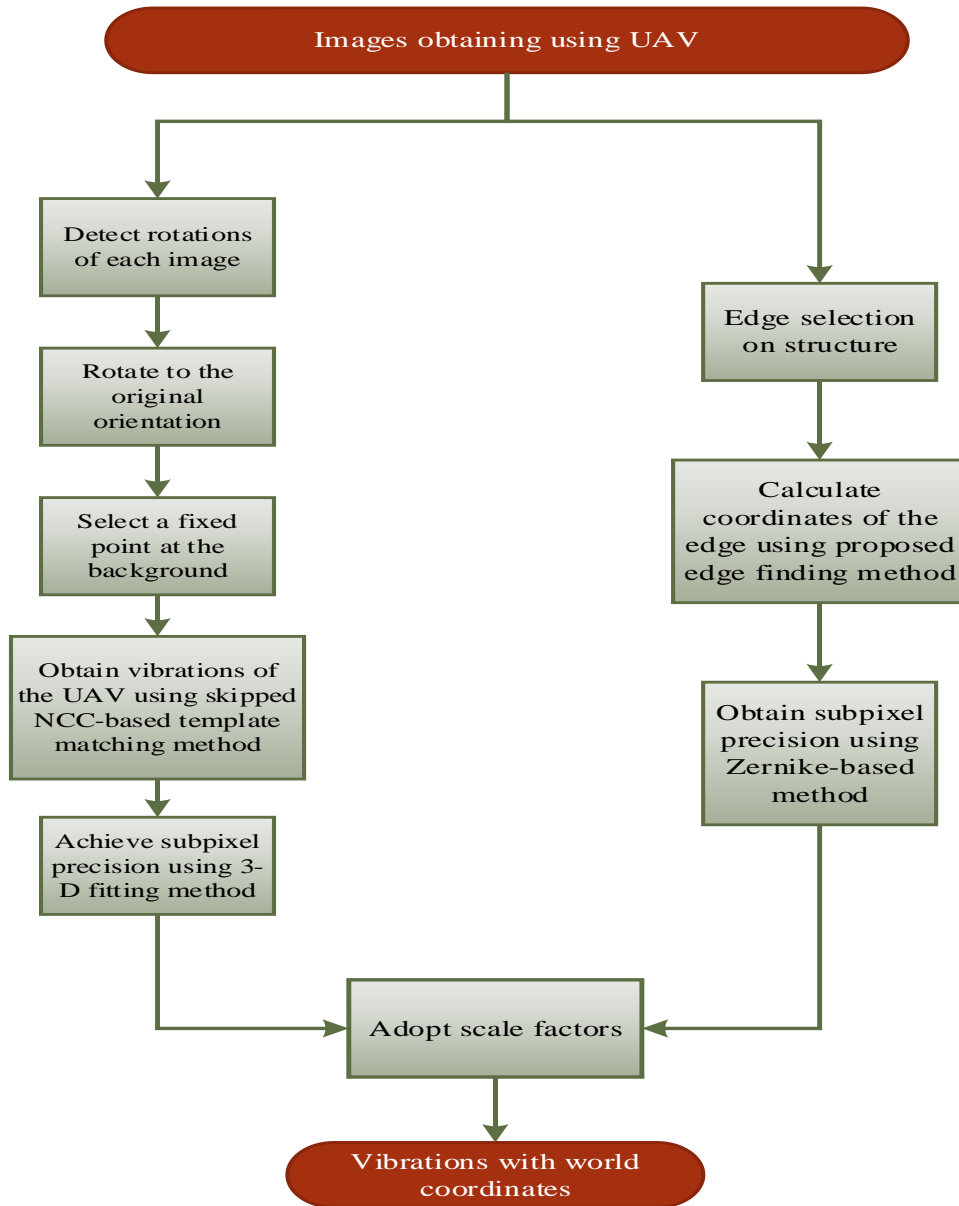
Figure 32. Flowchart of the UAV vision-based displacement monitoring.

5.2.3. Getting Rid of the Translations and Rotations of the UAV

Define x-direction and y-direction in the image plane, and z-direction perpendicular to the image plane. For the rotation of camera that is parallel to the structure plane, it would cause the rotation of the image. Typically, this rotation of the camera is small but may not be negligible.

Firstly, the real rotation (αI) at the first image is calculated by tracing some horizontal or vertical edges (*RE1*) as shown in Figure 33(a). Here *RE1* is the 1st frame image location of a real

horizontal edge in the target, and *REn* is the nth frame image location of the real edge. These horizontal or vertical edges could be edges of buildings, decks of bridges and columns of bridges. With the keystone effect considered, the horizontal lines in the image keep horizontal and straight. Due to the anti-shake system of UAVs, the keystone effect caused by rotations of the UAVs could be small.

Rotations of every image are obtained using the proposed SURF method by tracing background features or the whole captured images as shown in Figure 33(b). Finally, images are rotated to the correct position based on the angles calculated. Based on the rotation on the first image and the rotations between the current image frame and the first image frame, the true rotation angel can be calculated as Eq. (28):

$$\alpha'_n = \alpha_1 + \alpha_n \tag{28}$$

n is the number of the frame; α_n' is the real rotation angle of the frame n. α_1 is the rotation angle of the first frame. α_n is the rotation angle of the frame *n* between the current image frame and the first image frame.

(a) 1st frame (b) nth frame
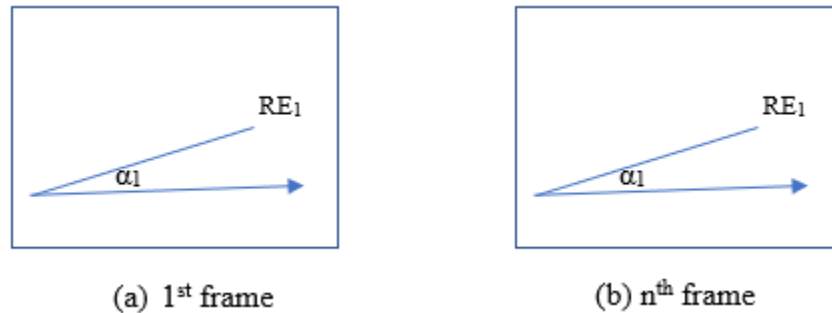
Figure 33. Schematic plots for calculation of rotation angles. (a) First frame, (b) The nth frame.

Sometimes, the keystone effect could not be ignored. To overcome the keystone effect, two elements (E_1 and E_2) at the same height on the structure are selected, such as bolts groups on the surface of bridges. The coordinates and scale factors of these two elements are first obtained. Then

the rotation angle could be calculated as shown in Eq. (29), based on their coordinates and scale factors.

$$\alpha_1 = \arctan \frac{Y_{E1} \times SF_{E1} - Y_{E2} \times SF_{E2}}{X_{E1} \times SF_{E1} - X_{E2} \times SF_{E2}} \tag{29}$$

X_{E1}, Y_{E1}, X_{E2}, and Y_{E2} are the coordinates of elements E_1 and E_2 in the image, respectively; SF_{E1} and SF_{E2} are scale factors of E_1 and E_2. The scale factor is defined as the ratio of real distance of targets over the image distance of the targets in the photo, which can be calculated when the physical coordinates of elements and coordinates of them in the image are known.

5.2.4. Skipped NCC Template Matching

Original template matching methods always need much time to execute, because the NCC at every point needs to be calculated and the position with the highest NCC is the position of the template in the target image. The method named the skipped NCC template matching method was proposed in this paper to accelerate the execution of the original NCC method. Only one peak value of NCC exists in the target image, thus a modified bisect method could be adopted, which will accelerate the NCC process significantly.

The skipped fast NCC template matching method is conducted in multiple stages. At the first stage, the NCCs at the designated points are calculated using Eq. (30). Eq. (31) is then used in the second step to search locally the potential template location.

$$NCC(sh+1, sk+1) = \frac{\sum_{x,y}[f(x,y) - \bar{f}_{sh+1,sk+1}][t(x-sh-1, y-sk-1) - \bar{t}]}{\{\sum_{x,y}[f(x,y) - \bar{f}_{sh+1,sk+1}]^2[t(x-sh-1, y-sk-1) - \bar{t}]^2\}^{0.5}} \tag{30}$$

$$NCC_{\sum_{u\max-s}^{u\max+s} u \sum_{v\max-s}^{v\max+s} v} = \frac{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}][t(x-u, y-v) - \bar{t}]}{\{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}]^2[t(x-u, y-v) - \bar{t}]^2\}^{0.5}} \tag{31}$$

Where f is the pixel value of the image; $\bar{f}_{u,v}$ is the mean of the pixel values of the image; t is the pixel value of the template; $\bar{t}$ is the mean of the pixel values of the template; u and v are the

positions in the image; s is the skipped distance at the first step; $su+1$ and $sv+1$ are the positions in the target image and no more than the size of the target image; u and v are natural integer numbers starting from 0; $umax$ and $vmax$ are the coordinates of the highest NCC point found in the first stage.

Assume the size of the template is $n \times m$ and the size of a target image is $N \times M$. To obtain the position of the template in the target image, $(N1 \times M1)$ NCCs need to be calculated using the original NCC template matching method. $N1$ is equal to $N-n+1$ and $M1$ is equal to $M-m+1$. The method used in this paper skips the calculations of NCCs at some locations. If the skipped distance is 2 as shown in Figure 34 in the first stage, the NCCs at points which are 2 pixels skipped with each other are calculated. The "*1*" in Figure 7 shows the points where NCC is calculated in the first stage. The total calculations of NCCs at the first step is $(N1 \times M1)/9$. At the second stage, the NCCs of points surrounding the highest NCC point found in stage 1 are calculated. The '*2*' represents the calculations of NCCs at the second stage in Figure 38. When the skip is s, the total times of NCC calculations are $(N1 \times M1)/(s+1)^2+((2s+3)^2-9)$, which is always smaller than $(N1 \times M1)$ no matter what s is.

..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
..	1			1			1			1			1	..
..														..
..			3											..
..	1			1	2	2	1	2	2	1			1	..
..				2	2	2	2	2	2	2				..
..				2	2	2	2	2	2	2				..
..	1			1	2	2	1	2	2	1			1	..
..				2	2	2	2	2	2	2				..
..				2	2	2	2	2	2	2				..
..	1			1	2	2	1	2	2	1			1	..
..														..
..														..
..	1			1			1			1			1	..
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..

Figure 34. Calculations of NCC using the proposed skipped NCC method.

72

There is only one template in the target image and the NCC values in the whole searching region are approximately continuous. If the position of the largest value does not locate in both regions at the first stage and the second stage, there would be:

$$N_{max} > N_B < N_{1max} \tag{32}$$

N_{max} is the largest NCC value in the whole searching region which is located outside the selected points at the first stage and the second stage, as shown '*3*' in Figure 38. N_B is the NCC value of the points at the boundary between regions at the first stage and the second stage. N_{1max} is the highest NCC value in the searched regions at the first stage and the second stage. The more pixels overlapped between the template and the searching region in the target image, the larger NCC would be. If the distance between the point to the position of the largest NCC is decreasing, the NCC would be increasing. Thus, N_{max} should be less than N_{1max}, which is impossible. In conclusion, the position with the largest NCC value should be in the regions of the second stage when the skipped distance is much less than the template size.

Figure 35 (a) shows the NCC results using the original NCC template matching method. Other figures in Figure 39 show the NCC results using the proposed skipped fast NCC template matching method using skipped distance s from 1 to 19. The original NCC template matching method detected the highest value which is 0.9408 at location (76, 169) in the searching region with 78309 calculated numbers of NCCs. However, the proposed skipped fast NCC template matching method detected the same highest value of 0.9408 at location (76, 169) by calculating only 8750 numbers of NCCs. When the skipped distances are increased, the number of NCC calculations at stage 1 is decreasing but the number of NCC calculations at stage 2 is increasing.
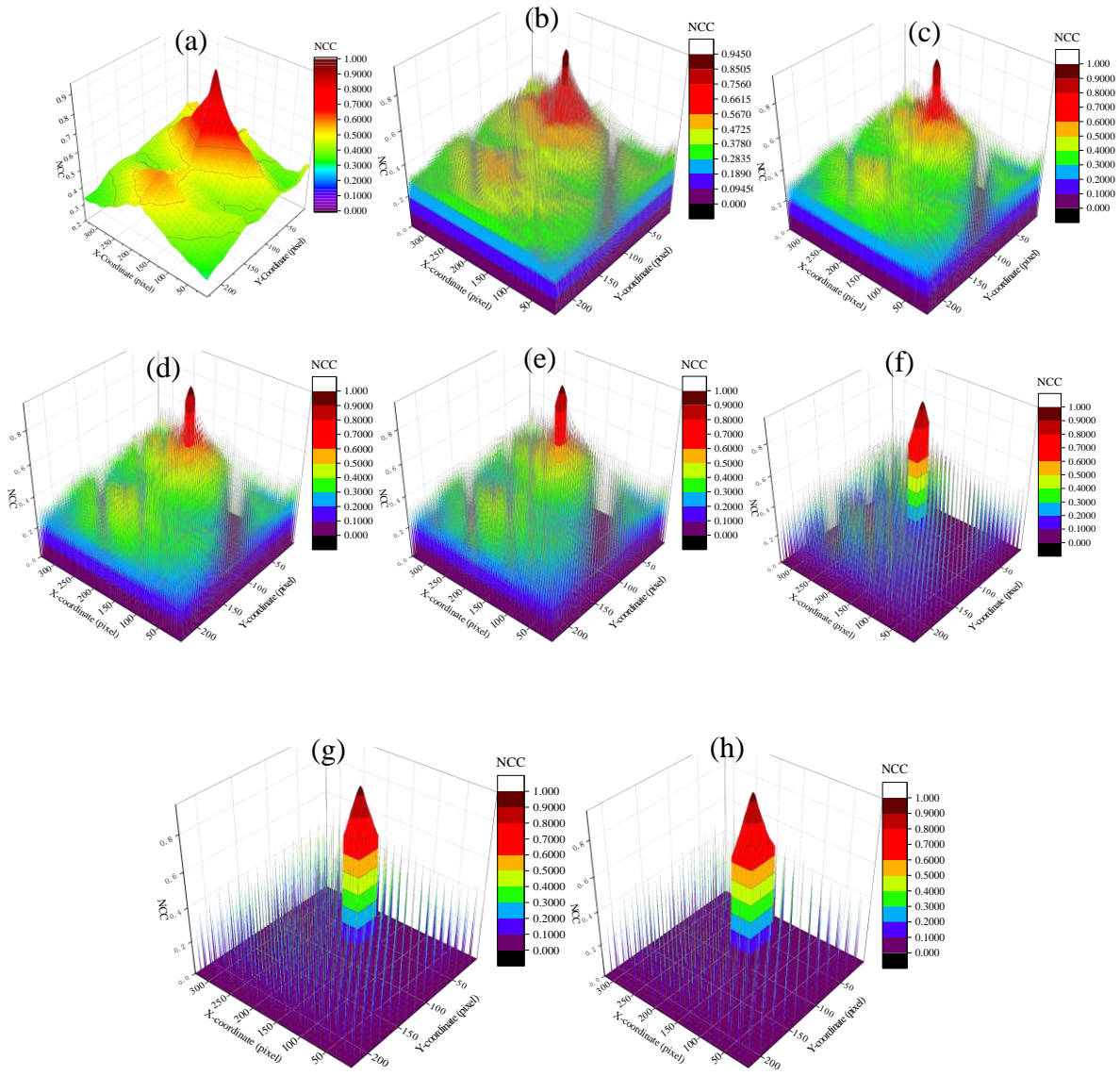
Figure 35. Results of calculation of the NCC using the proposed skipped NCC-based template matching with different skipped distances. (a) 0; (b) 1; (c) 2; (d) 3; (e) 5; (f) 10; (g) 15; (h) 20.

Figure 36 shows numbers of the NCC calculations when the skipped distance s is from 0 to 22. The searching region is an image with 231×339 pixels. A sharp drop in the number of the NCC calculations is founded when the skipped distance s is from 0 to 5. When the skipped distance s is from 5 to 10, the number of the NCC calculations is stable. The number of the NCC calculations begins increasing when the skipped distance s is larger than 10. For a searching region of 231×339

pixels, the skipped distance s could be selected as 5, which will need the least NCC calculations and avoid errors due to the skipped regions.
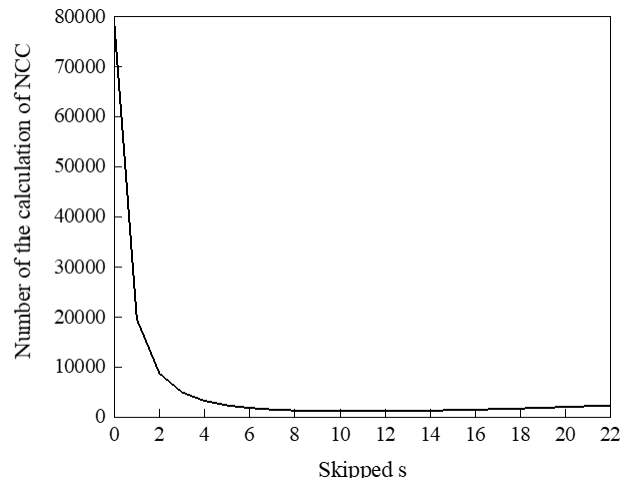
Figure 36. Numbers of the NCC calculations when the skipped distance s is from 0 to 22.

5.2.5. Rotation Angle and Scale Factor Calculations Based on SURF Method

To obtain the rotation angle and the change of the scale factor, the SURF method is used to detect, extract and match features using their descriptors in both the original image and the distorted image. The MSAC method is then used to remove outliers while computing the transformation matrix and detect the scale and the angle using the obtained geometric transformation. To verify accuracy of the calculated scale factors and rotation angles based on the SURF method and MSAC method, the whole images captured by the UAV could be used, since better accurate results for rotation angles and scale factors could be obtained if more image features are included. Figure 37(a) shows the original image captured by the UAV. Figure 37(b) shows the image after rotating 30ºC counterclockwise and scaling down to 70% of its original size. Figure 37(c) shows the results of calculated rotation angle and scale factor based on the original image and scaled image using the SURF method and MSAC method. The obtained rotation angle and the scale factor are 29.998º and 0.699 respectively, which matches well with the input values.
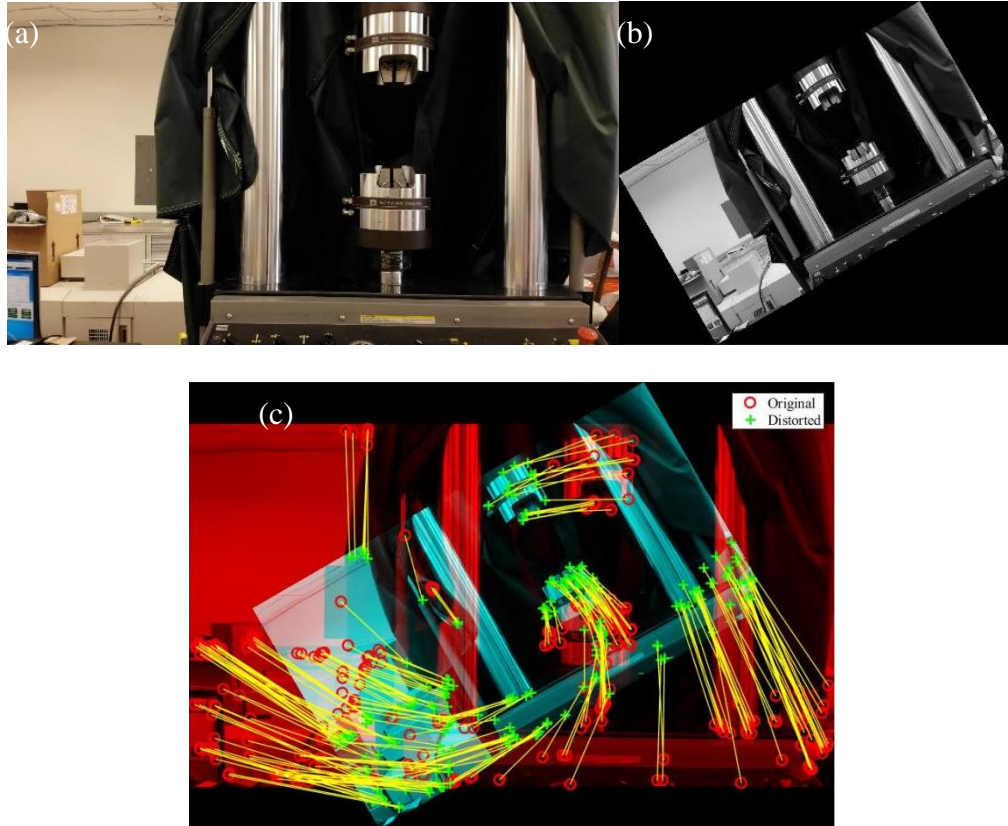
Figure 37. Calculation of the rotation angle and scale factor of images using the SURF and MSAC method. (a) Original image; (b) Rotated and scaled image; (c) Feature matching process through the SURF and MSAC method.

5.3. Performance Evaluation Through MTS Experiment in Lab

5.3.1. MTS Test Setup

As shown in Figure 38 and Table 1, the DJI Mavic drone was flying at 1.5 m away from the MTS and at a height of 1.7 m to record videos of vibrations of the MTS, which has been operated at frequencies of 0.5 Hz, 1 Hz, and 2 Hz with amplitudes of 10 mm, 5 mm, and 2.5 mm respectively. The displacement output error from the MTS is less than 0.5% with respect to the given input command, which is taken as the benchmark for the image-based displacement monitoring method. An iPhone 6s was used at 0.5m away from the MTS and at a height of 1.7m to record the vibration of the MTS too, as a comparison method with the proposed method. The recorded videos were processed using the proposed skipped template matching method to obtain

the translations and rotations of the UAV and the edge detection method developed by authors to obtain displacements of the measured points on the MTS.
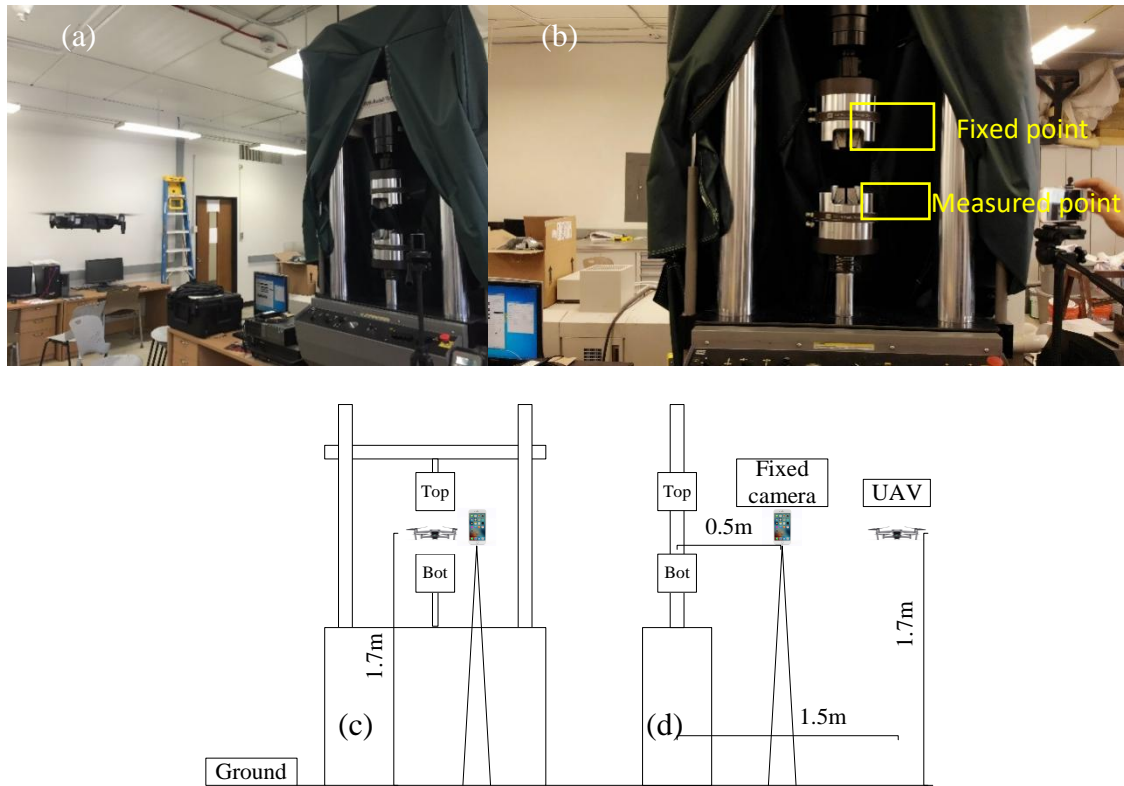
Figure 38. Setup of the experiment on MTS. (a) Front view; (b) UAV view; (c) Front view of the schematic diagram of the experiment setup; (d) Side view of the schematic diagram of the experiment setup.

The main device used in these experiments is the DJI Mavic Air with 4k resolution and 30 frames per second, whose 4k resolution is adequate to obtain high-precision results. The scale factor of experiments using the fixed camera at the first frame is 5/8 mm/pixel. The scale factor of experiments using UAV at the first frame is 0.486 mm/pixel. The iPhone 6s has 30 frames per second and resolution of 544x960 pixels.

5.3.2. Results of Experiments

Figure 39 shows the displacement monitoring results measured by the suggested subpixel image processing method for the cases of using the UAV and the fixed camera, when sinusoidal motions with amplitude of 10.0 mm and frequencies of 0.5 Hz, 1 Hz and 2Hz were applied to the MTS system.
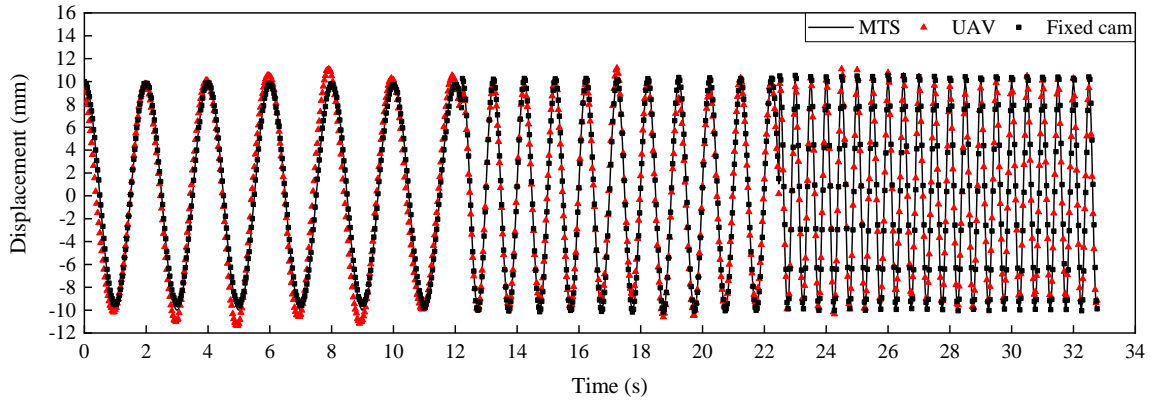
Figure 39. The vibrations obtained by MTS, UAV, and the fixed camera with amplitude of 10.0 mm and frequencies at 0.5 Hz, 1 Hz, and 2 Hz.

Figure 40 shows the original vibration, vibration of the UAV and corrected vibrations at 10.0 mm of amplitude and 0.5Hz, 1 Hz and 2 Hz of frequencies. To obtain correct vibrations, adaptive scale factors, rotations, and translations of the UAV are used to obtain the actual vibrations of the MTS through the image data acquired by the UAV.
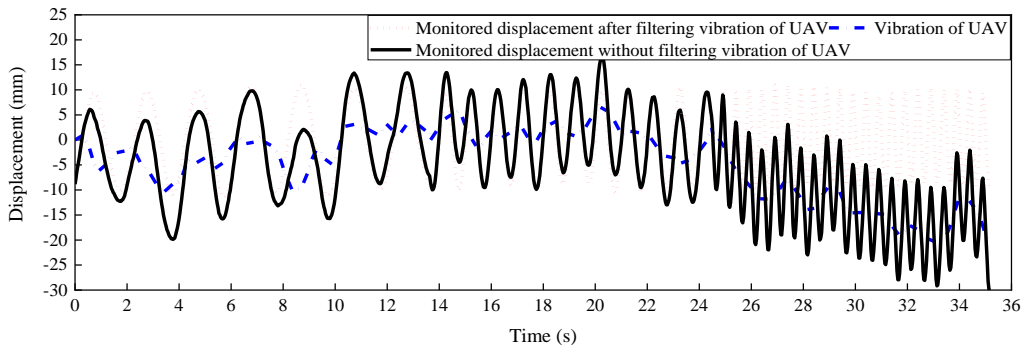
Figure 40. The vibrations of the MTS obtained after filtering vibration of the UAV, vibrations of the UAV, the vibrations of the MTS without filtering vibration of the UAV for the case of amplitude at 10.0 mm and frequencies at 0.5 Hz, 1 Hz, and 2 Hz.

Figure 41 shows the monitored displacement results measured by the suggested subpixel image processing method for the cases of using the UAV and fixed camera, when sinusoidal motions with amplitude of 5.0 mm and frequencies of 0.5 Hz, 1Hz, and 2 Hz were applied to the MTS system.

Figure 41. The monitored displacement obtained through MTS, UAV and the fixed camera with amplitude at 5.0 mm and frequencies at (a) 0.5 Hz, (b) 1 Hz, and (c) 2 Hz.

Figure 42(a) shows FFT results of the monitored MTS displacement after filtering vibrations of the UAV, vibration of the UAV, the monitored displacement without filtering vibrations of the UAV, and the monitored MTS displacement obtained by the fixed camera. The FFT results of the vibration of the UAV have two peaks, which are 0.366 Hz and 0.659 Hz. When the data was processed through the suggested method considering noise correction, only one peak at 0.5 Hz was obtained in the FFT results. Figure 42(b) shows FFT results of the monitored MTS displacement when the MTS is operated at 1 Hz and an amplitude of 10 mm. Two peaks were obtained at 0.184 Hz and 0.552 Hz for the vibration of the UAV. Two peaks at 0.184 Hz and 1.102 Hz were found for the monitored MTS displacement without filtering the vibration effect of the UAV. Only one peak was obtained at 1.012 Hz for the monitored MTS displacement after filtering the vibration of the UAV and a similar peak at 0.974 was found for the monitored MTS displacement using the fixed camera. Figure 42(c) showed FFT results for the monitored MTS displacement when it was operated at 2 Hz and amplitude of 10 mm. One peak at 0.956 Hz was detected for the vibration of the UAV and two peaks at 0.562 Hz and 1.969 Hz for the monitored

MTS displacement without filtering the vibration effect of the UAV. Only one peak was obtained at 1.969 Hz for the monitored MTS displacement after filtering the vibration of the UAV and a similar peak at 2.045 Hz was found for the monitored displacement using the fixed camera.
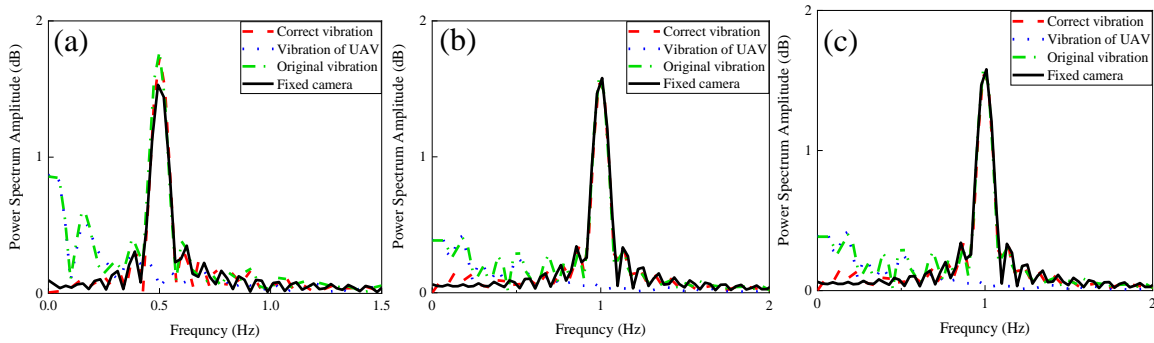
Figure 42. FFT results for the monitored MTS displacement through UAV images after filtering the vibration of the UAV, the vibration of UAV, the monitored MTS displacement without filtering the vibration of the UAV, and the monitored MTS displacement using the fixed camera for the excitations with an amplitude of 10.0 mm and frequencies at 0.5 Hz, 1 Hz, and 2 Hz.

Figure 43 shows the monitored displacement without filtering the vibration of the UAV, vibration of the UAV, and the monitored displacement after filtering vibration of the UAV for the case of excitations with an amplitude of 5.0 mm and frequency at 0.5 Hz, 1Hz, and 2 Hz. To obtain the correct monitored displacements, adaptive scale factors, filtration of translations and rotations of the UAV are used to enhance the image processing.
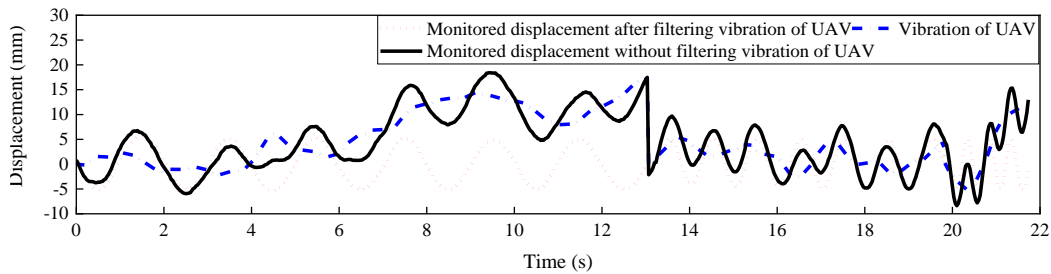
Figure 43. The monitored MTS displacement of the MTS after filtering vibration of the UAV, vibration of the UAV, and the monitored MTS displacement without filtering vibration of the UAV for the case of excitations with an amplitude of 5.0 mm and frequencies at 0.5 Hz, 1 Hz, and *2 Hz.*

Figure 44(a) shows FFT results of the monitored MTS displacement after filtering vibration

of the UAV, vibration of the UAV, the monitored MTS displacement without filter vibration of

the UAV, and the monitored MTS displacement obtained by the fixed camera. FFT results of the

vibration of the UAV have two peaks, which are 0.230 Hz and 0.382 Hz. Three peaks at 0.230 Hz,

0.459 Hz and 0.535 Hz were obtained for the FFT results of the monitored MTS displacement

without filtering vibration of the UAV. When the data was processed through the suggested

method after filtering vibration of the UAV, only one peak at 0.505 Hz was obtained in the FFT

results. Figure 44(b) shows FFT results for the case of excitations at 1 Hz. The FFT results of the

monitored displacement after filtering vibration of the UAV clearly indicated the corrected

frequency has been found at 1.04 Hz and without additional noise frequencies. The frequency

identified has the similar level of accuracy as the monitored displacement through the fixed

camera. Figure 44(c) found the similar trend and correctly identified the input frequency at 2.069

Hz.
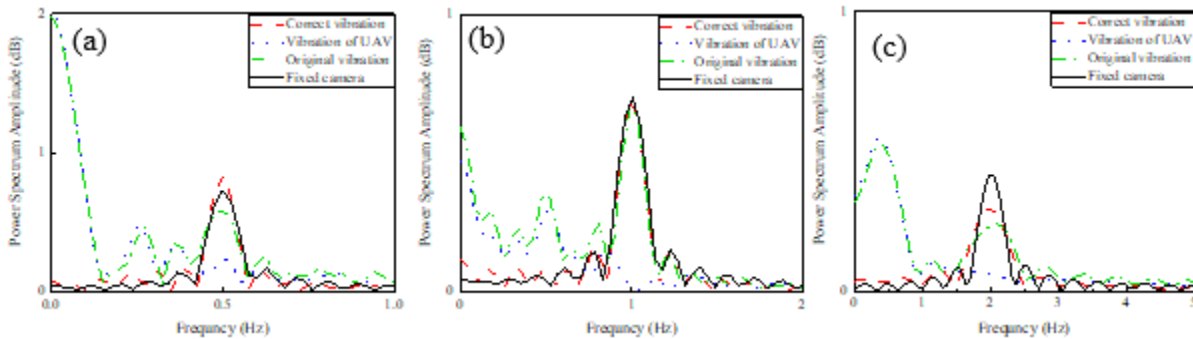
Figure 44. FFT results of the monitored MTS displacement after filtering vibration of the UAV,
the vibration of UAV, the monitored MTS displacement without filtering vibration of the UAV,
and the monitored MTS displacement through the fixed camera for the case of excitations at an
amplitude of 5.0 mm and frequencies at (a) 0.5 Hz, (b) 1 Hz and (c) 2 Hz.

Figure 45 shows the monitored MTS displacement measured by the suggested subpixel

image processing method for the cases of using the UAV and fixed camera, when sinusoidal

motions with an amplitude of 2.5 mm and frequencies at 0.5 Hz, 1Hz and 2 Hz were applied to the MTS system.
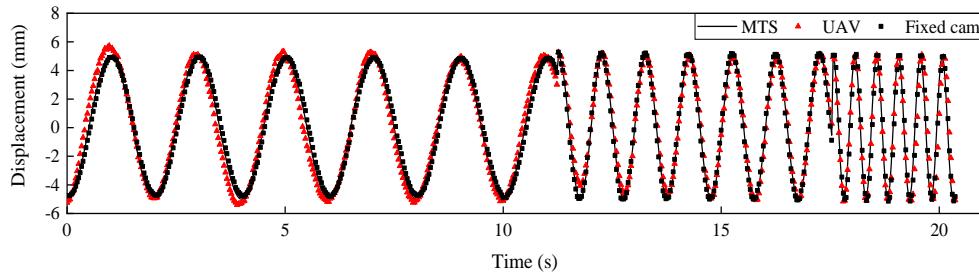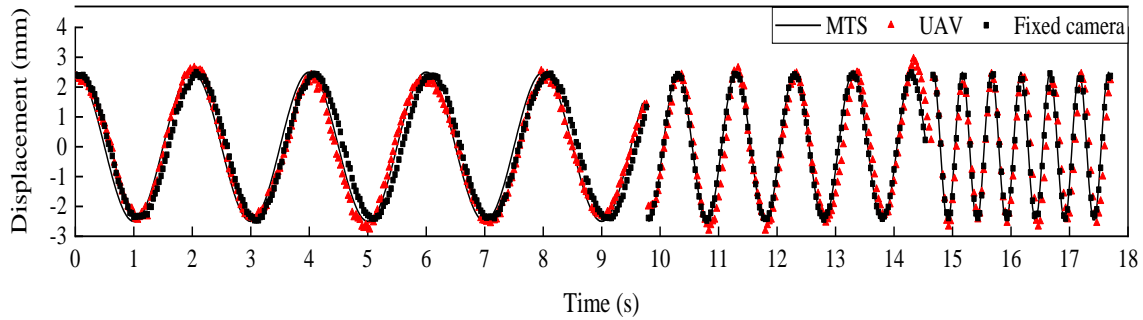

Figure 45. The monitored displacement obtained by MTS, UAV, and the fixed camera for the case of excitation at an amplitude of 2.5 mm and frequencies at 0.5 Hz, 1 Hz and 2 Hz.

Figure 46 shows the monitored displacement without filtering vibration of the UAV, vibration of the UAV, and the monitored displacement after filtering vibration of the UAV for the excitation at an amplitude of 2.5 mm and frequency at 0.5 Hz, 1Hz and 2 Hz. To obtain correct displacements, adaptive scale factors, filtration of rotations and translations of the UAV are used to improve the quality of image processing.
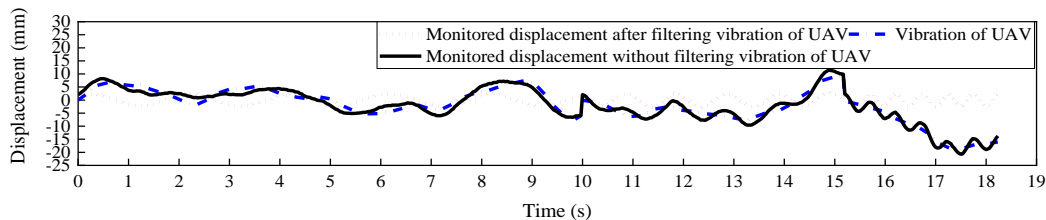
Figure 46. The monitored MTS displacement after filtering vibration of the UAV, vibration of the UAV, and the monitored MTS displacement without filtering vibration of the UAV for the excitation of an amplitude at 2.5 mm and frequencies at 0.5 Hz, 1 Hz and 2 Hz.

Figure 47(a) shows scale factors obtained by the UAV and the fixed camera at each frame using the proposed SURF method and MSAC method. Figure 51(b) shows the scaled plot of the region between frame 1200 and frame 1500. The 300 frames in each experiment are selected and the total number of selected frames is 2700. There are some gaps at the boundaries of each experiment because these experiments were conducted at different time. Scale factors of images

captured by the UAV are from 0.87 to 1.02 and larger than the scale factors obtained by the fixed camera, which are close to 1. As shown in Figure 47(b), the range of the scale factor is the largest between frame 1200 and frame 1500 which may be caused by interferences due to person walking-by or obstacles nearby.
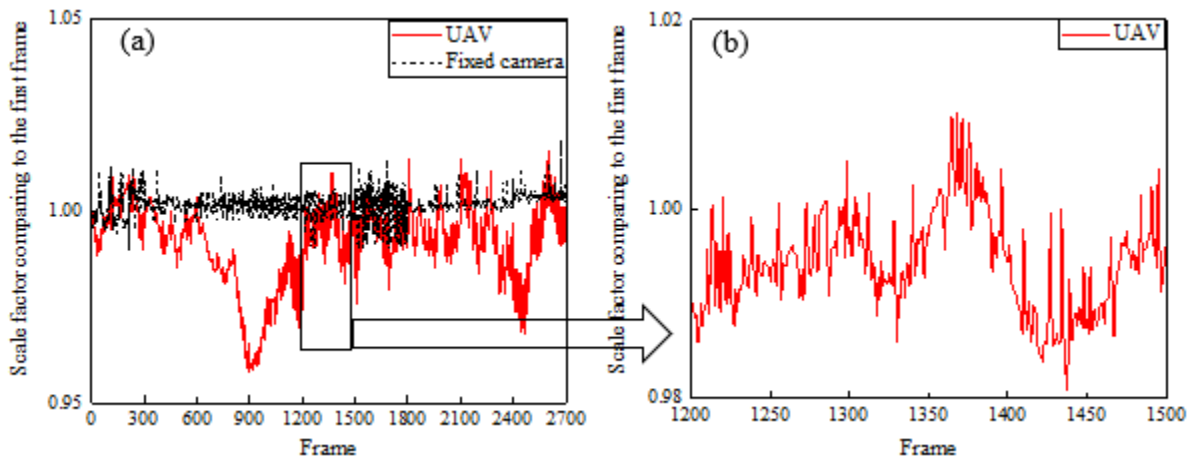
Figure 47. (a) Scale factors of the captured images obtained by UAV and fixed camera, (b) Zoom-in plot of scale factors of the images captured by UAV.

Figure 48(a) shows rotation angles around the z-direction that is vertically parallel to the image plane. Figure 48(b) shows the zoom-in plot of the region between frame 900 and frame 1200, because in this region, the amplitude of the changes of the rotation angles is biggest. Overall, the changes of the rotation angle obtained by UAV are bigger than those obtained by the fixed camera, but they are acceptable and make sense. The biggest change of the rotation angles by using both UAV and the fixed camera are less than 1 degree, whose effect could be negligible.
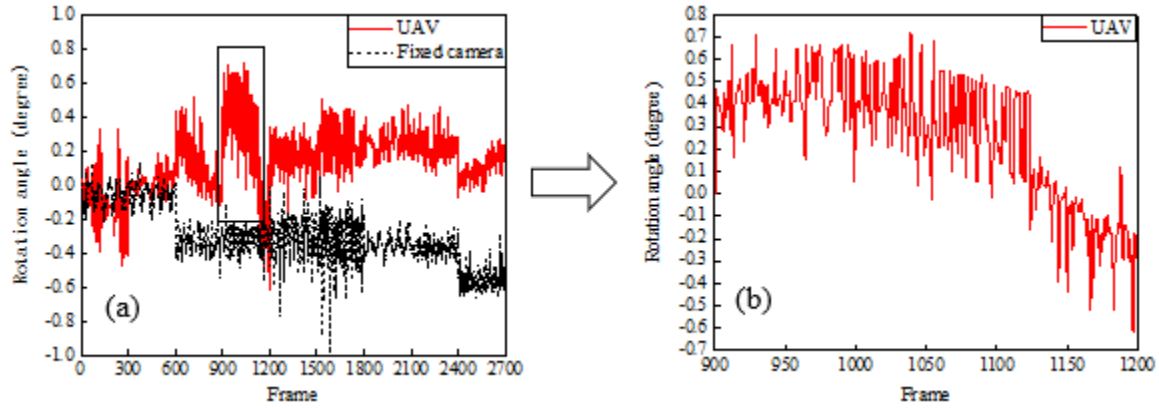
Figure 48. (a) Calculated rotation angles of different devices when capturing images, (b) Zoom-in plot of UAV rotations.

5.3.3. Error Quantification

Figure 49 shows NCCs between the monitored displacement and the benchmark MTS displacement using the image processing method after filtering vibration of the UAV, fixed camera, and the image processing method without filtering vibration of the UAV. The monitored displacement obtained using the fixed camera has the highest NCC compared with the other 2 methods. In contrast, the image processing method without filtering vibration of the UAV has the lowest NCC. The significant improvement of the NCC could be found after filtering vibration of the UAV. In terms of effect of the input frequencies, the monitored displacement has higher accuracy with lower input frequencies. In terms of effect of the input excitation amplitude, the monitored MTS displacement shows higher precision for higher input amplitudes, since the magnitude of error depends on the flight of UAV that is irrelevant to the input displacement amplitude but the ratio between the amplitude of error and the amplitude of the displacement becomes less for larger input displacement amplitudes.
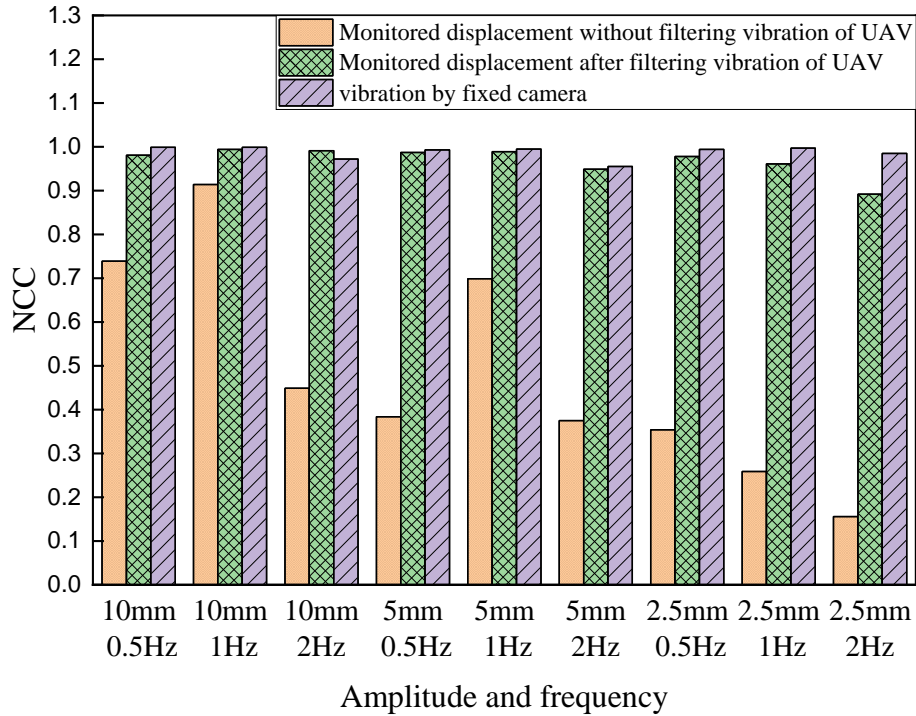
Figure 49. NCCs between the monitored displacement without filtering vibration of the UAV, the monitored displacement after filtering vibration of the UAV, and the monitored displacement captured by the fixed camera with the input excitation of the MTS at amplitude of 10 mm, 5 mm, and 2.5mm and frequencies of 0.5 Hz, 1 Hz and 2Hz.

Table 4 shows the identified frequencies for the monitored displacement for the case after filtering vibration of the UAV, the monitored displacement for the case without filtering vibration of the UAV, the monitored displacement obtained through the fixed camera, and vibrations of the UAV itself. The identified frequencies for the case after filtering vibration of the UAV and through the fixed camera have a unique peak which is very close to the true input frequency of the MTS. Multiple noise frequencies are found in vibrations of the UAV itself. When the vibration amplitude of the UAV is larger than the amplitude of the measured displacement, it will cause significant impacts on the accuracy of monitored displacements.

Table 4. Frequency peaks founded in the FFT results of filtered and unfiltered vibrations, vibrations of the UAV, and the vibrations of MTS captured by the fixed camera at amplitude of 10 mm, 5 mm, and 2.5mm and frequencies of 0.5 Hz, 1 Hz and 2Hz.

Identified frequencies through FFT analysis of the monitored displacement									
Experiment	1	2	3	4	5	6	7	8	9
Amplitude	10mm			5mm			2.5mm		
Frequency	0.5Hz	1Hz	2Hz	0.5Hz	1Hz	2Hz	0.5Hz	1Hz	2Hz
Monitored displacement after filtering vibration of UAV	0.51	1.01	1.97	0.51	1.04	2.07	0.5	0.96	1.96
Monitored displacement without filter vibration of UAV	0.51	0.18 1.10	0.56 1.97	0.23 0.46 0.54	0.45 1.04	0.52 2.07	0.10 0.40 0.70	0.19 0.96	1.96
Fixed camera	0.49	0.97	2.05	0.53	0.96	2.14	0.51	1.06	1.96
Vibration of UAV	0.37 0.67	0.18 0.55	0.96	0.23 0.38	0.45 0.89	0.52	0.2 0.4 0.7	0.42	NA

5.4. Performance Evaluation Through Shaking Table Test in Field

5.4.1. Shaking Table Test Setup

Figure 50 shows the setup of field experiments, in which a UAV was used to obtain horizontal displacements of a column fixed on a shaker. The shaker was used to provide excitations to the column with controlled frequencies up to 200 cycles per minute and an amplitude up to 9.35 mm. In the experiment, the frequency of the shaker was set to 1.367 Hz (82 cycles per minute). The UAV flied at 1.10 m away from the shaker and 80 cm of height from the ground. A fixed camera was also used to record the vibration of the column at 2.10 m away from the shaker and 80

cm of height. The experiment was conducted in open field on a windy day to verify whether the proposed method could monitor vibrations of structures correctly and promptly when there are large noises.
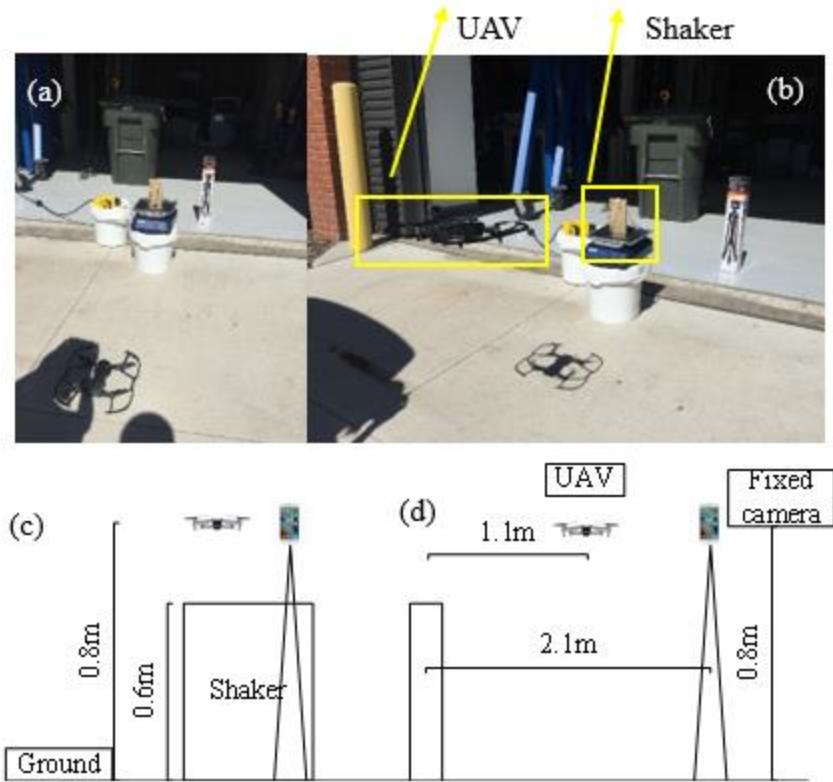
Figure 50. Setup of outdoor experiments. (a) Top view; (b) Front view; (c) Front view of the schematic diagram of the experiment setup; (d) Side view of the schematic diagram of the experiment setup.

5.4.2. Subpixel Resolution Performance

Figure 51(a) shows the column's vibration obtained by the UAV without and with consideration of scale factors, compared with the column's vibrations obtained by the fixed camera and the input of the shaker. Good agreement between them is founded and the NCCs between the UAV monitored displacement without considering scale factors, the UAV monitored displacement with considering scale factors, the fixed camera monitored displacement and the shaker input are 0.920, 0.975, and 0.990, respectively. Figure 51(b) shows the displacement of the column obtained

through the UAV after filtering vibration of the UAV, the vibration of the UAV by tracking the

template point on background, and the monitored displacement of the column without filtering the

vibration of the UAV. Motions of the UAV are large and reach to 20 cm due to the wind in the

windy day. True displacement of the column could be obtained through the proposed image-based

displacement monitoring method after filtering vibration of the UAV.
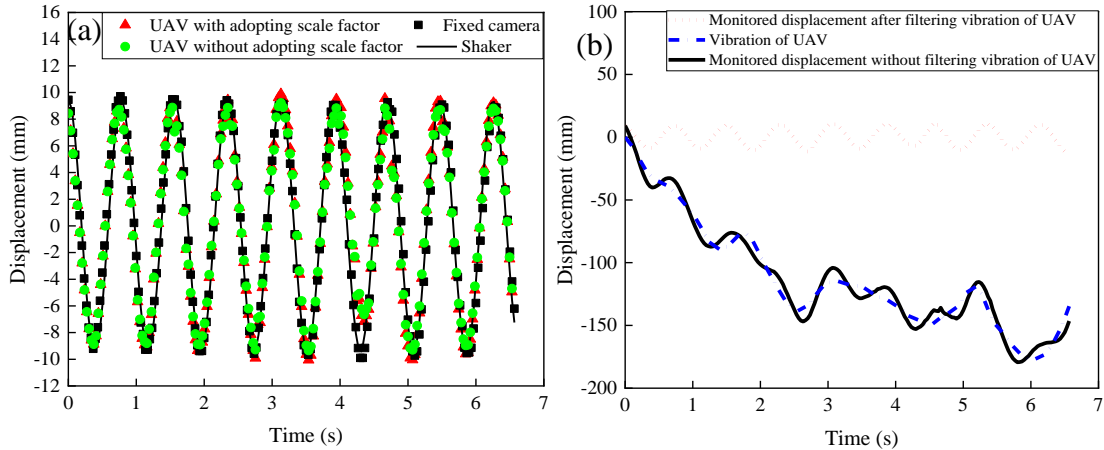
Figure 51. Results of the field experiment. (a) Vibrations captured by the UAV, the fixed camera, and the shaker, (b) The monitored displacement obtained after filtering vibration of the UAV, vibration of the UAV, and the monitored displacement obtained without filtering vibration of the UAV.

Figure 52 shows the FFT results of the monitored column displacement after filtering the

vibration of the UAV, the vibration of the UAV itself, the monitored column displacement without

filtering the vibration of the UAV, and the column displacement obtained by the fixed camera.

Multiple frequencies of 0.606 Hz, 1.061 Hz and 1.363 Hz are obtained in the FFT results of the

monitored column displacement without filtering the vibration of the UAV. Only one clear peak

exists for the FFT results of the monitored column displacement after filtering the vibration of the

UAV and those using the fixed camera, which is at 1.363 and 1.364, respectively, and close to the

true excitation frequency of 1.367. The proposed method could obtain the true frequency of the

structure successfully under aggressive windy environments.
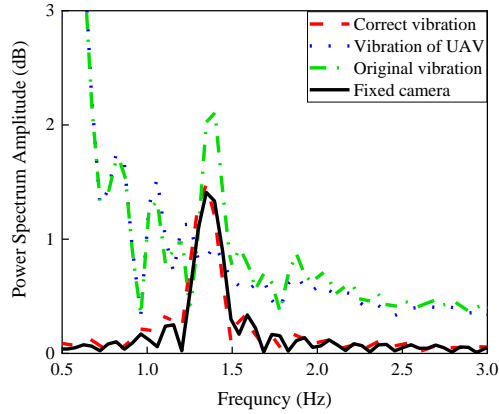
Figure 52. FFT results of the field monitored experiment through different methods.

Figure 53(a) shows variation of rotation angles of the UAV and the fixed camera. Figure 53(b) shows variation of the scale factors of the images captured by the UAV and the fixed camera. Calculated rotation angles of the UAV are larger than those of the fixed camera but still small, ranging from -0.15 degrees to 0.4 degrees. Scale factors of the images captured by the UAV are 1.0 to 1.1 times of the initial scale factor of the first frame, while the scale factors of the images captured by the fixed camera is stable and close to 1.0.
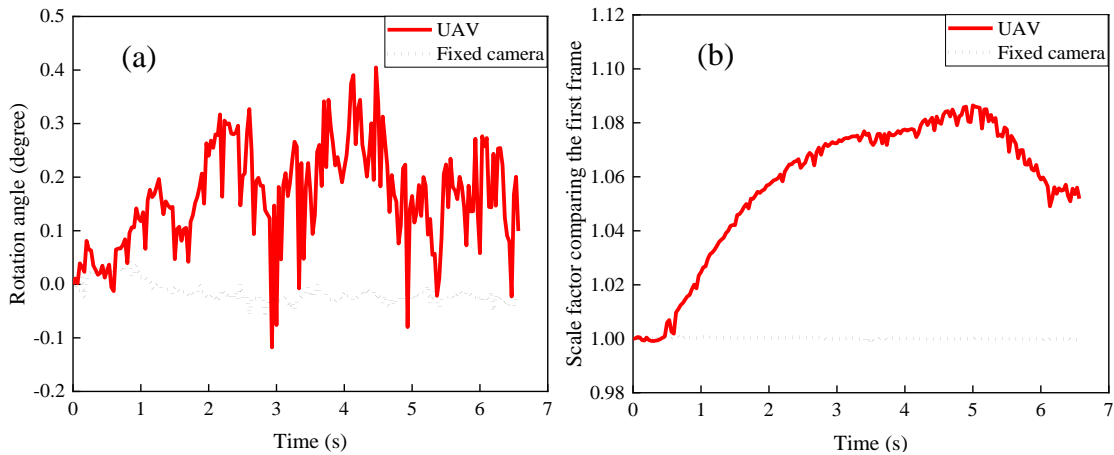
Figure 53. (a) Calculated rotation angles of different devices when capturing images, (b) Scale factors of the captured images by the UAV and the fixed camera.

5.5. Conclusions

The proposed vision-sensor system obtained vibrations of the UAV through examining fixed points at the structure background and derived displacements of measured points on the structure successfully and with high accuracy, through filtering the motions of the UVA. The accuracy of the monitored displacement was improved significantly by filtering vibration of the UAV and through applying the Zernike moment subpixel technique on the proposed edge-coordinate-finding method. 3-D fitting-based subpixel method on the proposed fast skipped NCC-based template matching method was adopted to improve the accuracy in determining camera's motions. Excellent agreements were obtained among displacements calculated by the proposed image-based displacement sensor, the outputs of MTS, and those calculated by the fixed camera.

In summary, the suggested image-based sensor system has many advantages compared to existing image sensor systems in literature, such as

- The proposed image-based sensor system could filter vibrations of the UAV, which improves the accuracy of monitored displacements significantly.

- UAV's rotations which are parallel to the image plane are calculated using the SURF method and MSAC method comparing with the original rotation at the first frame. Moreover, scale factors were calculating at the mean time with rotation angles. To obtain more accurate results, the whole template captured by the UAV are used to provide more features and more accurate results. Because larger template would contain more features which could be traced and used to calculate the scale factor and the rotation angle.

- The proposed skipped NCC-based template matching method is 10 times faster than the original NCC method and these two methods detect the same position of the template in the target images with similar accuracy.

- The proposed image-based displacement sensor could obtain accurate frequencies of measured structure displacements in their FFT results. These monitored displacements are in more than 99% accuracy and do not get contaminated by other frequencies from UAVs and/or environmental factors such as wind.

- Translations of the UAV could be large and up to 20 cm in field. The proposed method could filter these translations successfully. Rotation angles of the UAV are always small both in field and in lab, partially because most of drones have an embedded anti-shake system to control their vibrations.

5.6. Reference

1. Lee, J. J., and Shinozuka, M. (2006). A vision-based system for remote sensing of bridge displacement. Ndt & E International, 39(5), 425-431.

2. Sładek, J., Ostrowska, K., Kohut, P., Holak, K., and Gąska, A. (2013). Development of a vision-based deflection measurement system and its accuracy assessment. Measurement, 46, 1237–1249.

3. Wu, L.J., Casciati, F., and Casciati, S. (2014). Dynamic testing of a laboratory model via vision-based sensing. Eng. Struct. 60, 113–125.

4. Cigada, A., Mazzoleni, P., and Zappa, E. (2014). Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system. Exp Mech, 54(2), 255-271.

5. Henke, K., Pawlowski, R., Schregle, P., and Winter, S. (2015). Use of digital image processing in the monitoring of deformations in building structures. J. Civ. Struct. Health Monit, 5(2), 141-152.

6. Olaszek, P. (1999). Investigation of the dynamic characteristic of bridge structures using a computer vision method. Measurement, 25, 227–236.

7.   Ri, K., Wang, N., and Zhao, X. (2017). Experimental Verification of Vision-based Cable Force Evaluation using a Smartphone. Struct Health Monit.

8.   Pan, B., Qian, K., Xie, H. and Asundi, A. (2009). Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. Meas Sci Technol, 20(6), 062001.

9.   Yoon, H., Hoskere, V., Park, J.-W., and Spencer, B. (2017). Cross-correlation-based structural system identification using unmanned aerial vehicles. Sensors, 17(9), 2075.

10.   Yoon, H., Shin, J., and Spencer Jr, B. (2018). Structural displacement measurement using an unmanned aerial system. COMPUT-AIDED CIV INF, 33(3).

6. DEEP LEARNING-BASED STRUCTURAL HEALTH MONITORING USING BOTH DEFECT IMAGES AND VIBRATION DATA

6.1. Introduction

Computer vision-based sensors combined with deep learning techniques were conducted to detect structural damage automatically, which could overcome the limitations of visual inspection. In the last several decades, visual inspection by trained workers is the main approach to detect structural damage. However, it is subjective, time-consuming, and costly.

Some researchers have proposed deep learning-based defect detection methods using images as input data [1-10] and some other researchers have validated the performance of deep learning-based damage detection methods using vibration data [11-18]. However, there is a challenge when evaluating a civil structure comprehensively by using both defect images and vibration data. In literatures, the vibration data could be used to train CNNs and detect damage locations on structures. Images could be adopted to detect defects as well through the conventional photos collected by UAVs and sent to train CNNs. In this chapter, a novel deep learning-based method was developed, which could use combined images of defects and vibration images as input and improve the accuracy of defect identification and classification significantly. To verify this method, 8000 combined images are generated as the input of the proposed deep learning algorithm.

6.2. Novel Damage Evaluation Method

6.2.1. Process of the Novel Damage Evaluation Method

As shown in Figure 54, first, a structure is selected to be inspected and the potential sensor locations are determined. Then, the video of vibrations of the structure at measure points was recorded through cameras or UAVs. The proposed image-based displacement monitoring method will be used to analyze the recorded video and obtain the curves of vibrations at the measured

points. These vibration data were sent to the proposed CNN-based damage detection algorithm to obtain the location of the damage. Then these damaged locations on the structure will be scanned carefully using the camera or the UAV. These images are trained using the proposed transfer learning method to detect the defects at the damage location. Finally, the evaluation of this structure will be completed according to the damage location and defects at this location.
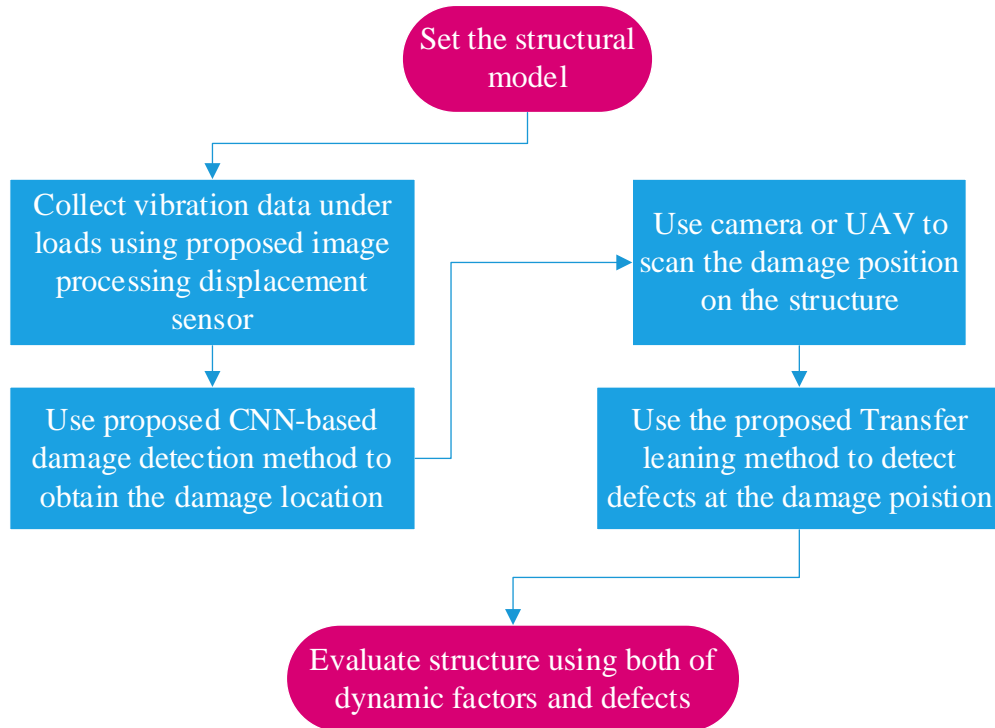
Figure 54. Process of the proposed method.

The CNNs is one of the methods of artificial neural networks (ANNs), which is very popular because of their applications in classification of images, audios, and texts. The main difference between a typical neural network (NN) and a CNN is that NN uses general multiplication while CNN uses convolution which makes it a perfect tool for processing data in a vector or grid form such as images. For SHM, it is very important that some algorithms could use raw data of the structure as the input such as the vibration data. Time history of vibration data obtained from sensors like cameras and accelerometers could be treated as distinct images. The

CNN model is a subset of machine learning, and it is the crucial part of the deep learning algorithms. The CNN structure contains input layers, one or more hidden layers, and an output layer. And there are three main types of layers, which are: convolutional layer, pooling layer, and fully connected layer. The convolutional layer is the core building block of the CNN model and contains most of the computation of the whole CNN structure. The size of the convolutional layer could vary, and the typical size is 3x3. The next step is filtering on an area of the image and a dot product is calculated between the input pixels and the filter. This dot product is then fed into an output array.

6.3. Deep Learning-Based Damage Detection Using Vibration Data

6.3.1. Training the Deep-Learning Damage Detection Model

The convolutional neural network was involved to identify the structural damage through vibration data at measure points. There are three main sections in the progress as shown in Figure 55, which are the dataset collection section, the CNN model training section, and the validation and prediction section. In this chapter, vibrations at measure points under dynamic loads are obtained in the ABAQUS model, and.these vibration data were enlarged by adding Gaussian white noises with different Signal-To-Noise Ratio (SNR) at different cases. The calculation of the *SNR* is shown blow:

$$SNR = \frac{Signal\ Power}{Noise\ Power} \tag{33}$$

$$SNR(dB) = 10\ log_{10} \frac{Signal\ Power}{Noise\ Power} \tag{34}$$

As shown in Figure 55, the vibration pattern will be classified according to different damage case. The vibration data are divided into three groups randomly, which are 70% of the data for training, 15% of the data for the validation, and 15% of the data for the test. The training data is sent to the proposed CNN model as the input data to train this system. Training section and

validating section are necessary for the CNN model to build and improve the sensitivity of the feature learning and the performance of classification. Finally, the test data is sent to the model to predict and obtain the accuracy of this model.
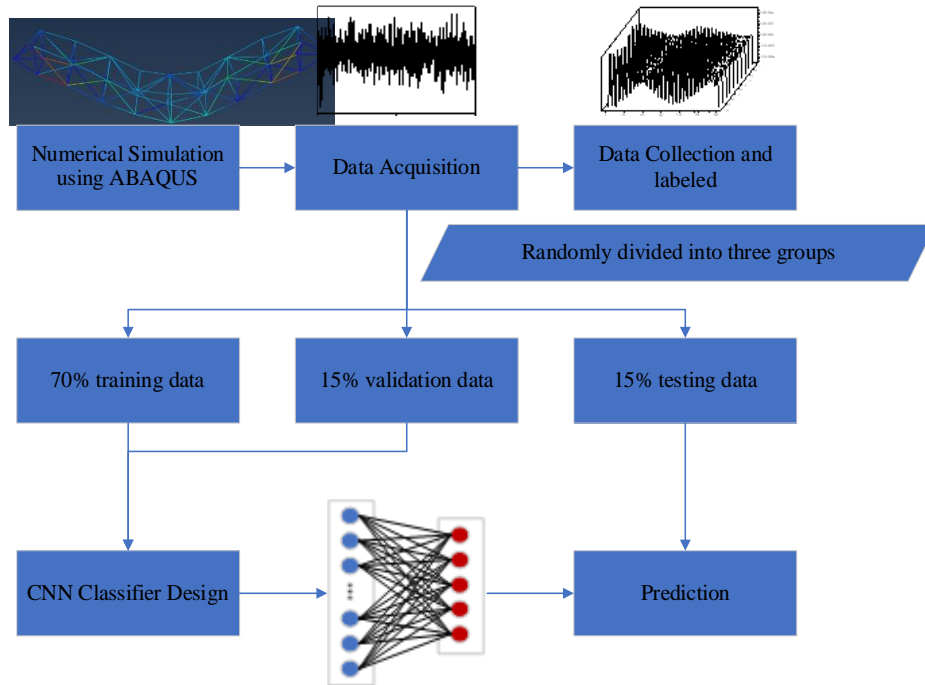
Figure 55. Flow chart of the intended CNN damage detection model.

6.3.2. Proposed CNN Structure for Analyzing Vibration Data

The CNN model is composited of layers that contain artificial neurons arranged in three dimensions including width, height, and depth. In this damage detection chapter, the input data is time-history vibrations at 10 points on the structure, which is a two-dimensional matrix. Stepping through each layer of the CNN model, the matrix is converted into a one-dimensional vector corresponding to the category. Figure 56 illustrates the whole process of the proposed CNN system. Three main layers were involved in the CNN architecture, including the convolution layer, the pooling layer, and the fully connected layer.
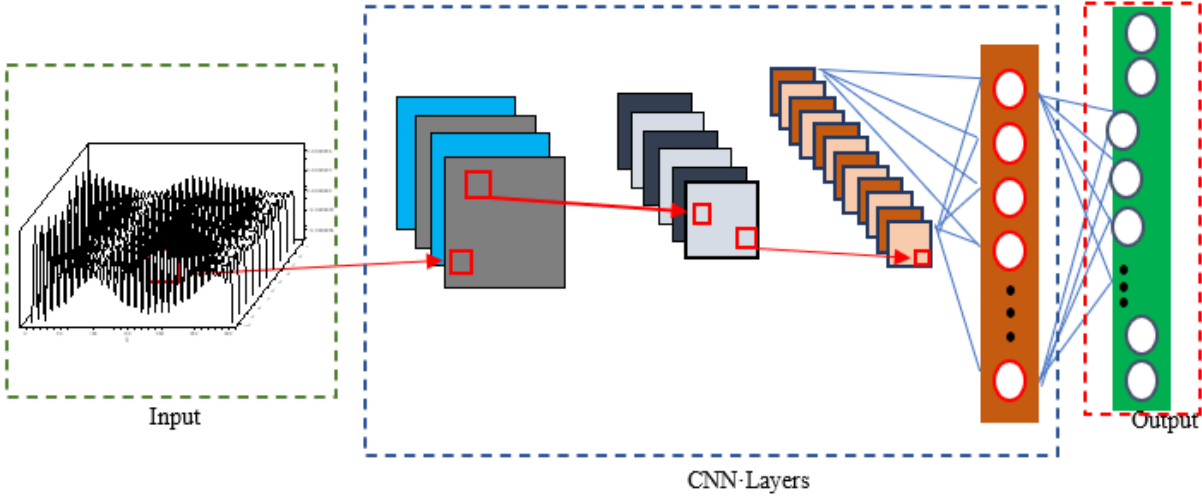
Figure 56. The proposed CNN model structure.

6.3.3. Convolutional Layer

The convolutional layer is the main part of the CNN model. Each convolutional block contains learnable parameters such as weights (filters) and biases. Compared with the input, the width and the height of the filter are spatially smaller, but the depth is the same. For example, the RGB image has 3 layers, so, the convolutional layer should have 3 depths to analyze these 3 layers in the RGB image respectively. The feature maps from the previous layer are convolved with filters and formed the output through the activation function. The formula of the convolutional layer for a pixel is

$$C_{xy} = \sigma \left(\sum_{i=1}^{h} \sum_{j=1}^{w} \sum_{k=1}^{d} f_{ijk} * X_{x+i,y+j-1,k} + b \right) \qquad (35)$$

where *h* and *w* are the sizes of the filter and *d* is the number of channels of the input, *X* is the input vector, *b* is the bias vector, σ is the activation function, and *f* is the weight matrix.

Figure 57 shows the convolution operation on a 4×4 matrix randomly selected by the researcher. A 3×3 matrix is selected as the filter which is generated randomly at the initial state and updated from the model by a backward propagation algorithm. The stride is equal to 1 and

97

four sub-arrays of the same size is generated by sliding along the width and height of the input matrix. Each sub-array is multiplied with the filter matrix. Then, the output value is obtained, which is the sum of the multiplied values and the bias. The size of the output is smaller than the previous layer because of the stride.

After the convolution layer, a nonlinear activation function is followed, which is used for introducing non-linearity into the network. Two of the most common activation functions used in neural networks are rectifier linear unit (ReLU) and softmax. The function of ReLU and Softmax are expressed as follow:

$$f(x) = \max(0, x) \tag{36}$$

$$f(x)_i = \frac{e^{xi}}{\sum_{j=1}^{K} e^{xj}} \tag{37}$$

Where $f(x)$ is the parameter which could represent probability for each group, notation x_i and x_j denotes the membrane potential of neuron, K is the step number.
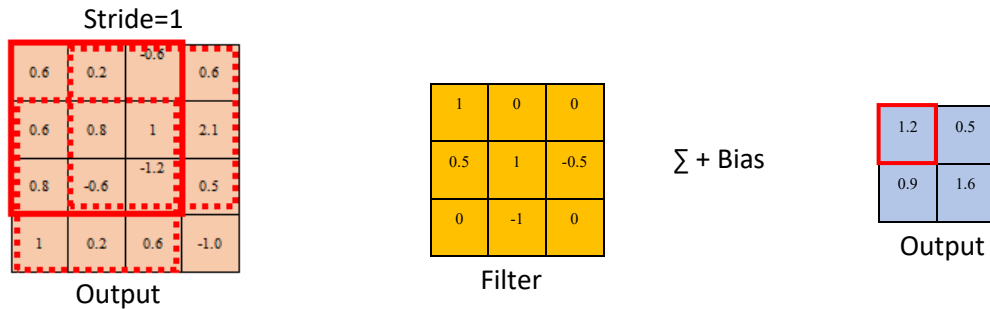
Figure 57. Process of the convolution layer.

6.3.4. Pooling Layer

The pooling layer is used to reduce the spatial size of the feature maps to speed up the computation and increase the robustness of the feature detection. The most common ways are max pooling and average pooling. In max pooing, the maximum values in the filter area are chosen as outputs. For example, as shown in Figure 58, an input layer with the size of 4×4 is operated by a 2×2 max-pooling filter. The stride is designed as 2 which means the next filter should move by

two positions along the right and down, which causes the dimension of the output is declined to 2×2, and the value is the maximum elements in the response field. For the average pooling, average values would be selected in the filter area.
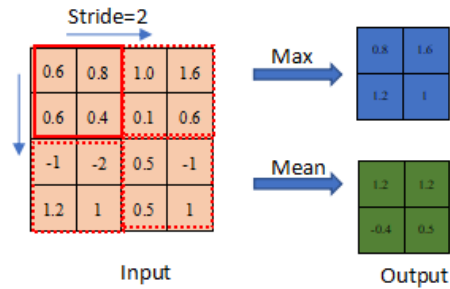
Figure 58. Process of the pooling layer.

6.3.5. Fully Connected Layer

The fully connected layer is the last layer before the output layer of the whole network. In this layer, all the neurons are related to the features generated in the previous layer. Weights and biases in this layer convert the generated features into correspondent categories. The equation of the output y^l is shown:

$$y^l = \sigma(y^{l-1} * w + b) \tag{38}$$

where, σ is the activation function; w and b represent the weights and bias vectors in this layer respectively.

6.4. Different Transfer Learning Methods Used

6.4.1. Alex Transfer Learning

AlexNet is a very famous technique whose image classification performance is significantly accurate. AlexNet is a huge network with 60 million parameters and 650,000 neurons. Table 4 shows the architecture of the AlexNet algorithm. The AlexNet has 8 layers. Its first 5 layers are convolutional layers, and the last 3 layers are fully connected. Between these two groups of layers, there are some layers called polling and activation.

Table 5. The architecture of the AlexNet algorithm.

1	'data'	Image Input	227 × 227 × 3 images with 'zero center' normalization
2	'conv1'	Convolution	96 11 × 11 × 3 convolutions with stride [4 4] and padding [0 0 0 0]
3	relu1'	ReLU	ReLU
4	'norm1'	cross Channel Normalization	cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]
6	'conv2'	Convolution	256 5 × 5x48 convolutions with stride [1 1] and padding [2 2 2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]
10	'conv3'	Convolution	384 3 × 3x256 convolutions with stride [1 1] and padding [1 1 1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3 × 3x256 convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu4'	ReLU	ReLU
14	'conv4'	Convolution	256 3 × 3x192 convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]
17	'fc6'	Fully Connected	4096 fully connected layer
18	'relu6'	ReLU	ReLU
19	'relu6'	Dropout	50% dropout
20	'fc7'	Fully Connected	4096 fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% dropout
23	'fc8'	Fully Connected	1000 fully connected layer
24	'prob'	Softmax	Softmax
25	'output'	Classification Output	crossentropyex with 'tench' and 999 other classes

Figure 59 presents the diagram of the AlexNet. The 5 convolutional layers and 3 connected layers are shown in this figure with data sizes for each layer. The size of input images is 227×227 pixels. The number of categories in the final output is up to 1000.
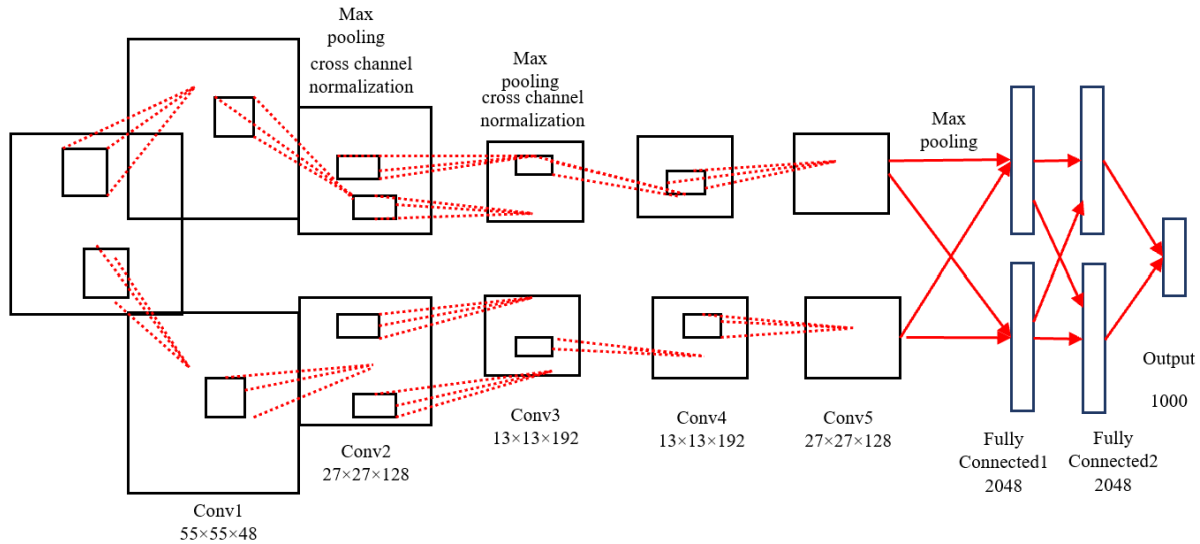
Figure 59. The diagram of AlexNet.

6.4.2. Vgg16 and Vgg19

VGG-16 is a convolutional neural network that is 16 layers deep. The model loads a set of weights pre-trained on ImageNet. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. The default input size for the VGG16 model is 224 x 224 pixels with 3 channels for RGB images.

The concept of the VGG19 model is the same as the VGG16 model except that it supports 19 layers. The "16" and "19" stand for the number of weight layers in the model (convolutional layers), which means that VGG19 has three more convolutional layers than VGG16 does.

6.4.3. GoogLeNet

In the GoogLeNet model, the different part from AlexNet is that it has inception modules. In the inception module, 1×1, 3×3, 5×5 convolution layers and 3×3 max pooling layers are

101

performed at the meantime and the output of these layers is stacked together to generate final output, which could handle objects at multiple scales better.

Table 6. The architecture of the GoogLeNet algorithm.

type	patch size stride	outputsize	depth
convolution	7×7/2	112X112x64	1
max pool	3×3/2	56X56×64	0
convolution	3x3/1	56×56×192	2
max pool	3x3/2	28×28×192	0
inception(3a)	/	28×28×256	2
inception(3b)	/	28×28×480	2
max pool	3x3/2	14×14 ×480	0
inception(4a)	/	14×14×512	2
inception(4b)	/	14×14 ×512	2
inception(4c)	/	14×14×512	2
inception(4d)	/	14×14×528	2
inception(4e)	/	14×14×832	2
max pool	3×3/2	7×7×832	0
inception(5a)	/	7×7×832	2
inception(5b)	/	7×7×1024	2
avg pool	7×7/1	1×1×1024	0
Dropout (40%)	/	1×1×1024	0
linear	/	1×1×1000	1
soft max	/	1×1×1000	0

6.4.4. InceptionResNetv2

The InceptionResNetv2 is a kind of convolutional neural network which has 164 layers and could classify images into up to 1000 categories.

6.4.5. ResNet50 and ResNet101

The ResNet-50 model consists of 5 stages each with a convolution and Identity block. Each convolution block has 3 convolution layers, and each identity block also has 3 convolution layers as shown in Figure 60. The ResNet-50 has over 23 million trainable parameters. The difference between ResNet50 and ResNet101 is that the ResNet-50 model has 101 layers.
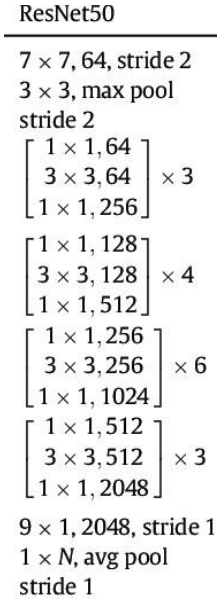
ResNet50

$7 \times 7, 64,$ stride 2
$3 \times 3,$ max pool
stride 2

$$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$$

$$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$$

$$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$$

$$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$$

$9 \times 1, 2048,$ stride 1
$1 \times N,$ avg pool
stride 1

Figure 60. The architecture of the ResNet50 algorithm.

6.5. Abaqus Modeling of a Cantilever Beam

6.5.1. Abaqus Modeling of a Cantilever Beam1 With 10 Elements and 16 Different Cases

Figure 61(a) shows a schematic diagram of the beam modeled using ABAQUS. There are 10 segments (*S1* to *S10*) and 10 measured points (*MP1* to *MP10*). For each case, different segments would be damaged by reducing their Elastic Modulus by 30%. Then a concentrated force (*CF*) is applied to the free end of the cantilever beam. Then the *CF* is released to let the cantilever beam perform free vibration. Thus, the vibrations at each measured point can be obtained. These vibrations need to be labeled and included in the training set of the CNN model. Figure 62(b) shows the beam model in ABAQUS, in which one end is fixed, and the other end is free.
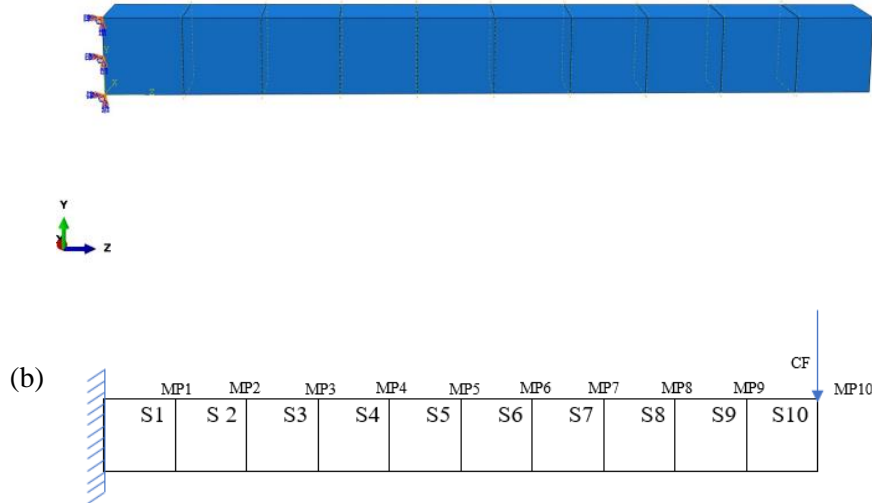
103

Figure 61. (a) The beam model in ABAQUS, (b) The applied load, boundary condition, and the measurement points in the beam model.

Table 6 shows the different damage locations of each case. There are 10 cases with 1 damage location (2 to 11), 1 healthy beam (1), and 5 cases (12 to 16) with 2 damage locations. For example, the No.14 case represents damage locations at element 1 and element 5.

Table 7. The damage location of each case.

Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Damage	NO	1	2	3	4	5	6	7	8	9	10	1 4	3 6	1 5	4 8	5 9

6.5.2. Abaqus Modeling of a Cantilever Beam2 With 10 Elements and 16 Different Cases

As shown in Figure 62, another modeling of a cantilever beam in ABAQUS is conducted. The beam is divided into 256 elements and there are 15 cases with different damage locations. There are one case with a health beam, seven cases with one damage location, four cases with two damage locations, two cases with three damages locations, and one case with four damage locations. The damage of the element was performed by reducing the young's modular to 70%. Vibrations of the beam were triggered by the release of a concentrated force at the free end of the beam. Then these vibrations at 16 measured points which are shown as red points in Figure 62(15).

Then Gaussian noise was used in this chapter to enlarge the training dataset and finally, there are
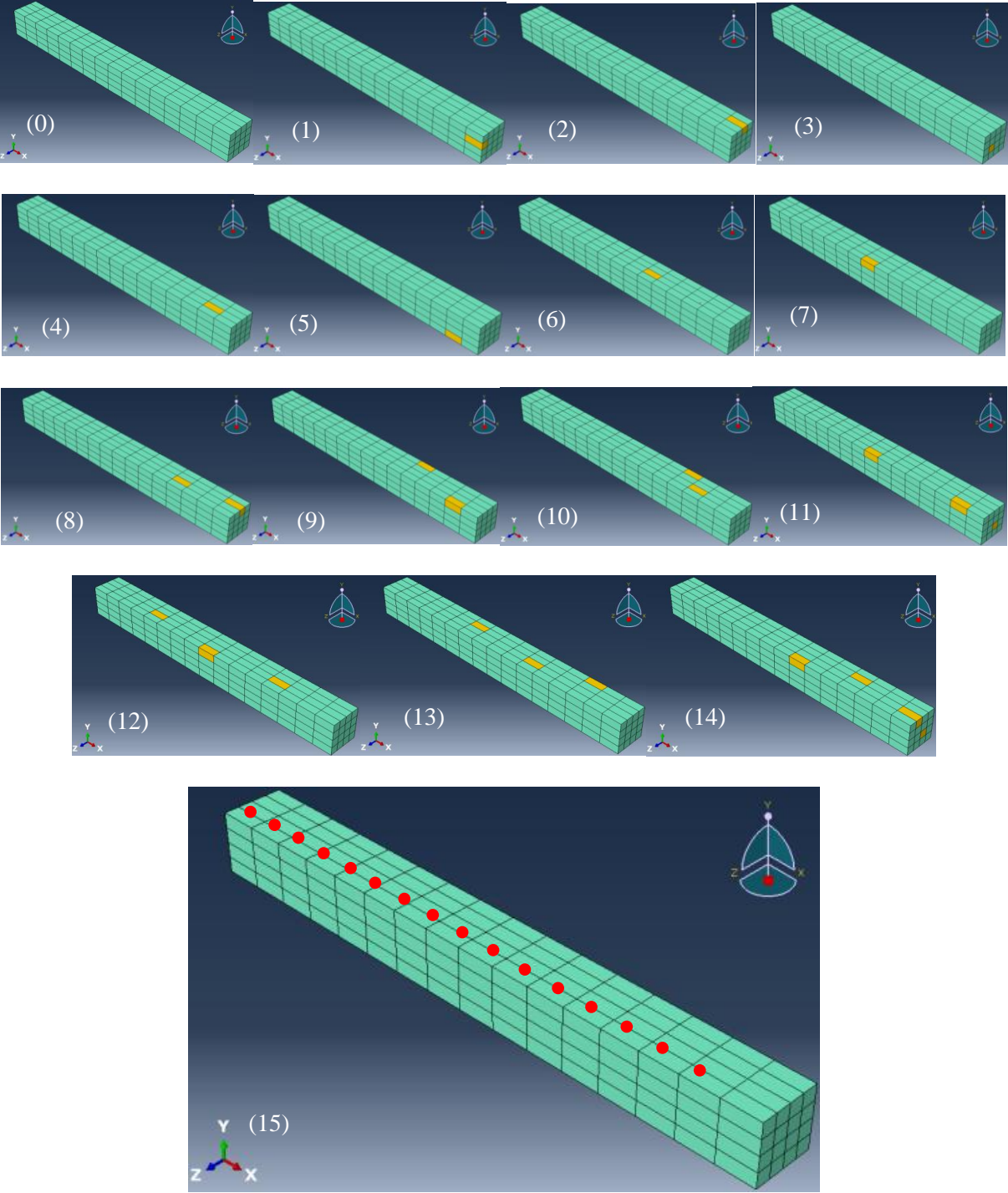
7,500 samples in total.

Figure 62. (0)-(14): All cases in ABAQUS modeling with different damages, (15) 16 measured points.
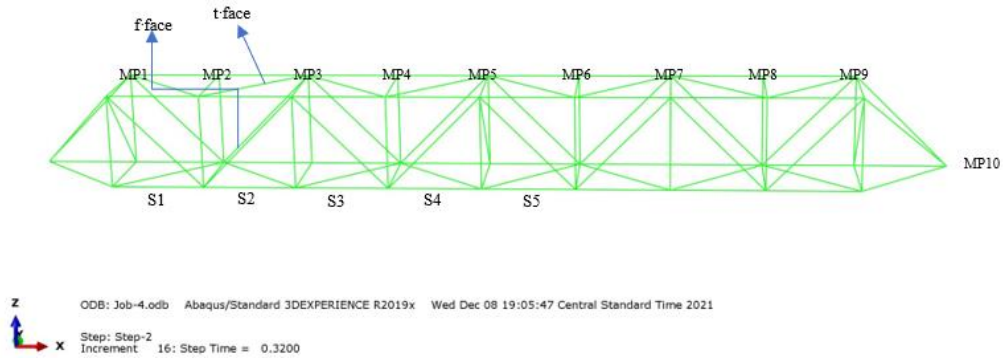
6.6. Abaqus Modeling of a Steel Truss

Figure 63. Modeling of the truss in ABAQUS.

Figure 63 shows the truss model in ABAQUS. The truss in this study has 8 spans and there are 10 measured points on the truss. Table 7 shows the damage locations of different cases and there are 16 cases in total. f represents the front face and t represents the top face of the truss. t, m, and r represent top element, middle element, and right element respectively. There are 5 spans as shown in Figure 63. For example, for the number 8 case, the sloping rod on the front face is damaged by reducing 30% of its Elastic modulus. No. 5 case represents the damage is on the top element of the front face at the fourth span of the truss.

Table 8. Damage positions for different cases.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
span	/	S1	S2	S3	S4	S1	S2	S1	S2	S3	S4	S1	S2	S3	S4	S5
face	/	f	f	f	f	t	t	f	f	f	f	f	f	f	f	f
element	/	T	T	T	T	m	m	m	m	m	m	r	r	r	r	m

106

6.7. Results of Cases on Beam 1

Figure 64(a) shows the original vibration data and Figure 68(b) shows the vibration data with inclusion of the noise in the truss study. The changes are obvious after applying noise, which could enlarge the data set tremendously.
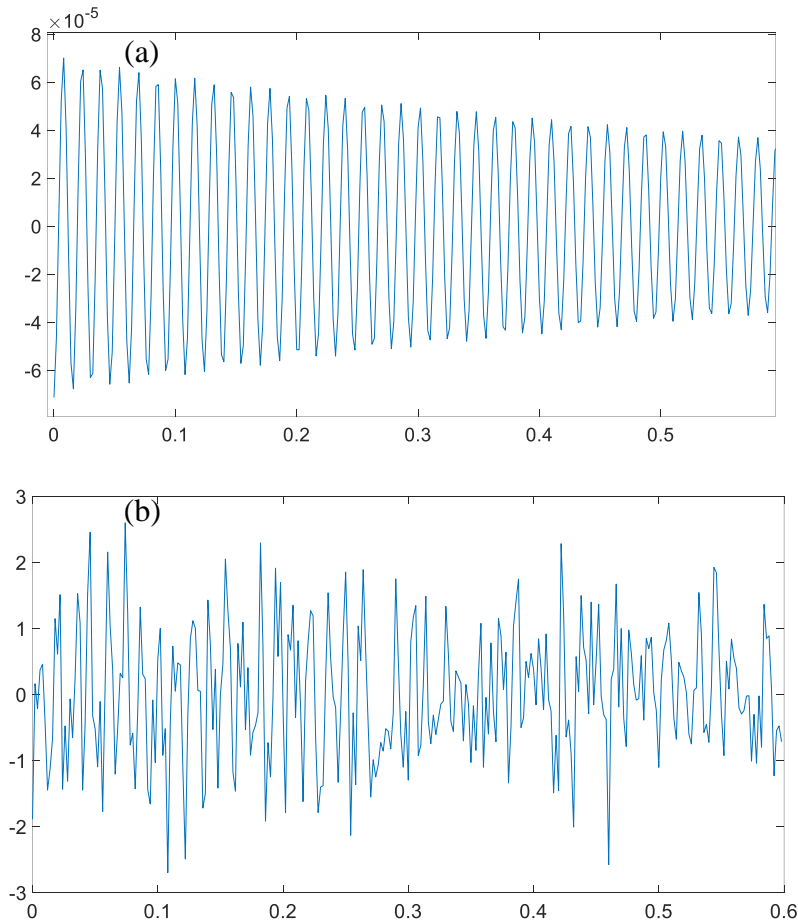
Figure 64. (a) The original vibration data; (b) The vibration data after applying noise.

As shown in Figure 65, the accuracy is 73.88% for the case that the SNR is 90dB. Moreover, the main error exists at point 8, point 9 and point 10, which is partially because the damage at a location far from the fixed end will have a small influence on the dynamic characters of a cantilever beam. And the 2-damage case would be affected by the cases which has at least one same damage location, such as case 4 and case 48.
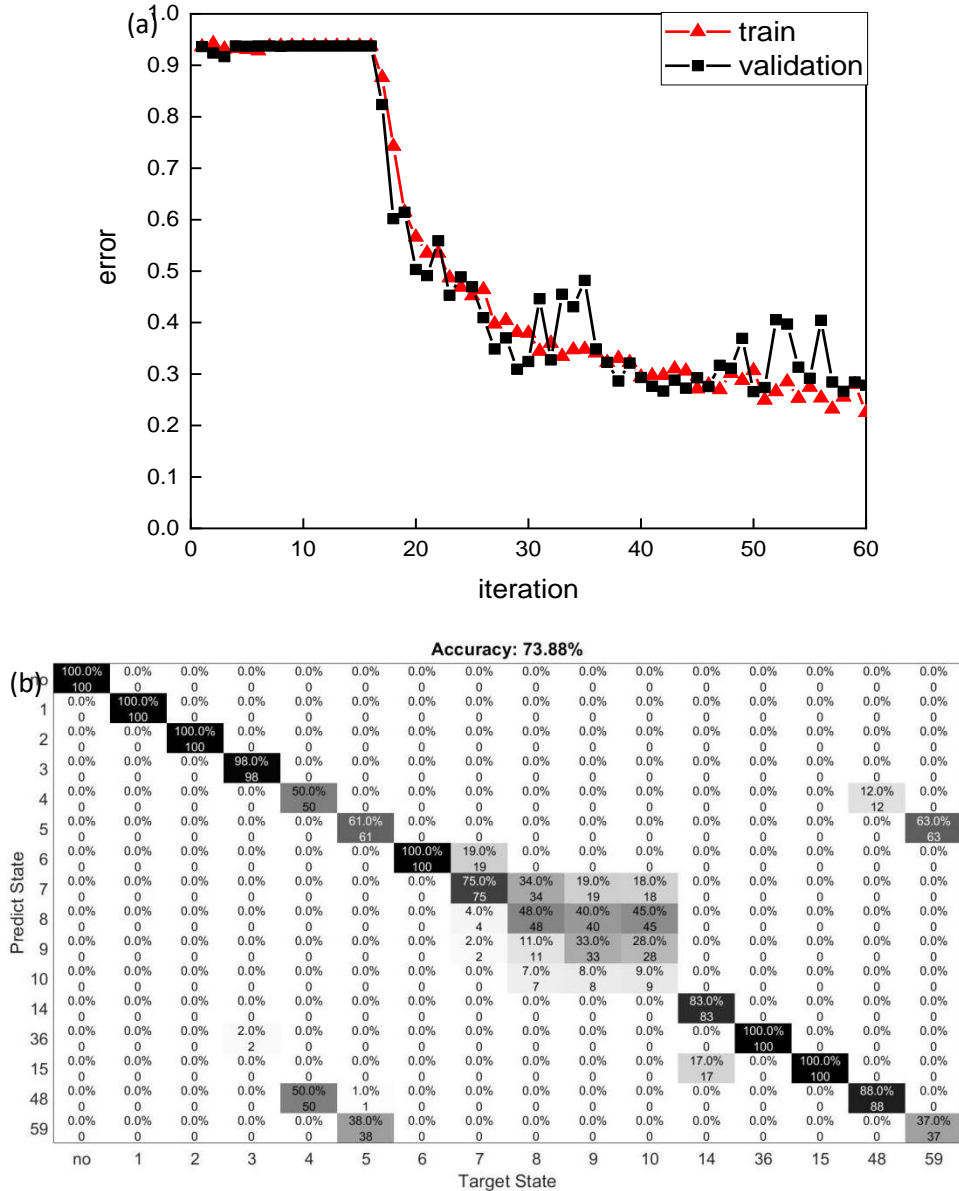
Figure 65. The training results when the noise is 90dB. (a) The training error and validation error, (b) Prediction results.

As shown in Figure 66, the accuracy is 90.44% for the case that the SNR is 100dB. Moreover, the main error also exists at point 8, point 9 and point 10. And the 2-damage case would be affected by the case which has at least one same damage location, such as case 5 and case 59. For this case, the accuracy is enough for using in field.
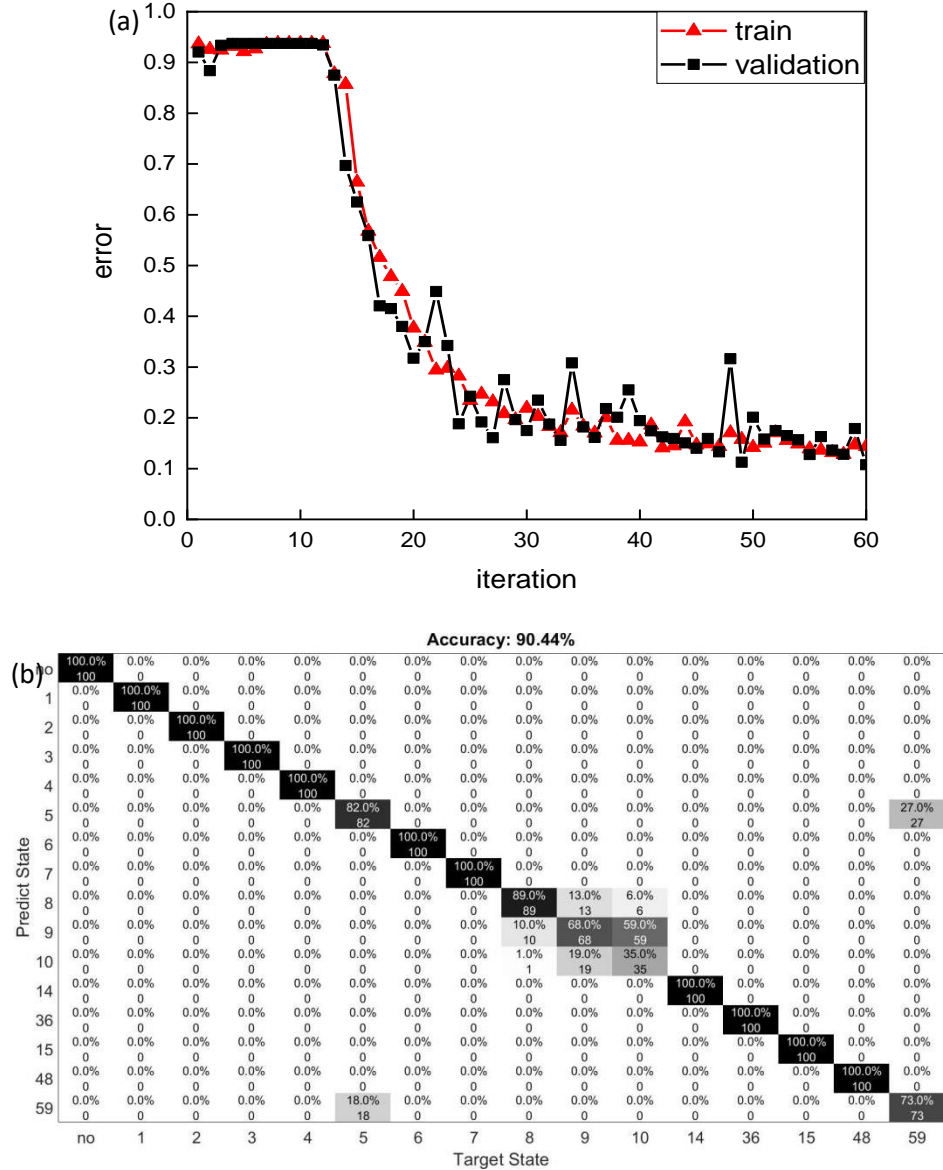
(a) 1.0

error

train
validation

iteration

Accuracy: 90.44%

(b)

Predict State

Target State

Figure 66. The training results when the noise is 100dB. (a) The training error and validation error, (b) Prediction results.

As shown in Figure 67, the accuracy is 93.63% for the case that the SNR is 110dB. In the training process, both training error and validation error are decreasing quickly to 0.1. For the prediction process, the main error is between case 9 and case 10, which are the two elements close to the free end of the beam.
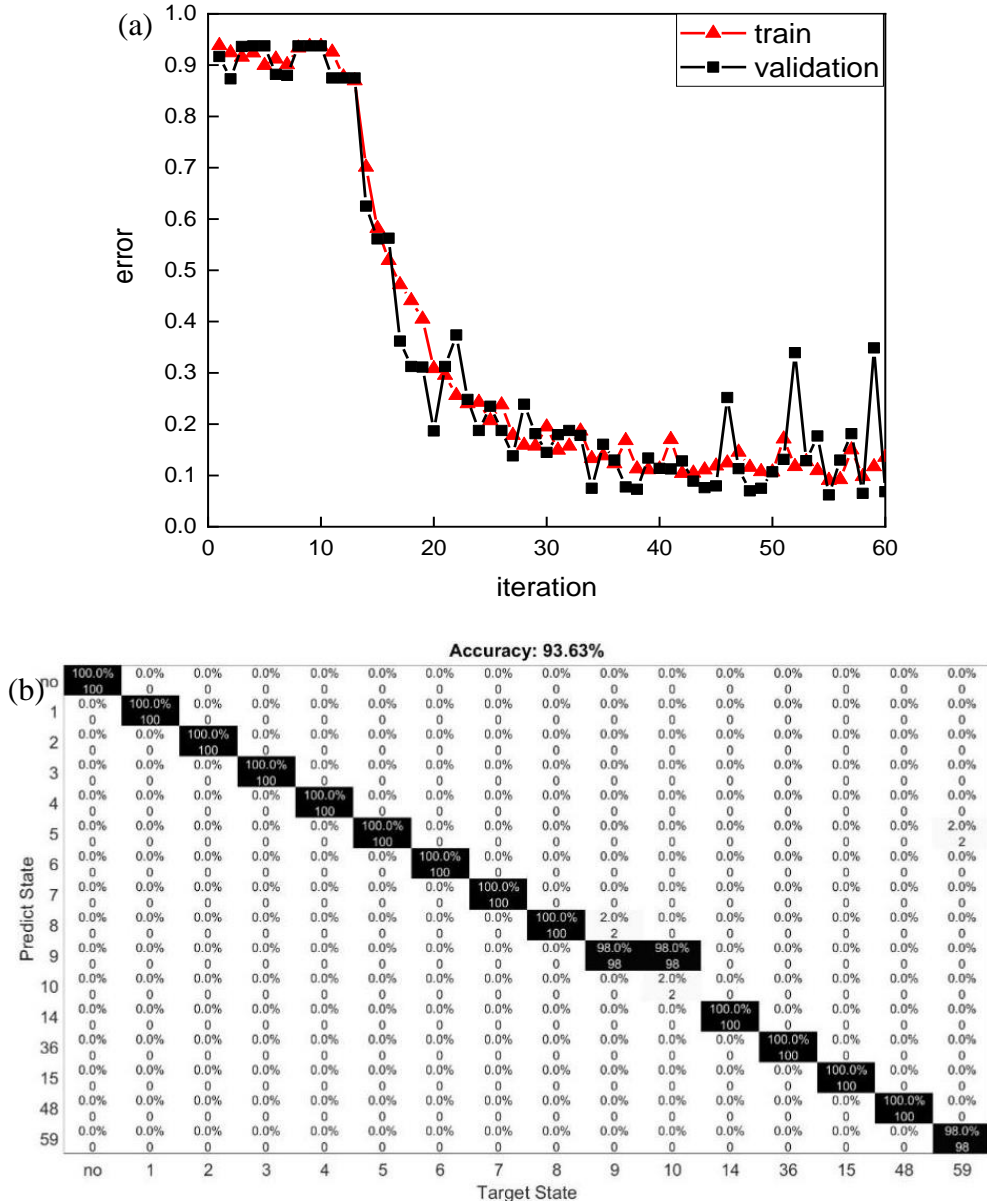
Figure 67. The training results when the noise is 110dB. (a) The training error and validation error, (b) Prediction results.

As shown in Figure 68, the accuracy is 94.13% for the case where the SNR is 120dB. In the training process, both training error and validation error are decreasing quickly to about 0.05. For the prediction process, the main error is between case 9 and case 10, which are the two elements close to the free end of the beam.

Figure 68. The training results when the noise is 120dB. (a) The training error and validation error, (b) Prediction results.

Figure 69(a) shows the training process and both the training error and validation error decease to zero at epoch 60. Figure 69(b) shows prediction results which has a small error between case 9 and case 10 with a satisfied accuracy close to 99%.
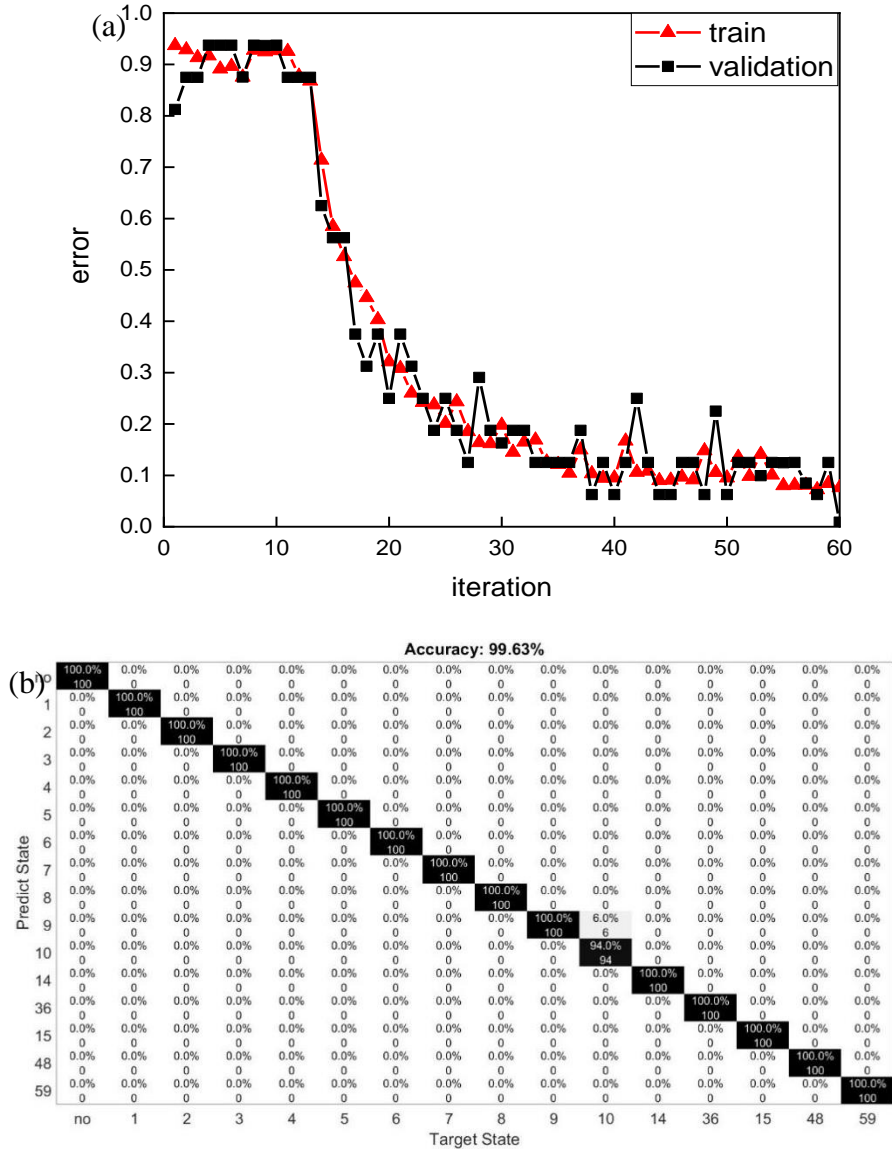
Figure 69. The training results when the noise is 130dB. (a) The training error and validation error, (b) Prediction results.

6.8. Results of Cases on Beam 2

Figure 70 shows the vibration curves after applying noises at different levels. Applying noises could significantly affect vibration curves which could help to enlarge the database.
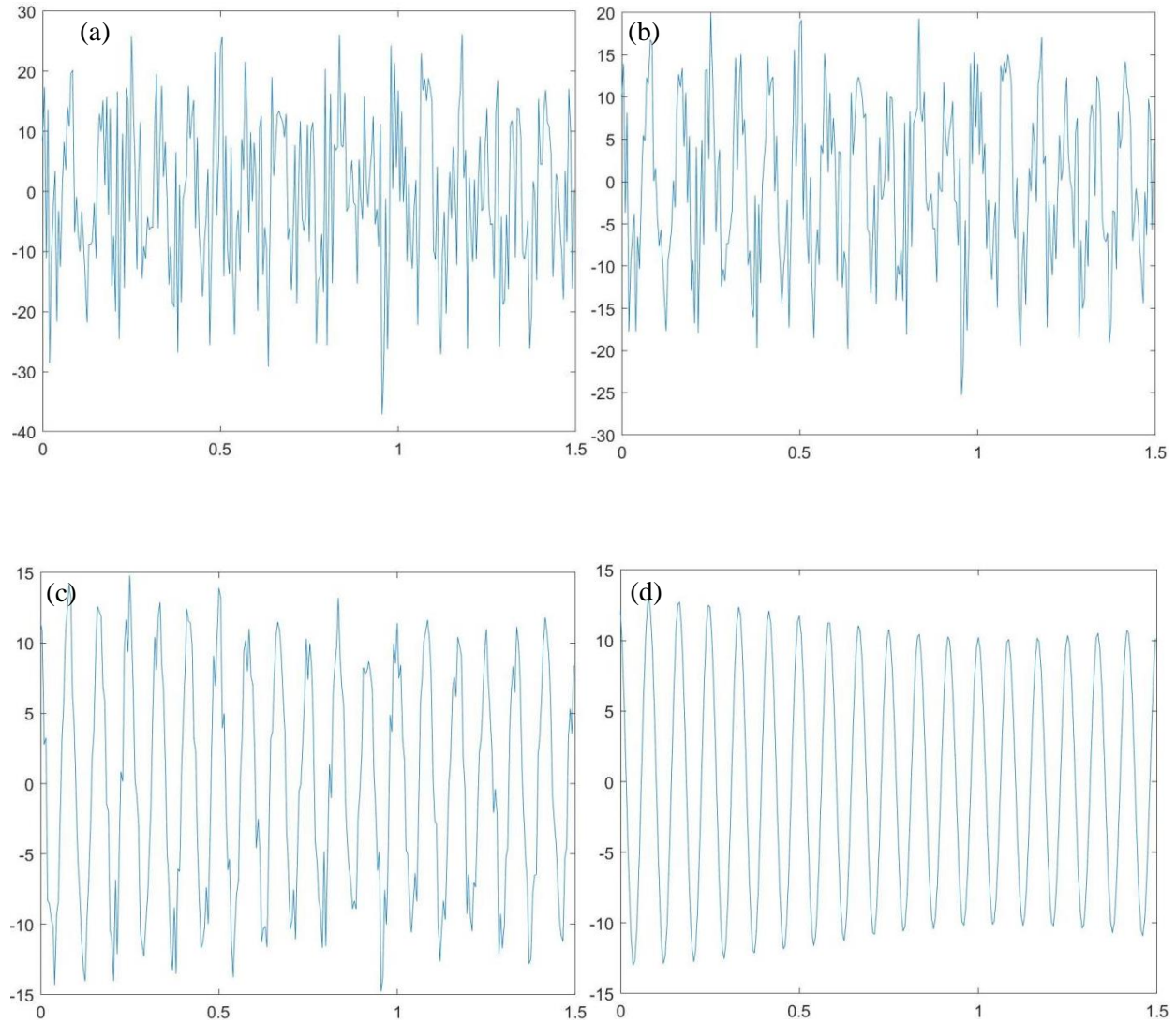
Figure 70. Vibration curves after applying different noises comparing with healthy vibrations. (a) 80dB; (b) 85dB; (c) 100dB; (d) No noises.

Figure 71 shows the training processes of the proposed CNN structure at different noise levels. When the SNR is below 90dB, the accuracies are between 80% to 90%. When the SNR is from 90dB to 120dB, the accuracy is near 100%. For cases with SNR higher than 120dB, the accuracy is 100%. Cases with one damaged location have more errors. However, cases with more than one damaged location have higher accuracy.
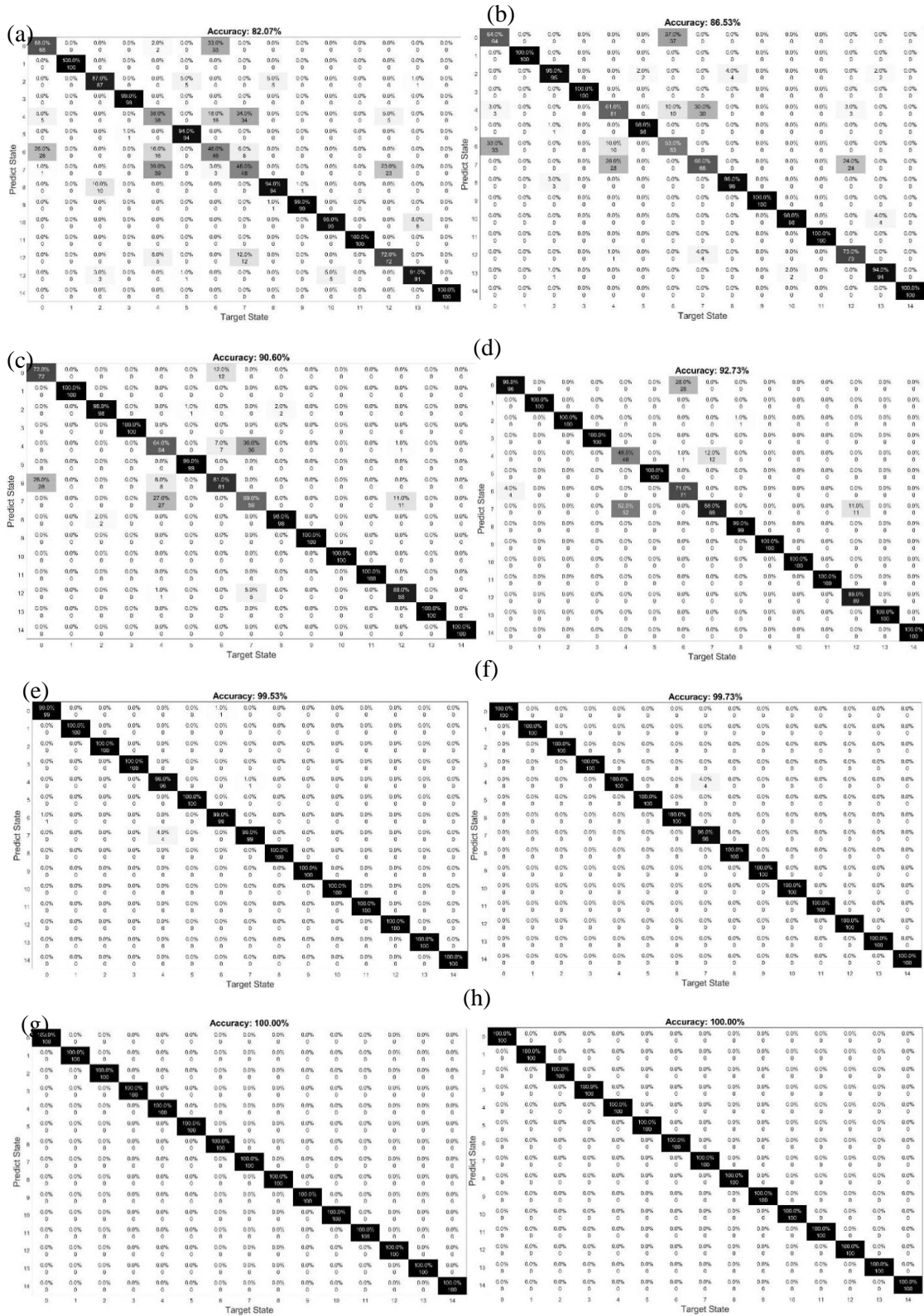
Figure 71. Testing results of the training process with noise at (a) 80dB; (b) 85dB; (c) 90dB; (d) 95dB; (e) 100dB; (f) 110dB; (g) 120dB; (h) 130dB.

Figure 72 shows the training processes with different SNR level. Smaller SNR would need more iterations to achieve stable for the training curves. As the SNR become smaller, the accuracy would be lower and there are bigger gapes between training loss and validation loss.
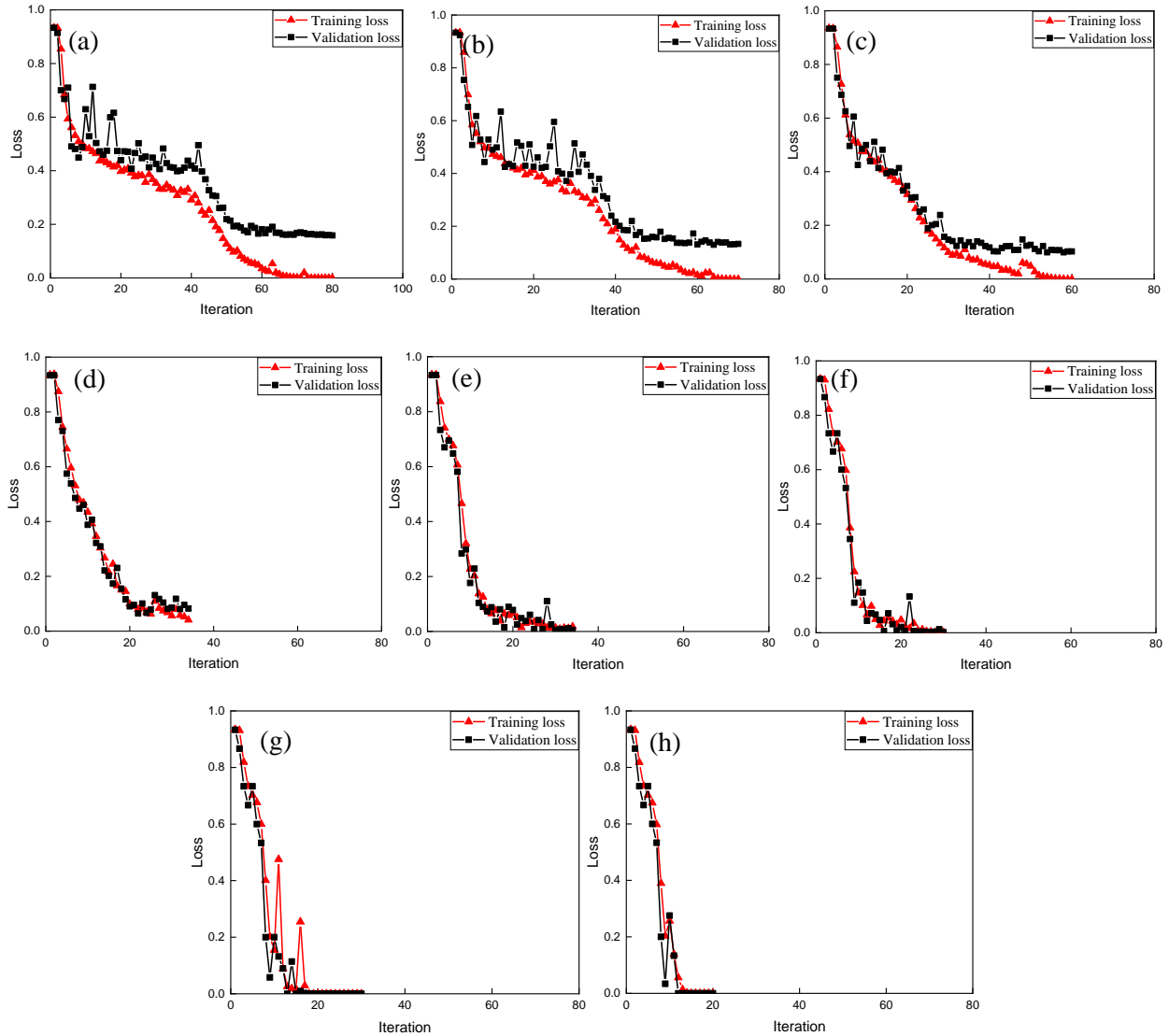

Figure 72. Training processes with noise at (a) 80dB; (b) 85dB; (c) 90dB; (d) 95dB; (e) 100dB; (f) 110dB; (g) 120dB; (h) 130dB.

6.9. Results of the Case of the Truss

Figure 73(a) shows the original vibration data and Figure 73(b) shows the vibration data after applying the noise in the truss study.
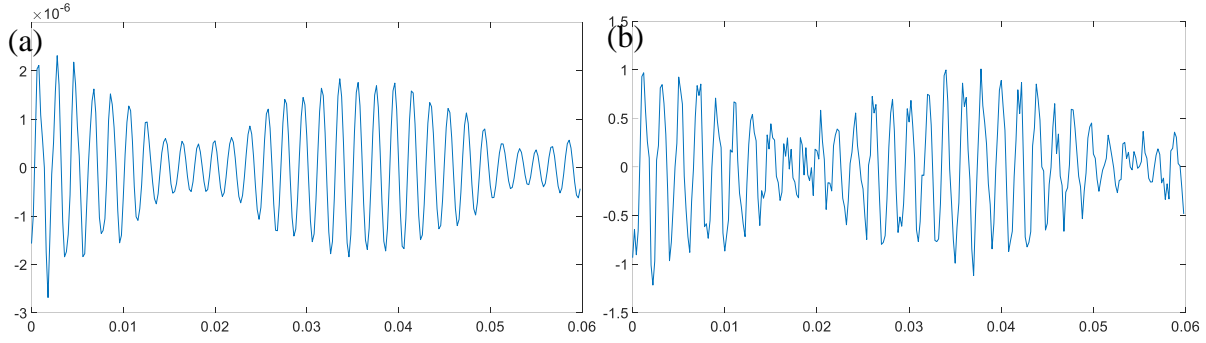
Figure 73. (a) The original vibration data, (b) The vibration data after applying noise.

As shown in Figure 74, the overall accuracy is 74.81% for the case where the SNR is 105dB. Moreover, the main error exists between case 12 and case 15 whose damages are on the vertical elements in the front face. Because those damages cause small change in the stiffness and dynamic characters of the truss and those effect on the vibration is similar.
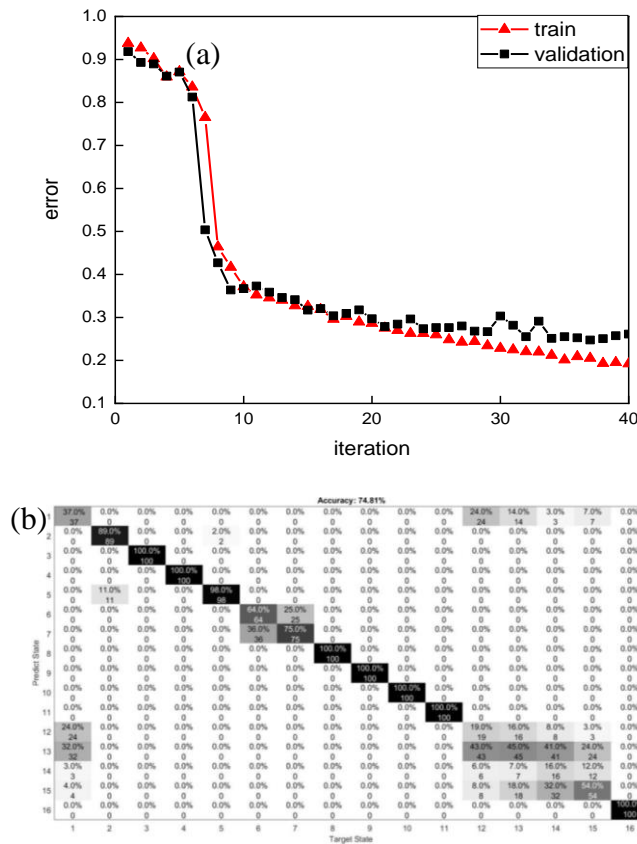

Figure 74. The training results when the noise is 105dB. (a) The training error and validation error, (b) Prediction results.

As shown in Figure 75, the overall accuracy is 84.44% for the case where the SNR is 110dB.

Moreover, the main error also exists between case 12 and case 15 with less error than that at 100dB.
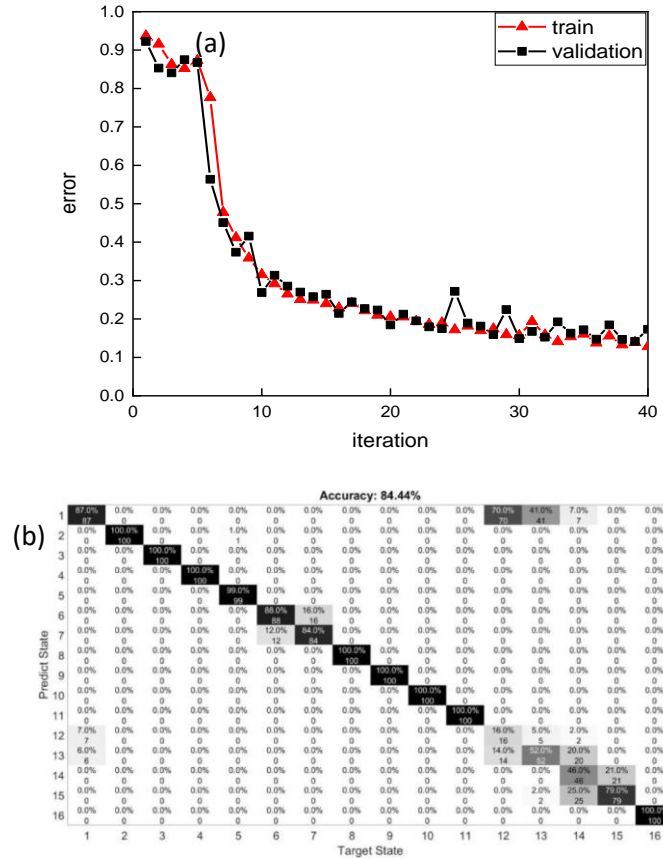

Figure 75. The training results when the noise is 110dB. (a) The training error and validation error, (b) Prediction results.

As shown in Figure 76, the overall accuracy is 93.50% for the case where the SNR is 115dB.

Moreover, the main error exists between case1 and case 12. Because the change of stiffness in case

12 is the smallest and similar to that in case 1 (no damage).

Figure 76. The training results when the noise is 115dB. (a) The training error and validation error, (b) Prediction results.

Figure 77 shows the training process and prediction results when the noise level is 120dB. Moreover, the main error is observed between case 1 and case 12. The prediction accuracy is 96.75%.
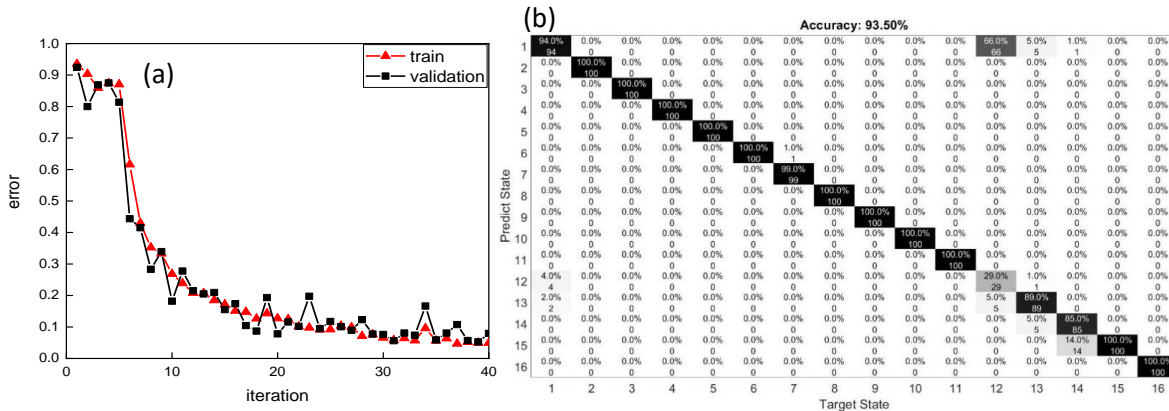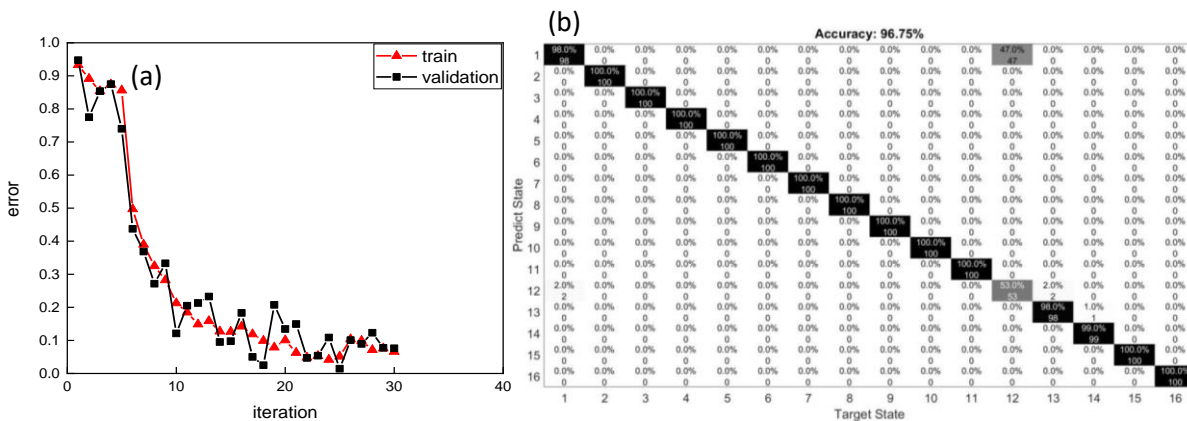

Figure 77. The training results when the noise is 120dB. (a) The training error and validation error, (b) Prediction results.

Figure 78 shows the training process and prediction results when the noise level is 130dB. The error and validation errors are decreasing significantly and close to zero after about 17 epochs. The prediction accuracy is 100%.

Figure 78. The training results when the noise is 130dB. (a) The training error and validation error, (b) Prediction results.

6.10. Results of Cracks Detection Using the Proposed CNN Structure

Figure 79 shows two groups of labeled images that are with-crack images and no-crack images. There are 20000 with-crack images and 20000 no-crack images for training the proposed transfer learning method. Some with-crack images of different widths, different lengths, different directions, and different image brightness are collected for training the model.
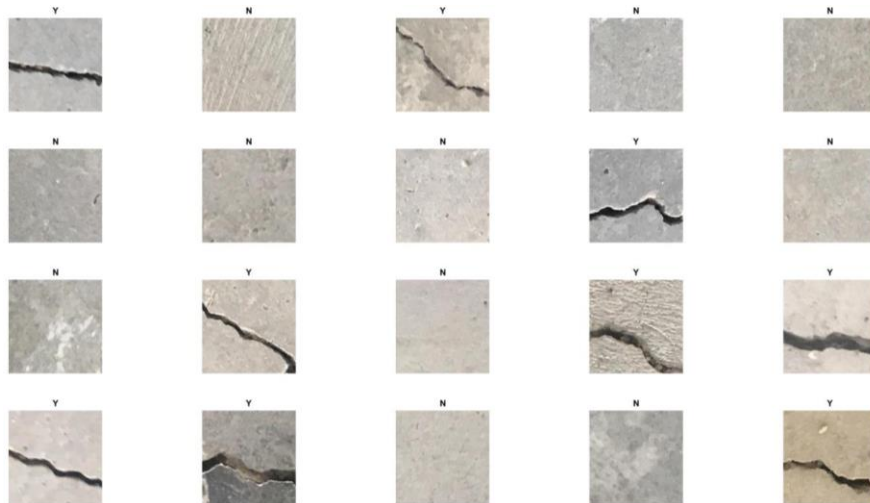
Figure 79. The part of training images.

Figure 80 shows the training process of the proposed method. The method is very fast and effective because after only several epochs, both testing accuracy and validation accuracy are close to 100%. Some sharp changes in the training accuracy are founded whose reason is overfitting. The features of cracking in images are simple and do not need too many training circles to find them. The CNN is prone to overfitting because of the millions or billions of parameters it encloses. A model with these many parameters can be overfitted on the training data because it has sufficient capacity to do so. However, in the next epoch, the testing accuracy returns to nearly 100%.
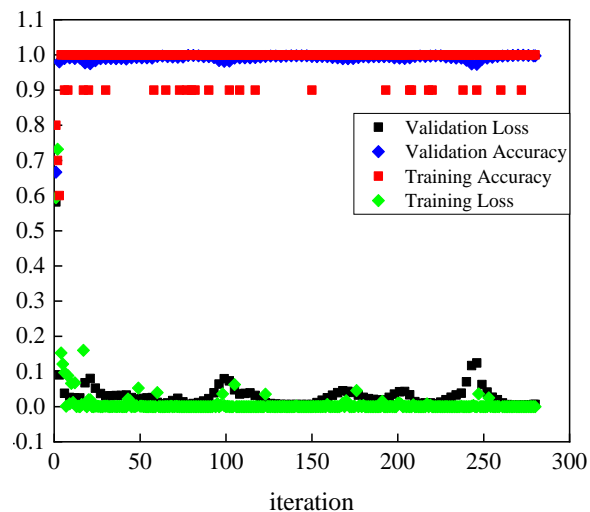

Figure 80. The training process with training accuracy and validation accuracy using AlexNet.

Figure 81 shows the training process of the proposed method. The method is very fast and effective, because after only several epochs, both testing accuracy and validation accuracy are close to 100%.
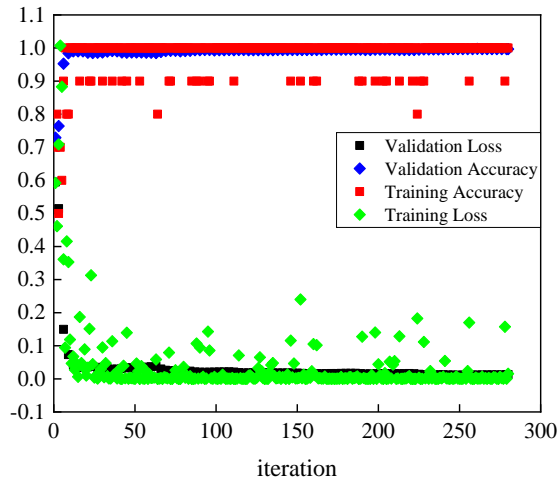

Figure 81. The training process with training accuracy and validation accuracy using GoogLeNet.

Figure 82 shows the training process with training and validation accuracy using the InceptionResNetv2 model. The training accuracy is more stable, but the validation accuracy fluctuates and over 90%. The training loss is very unstable, and validation is better for comparing. In the end, the validation loss is close to 0.3 which is not satisfactory.
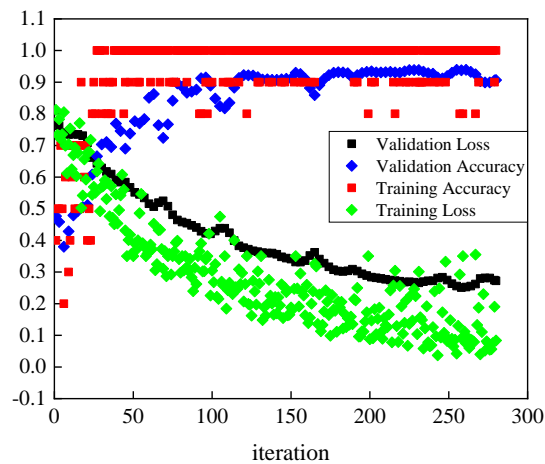

Figure 82. The training process with training accuracy and validation accuracy using InceptionResNetv2.

Figure 83 shows the training process with training and validation accuracy using the ResNet50 model. The training accuracy is stable, but the validation accuracy fluctuates and over 90%. The training loss is very unstable, and validation is better in comparison. In the end, the validation loss is close to 0.2 which is not satisfactory.

Figure 83. The training process with training accuracy and validation accuracy using ResNet50.

Figure 84 shows the training process with training and validation accuracy using the ResNet101 model. The training accuracy is stable, but the validation accuracy fluctuates and over 90%. The training loss is very unstable, but validation is better in comparison. In the end, the validation loss is close to 0.1 which is not satisfactory.
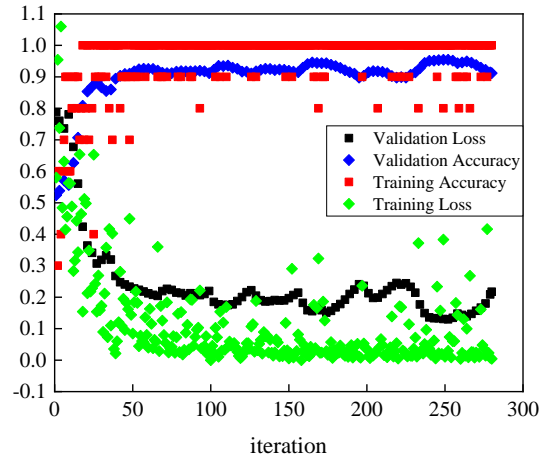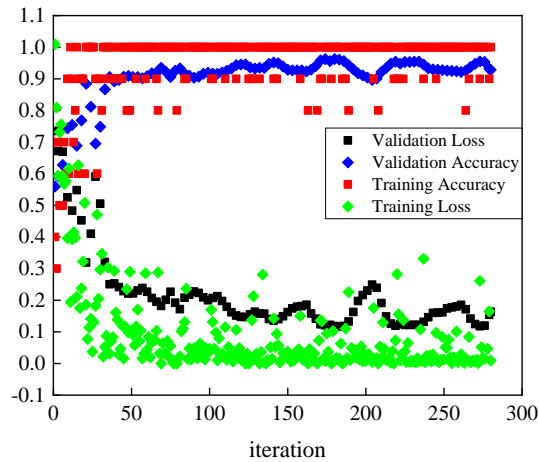

Figure 84. The training process with training accuracy and validation accuracy using ResNet101.

Figure 85 shows the training process with training and validation accuracy using VGG16. The training and validation accuracy reach near 100% rapidly after only several iterations. The training loss and validation loss decrease quickly to zero after 50 iterations. There are some peaks for all these four curves, which are due to the overfitting errors. Moreover, after these peaks, these curves return to be stable quickly.

Figure 85. The training process with training accuracy and validation accuracy using VGG16.

Figure 86 shows the training process with training and validation accuracy using VGG19. The training and validation accuracy reach near 100% rapidly after only several iterations. The training loss and validation loss decrease quickly to zero after 50 iterations. There are some peaks for all these four curves, which are due to the overfitting errors. Moreover, a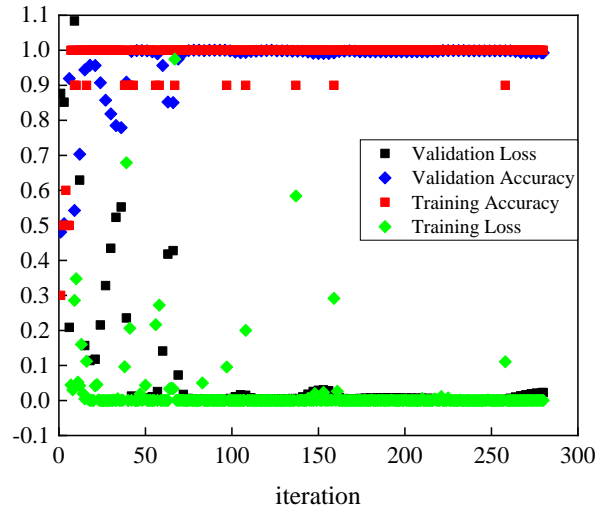fter these peaks, these curves return to be stable quickly. From No.170 iteration to the end, these four curves are very stable.
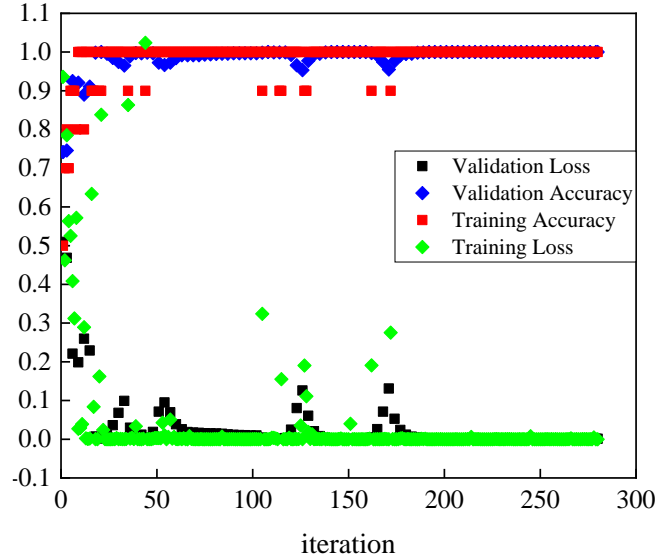
Figure 86. The training process with training accuracy and validation accuracy using VGG19.

Figure 87(a) shows the with-crack image captured in the field and Figure 87(b) shows the results of all proposed methods. Moreover, each segment of the crack in the with-crack image was detected successfully.

Figure 87. The testing results using an image captured in field: (a) The captured image, (b) The output image using the proposed method.

Figure 88 shows the testing results of the proposed method using images in the same dataset with training data. '*N*' represents the label of no-crack images and '*Y*' represents the label of with-crack images. The accuracy of the testing results for these 20 images is 100%.
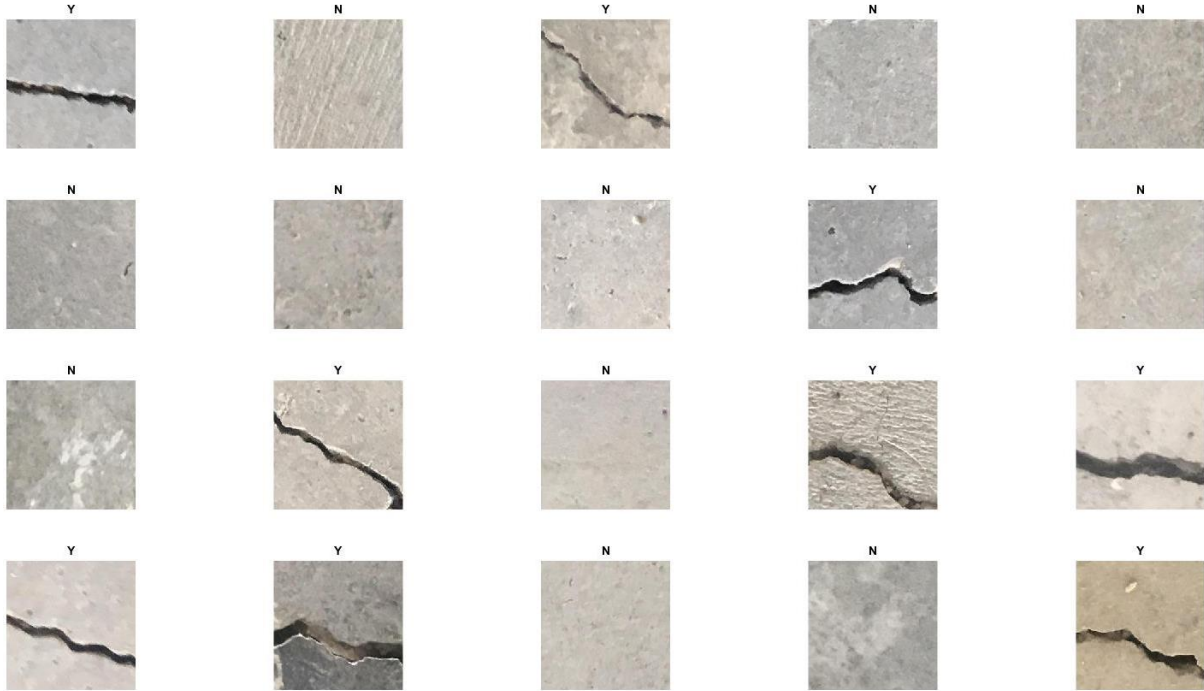
Figure 88. The testing results using images in the same dataset with training data.

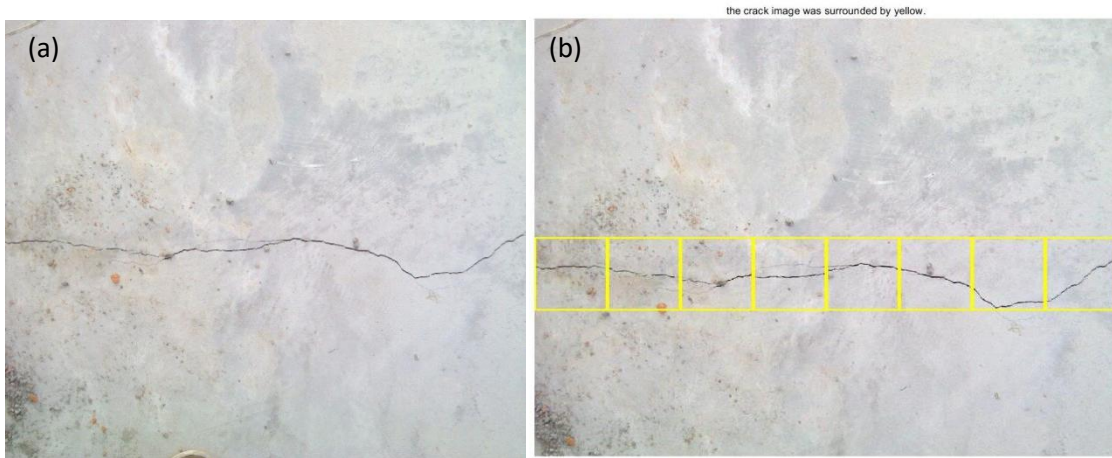Table 8 shows the running time and accuracy for each deep learning method trained by the crack dataset. The AlexNet model has the smallest running time (2m58s) and the InceptionResNetv2 model has the most running time (62m55s). The ResNet101, Vgg16, and Vgg19 models have similar running times which is about 20 minutes. The GoogleNet and ResNet50 model have 5m24s and 14m15s on the running time respectively. The accuracy of the RetNet101 model is less than 99%, and the accuracies of the rest of those methods are more than 99%.

Table 9. Running times and accuracy for the 7 different transfer learning methods.

	AlexNet	GoogLeNet	InceptionResNetv2	ResNet50	ResNet101	Vgg16	Vgg19
Running Time	2m58s	5m24s	62m55s	14m12s	27m15s	20m37s	23m50s
Accuracy	99.33%	99.33%	99.08%	99.00%	95.92%	99.50%	100%

6.11. Structural Evaluation Using a Novel Deep-Learning-Based Method With Both Defect Image and Vibration Data

6.11.1. Process of the Proposed Method Using Images of Vibrations Calculated by GASF and Raw Images of Defects

As shown in Figure 89, images of defects at important positions of a structure are captured by cameras or UAVs. The vibration data of the structure is obtained by the image-based displacement sensor discussed in Chapters 3-5. Then these one-dimension vibrations are converted into two-dimension images. These two groups of images are then combined to provide input data for the suggested deep-learning model.
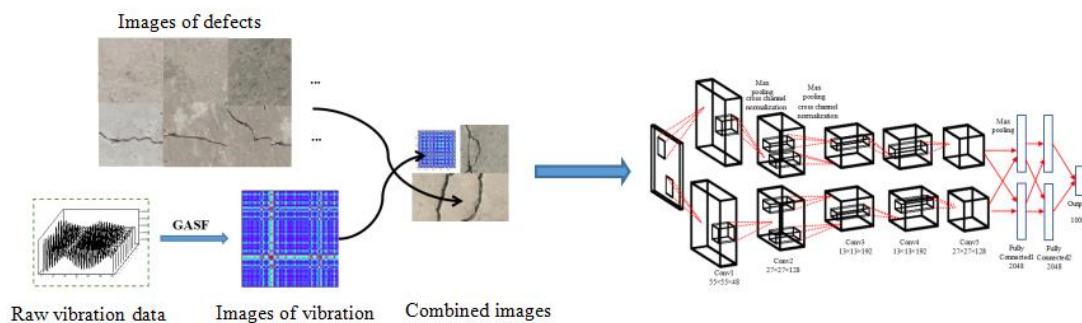
Figure 89. The process of the proposed damage evaluation method using images of vibration and raw images of defects.

6.11.2. Gramian Angular Field

In the field of time series data, the GAF algorithm was created to encode 1D time series into 2D images without losing any features. GAF can be completed through two different algorithms: grammar angular sum field (GASF) and gramian angular difference field (GADF). The vibration data is $X = \{x1, x2, ..., xi, ..., xn\}$, where n is the total number of displacement points and xi is the value of the displacement. The second step is to normalize the spectral data and scale it to the range of [0, 1], and record it as $\tilde{x}i\_n$.

$$x_{i_n} = \frac{x_i - min(X)}{max(X) - min(X)} \tag{39}$$

Next, the 1D spectral sequence in the Cartesian coordinate system is transformed into a polar coordinate system, the formula is as follows:

$$\varphi = arccos\ (x_i, n) \tag{40}$$

The Gram matrix retains a time dependency. Since time increases as the position moves from the upper left corner to the lower right corner, the time dimension is encoded into the geometry of the matrix. It can be seen from Eq. (39) that the value range of the converted angle ϕ is $[0, \pi]$, and the cosine value decreases monotonically within this range. With the increase of wavelength, x_i in each Cartesian coordinate system only corresponds to the angle value in the polar coordinate system, time T corresponds radius in the polar coordinate system, and the corresponding bending occurs between different angle points on the polar coordinate circle. By calculating the cosine value of the sum of angles between different points, we can get the following formula for the calculation of the gramian angular summation field (GASF) as shown in Figure 90. Where I is the unit vector. $\tilde{X}_i$ is the normalized vibration data. φ is the angle of the point.

$$GASF = [cos(\phi_i + \phi_j)]$$

$$= \begin{bmatrix} cos(\phi_1 + \phi_1) & cos(\phi_1 + \phi_2) & \cdots & cos(\phi_1 + \phi_n) \\ cos(\phi_2 + \phi_1) & cos(\phi_2 + \phi_2) & \cdots & cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \vdots & \vdots \\ cos(\phi_n + \phi_1) & cos(\phi_n + \phi_2) & \cdots & cos(\phi_n + \phi_n) \end{bmatrix} = cos\left(arccos\left(\tilde{x}_i \right) + arccos\left(\tilde{x}_j \right) \right)$$

$$= cos\left(arccos\left(\tilde{x}_i \right) \right) \cdot cos\left(arccos\left(\tilde{x}_j \right) \right) - sin\left(arccos\left(\tilde{x}_i \right) \right) \cdot sin\left(arccos\left(\tilde{x}_j \right) \right)$$

$$= \tilde{x}_i \cdot \tilde{x}_j - \sqrt{1 - \tilde{x}_i^2}\sqrt{1 - \tilde{x}_j^2} = \tilde{X}^T\tilde{X} - \sqrt{I - \tilde{X}^2}^T\sqrt{I - \tilde{X}^2}$$

Figure 90. Formula for the calculation of the gramian angular summation field (GASF).

Figure 91 shows an example using vibration data obtained from the ABAQUS model of the steel truss. The vibration is converted into images using the proposed GASF and there are many obvious features in the image, such as the highlighted points and lines in the Figure 90(b).
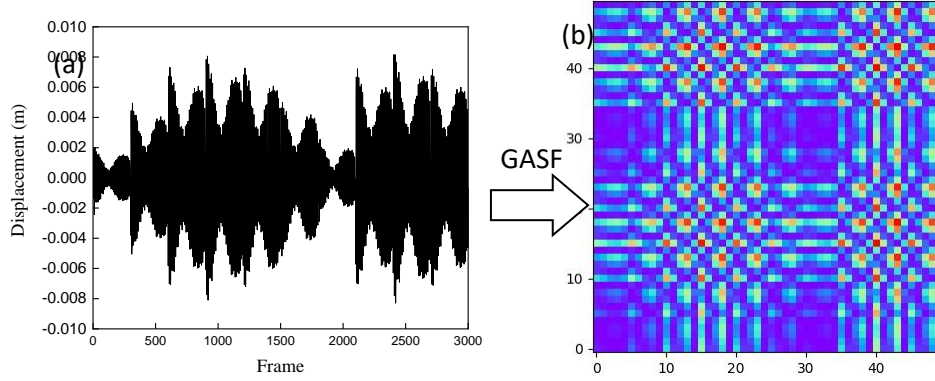
Figure 91. Example of GASF. (a) Raw vibration obtained in ABAQUS model, (b) Images obtained by GASF.

6.12. Gaussian Noise

The Gaussian noise, typically called normal noise, is used to test the robustness of the method. The Gaussian noise could be calculated as:

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\bar{z})^2}{2\sigma^2}} \tag{41}$$

Where z is the intensity, $\bar{z}$ represents the mean value of z, and the σ is the standard deviation of z. σ^2 is the variance.

The Matlab was used in this study to apply the Gaussian noises. $0.01, 0.1$, and 1 were used in codes and represent the mean value in the Gaussian noise function after multiplying 255. So, the mean pixel values in the image of the Gaussian noises were 2.55, 25.5, and 255. The variance was set at 0.1.

6.13. Results of the Proposed Method

To verify this method, vibrations obtained from the ABAQUS models of the steel truss are used to be converted into images using GASF. There are 8 different groups of combined images with different damage locations (no-damage, No.3 damage, and No.5 damage in the ABAQUS model of the steel truss) and different defects (no-crack, small crack, and large crack). There are 9000 images in total and 1000 images for each group.

128

Figure 92 shows combined images after applying different Gaussian noises (0.01, 0.1, 1) when damage is No. 3 and crack is large. Significant effect on images by Gaussian noise could be seen here and they are unable to classify them by humans.

Figure 92. Input images for cases of with no-image, with 0.01 Gaussian noise, with 0.1 Gaussian noise, with 1 Gaussian noise when damage is No. 3.

Figure 93 shows combined images after applying different Gaussian noises (0.01, 0.1, 1) when damage is No. 5 and crack is small. Significant effect on images by Gaussian noise could be seen here and they are unable to classify them by humans.
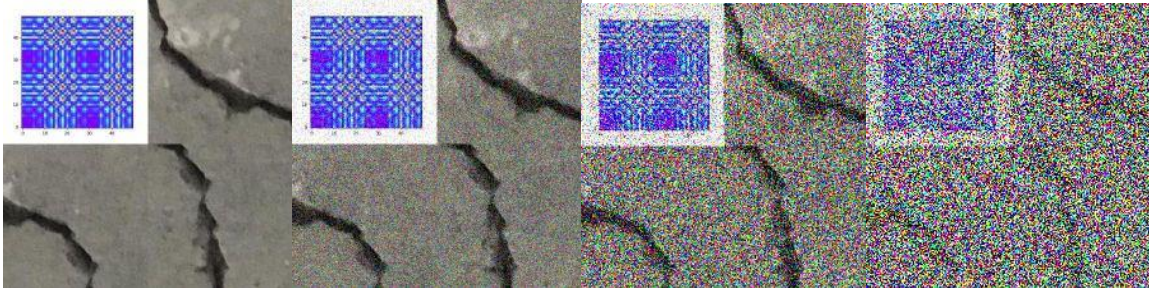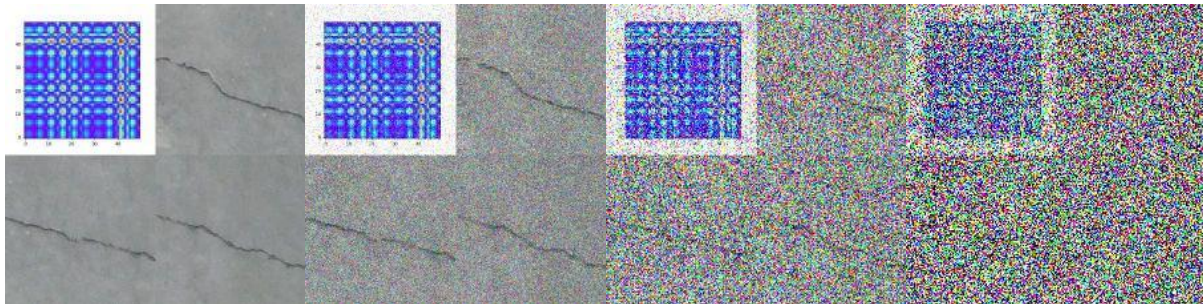

Figure 93. Input images for cases of with no-image, with 0.01 Gaussian noise, with 0.1 Gaussian noise, with 1 Gaussian noise when the damage is No. 5.

Figure 94 shows the training process with training and validation accuracy using AlexNet when there is no noise applied to images. The training accuracy and validation accuracy reach near 100% after 160 iterations. The training loss and validation loss decrease to zero after about 160 iterations. From 160 iteration to the end, these four curves are very stable.

Figure 94. Training results for the case without Gaussian noise.

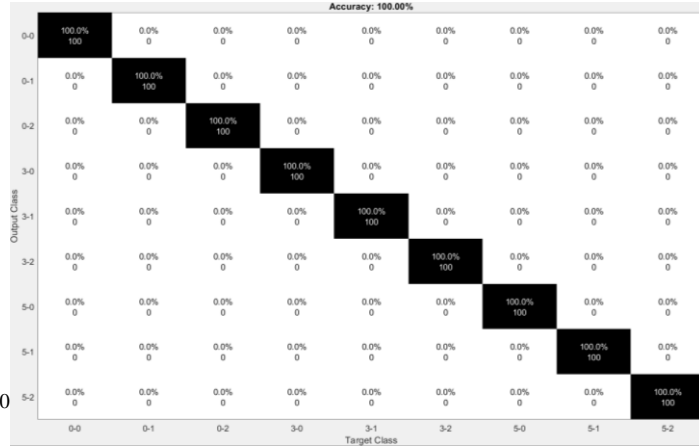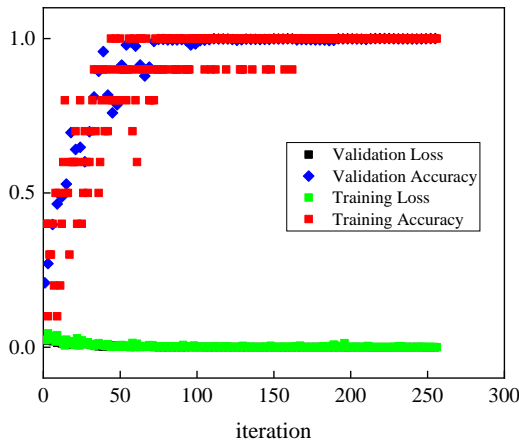Figure 95 shows the training process with training and validation accuracy using AlexNet when 0.01 Gaussian noises are applied to images. The training accuracy and validation accuracy reach near 100% after 180 iterations and the training loss and validation loss decrease to zero after about 180 iterations. From the 180th iteration to the end, these four curves are very stable which are similar to the results from the case of no noise.

Figure 95. Training results for the case with applied Gaussian noise at 0.01.

Figure 96 shows the training process with training and validation accuracy using AlexNet when 0.1Gaussian noises are applied to images. The training accuracy and validation accuracy reach near 100% after 200 iterations and the training loss and validation loss decrease to zero after

about 200 iterations. From the 200th iteration to the end, these four curves are stable but at some

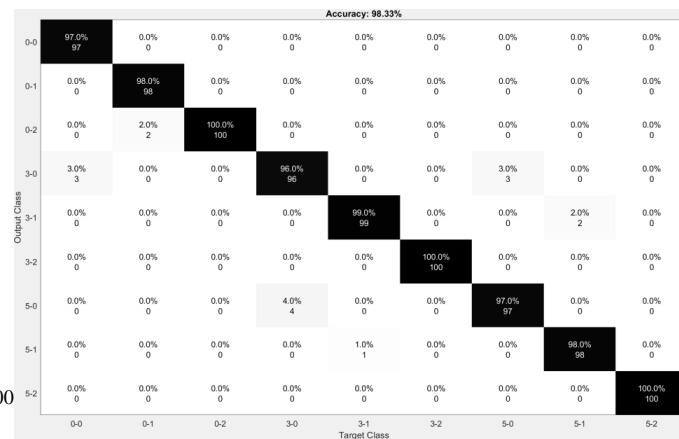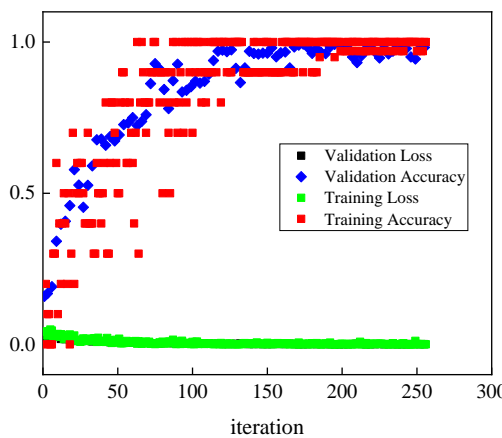positions, there are some small peaks due to over-fitting.
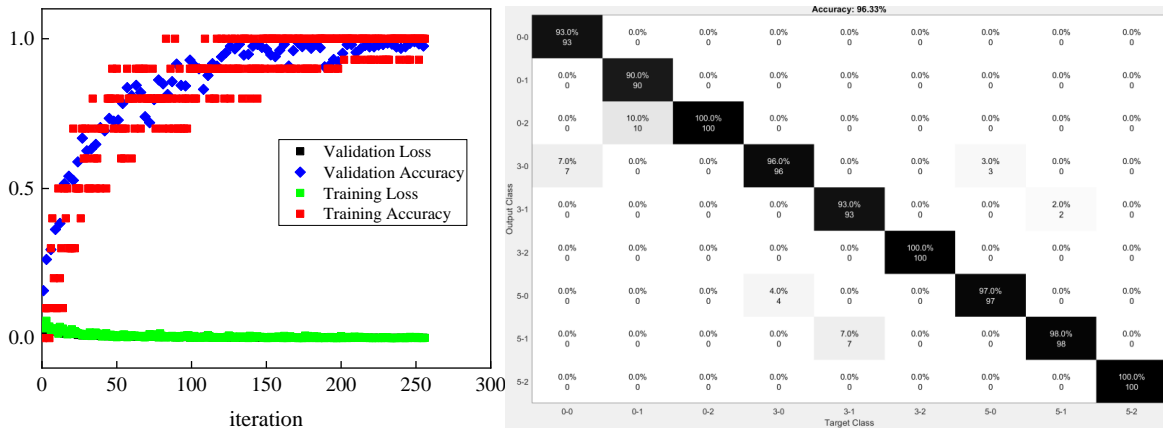
Figure 96. Training results for the case with applied Gaussian noise at 0.1.

Figure 97 shows the training process with training and validation accuracy using AlexNet

when 1 Gaussian noise is applied to images. Two accuracy curves increase slowly and are waved

during the whole training process because the Gaussian noise is too big.
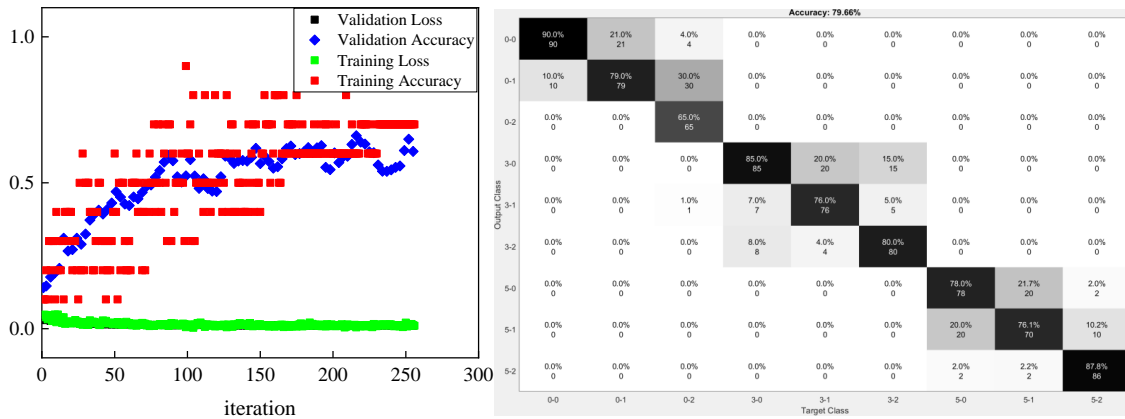

Figure 97. Training results for the case with applied Gaussian noise at 1.

Table 9 shows accuracies at different cases with different Gaussian noises applied to

images. When the noise is from 0.01 to 0.1, the accuracy only has a small decrease from 98.33%

to 96.33%, which is satisfactory to classify these combined images. When the noise is 1, the

accuracy decreases to 79.66%.

131

Table 10. The accuracy at different cases with different Gaussian noise applied to images.

Case	No noise	0.01 noise	0.1 noise	1 noise
Accuracy	100%	98.33%	96.33%	79.66%

The transfer learning method always has a large set of parameters, which would cause over-fitting. Also, overfitting occurs when a model tries to predict a trend in data that is too noisy. There are some ways to prevent the over-fitting, including simplifying the model, early stopping, using data augmentation, regularization, and dropouts.

6.14. Process of the Proposed Method Using Images of Vibration Calculated by the GASF and Images of Defect Maps

As shown in Figure 98, images at the damage location of the structure are sent to the transfer learning algorithm to detect different defects with different severities. The defect map would be generated based on results from the transfer learning model. Different colors represent different types of defects and different color depths represent how serious the defects are. For example, the white color represents no damage and the yellow color represents crack. At the same time, vibration data obtained by the proposed displacement sensor is converted to RGB images and combined with defect maps. These combined images are sent to the CNN structure to train the CNN model and obtain the final classification and evaluation of the structure comprehensively.
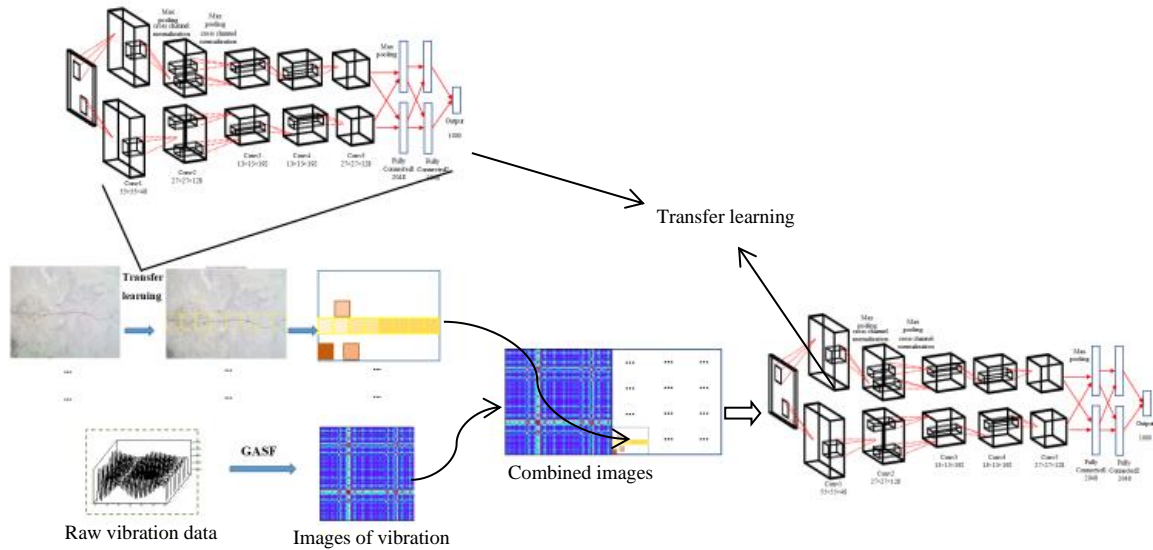
Figure 98. The process of the proposed damage evaluation method using images of vibration and images of defect maps.

6.15. Conclusions

● Modeling of a steel truss and a beam using ABAQUS were conducted. Free vibration after release of a concentrated force was conducted. The model provides vibration data of any point on the structure successfully.

● The proposed CNN-based method was trained by raw vibration data obtained in the ABAQUS models and had very high accuracy. This method classified different damages in the truss and beam case accurately and respectively.

● The seven transfer learning methods used in this chapter had high accuracy, in which the ResNet101 model had the lowest accuracy at 95%. The Alex transfer learning method is more effective because it has higher accuracy and less running time compared to the other six transfer learning methods when they are trained by local crack photos and the converted vibration mages.

● Images of vibrations obtained by GASF are able to train the CNN algorithm and identify damages accurately.

● Higher Gaussian noises would generate lower accuracy, but the proposed damage detection method is accurate to classify cases with different damages and different defects under high noise levels.

● Theoretically, combined images using a defect map and a vibration map are more sensitive and would generate more accurate results. Because different colors and their depth could be selected manually to make them easier to be classified.

6.16. Reference

1. Ebrahimkhanlou, A., Farhidzadeh, A., and Salamone, S. (2016) Multifractal analysis of crack patterns in reinforced concrete shear walls. Struct. Health Monit. 15, 81–92.

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

3. Rawat, W., and Wang, Z. (2017) Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput. 29, 2352–2449.

4. Chen, F., and Jahanshahi, M.R. (2018) Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion. IEEE Trans. Ind. Electron. 65, 4392–4400.

5. Dorafshan, S., Maguire, M., Hoffer, N.V., and Coopmans, C. (2017) Fatigue Crack Detection Using Unmanned Aerial Systems in Under-Bridge Inspection. Ida. Transp. Dep. 2, 1–120.

6. Li, G., Ma, B., He, S., Ren, X., and Liu, Q. (2020) Automatic Tunnel Crack Detection Based on U-Net and a Convolutional Neural Network with Alternately Updated Clique. Sensors. 20, 717.

7.   Soloviev, A., Sobol, B., and Vasiliev, P. (2019) Identification of Defects in Pavement Images Using Deep Convolutional Neural Networks. Adv. Mater. 4, 615–626.

8.   Li, B., Wang, K.C., Zhang, A., Yang, E., and Wang, G. (2020) Automatic classification of pavement crack using deep convolutional neural network. Int. J. Pavement Eng. 21, 457–463.

9.   Tong, Z., Gao, J., Han, Z., and Wang, Z. (2018) Recognition of asphalt pavement crack length using deep convolutional neural networks. Road Mater. Pavement Des. 19, 1334–1349.

10.   Fan, Z., Li, C., Chen, Y., Mascio, P.D., Chen, X., Zhu, G., and Loprencipe, G. (2020) Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement. Coatings 10, 152.

11.   Adeli, H., and Jiang, X. (2008). Intelligent infrastructure: Neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures Boca Raton, FL: CRC Press. pp. 305–332.

12.   Cawley, P., and Adams, R. D. (1979). The location of defects in structures from measurements of natural frequencies. J Strain Anal Eng Des, 14(2), 49–57.

13.   Pandey, A. K., Biswas, M., and Samman, M. M. (1991). Damage detection from changes in curvature mode shapes. J SOUND VIB, 145(2), 321–332.

14.   Chang, K. C., and Kim, C. W. (2016). Modal-parameter identification and vibration-based damage detection of a damaged steel truss bridge. Eng. Struct, 122, 156–173.

15.   Reynders, E., Wursten, G., and De Roeck, G. (2014). Output-only fault detection in structural engineering based on kernel PCA. BIL2014 workshop on data-driven modeling methods and applications, Leuven, Belgium, 14–15 July.

16.   Yan, A. M., Kerschen, G., De Boe, P., and Golinval, J. (2005). Structural damage diagnosis under varying environmental conditions. Part I: A linear analysis. Mech Syst Signal Process 19(4), 847–864.

17.   Lin, Y. Z., Nie, Z. H., and Ma, H. W. (2017). Structural damage detection with automatic feature extraction through deep learning. COMPUT-AIDED CIV INF, 32(12), 1025–1046.

18.   Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., and Inman, D. J. (2018). Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks. J SOUND VIB, 424, 158–172.

7. CONCLUSIONS AND FUTURE WORK

7.1. Main Contributions and Concluding Remarks

- The computer-vision based displacement sensor was developed using the edge detection and the Zernike moment method, which does not need any manmade targets or artificial targets on structures. This method is faster than the template matching-based displacement sensor and has very high resolution by adopting the Zernike moment method for subpixel resolution precision.

- The UAV-based displacement monitoring system gets rid of self-vibration of UAV itself and environment-induced vibration at all six directions (3 translations and 3 rotations) successfully. The breaking improvement for the industry is that this system does not need any manmade targets on the background of the structure.

- The skipped-NCC template matching method was developed, which could improve the computational efficiency of the displacement monitoring method significantly without affecting its accuracy. The 3-D polynomial fitting method was used to achieve subpixel precision using the NCC results.

- Two ABAQUS models on a cantilever beam and a truss were conducted and a free vibration simulation was modeled. The vibration data was collected and sent to the proposed CNN model to classify different damages.

- A novel structural damage evaluation method based on the transfer learning and CNN was developed, which could analyze both vibration data and defect images. By adopting this method, the damage location would be detected first using CNN with the vibration data as the input. The UAV will be then used to scan the area of the damage on the structure. These images are further used as the input data of the proposed CNN to detect defects. Another novel structural damage

137

evaluation method based on the transfer learning and CNN was also developed, which could use combined images of defects' photos and vibration data.

7.2. Future Research Directions

- Resolution of structural defects that could be detected using the deep learning methods depends on the differences between different groups of defects in labeling process. The number of the groups of defects would determine the resolution of the defect's detection in deep learning. Collect more images with different damages and different severities. The system based on the AlexNet transfer learning is completed but the damage detection is only completed for the crack's detection. In the future, images with more damage types need to be collected such as corrosion, losing bolts, and distortion of joints. When this is achieved, the developed system could be used for wider applications in structural condition assessment.

- Conduct field inspection of a real bridge using the proposed damage detection system in a real-time setting, including vibration-based damage detection, photo-based defect detection, and infused vibration-image-based defect detection. High speed data process and transmission through 5G and Raspberry Pi would help to achieve data transfer in real-time fast and accurately.