VIDEO-SENSOR DATA FUSION FOR ENHANCED STRUCTURAL MONITORING

This document is a technical summary of the project report, *Video-sensor Data Fusion for Enhanced Structural Monitoring*, prepared by David Lattanzi for the U.S. DOT Region 3 University Transportation Research Center

INTRODUCTION

Engineers have used sensor arrays to monitor the behavior and condition of infrastructure systems for decades. These arrays are typically attached or embedded within a structure and provide localized measurements of a system's response. While these sensors are highly accurate, they have well-known practical limitations. One drawback is that most sensors only provide optimal measurements near defects, locations that are difficult to know at the time of installation [1]–[3]. This is compounded by the fact that the costs and practicalities of sensor array maintenance limit the implementation of the widespread and dense sensor networks necessary to overcome this issue.

In response, researchers are actively working on the development of video-based monitoring methods that do not require the installation of dense sensor arrays. The general concept is to leverage concepts from computer vision to quantify detected motion in a video and then relate per-pixel motions to infrastructure integrity. There are now a suite of computer vision methods and several commercially available monitoring systems that employ these technologies [3], [4]. While computer vision methods have distinct advantages over installed sensors, they have several key downsides [5]. Rather than as a full replacement for installed sensor systems, it is perhaps more reasonable to consider computer vision methods as a complementary technology. The objective of this study was to develop and implement a procedure for fusing video-based deformation measurements with those from an installed sensor array in order to capture the best from both sensing modalities.

METHODOLOGY

Computer vision methodology
Among different vision-based displacement methods, dense optical flow and phase-based flow algorithms are considered to be the most effective techniques [3], [6]. Each of these techniques measures displacements based on a different unique aspect of the recorded videos, and one can assume that each technique acts as an independent measurement of a 2D pixel field. This led to one of the key findings of this research program: vision methods can be combined into an ensemble measurement that is superior to any single video analysis method in isolation. This ensemble approach was evaluated for combinations of the dense flow and phase-based displacement methods.

The fusion of the two video measurement methods was achieved through a deep neural network [7]. One way to implement this concept is to "stack" 1D signal waveforms from different sensors to create a 2D signal, then use this 2D signal as input for a convolutional neural network (CNN) [8], [9]. An alternative is to use a generative adversarial network (GAN), exploiting sensor data as the training set for the generative network, and a ground truth signal as input of the discriminator network [10]. Both of these related deep learning approaches were evaluated for use in ensemble video analysis.

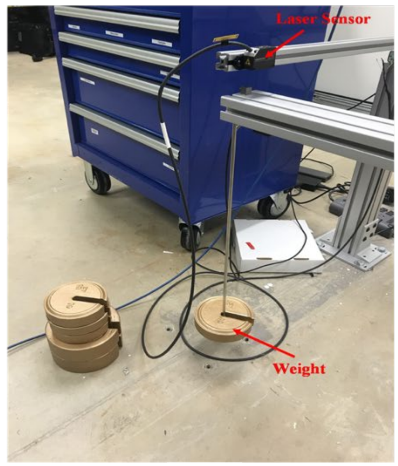
Image-sensor fusion methodology
The second aspect of the project involved creating a process for fusing image measurements with sensor measurements for enhanced displacement measurement. While data fusion is an active and well-defined research domain [11], previous studies in video fusion used a "higher-level" data fusion that produces decision support information rather than improved data fidelity, as was the goal here [12].

For this work, the goal was to create a Kalman filter designed to combine image measurements from a region of interest in a video and combine them with accelerometer measurements. Accelerometers are well-established as providing highly accurate measurement of acceleration response. This led to the design of a multi-rate Kalman filter, in order to accommodate discrepancies in sample rates between the accelerometer and the video recordings [13].

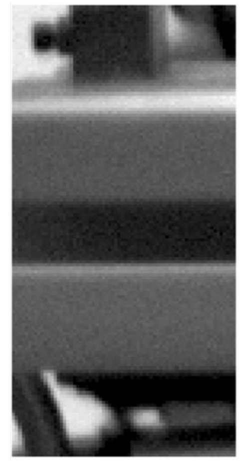
DATA SUMMARY

To generate the dataset for this research project, a series of experiments were performed at George Mason University. Structural aluminum was used to make a free vibrating cantilever beam. This structural beam configuration was chosen because the static and dynamic responses of cantilever systems are well understood, providing a reasonable basis for experimental analysis and comparison. Simple loadings were applied as lumped mass at the tip of the beam. These loadings were varied to induce various degrees of flexure in the beam (Figure 1).

(a)



(b)



(c)

Figure 1. Experimental setup: (a) camera placement, (b) loadings and displacement sensor, and (c) region of interest for video analysis.

EVALUATION RESULTS

Ensemble learning analysis

The key metric for the accuracy of the ensemble measurement methods was training and testing loss. This loss is a representation of model measurement accuracy. The losses for the individual vision-based displacement measurement techniques were 0.40 mm and 0.74 mm for the dense flow and phase-based methods, respectively. The loss for the CNN was 0.22 mm, and the loss for the GAN fluctuated between 0.49 mm and 0.55 mm (Figure 2). Overall, the CNN approach yielded a significant improvement over either of the computer-vision methods in isolation, as well as over the GAN, and reduced measurement errors by 0.18 mm.

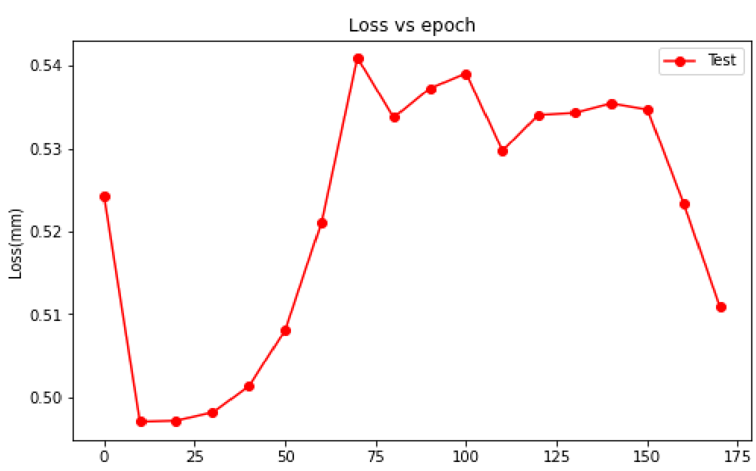


Figure 2. Test loss for GAN ensemble measurements.

Sensor-video fusion analysis

The results of the data fusion study showed that the Kalman filter successfully corrected for inaccuracies in phase-based computer vision measurements (Figure 3). However, a further analysis of the quasi-static tests indicated that the data fusion actually reduced measurement accuracy immediately after a load increment was applied. This effect was most noticeable for the increment that was applied at about 5.5 seconds into the test. In the context of a Kalman filter, these rapid loadings created regions of high nonlinearity in the response signals, a well-known challenge with Kalman filters [14]. An analysis of the dynamic test did not indicate the same sort of distortion as was observed for the quasi-static tests. This further reinforces the idea that the issue was largely due to the nonlinear change in system response during static testing and the Kalman filter's overcompensation for this change. For the dynamic test, both the dense flow and Kalman filtered approaches yielded a slight phase lag and underestimation of system response.

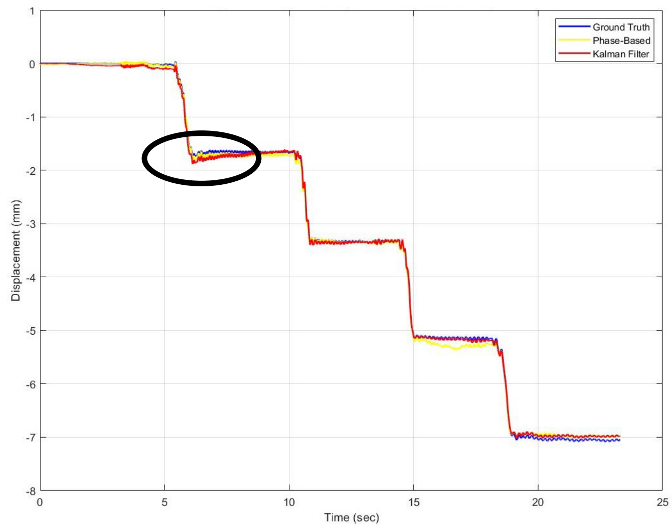


Figure 3. Quasi-static test results for fusing phase-based video measurements with accelerometer data. *The circled region illustrates Kalman filter distortions.*

CONCLUSION AND IMPLEMENTATION

This project investigated how to combine video measurements with sensor data. A significant and unanticipated contribution of this work was a new process for combining distinct computer-vision measurements together into an enhanced displacement measurement. This ensemble approach to video analysis uses a deep convolutional neural network to combine the video measurements together, and reduced measurement errors by approximately 50%. For sensor-video fusion, a multi-rate Kalman filter was designed to provide data fusion between accelerometer measurements and video measurements. The resulting data fusion improved measurement accuracy, but it also resulted in minor signal distortions for highly nonlinear regions in the test structure's dynamic response. Overall, the project proved the feasibility of data fusion in enhancing computer vision measurements. While the desired image-sensor fusion was achieved, it is the ensemble approach to video analysis that shows the most potential for future work and development. However, the deep learning approach implemented for ensemble learning should be investigated across a larger range of structural systems and application scenarios. This will be essential for evaluating the generalizability and practical utility of the method. And, while a Kalman filter was able to provide data fusion, the resulting fused signal was not ideal. Additional research is necessary in order to reduce these distortions, likely through investigation of more sophisticated nonlinear Kalman filter methods.

References

[1] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, “Modal identification of simple structures with high-speed video using motion magnification,” *J. Sound Vib.*, vol. 345, pp. 58–71, Jun. 2015, doi: 10.1016/j.jsv.2015.01.024.

[2] P. Poozesh, A. Sarrafi, Z. Mao, P. Avitabile, and C. Niezrecki, “Feasibility of extracting operating shapes using phase-based motion magnification technique and stereo-photogrammetry,” *J. Sound Vib.*, vol. 407, pp. 350–366, Oct. 2017, doi: 10.1016/j.jsv.2017.06.003.

[3] J. G. Chen, A. Davis, N. Wadhwa, F. Durand, W. T. Freeman, and O. Büyüköztürk, “Video Camera-Based Vibration Measurement for Civil Infrastructure Applications,” *J. Infrastruct. Syst.*, vol. 23, no. 3, p. B4016013, Sep. 2017, doi: 10.1061/(ASCE)IS.1943-555X.0000348.

[4] A. Khaloo and D. Lattanzi, “Pixel-wise structural motion tracking from rectified repurposed videos: Pixel-wise structural motion tracking from rectified repurposed videos,” *Struct. Control Health Monit.*, vol. 24, no. 11, p. e2009, Nov. 2017, doi: 10.1002/stc.2009.

[5] D. Sun, S. Roth, and M. J. Black, “Secrets of optical flow estimation and their principles,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 2432–2439. doi: 10.1109/CVPR.2010.5539939.

[6] T. Gautama and M. A. Van Hulle, “A phase-based approach to the estimation of the optical flow field using spatial filtering,” *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1127–1136, Sep. 2002, doi: 10.1109/TNN.2002.1031944.

[7] J. Gao, P. Li, Z. Chen, and J. Zhang, “A Survey on Deep Learning for Multimodal Data Fusion,” *Neural Comput.*, vol. 32, no. 5, pp. 829–864, May 2020, doi: 10.1162/neco_a_01273.

[8] G. Fan, J. Li, and H. Hao, “Dynamic response reconstruction for structural health monitoring using densely connected convolutional networks,” *Struct. Health Monit.*, vol. 20, no. 4, pp. 1373–1391, Jul. 2021, doi: 10.1177/1475921720916881.

[9] B. K. Oh and J. Kim, “Optimal architecture of a convolutional neural network to estimate structural responses for safety evaluation of the structures,” *Measurement*, vol. 177, p. 109313, Jun. 2021, doi: 10.1016/j.measurement.2021.109313.

[10] G. Fan, J. Li, and H. Hao, “Lost data recovery for structural health monitoring based on convolutional neural networks,” *Struct. Control Health Monit.*, vol. 26, no. 10, p. e2433, 2019, doi: 10.1002/stc.2433.

[11] R.-T. Wu and M. R. Jahanshahi, “Data fusion approaches for structural health monitoring and system identification: Past, present, and future,” *Struct. Health Monit.*, vol. 19, no. 2, pp. 552–586, Mar. 2020, doi: 10.1177/1475921718798769.

[12] T. Liu, S. Du, C. Liang, B. Zhang, and R. Feng, “A Novel Multi-Sensor Fusion Based Object Detection and Recognition Algorithm for Intelligent Assisted Driving,” *IEEE Access*, vol. 9, pp. 81564–81574, 2021, doi: 10.1109/ACCESS.2021.3083503.

[13] A. Smyth and M. Wu, “Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring,” *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 706–723, Feb. 2007, doi: 10.1016/j.ymssp.2006.03.005.

[14] F. Daum, “Nonlinear filters: Beyond the Kalman filter,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, pp. 57–69, Aug. 2005, doi: 10.1109/MAES.2005.1499276.

For More Information Principal Investigator: David Lattanzi Technical reports when published are available at **http://________________________________**	4400 University Drive, MS 6C1 Fairfax, VA 22030 Phone: 703.993.3695 E-mail: dlattanz@gmu.edu