

Automatic Displacement and Vibration Measurement in Laboratory Experiments with A Deep Learning Method

Yongsheng Bai, Ramzi M. Abdulllah, Halil Sezen, Alper Yilmaz

Department of Civil, Environmental and Geodetic Engineering

The Ohio State University

Columbus, OH, USA

bai.426@osu.edu, abdulllah.2@osu.edu, sezen.1@osu.edu, yilmaz.15@osu.edu

Abstract—This paper proposes a pipeline to automatically track and measure displacement and vibration of structural specimens during laboratory experiments. The latest Mask Regional Convolutional Neural Network (Mask R-CNN) can locate the targets and monitor their movement from videos recorded by a stationary camera. To improve precision and remove the noise, techniques such as Scale-invariant Feature Transform (SIFT) and various filters for signal processing are included. Experiments on three small-scale reinforced concrete beams and a shaking table test are utilized to verify the proposed method. Results show that the proposed deep learning method can achieve the goal to automatically and precisely measure the motion of tested structural members during laboratory experiments.

Index Terms—Mask R-CNN, displacement measurement, vibration measurement, structural experiments, shaking table.

I. INTRODUCTION

For safety assessment and structural health monitoring of infrastructures such as buildings and bridges and their components, their vibrations and deformations need to be recorded and evaluated. Traditional structural sensors such as Linear Variable Displacement Transducers (LVDTs), dial gauges, accelerometers and other advanced sensors are used to measure deformations and vibrations. However, in some cases, conventional sensors may not be a good option to access the desirable instrumentation locations and work in a timely and cost-efficient way. Most importantly, traditional sensors measure the displacement or vibration at a discrete location, i.e., one sensor is used to measure one quantity at a single point. On the other hand, a high definition video camera can record the movement of a component or an entire structure rather than a single point.

Recent developments on vision- and vibration-based technologies led to the measurement applications of low-cost and non-contact sensors on deformation and vibration monitoring of the infrastructures [1], especially because unusual and extreme deformations or vibrations in aging bridges and buildings may be an indication of significant serviceability or safety issues. Cameras can collect high-quality images or high-speed videos in lab or field tests as non-contact and non-destructive sensors. With computer vision technologies and various deep learning techniques, the vision sensors can not only be used as eyes for Artificial Intelligence (AI) vehicles and machines, but also provide opportunities for scientific measurements to researchers and engineers. In a laboratory

experiment, cameras can be fixed near the testing station and record live motion of the monitored structural members. They can be placed at a stationary location or some distance away from an in-service bridge or building to record its structural movements remotely in the field. The collected visual data can be processed further to identify potential damage and assess the motion of the observed structures precisely.

To better understand and efficiently use displacement and vibration measurement data from cameras, we used Mask R-CNN [2] with High-resolution network (HRNet) [3] to track a target attached on beam specimens in the laboratory and to measure their deflections in the first study. Then the Mask R-CNN was applied on a shaking table test to track the dynamic motion of four targets simultaneously in the second study.

II. LITERATURE REVIEW

Computer Vision (CV) and deep learning techniques are very useful to gain high-level understanding and extraction of desired information and precise motion measurements from images and videos.

Traditional CV techniques have been widely used by researchers for displacement or vibration measurements by cameras. These techniques include image processing technique [4], up-sampled cross correlation [5], adaptive Region of Interest algorithm [6], modified Taylor approximation [7], and contour extraction with Speeded-Up Robust Features (SURF) [8]. In addition, Lucas-Kanade template tracking algorithm [9], [10] and Digital Image Correlation (DIC) [11] are employed to track displacement and vibration of structural members. Furthermore, Hu and Pai [12] utilized a camera-based 3D motion analysis system to measure the resonant vibration of steel cables. Chen et al. [13] described an application with a video camera-based technique to test the vibration of an antenna tower on a tall building when a camera was placed 175 meters far from it. Hoskere et al. [14] used an Unmanned Aerial Vehicle (UAV) to measure the modal properties and dynamic response of a full-scale structure.

Deep learning is a relatively new research area for visual measurement applications. Dong et al. [15] implemented a full field optical flow algorithm named FlowNet2 to measure the displacement and vibration of structures. They also used the Spatio-Temporal Context Learning to track targets and utilized

a Taylor approximation to gain subpixel level precision for displacement measurement [16]. Also, Dong et al. [17] applied Visual Graph Visual Geometry Group (VGG) to extract features of the target for monitoring and measuring during the traffic time. These methods indicate how we can efficiently use cameras to monitor and measure the displacement and vibration of structural members or a structure in the laboratory or in the field.

III. METHODOLOGY

In our previous studies [18], [19], new variants of the latest Mask R-CNNs were successfully applied for structural damage detection with high accuracy. Therefore, we used one of the variants, Mask R-CNN with HRNet, to track and measure the displacement and vibration in this research. Fig. 1 shows the framework of this Mask R-CNN.

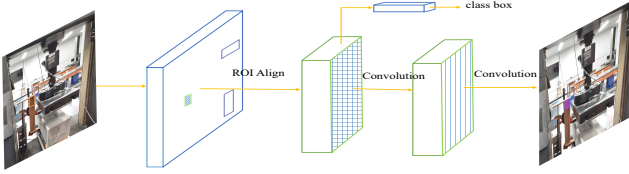


Fig. 1. The framework of Mask R-CNN for tracking and measuring displacement of a reinforced concrete beam

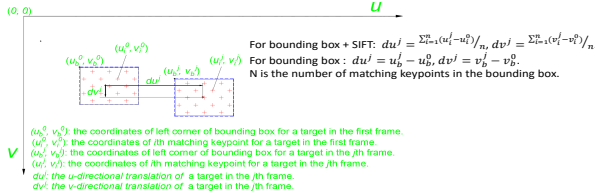


Fig. 2. Translation of a target measured by a bounding box or by matching keypoints.

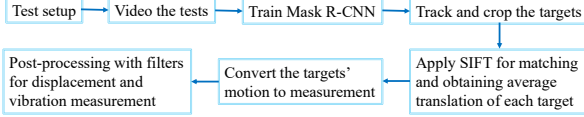


Fig. 3. A flowchart of Mask R-CNN and SIFT for automated displacement and vibration measurement in a lab experiment with a stationary camera.

As shown in Fig. 1 and 4, a wood frame is attached near the midspan of the tested beam so that it moves downward or upward when the beam is loaded or unloaded. The motion of the frame represents the deflection of the point where it is attached. Mask R-CNN is used to track the top of this wood frame (i.e., yellow dashed line on the left image in Fig. 1), which is marked by a bounding box and a mask in purple on the right image. Since the tracking target is a rigid body, its motion can be represented by any point on it or by the bounding box. On the other hand, as shown in Fig. 2 for the image plane of a stationary camera, translation of a target between the first frame and the j th frame, du^j and dv^j , can be calculated as the position change of the bounding box or the average motion of the matching keypoints with SIFT. We observed that the mask for a target may be not exactly the same as its real shape in some cases. Therefore, SIFT is introduced into the pipeline to eliminate this inadvertent drawback when the bounding box is inaccurate to represent the target. Furthermore, subpixel precision can be achieved by

matching and using the average motion of these keypoints on the target. In our pipeline (see Fig. 3), the ratio of good matching is restricted as SIFT commonly used in practice, but the range of coordinate change for each matching keypoint is also constrained such that the top good matching keypoints can be secured. Therefore, the mismatching is reduced dramatically and the accuracy of measurement is improved. Finally, the measurement is converted from pixel to length unit (inches or millimeters), which is also called a scalar, s . The horizontal and vertical displacement dx^j and dy^j of the target can be obtained as follows:

$$dx^j = s \times du^j \quad (1)$$

$$dy^j = s \times dv^j \quad (2)$$

As a comparison, an optical flow method called Lucas-Kanade (LK) tracker [20] was used to track and measure the same targets in our experiments. For the LK tracker, the relative displacement is measured between two adjacent frames in a video. The final displacement is the sum-up of all the relative measurements. In addition, Savitzky-Golay filter [21] and Butterworth filter [22] were employed to handle the inconsistency and noise of the measurements. Fast Fourier Transform (FFT) [23] was applied to extract the frequencies of the vibrating targets.

IV. IMPLEMENTATION

Two types of indoor experiments were utilized to verify our proposed methods. The first one is a small-scale experiment involving three reinforced concrete (RC) beams (see Fig. 4) conducted on the main campus of The Ohio State University in Columbus, Ohio. These beams are loaded and deflected until failure. Another test is an application on a video of a shaking table test (see Fig. 5) [24].

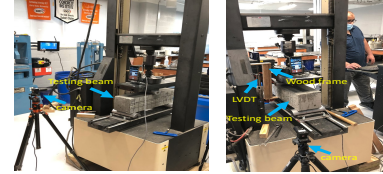


Fig. 4. A flexural test of a RC beam with a LVDT and a camera.

A. Deflection Measurement of RC Beams in Laboratory Tests

In this experiment, three RC beams were subjected to a monotonically increasing point load at the midspan. A displacement sensor (a LVDT or dial gauge) and a wood frame were used to measure the deflection near midspan of the beams, while a camera was placed 3-feet away from the midspan (see Fig. 4). Its definition was 1600×1200 and the frame rate was 15 per second. The wood frame was clamped on the beam to represent the deflection of the targeted point on the tested beam.

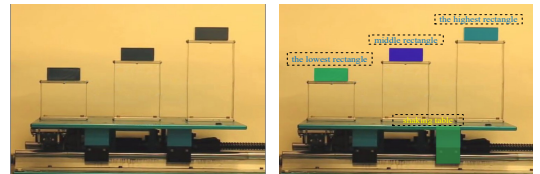


Fig. 5. Original image (left) and label (right) in the shaking table video [24].

The data process for the videos of three tests is like this: since this is a static test, which means the loading and deflection of the tested beams is slow, images from the video at each second are selected as the visual data. There are a total of 500 to 600 images for each test. To train the Mask R-CNN, only 50 images are randomly selected and labeled for detecting and tracking the top of wood frame (in purple mask) as shown in Fig. 1. The LK tracker is used to track the same object and measure the deflection. The testing results are shown in Fig. 8. Compared to the LK tracker, this Mask R-CNN with SIFT can provide a measurement closer to the ground truth of beam deflections by a dial gauge. In addition, Savitzky-Golay filter is applied to smooth the measurement such that it becomes consistent (see Fig. 9). Table I shows MAE (mean absolute error) for both methods. It can be inferred that SIFT and Savitzky-Golay filter can effectively readjust the position of the bounding box predicted from the Mask R-CNN and smooth the measurements, hence, the proposed method can outperform the LK tracker.

TABLE I
MAE FOR TWO METHODS (INCHES)

Methods	Test 1	Test 2	Test 3
Mask R-CNN + SIFT	0.005	0.005	0.005
LK tracker	0.030	0.012	0.012

B. Vibration Measurement of A Shaking Table Test

Our proposed method was also applied on a shaking table test [24] to check its applicability of monitoring dynamic movement of objects. In this test, there are three rectangles (masses) fixed on the shaking table at different heights (see Fig. 5). Each rectangle, which is supported by two sticks, like a structure has its unique resonant frequency in the horizontal direction. This is due to differences between the lateral stiffness of each pair of sticks. The frequencies of the applied shaking are increased from 4 Hz to 13.65 Hz to excite these masses and cause their harmonic vibrations. From the recorded video [24], 150 frames are randomly selected from a total of 6,674 frames and labeled for training the Mask R-CNN. The video has an image size of 640×480 and a frame rate of 30 per second. SIFT is not applied to smooth the measurements here, since the goal of this test is to detect the frequencies instead of accurate amplitudes of the vibration, which is in pixel unit in this test. Thus, the motion of the bounding box represents the translation of each object. The LK tracker is utilized to verify our method by tracking the same vibration of the shaking table. On one hand, all the raw data are processed by Butterworth filter, and FFT is applied to extract the frequencies for each tracking target. The filtered vibrations of the shaking table with the LK tracker and Mask R-CNN are shown in the left figures of Fig. 6. There are three frequencies of vibration at approximately 4 Hz, 6.35 Hz and 11.35 Hz excited by the table. Both methods capture these frequencies (yellow captions in the right figures of Fig. 6) with a less than 2.6% error. On the other hand, the vibrations of three rectangles are measured by the Mask R-CNN and raw data are processed like the procedures for the shaking table. As shown in Fig. 7, their resonant frequencies are very close

to the intended frequencies (i.e., 4 Hz, 6.35 Hz and 11.35 Hz). The error rate for this measurement is 0%, 0.3% and 2.6%, respectively. This indicates that the proposed Mask R-CNN can be used alone to track multiple objects and capture their vibrations characteristics precisely.

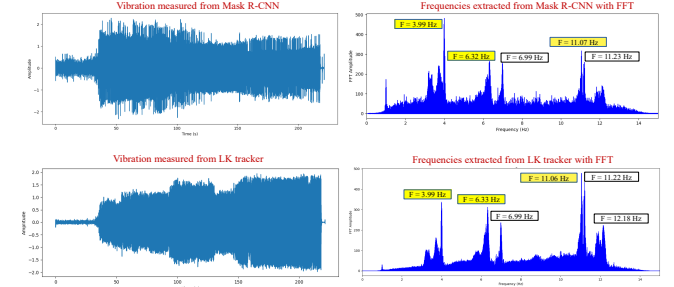


Fig. 6. Vibration and frequencies of the shaking table measured by Mask R-CNN and the LK tracker.

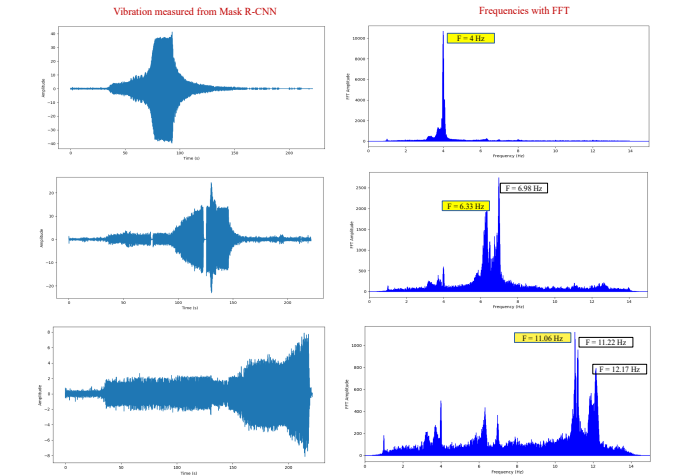


Fig. 7. Vibrations of three rectangles (the highest to the lowest one from top to bottom on the left hand side) measured by Mask R-CNN and the corresponding frequencies calculated by FFT (on the right hand side).

V. CONCLUSIONS

A deep learning method (i.e., Mask R-CNN with HRNet) and techniques such as SIFT and Savitzky-Golay filter are applied to automatically track the targets and provide the accurate measurement of their motions with a stationary camera. In our first experiment, Mask R-CNN and SIFT were used for precise deflection measurement of the tested RC beams, since SIFT can utilize the keypoints on the targets to refine the measurement. Our method can be closer to the measurement from traditional structural sensors and outperform the LK tracker. The Mask R-CNN was also used alone to track the vibration of multiple targets in a shaking table experiment and capture the resonant frequencies of these targets via Butterworth filter and FFT. These preliminary tests show that the proposed method is robust and has the potential for measuring displacements and vibrations of structural specimens precisely and automatically in laboratory experiments.

Our ongoing work involves application of the proposed method on actual buildings tested in the field to confirm its applicability for outdoor environments. Other deep learning methods are also being explored and tested.

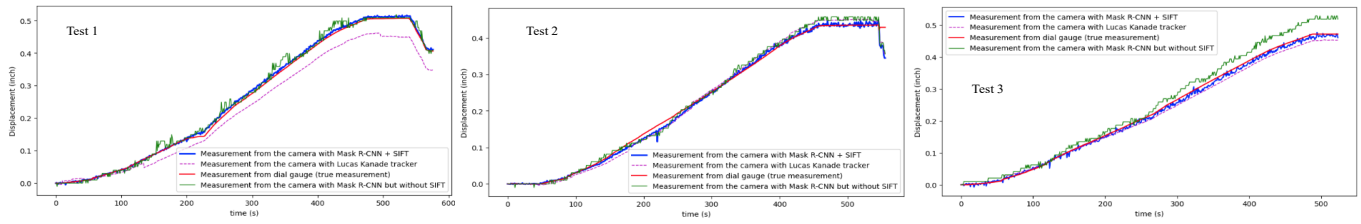


Fig. 8. The deflection-time relationship measured by a camera and a dial gauge for the three testing beams.

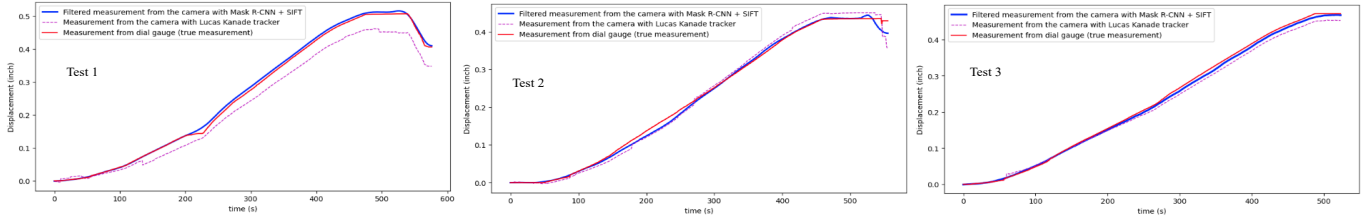


Fig. 9. The filtered deflection-time relationship measured by a camera and a dial gauge for the three testing beams.

REFERENCES

- [1] C.-Z. Dong and F. N. Catbas, "A review of computer vision-based structural health monitoring at local and global levels," *Structural Health Monitoring*, p. 1475921720935585, 2020.
- [2] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [3] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang, "High-resolution representations for labeling pixels and regions," *arXiv preprint arXiv:1904.04514*, 2019.
- [4] 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.
- [5] D. Feng, M. Q. Feng, E. Ozer, and Y. Fukuda, "A vision-based sensor for noncontact structural displacement measurement," *Sensors*, vol. 15, no. 7, pp. 16557–16575, 2015.
- [6] J. Lee, K.-C. Lee, S. Cho, and S.-H. Sim, "Computer vision-based structural displacement measurement robust to light-induced image degradation for in-service bridges," *Sensors*, vol. 17, no. 10, p. 2317, 2017.
- [7] B. Liu, D. Zhang, and J. Guo, "Vision-based displacement measurement sensor using modified taylor approximation approach," *Optical Engineering*, vol. 55, no. 11, p. 114103, 2016.
- [8] Z. Yin, C. Wu, and G. Chen, "Concrete crack detection through full-field displacement and curvature measurements by visual mark tracking: A proof-of-concept study," *Structural Health Monitoring*, vol. 13, no. 2, pp. 205–218, 2014.
- [9] J. Guo and C. Zhu, "Dynamic displacement measurement of large-scale structures based on the lucas-kanade template tracking algorithm," *Mechanical Systems and Signal Processing*, vol. 66, pp. 425–436, 2016.
- [10] C.-Z. Dong, O. Celik, and F. N. Catbas, "Marker-free monitoring of the grandstand structures and modal identification using computer vision methods," *Structural Health Monitoring*, vol. 18, no. 5-6, pp. 1491–1509, 2019.
- [11] G. Chen, Q. Liang, W. Zhong, X. Gao, and F. Cui, "Homography-based measurement of bridge vibration using uav and dic method," *Measurement*, vol. 170, p. 108683, 2021.
- [12] J. Hu and P. F. Pai, "Experimental study of resonant vibrations of suspended steel cables using a 3d motion analysis system," *Journal of Engineering Mechanics*, vol. 138, no. 6, pp. 640–661, 2012.
- [13] J. G. Chen, A. Davis, N. Wadhwa, F. Durand, W. T. Freeman, and O. Büyüköztürk, "Video camera-based vibration measurement for civil infrastructure applications," *Journal of Infrastructure Systems*, vol. 23, no. 3, p. B4016013, 2017.
- [14] V. Hoskere, J.-W. Park, H. Yoon, and B. F. Spencer Jr, "Vision-based modal survey of civil infrastructure using unmanned aerial vehicles," *Journal of Structural Engineering*, vol. 145, no. 7, p. 04019062, 2019.
- [15] C.-Z. Dong, O. Celik, F. N. Catbas, E. J. O'Brien, and S. Taylor, "Structural displacement monitoring using deep learning-based full field optical flow methods," *Structure and Infrastructure Engineering*, vol. 16, no. 1, pp. 51–71, 2020.
- [16] C.-Z. Dong, O. Celik, F. N. Catbas, E. O'Brien, and S. Taylor, "A robust vision-based method for displacement measurement under adverse environmental factors using spatio-temporal context learning and taylor approximation," *Sensors*, vol. 19, no. 14, p. 3197, 2019.
- [17] C.-Z. Dong and F. N. Catbas, "A non-target structural displacement measurement method using advanced feature matching strategy," *Advances in Structural Engineering*, vol. 22, no. 16, pp. 3461–3472, 2019.
- [18] Y. Bai, H. Sezen, and A. Yilmaz, "Detecting cracks and spalling automatically in extreme events by end-to-end deep learning frameworks," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. x, pp. xxx–xxx, 2021.
- [19] Y. Bai, H. Sezen, and A. Yilmaz, "End-to-end deep learning methods for automated damage detection in extreme events at various scales," *25th International Conference on Pattern Recognition (ICPR)*, pp. 6640–6647, 2021.
- [20] J.-Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel corporation*, vol. 5, no. 1-10, p. 4, 2001.
- [21] G. Bianchi and R. Sorrentino, *Electronic filter simulation & design*. McGraw Hill Professional, 2007.
- [22] W. H. Press and S. A. Teukolsky, "Savitzky-golay smoothing filters," *Computers in Physics*, vol. 4, no. 6, pp. 669–672, 1990.
- [23] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of computation*, vol. 19, no. 90, pp. 297–301, 1965.
- [24] "Missouri university of science and technology: Sdof system resonance vibration test," Apr 2017.