Hindawi
Shock and Vibration
Volume 2019, Article ID 9064830, 16 pages
https://doi.org/10.1155/2019/9064830

Research Article

Artificial Neural Network for Vibration Frequency Measurement Using Kinect V2

Jiantao Liu [1] **and Xiaoxiang Yang** [1,2]

[1] *School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China*
[2] *Quanzhou Normal University, Quanzhou, Fujian 362000, China*

Correspondence should be addressed to Xiaoxiang Yang; yangxx@fzu.edu.cn

Received 29 August 2018; Accepted 23 February 2019; Published 12 March 2019

Academic Editor: Huu-Tai Thai

Optical measurement can substantially reduce the required amount of labor and simplify the measurement process. Furthermore, the optical measurement method can provide full-field measurement results of the target object without affecting the physical properties of the measurement target, such as stiffness, mass, or damping. The advent of consumer grade depth cameras, such as the Microsoft Kinect, Intel RealSence, and ASUS Xtion, has attracted significant research attention owing to their availability and robustness in sampling depth information. This paper presents an effective method employing the Kinect sensor V2 and an artificial neural network for vibration frequency measurement. Experiments were conducted to verify the performance of the proposed method. The proposed method can provide good frequency prediction within acceptable accuracy compared to an industrial vibrometer, with the advantages of contactless process and easy pipeline implementation.

1. Introduction

Vibration measurement and analysis are important tools for monitoring and characterizing the physical property and fault diagnosis of structures and machinery. Measurement and analysis results, such as vibration frequency, are important for the predictive maintenance of civil or mechanical structures.

Traditional sensors such as accelerometers, gyroscopes, strain gauges, inclinometers, and global positioning systems (GPS) have been widely used in vibration measurement. However, many conventional vibration measurement methods are both labor intensive and expensive owing to complex wiring for power supply and signal transmission, as well as installation and deployment of sensors. In addition, since these types of sensors are physically attached to the target object, the physical properties of the object, such as stiffness, mass, or damping, may be altered, especially when the target object is relatively small compared to the sensor. Alternative noncontact measurement techniques, such as laser Doppler vibrometer (LDV) [1–4], and optical methods including optical flow [5, 6], marker tracking [7–10], digital image correlation (DIC) [11–13], and stereovision [11] are also used in practice. High cost of equipment and high requirement of the target surface limit the use of LDV. On the contrary, the use of optical methods in vibration frequency measurement has yielded promising results in laboratory and field experiments, providing data in temporal and spatial domains. However, optical methods require the use of a complicated image and signal analysis algorithm [14, 15] to obtain the vibration frequency, and lighting conditions are also critical in the measurement [8, 16].

The development of the depth sensor has unlocked new opportunities for researchers to utilize depth information to provide a device the capability to observe and detect real-world targets beyond human recognition; for instance, high-accuracy object recognition and tracking [17], SLAM application [18–20], high security level face recognition [21, 22], augmented reality [23], human postural recognition, and distant medic [24–26]. In recent years, the use of low-cost consumer level depth sensing input devices such as Intel RealSense and Microsoft Kinect have received significant research attention thereby extending the range of application of depth sensors. Vibration measurement is one of such applications [27, 28], and depth sensors are destined to play increasing important roles in the future kinematic measurement system.

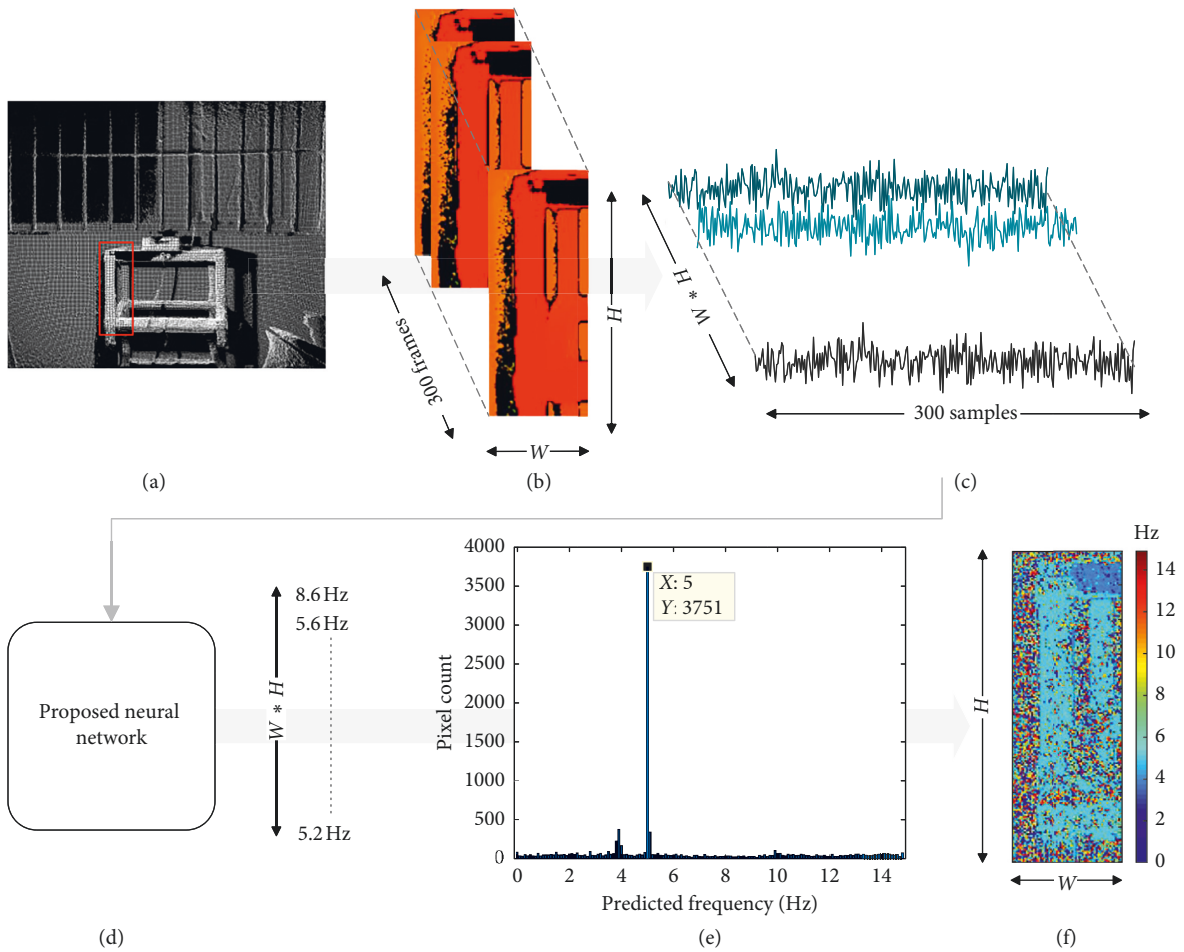


Figure 1: Pipeline illustration of proposed method. (a) Select ROI from the first frame of depth image. (b) Read in the depth image sequence from every frame. (c) Read out every spatial position depth value along time dimension separately. (d) Feed in proposed neural network and obtain predicted frequencies. (e) Interpret results as the statistical histogram of the predicted frequency distribution. (f) Predicted frequency reconstructed to its original spatial position for a better visualization result.

In this work, we proposed a method that utilizes depth information acquired from Microsoft Kinect v2 combined with an artificial neural network which is a further development of the method [29] to predict the vibration frequency of the target. This approach has the following advantages: (a) contactless and markerless; (b) the use of preprocessing such as denoising or other signal enhancement methods are not required; (c) it does not require performing extraction first for vibration signals, that is, it predicts vibration frequency directly, thereby ensuring simple pipeline and easy deployment in real application; and (d) models are trained using pure artificial data, making this method scalable. Experiments were conducted to evaluate the validity and accuracy of the proposed method. The results obtained from the proposed method were compared with those from a conventional contact-type industrial vibrometer.

2. Materials and Methods

This section describes the pipeline implementation of the proposed method. The neural network devised for vibration frequency prediction using depth information acquired from Kinect V2, as well as its corresponding training procedure and preparation of dataset for the network, is introduced. The experimental method is presented at the end of the section.

2.1. Pipeline. First, we read out the metadata recorded by using Kinect v2 and decoded the data as depth information. The ROI for the measurement was selected from the first frame of the reconstructed depth image, as shown in Figure 1(a). Next, we read in the depth image sequence from every frame within the ROI (Figure 1(b)) and extracted the depth value of every pixel within the ROI separately along the time dimension to obtain the $W \times H$ numbers of depth information instances as shown in Figure 1(c), where W and H are the width and height of the ROI (in pixels), respectively. Then, depth information of every pixel was directly fed into the trained network to determine the predicted vibration frequency of each pixel, as shown in Figure 1(d). A histogram of the predicted frequency distribution is plotted in Figure 1(e) to quantitatively evaluate the predicted vibration frequency result, which we can be employed to obtain the overall prediction of the vibration

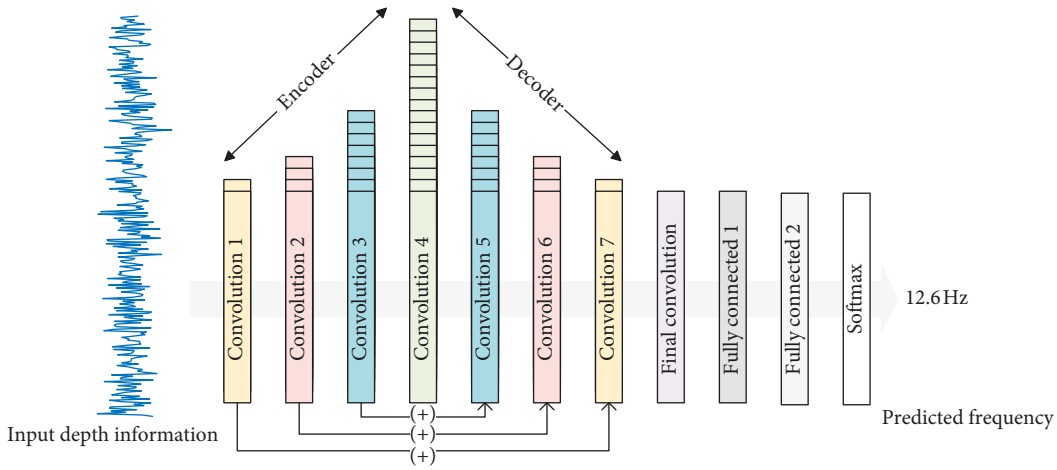


Figure 2: Proposed artificial neural network architecture.

frequency for the target ROI. In addition, we reconstructed the prediction result of each pixel to its original spatial location corresponding to the ROI as shown in Figure 1(f) for better visualization result and verification.

2.2. Network Architecture. In this work, we proposed an artificial neural network designed for vibration frequency prediction utilizing depth information. The input of the proposed network was a 1-dimensional vector with a length of 300, which is the product of the sampling rate of the Kinect depth sensor (30 Hz) and input data duration (10 s). The network was constructed using eight 1-dimensional convolution layers, and every convolution layer was followed by a batch-normalization (BN) [30] layer and rectified linear unit (ReLU) [31]. We mapped features to higher dimension as encoding and compressed them back to their original dimension as decoding. As shown in Figure 2, the number of filters in each convolution layer increased from 2 to 16 and decreased from 16 to 2 with a step of 2 such that the number of features was kept symmetrical for all convolution layers. The kernel size was 11 for all convolutional layers, and we zero-padded both sizes of the input by 5 points before every convolution layer so that the length was kept as 300. Fully connected layers were employed to capture activations from different parts of the input to 150 outputs, which is the number of possible frequency classes. The dropout [32] layer was followed by the first fully connected layer to prevent overfitting, and ReLU was also used as the activation function in all fully connected layers. The output of the fully connected layers was passed through the softmax layer, yielding probabilities associated with each possible frequency class. Skip connections [33, 34] were used in the proposed network architecture to speed up the optimization in the training stage and improve network performance.

2.3. Dataset. The proposed network was trained using a large amount of simulation depth signals to imitate real-world depth variation signals under a specific vibration frequency. An algorithm was used as the simulation signal generator, and the generator consists of two parts as shown in Figure 3: (1) generating standard sinewave with a specific frequency and (2) adding Gaussian distributed noise with varying standard deviation. The frequency of the simulation signals was in the range of 0–15 Hz with a step of 0.1, while the standard deviation of added noise was randomly selected in the range of 0.5–2.2. For each frequency step, we generated 12,000 simulation signals with a length of 10 s at a sampling rate of 30 Hz; randomly selected examples of generated simulation signals with frequencies of 3.6 Hz, 6.6 Hz, 9.6 Hz, and 13.6 Hz are plotted in Figure 4. We generated a dataset with a total of 1,800,000 instances, which was used for the training.

2.4. Training. We trained the network from scratch using the negative log likelihood loss and Adam optimizer [35], where the parameters β_1 and β_2 were set to 0.9 and 0.999, respectively. The input of the network was the generated simulation signals, while the ground truth was the corresponding vibration frequencies. The order of the generated simulation signals was shuffled before feeding into the network. The learning rate was set to a fixed value 1×10^{-5}. The training was implemented on a laptop with Nvidia GTX 1060 GPU with deep learning framework PyTorch [36], which usually yields a good model within 24 h.

2.5. Experimental Method. Experiments were conducted to evaluate the performance of the proposed method, namely, verification test, steel cantilever beam measurement, and simply supported carbon plate measurement. These three experiments were conducted under a controlled laboratory condition. The measurement targets were recorded using Kinect v2. In all experiments except the verification test, an industrial vibrometer DongHua DH5906 was used as reference for comparison. The sampling rate of the depth sensor was fixed at 30 Hz, and 10 s of metadata were recorded in each experiment. The distance between

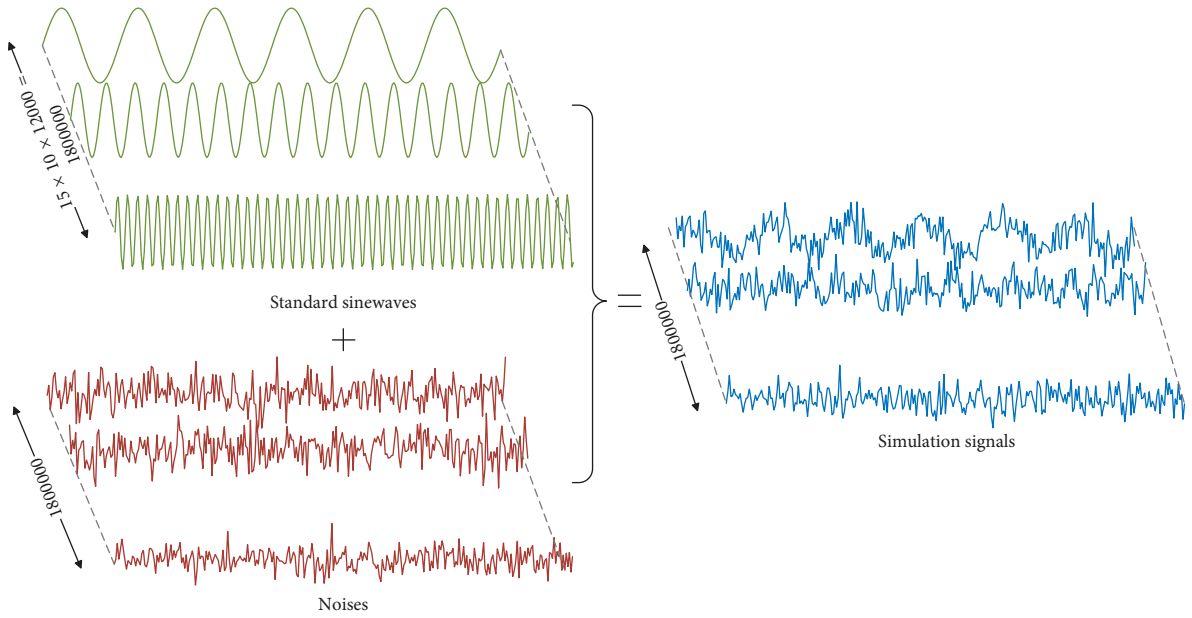


Figure 3: Simulation signals generator configuration.

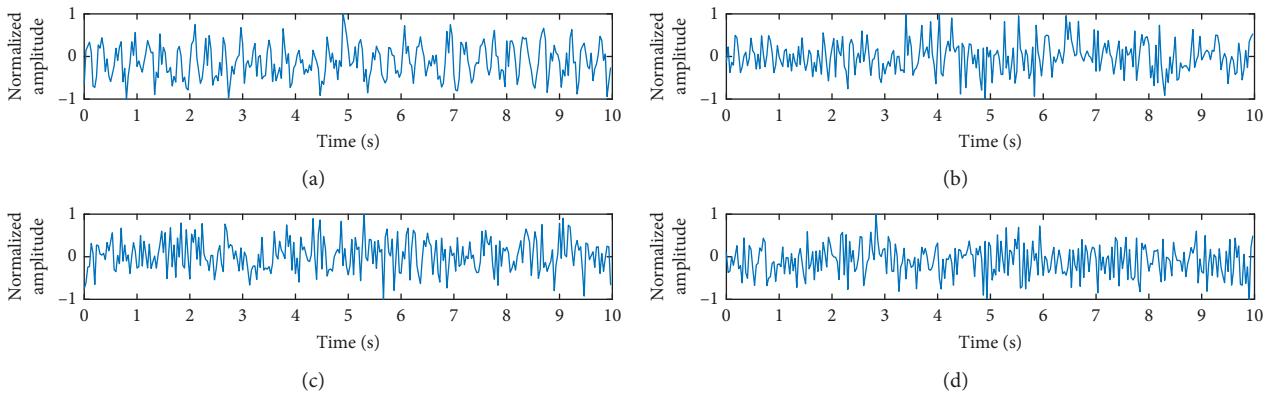


Figure 4: Examples of simulation signals with different vibration frequencies: (a) 3.6 Hz; (b) 6.6 Hz; (c) 9.6 Hz; (d) 12.6 Hz.

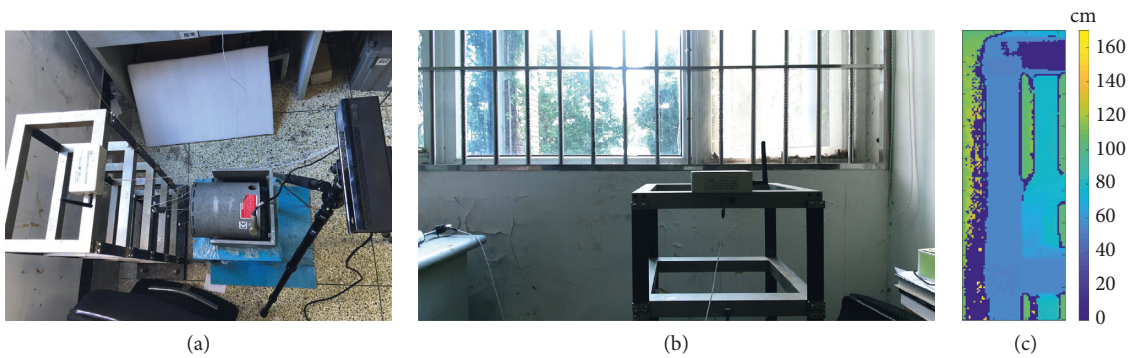
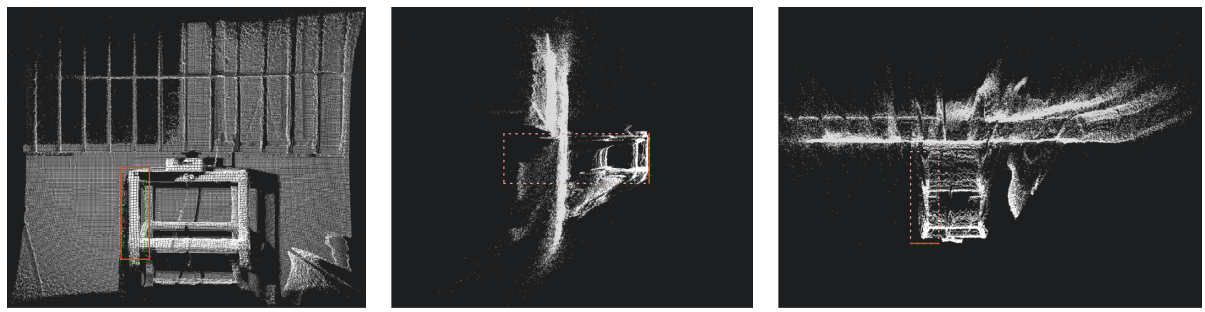


Figure 5: Verification test experimental setup. (a) Modular steel structure, Dynamics MODAL 50 exciter. (b) Field of view (FOV) of Kinect in the RGB image. (c) ROI raw depth information visualization.

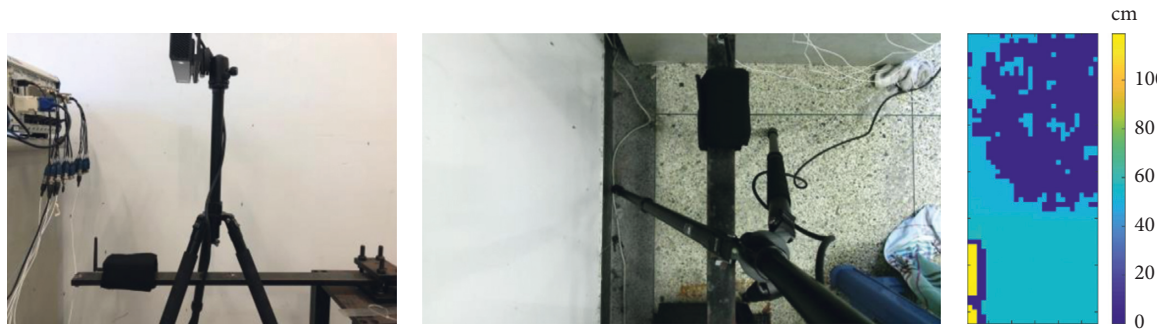
measurement target and Kinect v2 in all experiments is about 50 cm. In addition, the proposed method results will be compared with FFT peak picking results from raw depth information of Kinect v2 in Section 3.4.

2.5.1. Verification Test. To verify the performance of the proposed method, a verification test was conducted using controlled excitation in a vibration test system comprising an exciter (MB Dynamics MODAL 50), an arbitrary



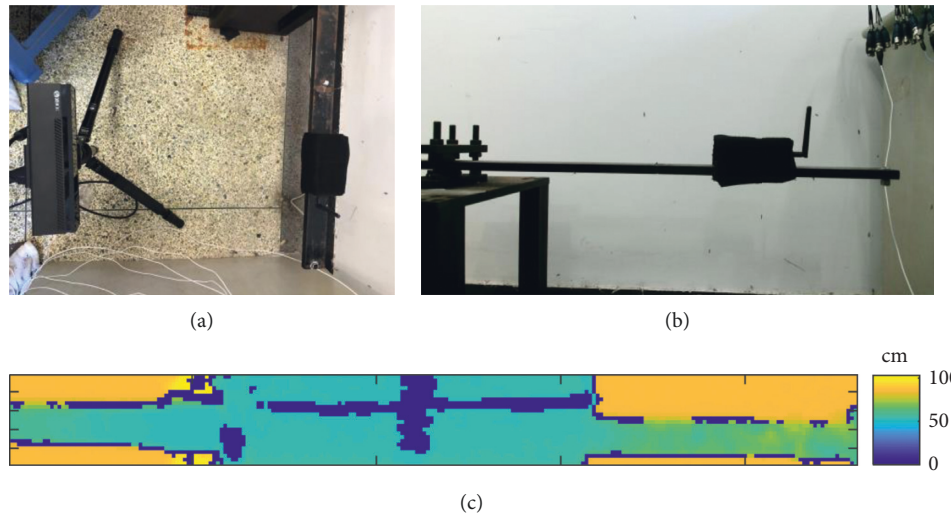
(a) (b) (c)

Figure 6: ROI for vibration measurement in the verification test, highlighted in red. 3D point cloud: (a) front view, (b) left view, and (c) top view.



(a) (b) (c)

Figure 7: Cantilever steel beam experimental setup for case 1. (a) Vibrometer attached at the end of the cantilever steel beam, Kinect pointing in a direction parallel to the direction of vibration. (b) Field of view (FOV) of Kinect in the RGB image. (c) ROI raw depth information visualization.

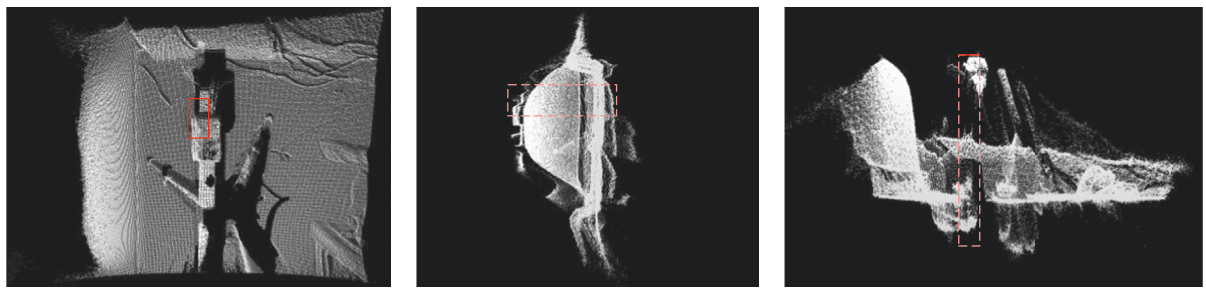


(a) (b)

(c)

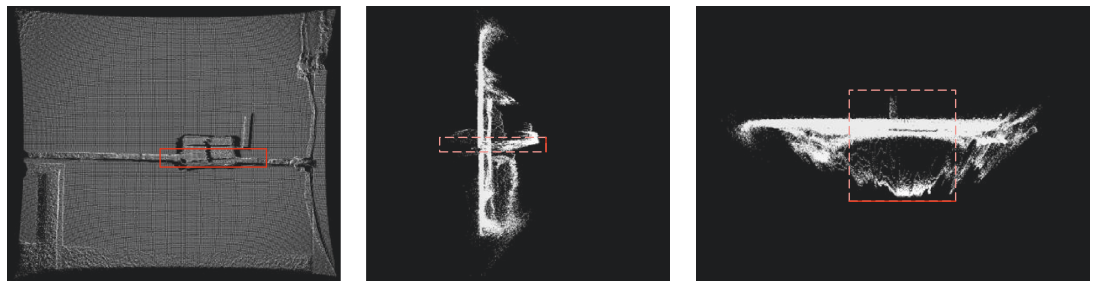
Figure 8: Cantilever steel beam experimental setup for case 2. (a) Vibrometer attached at the end of the cantilever steel beam, Kinect pointing in a direction perpendicular to the direction of vibration. (b) Field of view (FOV) of Kinect in the RGB image. (c) ROI raw depth information visualization.

waveform generator (RIGOL DG1022), and an amplifier (MB Dynamics SL500VCF). A modular steel structure was used as the measurement target, which was excited using precisely controlled vibration signals at different frequencies. The experimental setup is shown in Figure 5(a), and the Kinect field of view (FOV) of the RGB sensor and ROI raw depth information visualization are shown in Figures 5(b) and 5(c), respectively. Sine signals were generated every 2.5 Hz between 0 and 15 and passed through the amplifier to the exciter with minimum gain to excite the modular steel



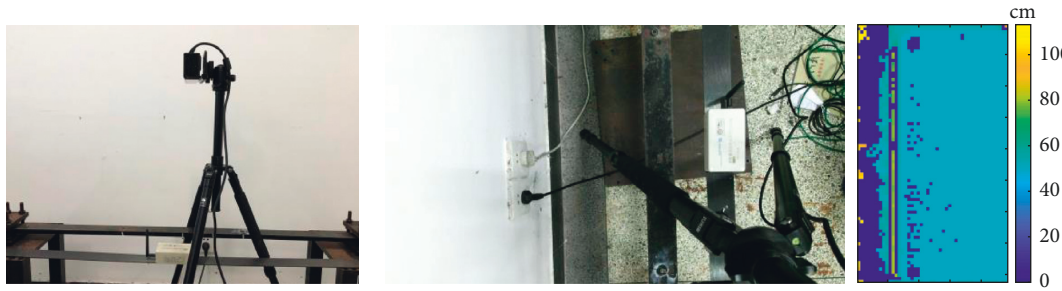
(a) (b) (c)

Figure 9: ROI (highlighted in red) for vibration measurement in the cantilever steel beam experimental case 1. 3D point cloud: (a) front view, (b) left view, and (c) top view.



(a) (b) (c)

Figure 10: ROI (highlighted in red) for vibration measurement in the cantilever steel beam experimental case 2. 3D point cloud: (a) front view, (b) left view, and (c) top view.



(a) (b) (c)

Figure 11: Simply supported carbon plate experimental setup for case 1. (a) Vibrometer placed at the midpoint of the carbon plate, Kinect pointing in a direction parallel to the direction of vibration. (b) Field of view (FOV) of Kinect in the RGB image. (c) ROI raw depth information visualization.

structure, and data were simultaneously recorded using Kinect. A part of the left column of the steel structure was selected as the ROI; for better illustration, three views of the ROI are shown in Figure 6. Then, the proposed method was applied to the Kinect recorded data and compared with result obtained with controlled excitation frequency.

2.5.2. Cantilever Steel Beam. A cantilever steel beam experiment was conducted to investigate the performance of the proposed method employed in vibration excited using real excitation. A steel beam was fixed at one end, while a vibrometer was attached at the free end as shown in Figures 7(a) and 8(a), which was struck at the free end using an impact hammer. The resultant vibration of the steel beam was recorded using the vibrometer and Kinect simultaneously. The sampling rate of the vibrometer was set to 30 Hz, which is consistent with the Kinect depth sensor. Two experimental cases were designed for the cantilever steel beam experiment to examine possibility of utilizing distance variation information between the depth sensor and test object and depth variation signal at the edge of test object. In case 1, the Kinect sensor was pointed in a direction parallel to the direction of vibration, while in case 2, the Kinect sensor was pointed in a direction perpendicular to the direction of vibration. In both experimental cases, the proposed method was applied to the data recorded by the Kinect, and the acceleration signals from

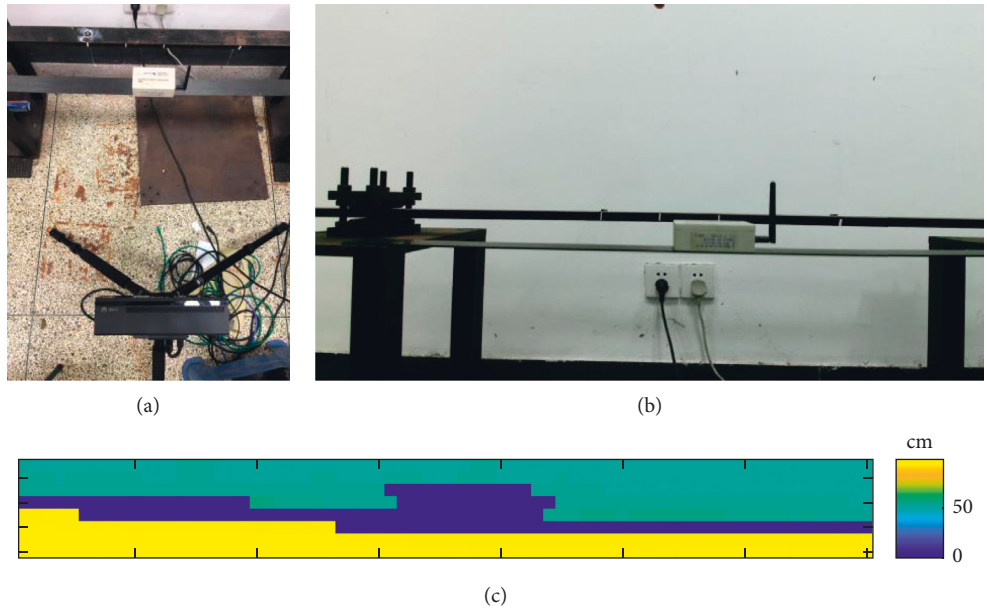


Figure 12: Simply supported carbon plate experimental setup for case 2. (a) Vibrometer placed at the midpoint of the carbon plate, Kinect pointing in a direction perpendicular to the direction of vibration. (b) Field of view (FOV) of Kinect in the RGB image. (c) ROI raw depth information visualization.

the vibrometer were transformed to the frequency domain via fast Fourier transformation (FFT). The vibration frequency components were examined in the frequency domain and then compared with the result from the proposed method.

Experimental Case 1. The experimental setup for case 1 is shown in Figure 7(a). The FOVs of the RGB sensor and ROI raw depth information visualization are shown in Figures 7(b) and 7(c), respectively. Three views of the ROI in this experimental case are shown in Figure 9.

Experimental Case 2. The experimental setup for case 2 is shown in Figure 8(a). The FOV of the RGB sensor and ROI raw depth information visualization are shown in Figures 8(b) and 8(c), respectively. Three views of the ROI for measurement in this experimental case are shown in Figure 10.

2.5.3. Simply Supported Carbon Plate. A simply supported carbon plate experiment was also conducted for the real excitation scenario. The carbon plate was supported at both ends, while the vibrometer was placed at the midpoint of the plate, as shown in Figures 11(a) and 12(a). We struck at the left support point using an impact hammer as excitation. The resulting vibrations were recorded using the vibrometer and Kinect simultaneously. The sampling rate of the vibrometer was set to 30 Hz. Two experimental cases were also designed for the simply supported carbon plate experiment to examine possibility of utilizing distance variation information between the depth sensor and test object and depth variation signals at the edge of test object. In case 1, the Kinect sensor was pointed in a direction parallel to the direction of vibration, while in case 2, the Kinect sensor was pointed in a direction perpendicular to the direction of vibration. The acceleration signals from the vibrometer were transformed to the frequency domain using FFT and were compared with the results from the proposed method.

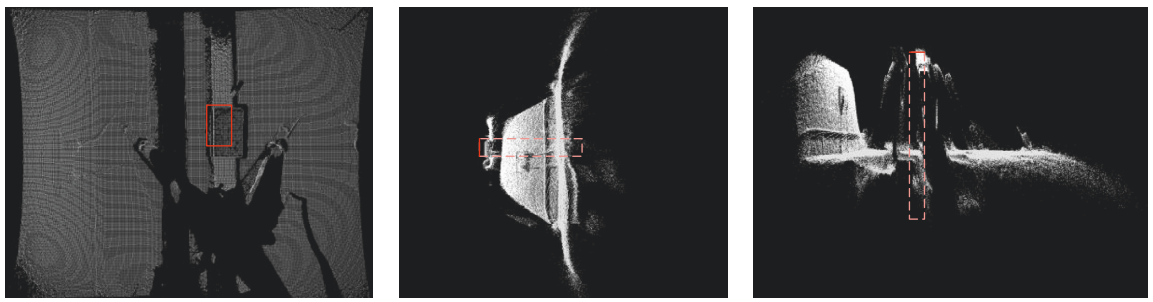
Experimental Case 1. The experimental setup for case 1 is shown in Figure 11(a). The FOV of the RGB sensor and ROI raw depth information visualization are shown in Figures 11(b) and 11(c), respectively. Three views of the ROI for measurement in this experimental case are shown in Figure 13.

Experimental Case 2. The experimental setup for case 2 is shown in Figure 12(a). The FOV of the RGB sensor and ROI raw depth information visualization are shown in Figures 12(b) and 12(c), respectively. Three views of the ROI for measurement in this experimental case are shown in Figure 14.

3. Results

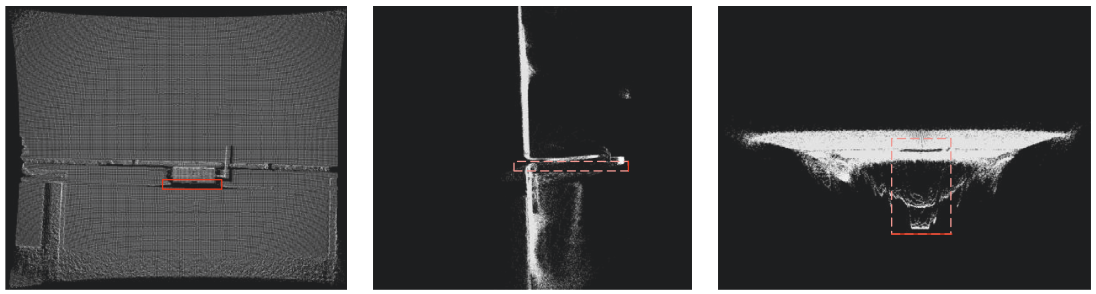
3.1. Verification Test. We selected an excitation frequency of 5 Hz as an example; the histogram of predicted frequency distribution is plotted in Figure 15(a), the visualized predicted frequency distribution over the spatial dimension is shown in Figure 15(b), and the result with a value of 5 Hz is highlighted in Figure 15(c). The predicted frequency distribution histogram of the remaining four excitation frequencies of 2.5 Hz, 7.5 Hz, 10.0 Hz, and 12.5 Hz are plotted in Figure 16. The results of all the excitation frequencies are summarized in Table 1.

The results of the verification test indicate that the proposed method can accurately predict the vibration frequency using the Kinect depth data.



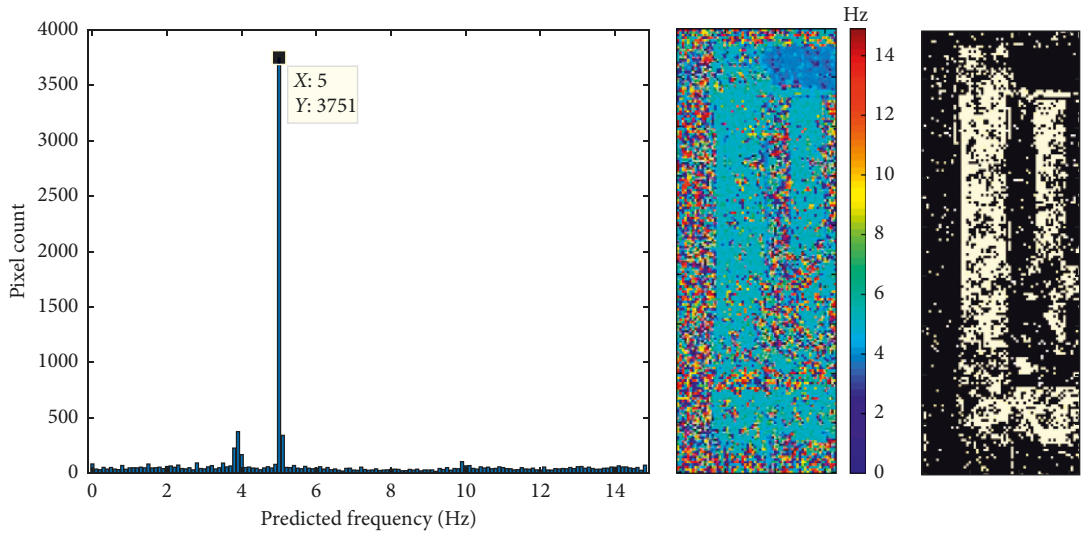
(a) (b) (c)

Figure 13: ROI (highlighted in red) for vibration measurement in the simply supported carbon plate experimental case 1. 3D point cloud: (a) front view, (b) left view, and (c) top view.



(a) (b) (c)

Figure 14: ROI (highlighted in red) for vibration measurement in the simply supported carbon plate experimental case 2. 3D point cloud: (a) front view, (b) left view, and (c) top view.



(a) (b) (c)

Figure 15: Verification test results for an excitation frequency of 5 Hz. (a) Histogram of predicted frequency distribution. (b) Prediction outputs visualization in their original spatial position. (c) Highlighted pixels predicted as 5 Hz for verification.

3.2. Cantilever Steel Beam. Histograms of predicted frequency result distributions for case 1 and case 2 are plotted in Figures 17(a) and 18(a), respectively. The results are visualized in their original spatial position in the ROI as shown in Figures 17(b) and 18(b), respectively, while the results with the value of 9.4 Hz and 9.5 Hz are highlighted and plotted in Figures 17(c) and 18(c), respectively. The raw vibration signals of the cantilever steel beam recorded by using the vibrometer were first normalized and plotted as time histories in Figures 19(a) and 20(a); then, the normalized power spectral density (PSD) obtained from FFT are plotted in Figures 19(b) and 20(b), respectively, to compare the peak frequency components of the predicted results.

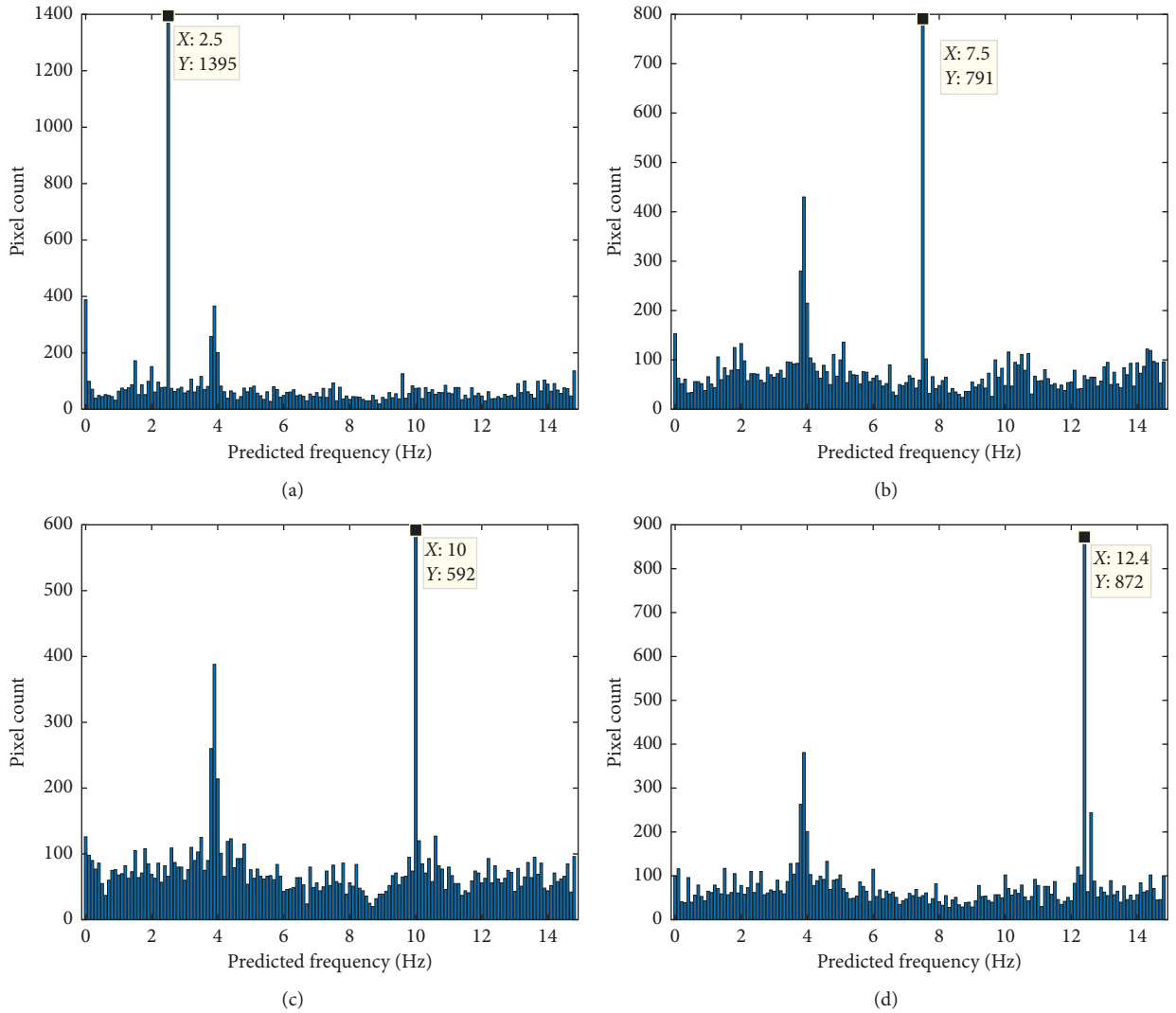


Figure 16: Histogram of predicted frequency distribution from the verification test. Results at different excitation frequencies: (a) 2.5 Hz; (b) 7.5 Hz; (c) 10 Hz; (d) 12.5 Hz.

It can be observed from the results of the experimental cases that the proposed method successfully predicted the vibration frequency and closely matched the results from the vibrometer, namely, 9.4 Hz and 9.5 Hz in case 1 and 9.5 Hz and 9.5 Hz in case 2. In addition, it can be observed that, in case 1, the pixels with the correct prediction result were from almost the entire ROI, while in case 2, the pixels with the correct prediction results were only distributed around the edge of the vibrometer and the steel beam.

3.2.1. Experimental Case 1. The cantilever steel beam experimental case 1 results from the proposed method and vibrometer are shown in Figures 17 and 18, respectively.

3.2.2. Experimental Case 2. The cantilever steel beam experimental case 2 results from the proposed method and vibrometer are shown in Figures 19 and 20, respectively.

Table 1: Summary of the verification test result.

Experimental cases	Excitation frequency (Hz)	Predicted frequency (Hz)
1	2.5	2.5
2	5.0	5.0
3	7.5	7.5
4	10.0	10.0
5	12.5	12.4

3.3. Simply Supported Carbon Plate. The prediction result histograms of the simply supported carbon plate experimental case 1 and case 2 are shown in Figures 21(a) and 22(a), respectively and the corresponding result visualizations are shown in Figures 21(b) and 22(b), respectively. Pixels predicted as 6.8 Hz are highlighted in Figures 21(c) and 22(c), respectively, for both cases. Figures 23 and 24 show the result from the contact vibrometer; the normalized time histories of the acceleration signal for case 1 and case 2 are plotted in Figures 23(a) and 24(a), respectively, while the

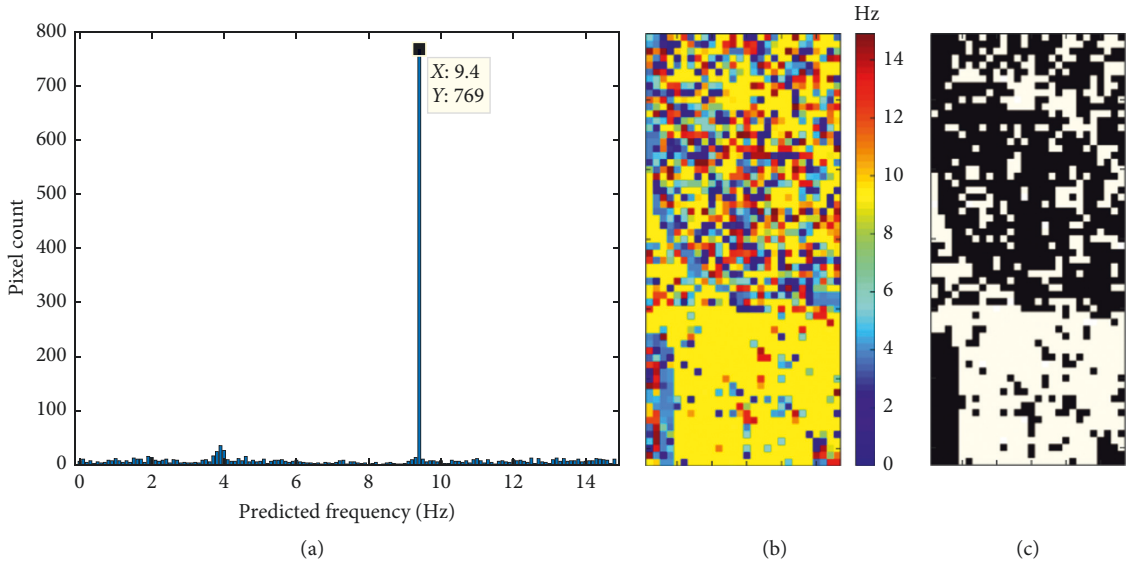


Figure 17: Cantilever steel beam experimental case 1 result from the proposed method. (a) Histogram of predicted frequency distribution. (b) Predicted outputs visualization in their original spatial position. (c) Highlighted pixels predicted as 9.4 Hz for verification.

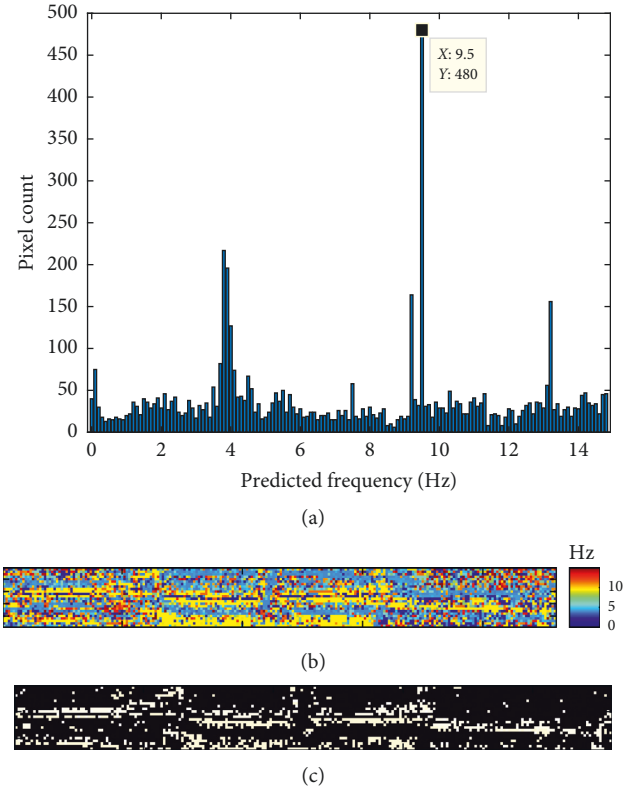


Figure 18: Cantilever steel beam experimental case 2 result from the proposed method. (a) Histogram of predicted frequency distribution. (b) Predicted outputs visualization in their original spatial position. (c) Highlighted pixels predicted as 9.5 Hz for verification.

corresponding PSDs obtained by using FFT with peak picking are shown in Figures 23(b) and 24(b), respectively.

The results indicate that the proposed method can effectively predict the vibration frequency of the target object ROI. Furthermore, it can be observed that, in case 1, almost the entire ROI pixels were predicted as 6.8 Hz, while in case 2, only the pixels around the test object and vibrometer edge were predicted as 6.8 Hz.

3.3.1. Experimental Case 1. The simply supported carbon plate experimental case 1 results from the proposed method and vibrometer are shown in Figures 21 and 22, respectively.

3.3.2. Experimental Case 2. The simply supported carbon plate experimental case 2 results from the proposed method and vibrometer are shown in Figures 23 and 24, respectively.

3.4. FFT Peak Picking Comparison. To compare the result of the proposed method with that of the FFT peak picking method with raw depth information from Kinect, the peaks of the frequency-domain results obtained via FFT of raw distance signals in each pixel are picked as frequency prediction results for each pixel. Histogram of the predicted frequency distribution via FFT is also used to quantitatively evaluate the predicted vibration frequency result within ROI, and the highest counted prediction result will be considered to be the overall prediction of the vibration frequency for the target ROI.

The summary of the result comparison of the proposed method and FFT peak picking is shown in Table 2. For better illustration, the simply supported carbon plate experimental case 1 result is used as a representative example, the time history of raw distance from Kinect depth information of each pixel in this experiment case is shown in Figure 25(a), and the corresponding frequency domain result obtained by FFT is shown in Figure 25(b). The vibration frequency result distribution histogram is shown in Figure 26(a), and the corresponding result visualization is shown in Figure 26(b), and the pixels

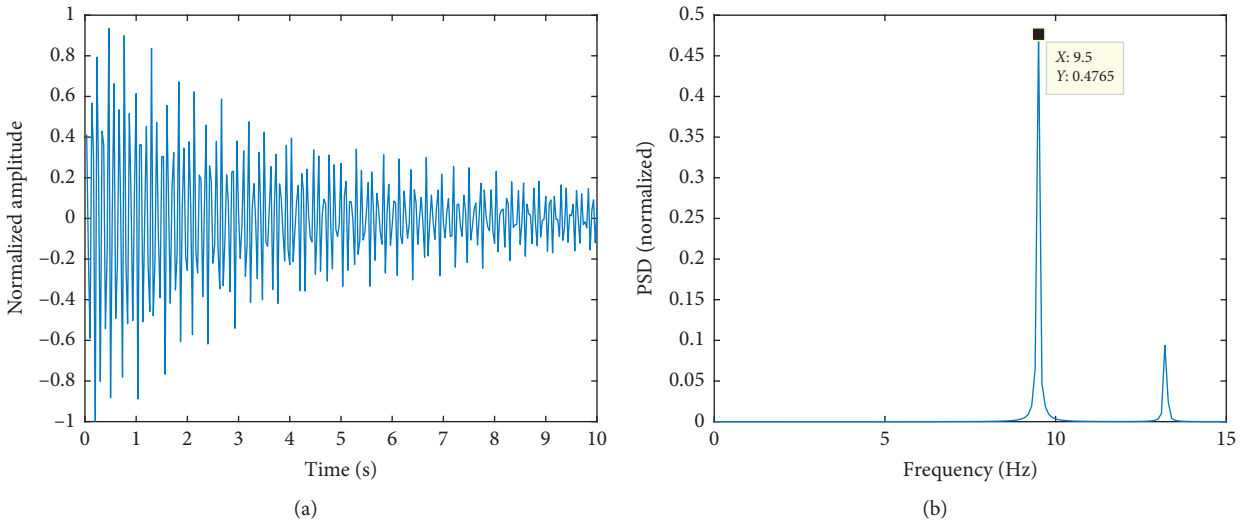


Figure 19: Cantilever steel beam experimental case 1 result from the vibrometer. (a) Normalized acceleration signal. (b) Normalized power spectral density.

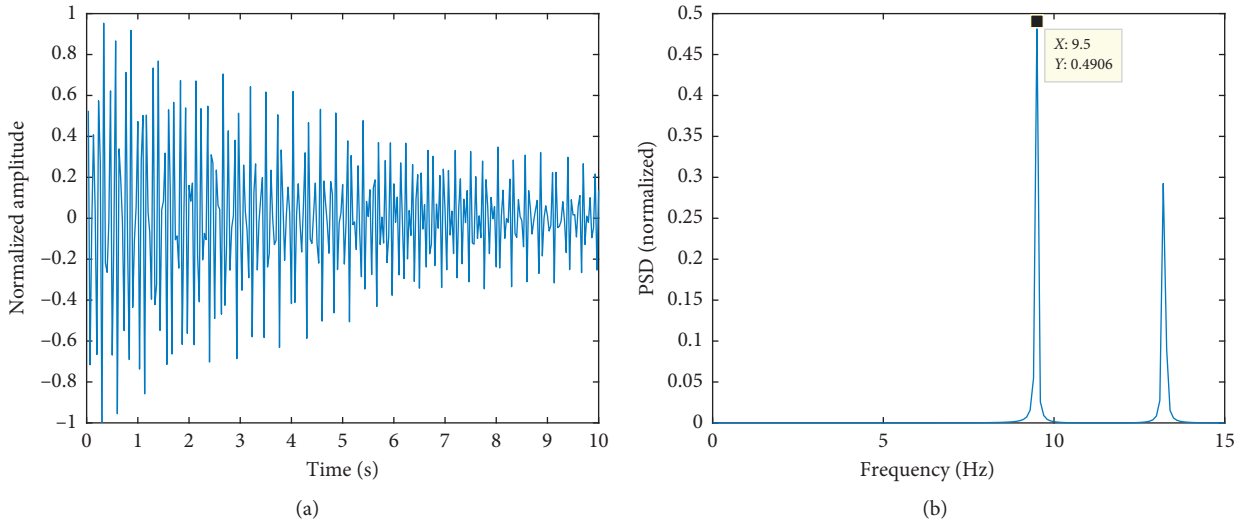


Figure 20: Cantilever steel beam experimental case 2 result from the vibrometer. (a) Normalized acceleration signal. (b) Normalized power spectral density.

predicted as 6.8 Hz via the FFT peak picking method are highlighted in Figure 26(c).

It can be observed that the frequency-domain result obtained by using FFT and pixel level frequency prediction result obtained by peak picking are dominated by trivial noisy low-frequency results, as shown Figures 25(b) and 26(a). It is also noted that there are still some pixels correctly predicted as 6.8 Hz using the FFT peak picking method as shown in Figure 26(c), but when compared with result of the proposed method result as shown in Figure 21(c), the proposed method is superior to the FFT peak picking method since almost all the pixels within ROI have correct predication, while the FFT peak picking method only has correct prediction of pixels on the top left area of the ROI. The prediction result distribution histograms of the proposed method and FFT peak picking method can also quantitatively confirm it, as shown in Figures 21(a) and 26(a), respectively.

4. Discussion

The results of the experiments conducted using different excitation sources and different test objects demonstrate the performance of the proposed method. The proposed method can utilize meta depth information acquired from Kinect V2 to predict the vibration frequency of a target ROI with minor errors. A significant finding of this study is that when the Kinect was pointed in a direction parallel to the vibration direction, the depth variation signals utilized by using the proposed method were from distance variation

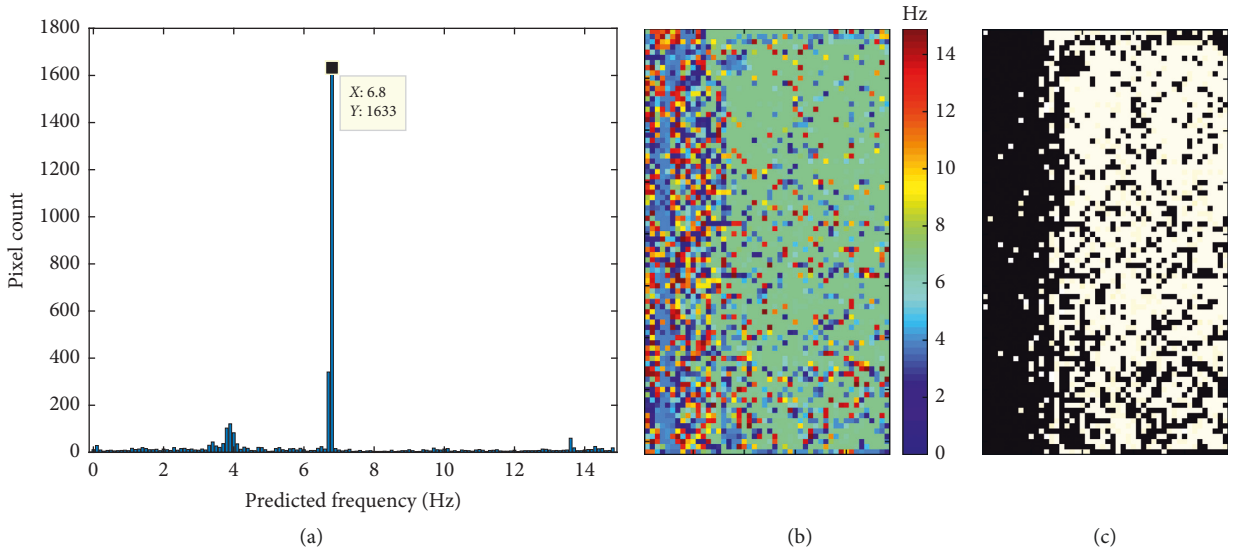


Figure 21: Simply supported carbon plate experimental case 1 result from the proposed method. (a) Histogram of predicted frequency distribution. (b) Predicted outputs visualization in their original spatial position. (c) Highlighted pixels predicted as 6.8 Hz for verification.

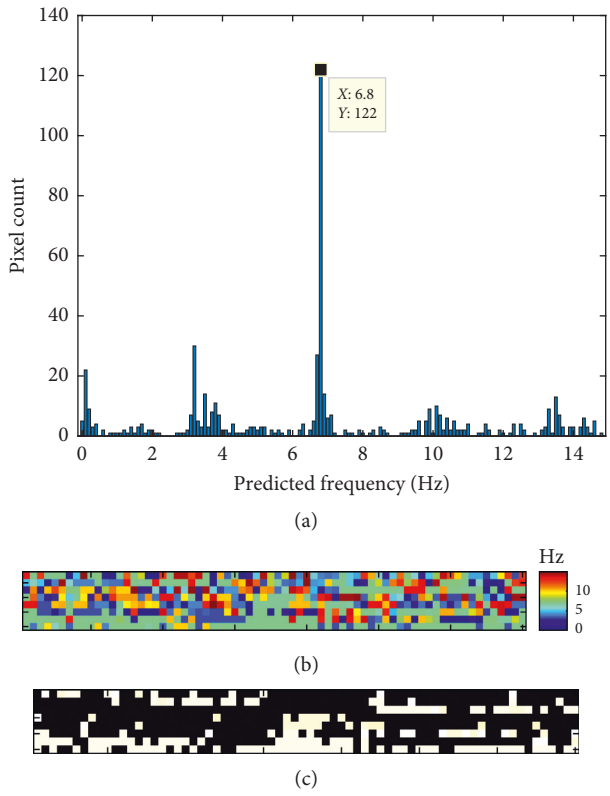


Figure 22: Simply supported carbon plate experimental case 2 result from the proposed method. (a) Histogram of predicted frequency distribution. (b) Predicted outputs visualization in their original spatial position. (c) Highlighted pixels predicted as 6.8 Hz for verification.

between the test object and depth sensor of the Kinect, as demonstrated in the steel beam experimental case 1 and carbon plate experimental case 1. When the Kinect was pointed in a direction perpendicular to the vibration direction, the proposed method can still provide useable result; in this case, it utilized depth variation signals at the edge of the test object rather than distance variations between the depth sensor and test object itself, as demonstrated in the steel beam experimental case 2 and carbon plate experimental case 2. These findings confirm the reliability and applicability of the proposed method for vibration frequency measurement.

Unlike traditional optical-based noncontact vibration measurements, we used Kinect v2 and feed-forward CNN to conduct vibration frequency measurement directly, and the use of additional signal processing or image processing algorithms is not required. Furthermore, the proposed method is fast and easy to deploy in applications as it does not require the explicit extraction of vibration signals and incorporation of denoise processing into the proposed artificial neural network for meta noisy depth signals. Besides, the proposed network is trained entirely using simulation signals, which indicates that the proposed network can be easily scaled for a larger measurement range and higher measurement precision.

This method also has some drawbacks and limitations. Interference from sunlight can occur as the depth sensor of Kinect V2 is based on infrared technology. Therefore, the proposed method is limited to indoor applications. Other inherent drawbacks of the Kinect depth sensor are that the measurement distance range is restricted to 0.4–4.5 m, and the frequency measurement range is limited to within 15 Hz since the sampling rate of the Kinect depth senor is fixed at 30 Hz. Furthermore, the measurement precision of the proposed neural network is controlled by the network configuration and the dataset it trained with, and the proposed network can only detect the resonant or peak frequency, while other frequency components are undetected. The proposed method is also vulnerable to camera shake, lighting condition vibration, and other types of electrical or mechanical noise, just like all traditional optical-based methods.

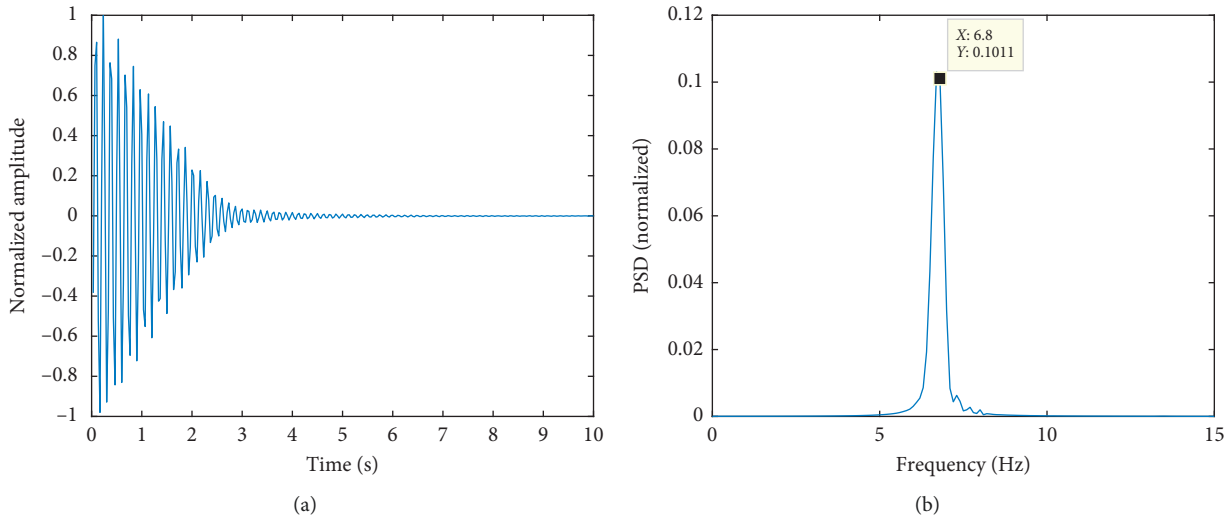


Figure 23: Simply supported carbon plate experimental case 1 result from the vibrometer. (a) Normalized acceleration signal. (b) Normalized power spectral density

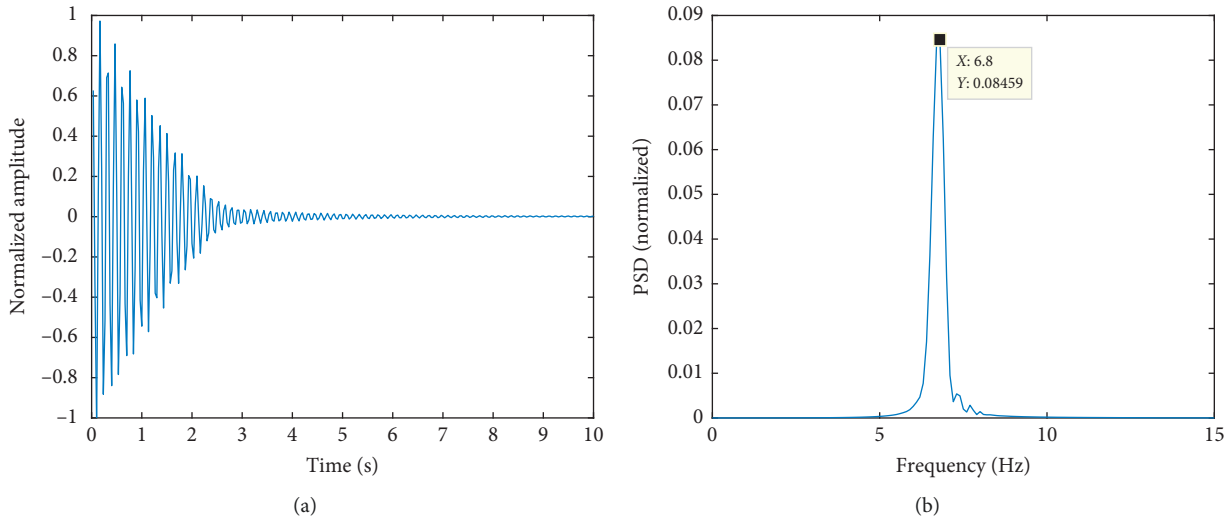


Figure 24: Simply supported carbon plate experimental case 2 result from the vibrometer. (a) Normalized acceleration signal. (b) Normalized power spectral density.

Table 2: Summary of result comparison.

Experimental scenario	Excitation/measured frequency	Proposed method (Hz)	FFT peak picking (Hz)
1	2.5	2.5	0.1
2	5.0	5.0	0.1
3	7.5	7.5	0.1
4	10.0	10.0	0.1
5	12.5	12.4	0.1
6	9.5	9.4	0.1
7	9.5	9.5	0.1
8	6.8	6.8	0.09
9	6.8	6.8	0.09

The proposed method offers a new possibility for future research on optical-based vibration measurement. We can further utilize the feature extraction capability of deep neural network for optical-based vibration signal extraction and processing. A more advanced network architecture specifically designed for vibration measurement could be used in future work. The dataset of the signal generator algorithm can also be improved and can be

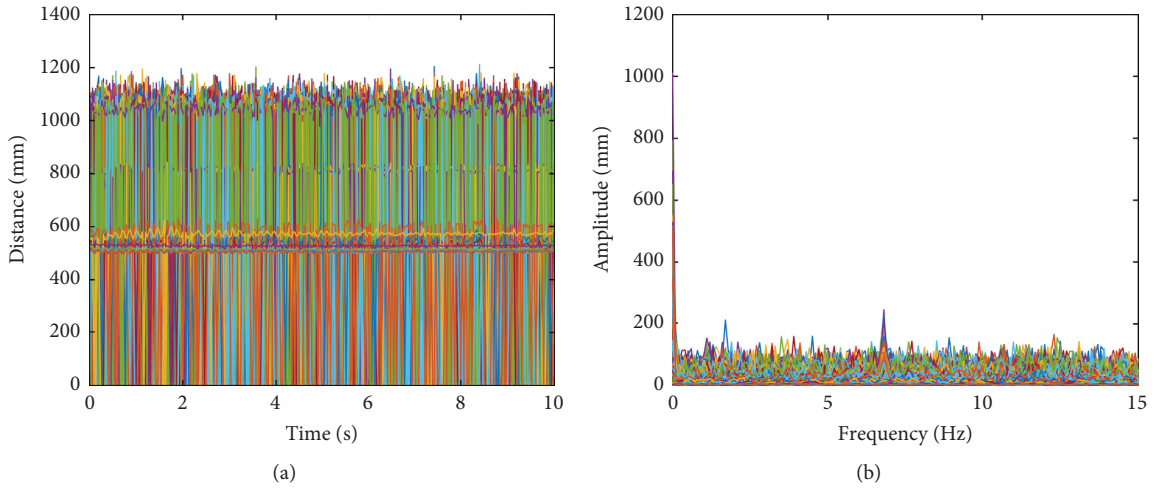


Figure 25: (a) Raw distance time history of each pixel (overlaying plot) from Kinect depth information. (b) Corresponding frequency-domain result obtained by FFT.

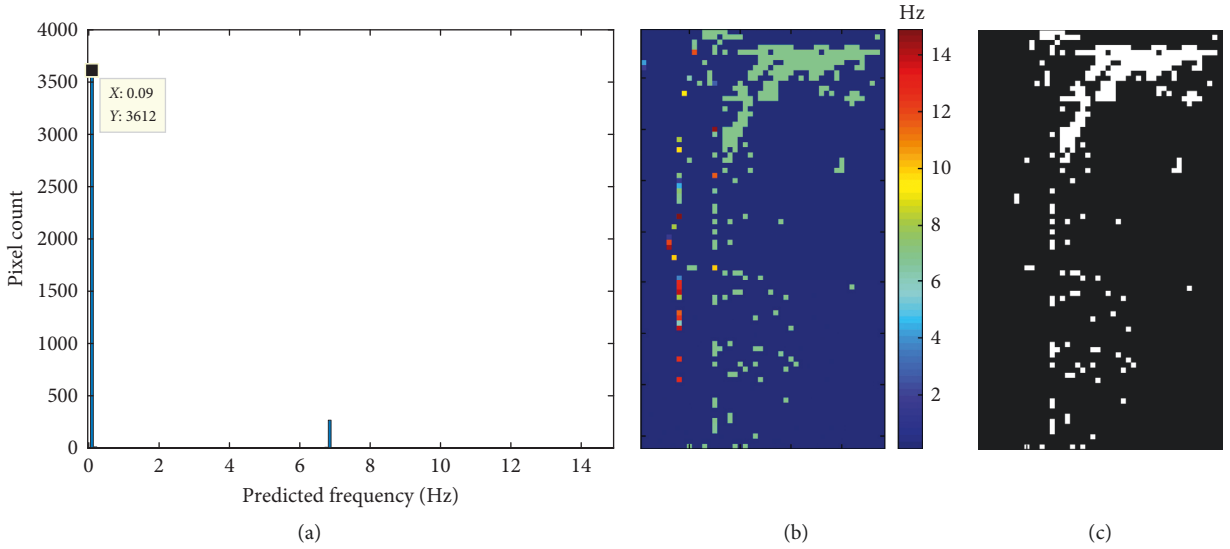


Figure 26: Simply supported carbon plate experimental case 1 result from FFT peak picking. (a) Histogram of peak picking result distribution. (b) Result visualization in their original spatial position. (c) Highlighted pixels predicted as 6.8 Hz for verification.

incorporated and augmented with real sampled depth signals for better generalization.

5. Conclusions

In this paper, we proposed a method for vibration frequency measurement using Kinect V2 and artificial neural network. Experiments were conducted to evaluate the performance of the proposed method, and results show that the proposed method can provide good vibration frequency measurement results compared to those from an industrial vibrometer.

This method is limited by the inherent drawbacks of the Kinect depth sensor and the architecture of the proposed network, as it cannot detect all the frequency components of the measurement target. In the future, we will redesign and further improve the network architecture, dataset preparation process, and workflow of the proposed method to address these limitations.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank Professor Zhiwei Chen and Professor Ying Lei from Department of Civil Engineering, Xiamen University, for the support of the experiments. This work was supported

by the National Natural Science Foundation of China (11372074).

References

[1] A. Z. Khan, A. B. Stanbridge, and D. J. Ewins, "Detecting damage in vibrating structures with a scanning LDV," *Optics and Lasers in Engineering*, vol. 32, no. 6, pp. 583–592, 2000.

[2] A. B. Stanbridge, I. A. Sever, and D. J. Ewins, "Vibration measurements in a rotating blisk test rig using an LDV," in *Proceedings of Fifth International Conference on Vibration Measurements by Laser Techniques: Advances and Applications*, pp. 1–8, Ancona, Italy, May 2002.

[3] P. L. Reu, D. P. Rohe, and L. D. Jacobs, "Comparison of DIC and LDV for practical vibration and modal measurements," *Mechanical Systems and Signal Processing*, vol. 86, pp. 2–16, 2017.

[4] L. Bin, L. Jing-Wen, and Y. Chun-Yong, "Study on the measurement of in-plane displacement of solid surfaces by laser doppler velocimetry," *Optics & Laser Technology*, vol. 27, no. 2, pp. 89–93, 1995.

[5] D. H. Diamond, P. S. Heyns, and A. J. Oberholster, "Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence," *Measurement*, vol. 95, pp. 166–172, 2017.

[6] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.

[7] J. J. Lee and M. Shinozuka, "Real-time displacement measurement of a flexible bridge using digital image processing techniques," *Experimental Mechanics*, vol. 46, no. 1, pp. 105–114, 2006.

[8] L. Tian and B. Pan, "Remote bridge deflection measurement using an advanced video deflectometer and actively illuminated led targets," *Sensors (Basel)*, vol. 16, no. 9, p. 1344, 2016.

[9] D. Kalpoe, K. Khoshelham, and B. Gorte, "Vibration measurement of a model wind turbine using high speed photogrammetry," in *Proceedings of International Society for Optics and Photonics Videometrics, Range Imaging, and Applications XI*, p. 80850, Munich, Germany, June 2011.

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

[11] L. Yu and B. Pan, "Single-camera high-speed stereo-digital image correlation for full-field vibration measurement," *Mechanical Systems and Signal Processing*, vol. 94, pp. 374–383, 2017.

[12] M. N. Helfrick, C. Niezrecki, P. Avitabile, and T. Schmidt, "3D digital image correlation methods for full-field vibration measurement," *Mechanical Systems and Signal Processing*, vol. 25, no. 3, pp. 917–927, 2011.

[13] T. Beberniss, T. Eason, and S. Spottswood, "High-speed 3D digital image correlation measurement of long-duration random vibration; recent advancements and noted limitations," in *Proceedings of 25th International Conference on Noise and Vibration engineering (ISMA2012)*, Leuven, Belgium, September 2012.

[14] D. Garcia, J. J. Orteu, and L. Penazzi, "A combined temporal tracking and stereo-correlation technique for accurate measurement of 3d displacements: application to sheet metal forming," *Journal of Materials Processing Technology*, vol. 125-126, pp. 736–742, 2002.

[15] T. Schmidt, J. Tyson, and K. Galanulis, "Full-field dynamic displacement and strain measurement-specific examples using advanced 3D image correlation. Photogrammetry: part II," *Experimental Techniques*, vol. 27, no. 4, pp. 22–26, 2003.

[16] D. Ribeiro, R. Calçada, J. Ferreira, and T. Martins, "Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system," *Engineering Structures*, vol. 75, pp. 164–180, 2014.

[17] T. Nakamura, "Real-time 3D object tracking using kinect sensor," in *Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 784–788, Mueang Phuket, Thailand, December 2011.

[18] A. Oliver, S. Kang, B. C. Wünsche, and B. MacDonald, "Using the kinect as a navigation sensor for mobile robotics," in *Proceedings of 27th Conference on Image and Vision Computing New Zealand*, pp. 509–514, ACM, Dunedin, New Zealand, November 2012.

[19] O. Wasenmüller, M. Meyer, and D. Stricker, "Corbs: Comprehensive RGB-D benchmark for slam using kinect v2," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–7, Lake Placid, NY, USA, March 2016.

[20] N. Engelhard, F. Endres, J. Hess, J. Sturm, and W. Burgard, "Real-time 3D visual slam with a hand-held RGB-D camera," in *Proceedings of RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, pp. 1–15, Vasteras, Sweden, April 2011.

[21] B. Y. Li, A. S. Mian, W. Liu, and A. Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 186–192, Tampa, FL, USA, January 2013.

[22] R. Min, N. Kose, and J.-L. Dugelay, "KinectFaceDB: a kinect database for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 11, pp. 1534–1548, 2014.

[23] A. P. Placitelli and L. Gallo, "Low-cost augmented reality systems via 3D point cloud sensors," in *Proceedings of Seventh International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*, pp. 188–192, Dijon, France, November 2011.

[24] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using kinects," *IEEE transactions on visualization and computer graphics*, vol. 18, pp. 643–650, 2012.

[25] R. A. Clark, Y.-H. Pua, K. Fortin et al., "Validity of the microsoft kinect for assessment of postural control," *Gait & Posture*, vol. 36, no. 3, pp. 372–377, 2012.

[26] A. P. Placitelli and L. Gallo, "3D point cloud sensors for low-cost medical in-situ visualization," in *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, pp. 596-597, Atlanta, GA, USA, November 2011.

[27] T. Dutta, "Evaluation of the kinect sensor for 3-D kinematic measurement in the workplace," *Applied Ergonomics*, vol. 43, no. 4, pp. 645–649, 2012.

[28] H. Shen, B. He, J. Zhang, and S. Chen, "Obtaining four-dimensional vibration information for vibrating surfaces with a kinect sensor," *Measurement*, vol. 65, pp. 149–165, 2015.

[29] J. Liu and X. Yang, "Learning to see the vibration: a neural network for vibration frequency prediction," *Sensors*, vol. 18, no. 8, p. 2530, 2018.

[30] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of International Conference on Machine Learning*, pp. 448–456, Lille, France, July 2015.

[31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of 27th International Conference on Machine Learning (ICML-10)*, pp. 807–814, Haifa, Israel, June 2010.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[33] X. Mao, C. Shen, and Y. Yang, "Image denoising using very deep fully convolutional encoder-decoder networks with symmetric skip connections," March 2016, https://arxiv.org/abs/1603.09056.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[35] D. P. Kingma, J. Ba, and Adam, "A method for stochastic optimization," December 2014, https://arxiv.org/abs/1412.6980.

[36] A. Paszke, S. Gross, S. Chintala et al., "Automatic differentiation in PyTorch," in *Proceedings of 31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, October 2017.