CIAMTIS

U.S. DOT Region 3 University Transportation Center

Enter Video-Sensor Data Fusion for Enhanced Structural Monitoring

April 28, 2022

Prepared by:

D. Lattanzi and M. Ghyabi,
George Mason University

r3utc.psu.edu

PennState College of Engineering

LARSON TRANSPORTATION INSTITUTE

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
CIAM-COR-R25		

4. Title and Subtitle	5. Report Date
Video-Sensor Data Fusion for Enhanced Structural Monitoring	April 28, 2022
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
D. Lattanzi https://orcid.org/0000-0001-9247-0680 and M. Ghyabi https://orcid.org/0000-0003-2652-426X	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
4614 Nguyen Engineering Building 4 400 University Drive, MS 6C1 Fairfax, VA 22030	**11. Contract or Grant No.** 69A3551847103

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
U.S. Department of Transportation Research and Innovative Technology Administration 3rd Fl, East Bldg E33-461 1200 New Jersey Ave, SE Washington, DC 20590	Draft Final Report xx/xx/20xx – xx/xx/20xx
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

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 a priori [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.

17. Key Words	18. Distribution Statement
Sensor, video based monitoring, data fusion	No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	12	

Form DOT F 1700.7 **(8-72) Reproduction of completed page authorized**

[Add Table of Contents, List of Figures, List of Tables]

CHAPTER 1

Introduction

BACKGROUND

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 a priori [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 system dynamics through a series of dimensional scaling transforms. The result is the ability to measure displacement as a 2D or 3D *field*, rather than the 1D measurements that most sensors produce. There are now a suite of candidate 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. The most notable is that image-based measurements are noisy and highly uncertain when compared to traditional sensor arrays [4]. They can also be sensitive to small changes in signal processing [5]. Rather than as a full replacement for installed sensor systems, it is perhaps more reasonable to consider computer vision methods as a complimentary technology. This project explored how to combine sensor and video measurements together to overcome the limitations of each measurement modality and improve the accuracy and certainty of structural health monitoring systems.

OBJECTIVES

The objective of this research program was to develop and implement a procedure for fusing video-based measurements with those from an installed sensor array. While data fusion is an active and well-defined research domain [6], 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 [7].

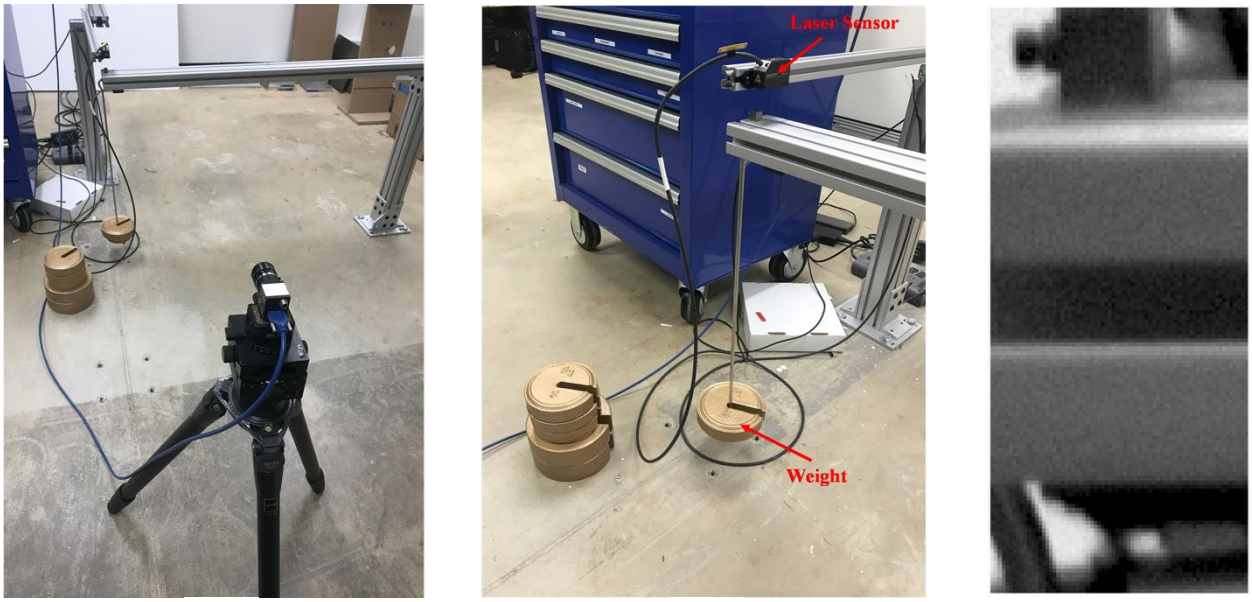
Specific sub-objectives of this project included:

1. Identification and refinement of a computer vision method for video-based infrastructure monitoring and data fusion
2. Creation of a data fusion algorithm that combines video measurements and sensor data
3. Experimental evaluation of all algorithms

DATA AND DATA STRUCTURES

Dataset Generation

To generate the dataset for this research project, a series of experiments were performed in the Advanced Infrastructure Monitoring Lab 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

An Edmund Optics EO-2323 industrial camera was used to record the videos. This is a monochrome machine vision sensor, with a sensor size of 9.22x5.76 mm (1920x1200 pixels). To achieve higher data sampling rates, only a small region of interest (ROI) containing the tip of the cantilever beam was recorded. By recording this 444x160 pixel area (Figure 1c), and setting the pixel clock and exposure time to 200 MHz and 1.19 ms respectively, a frame rate of 1000 fps was achieved. A pixel depth of 8 bits was used. An accelerometer was installed on the cantilever tip as well. This accelerometer was a PCB brand accelerometer, sampled at 50 Hz.

To record the ground truth displacement data, a Micro-Epsilon optoNCDT 1320-10 laser sensor was installed above the cantilever beam. The precision of this device is 10 µm, and the measuring range of this model is 10 mm, sufficient for the range of displacement of the cantilever beam in this experimental setup. This laser triangulation displacement meter is able to record data with acquisition frequency up to 2 kHz. However, to improve synchronization with the video recording, the data sampling frequency was set to 1 kHz.

Since the aluminum beam had negligible mass in comparison to the weights used for loading, the behavior of the system could be simplified to that of a lumped mass-spring-damper model (single degree of freedom). The benefit of this simplification for dataset generation was that it simplified both system dynamic modeling and also made it easier to quickly change the parameters of the dynamical system under observation. This change was achieved by varying the lump mass from 0 kg to 8 kg in 2 kg increments. At each step, 10 dynamic tests were recorded as damped free vibrations with an initial displacement. A series of quasi-static recordings of incremental loadings were recorded as well. A total of 48 videos, along with the laser gauge readings, were used for the final data set. The first 27 seconds of each test recording were synchronized against the laser ground truth. A total of 48x27(s)x1000(fps)=1,296,000 frames were used to build the dataset. Each 27 second signal was further subdivided into segments of 0.338s in length, with 0.238 overlap, resulting in 12,957 signal samples for the complete dataset. This subdivision provided sufficient data for machine learning algorithm development, as will be discussed.

RESEARCH PRODUCTS

Data and programming scripts

This project resulted in a variety of data types, as delineated in the Center's data management plan. The majority of these files are videos and sensor data of collected during experimental testing. Experimental sensor data as described was stored in comma delimited text files and standard video file formats for dissemination and data transfer. This data set also includes the processed and subdivided signal recordings. Analytical and numerical code was written in both the MATLAB and Python programming languages, and is stored as scripting files. All data will be deposited in the Center's data repository within 30 days after submission of this report. The data will also be transferred to Penn State for storage on their cloud-based storage box.psu.edu. The research team did not collect any personally-identifiable information (PII), confidential business information, or national security information as a result of this work.

Other research products

The research project also resulted in several other products. The work is associated with a conference publication and one additional conference presentation. The work has also resulted in a journal manuscript, to be submitted for publication in the summer of 2022. Two graduate students were supported by this research effort.

CHAPTER 2

Methodology

INTRODUCTION

The overall goal of this project was to develop a process for fusing video data with sensor array data, in order to improve measurement of structural deformations. Achieving this required the development of a computer vision methodology for quantifying measurements from images, as well as a methodology for fusing images and sensor data together. Both aspects of the research program are presented here.

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], [8]. The dense flow technique measures displacements by solving an optimization problem based on pixel intensities comparisons across frames of video. The phase-based algorithm finds displacements by measuring the phase of the video signal in the proper spatial frequencies. Since each of these techniques measures displacements based on a different unique aspect of the recorded videos, 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. Feature-tracking, which also a viable measurement approach, was not considered due to the nature of the experimental test setup and the cantilever region of interest.

The dense flow and phase-based methods used in this work are well-established techniques that have been shown effective in prior work. The PI's team also performed an exhaustive series of evaluations through other UTC supported projects. These evaluations showed that the dense flow method presented in [9] and the phase-based method of [8] were the most consistently accurate measurement approaches. For brevity and clarity, the details of these computer vision methods are not delineated here. Recorded videos were processed in Matlab for both computer vision methods. An example output, and comparison against the ground-truth laser scanner, is shown in Figure 2.

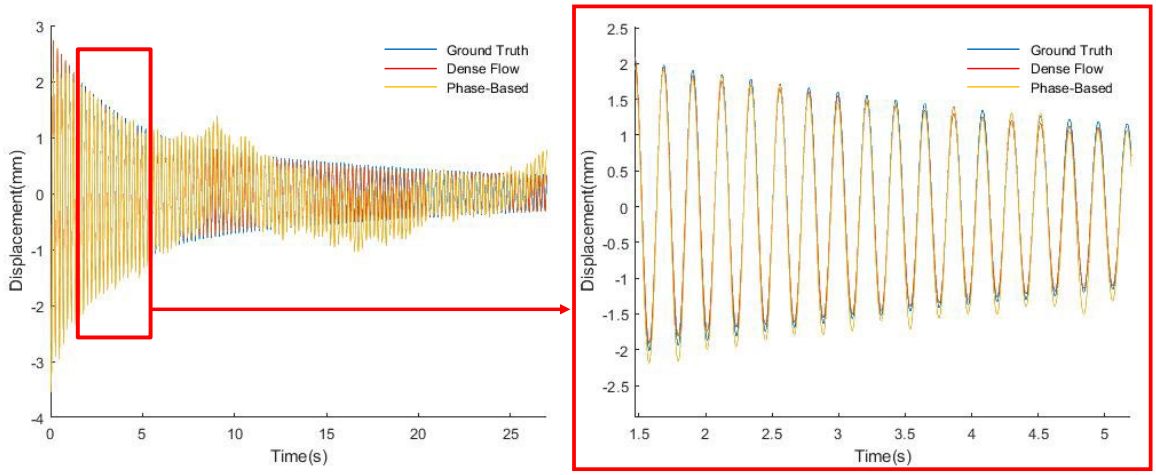


Figure 2- Sample signals over a 27 second interval, and over a 3.5 second interval

Ensemble of video methods

In recent years, deep neural networks have increasingly been used as reliable tools for ensemble analysis and data fusion [10]. 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) [11], [12]. 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 [13]. Both of these related deep learning approaches were evaluated for use in ensemble video analysis. For each network, 11,000 signal samples from the experimental tests were used for machine learning training and 1,957 were used for testing.

The CNN network was designed based on the previous research in [10]. This is a low-complexity network with 2 convolutional layers and one fully connected layer. Layer size, kernel size, and stride of the convolutional layers were set to [10,10], [5,1], [1,1] respectively, based on a previous parameter optimization study [11]. The fully connected layer size was set to 360, and the output layer has the size of 338, equal to the length of signal segments. Minibatch stochastic gradient descent (SGD) was selected as the optimizer. The learning rate was set to 0.1 and the mini batch size was set to 1000. Root mean square error (RMSE), was selected as the loss function. Train loss and test loss were recorded every 10 epochs.

The GAN network architecture required the design of both generator and discriminator networks. The input of the generator is a 26x26 signal created by combining the 1D waveforms of the two image analysis methods. Each signal segments from the dense flow and phase-based techniques were reshaped into 13x26 images and vertically stacked to form 26x26 pixel images (Figure 3). The input of the discriminator is a 1x338 signal taken directly from the laser gauge used for ground truth. The discriminator consists of three hidden fully connected layers. A least square approach was used for generator and discriminator losses.

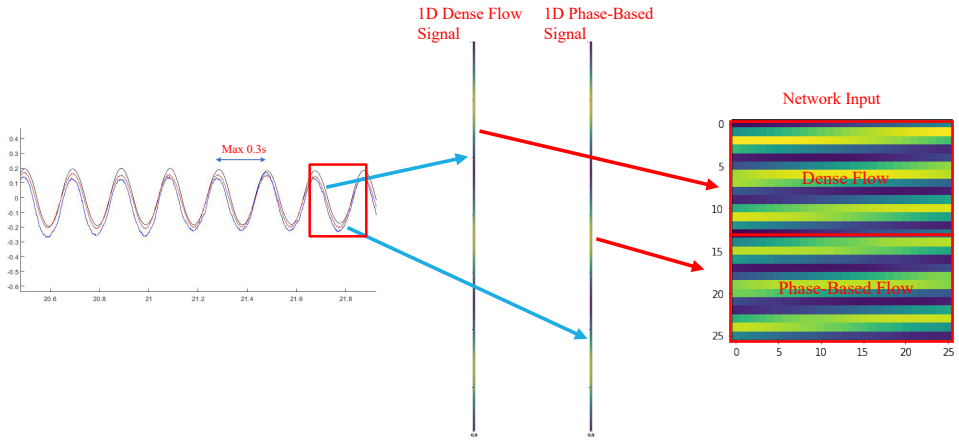


Figure 2- Segmentation and decomposition of the dataset for GAN development

IMAGE-SENSOR FUSION

The second aspect of the project involved creating a process for fusing image measurements with sensor measurements for enhanced displacement measurement. The Kalman filter is a widely used approach for data fusion, and was the focus of this project [14], [15]. In particular, prior research considered the concept of a Kalman filter for related work in combining velocity and laser displacement measurements [16], and for combining vision-based displacement measurements with accelerometer data [17]. Similar methods were also demonstrated in the context of monitoring a short-span railway bridge [18].

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. Conceptually, double integration of accelerometer measurements should provide accurate displacement responses, however this approach is known to suffer from low-frequency signal drift [19]. Hypothetically, this error could be compensated for through fusion with video measurement data.

This led to the design of a multi-rate Kalman filter, in order to accomodate discrepancies in sample rates between the accelerometer and the video recordings. For brevity and clarity, the details of the Kalman filter are not provided here. The reader is referred to [14] for more details on this process. The research team also investigated the potential for a smooth Kalman filter [20] designed to further reduce measurement noise. However, this process proved unsuccessful. While the resulting filter did smooth the measured response, it also resulted in low-frequency signal distortions and unacceptable results overall.

CHAPTER 3

Experimental 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 0.3985 mm and 0.7370 mm for the dense flow and phase-based methods, respectively. For the ensemble method, the loss was after iterations of neural network training and testing, referred to as epochs. After the first epoch, test losses for the CNN were 0.45 mm. After further iteration, the test loss declined to 0.22 mm (Figure 4). The behavior of the GAN was not as convergent as the CNN due to the additional complexity of the GAN learning approach. The test loss for the GAN fluctuated between 0.49 mm and 0.55mm, suggesting that the CNN ensemble method is the superior approach to combining computer vision measurements (Figure 5). Overall, the CNN approach yielded a significant improvement over either of the computer-vision methods in isolation, and reduced measurement errors by 0.18 mm.

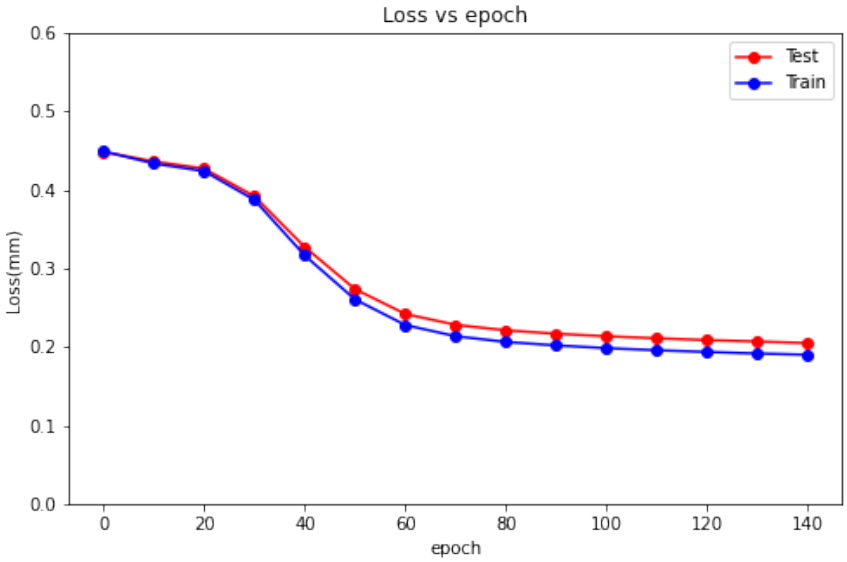


Figure 3- Test and train losses for CNN ensemble measurements

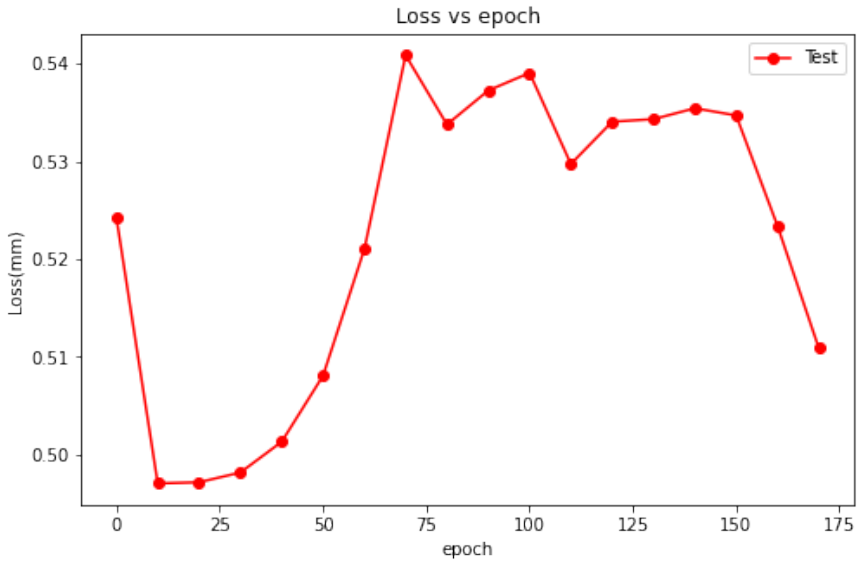


Figure 4- Test loss for GAN ensemble measurements

Sensor-video fusion analysis

The results of the quasi-static data fusion tests are shown in Figure 6. The results from a dynamic response test are shown in Figure 7. The results show that the Kalman filter successfully corrected for inaccuracies in phase-based computer vision measurements. 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 [21].

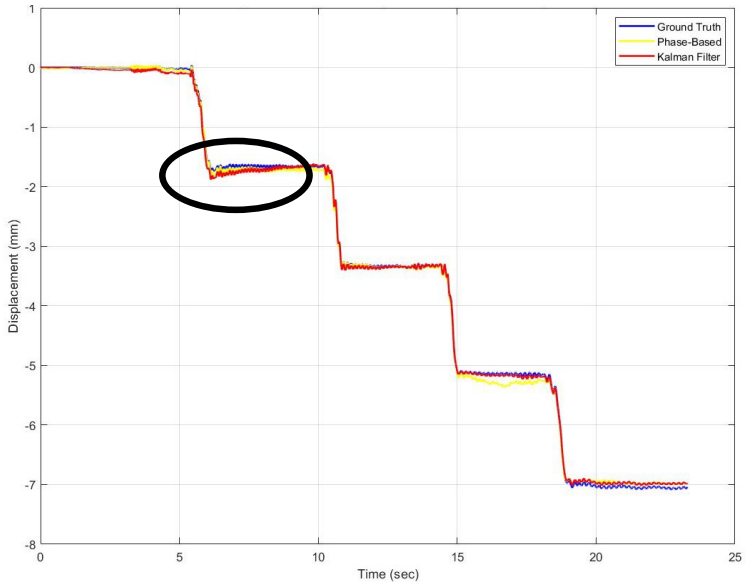


Figure 6- Quasi-static test results for fusing phase based video measurements with accelerometer data. Circled region illustrates Kalman filter distortions.

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.

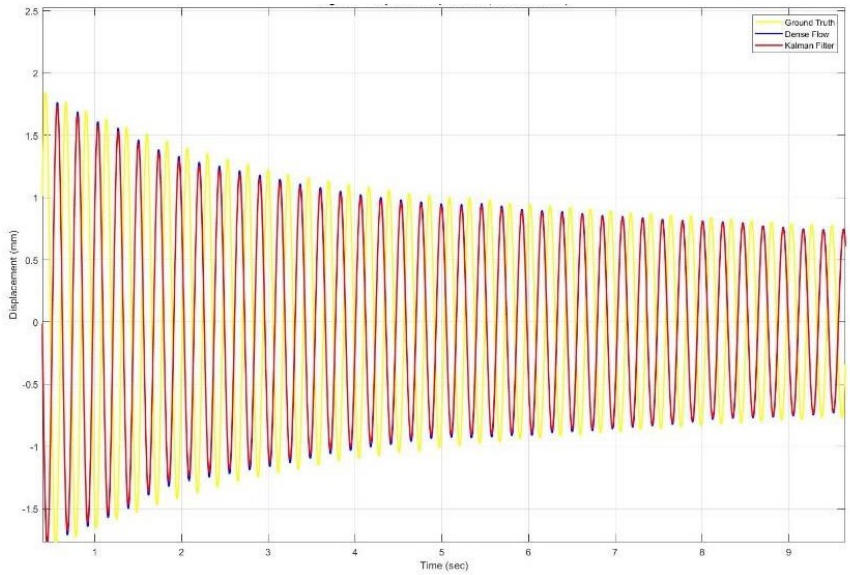


Figure 7- Dynamic test results for fusing phase based video measurements with accelerometer data

CHAPTER 4

Recommendations

Summary and Conclusion

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 errors by approximately 50%. A similar approach using a GAN network proved less successful for ensemble measurement. 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.

AVENUES FOR FUTURE WORK

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," *Journal of Sound and Vibration*, 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," *Journal of Sound and Vibration*, 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," *Journal of Infrastructure Systems*, 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]R.-T. Wu and M. R. Jahanshahi, "Data fusion approaches for structural health monitoring and system identification: Past, present, and future," *Structural Health Monitoring*, vol. 19, no. 2, pp. 552–586, Mar. 2020, doi: 10.1177/1475921718798769.
[7]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.
[8]T. Gautama and M. A. Van Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1127–1136, Sep. 2002, doi: 10.1109/TNN.2002.1031944.
[9]G. Farnebäck, "Two-Frame Motion Estimation Based on Polynomial Expansion," in *Image Analysis*, Berlin, Heidelberg, 2003, pp. 363–370. doi: 10.1007/3-540-45103-X_50.
[10]J. Gao, P. Li, Z. Chen, and J. Zhang, "A Survey on Deep Learning for Multimodal Data Fusion," *Neural Computation*, vol. 32, no. 5, pp. 829–864, May 2020, doi: 10.1162/neco_a_01273.
[11]G. Fan, J. Li, and H. Hao, "Dynamic response reconstruction for structural health monitoring using densely connected convolutional networks," *Structural Health Monitoring*, vol. 20, no. 4, pp. 1373–1391, Jul. 2021, doi: 10.1177/1475921720916881.
[12]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.
[13]G. Fan, J. Li, and H. Hao, "Lost data recovery for structural health monitoring based on convolutional neural networks," *Structural Control and Health Monitoring*, vol. 26, no. 10, p. e2433, 2019, doi: 10.1002/stc.2433.
[14]A. Smyth and M. Wu, "Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 706–723, Feb. 2007, doi: 10.1016/j.ymssp.2006.03.005.
[15]S. Cho, J.-W. Park, R. P. Palanisamy, and S.-H. Sim, "Reference-Free Displacement Estimation of Bridges Using Kalman Filter-Based Multimetric Data Fusion," *Journal of Sensors*, vol. 2016, p. e3791856, Sep. 2016, doi: 10.1155/2016/3791856.

[16]K. Kim and H. Sohn, “Dynamic displacement estimation by fusing LDV and LiDAR measurements via smoothing based Kalman filtering,” *Mechanical Systems and Signal Processing*, vol. 82, pp. 339–355, Jan. 2017, doi: 10.1016/j.ymssp.2016.05.027.
[17]J.-W. Park, D.-S. Moon, H. Yoon, F. Gomez, B. F. Spencer Jr., and J. R. Kim, “Visual–inertial displacement sensing using data fusion of vision-based displacement with acceleration,” *Structural Control and Health Monitoring*, vol. 25, no. 3, p. e2122, 2018, doi: 10.1002/stc.2122.
[18]Y. Xu, J. M. W. Brownjohn, and F. Huseynov, “Accurate Deformation Monitoring on Bridge Structures Using a Cost-Effective Sensing System Combined with a Camera and Accelerometers: Case Study,” *Journal of Bridge Engineering*, vol. 24, no. 1, p. 05018014, Jan. 2019, doi: 10.1061/(ASCE)BE.1943-5592.0001330.
[19]G. W. Roberts, X. Meng, and A. H. Dodson, “Integrating a Global Positioning System and Accelerometers to Monitor the Deflection of Bridges,” *Journal of Surveying Engineering*, vol. 130, no. 2, pp. 65–72, May 2004, doi: 10.1061/(ASCE)0733-9453(2004)130:2(65).
[20]S. A. Gadsden, S. Habibi, and T. Kirubarajan, “Kalman and smooth variable structure filters for robust estimation,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1038–1050, Apr. 2014, doi: 10.1109/TAES.2014.110768.
[21]F. Daum, “Nonlinear filters: beyond the Kalman filter,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pp. 57–69, Aug. 2005, doi: 10.1109/MAES.2005.1499276.