

2019

Investigation of Computer Vision Concepts and Methods for Structural Health Monitoring and Identification Applications

Chuanzhi Dong
University of Central Florida

 Part of the [Civil Engineering Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Dong, Chuanzhi, "Investigation of Computer Vision Concepts and Methods for Structural Health Monitoring and Identification Applications" (2019). *Electronic Theses and Dissertations*. 6867.
<https://stars.library.ucf.edu/etd/6867>

**INVESTIGATION OF COMPUTER VISION CONCEPTS AND METHODS
FOR STRUCTURAL HEALTH MONITORING AND IDENTIFICATION
APPLICATIONS**

by

CHUANZHI DONG

B.Sc. Zhejiang Gongshang University, 2011

M.Sc. Zhejiang University, 2016

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term

2019

Major Professor: F. Necati Çatbaş

© 2019 Chuanzhi Dong

ABSTRACT

This study presents a comprehensive investigation of methods and technologies for developing a computer vision-based framework for Structural Health Monitoring (SHM) and Structural Identification (St-Id) for civil infrastructure systems, with particular emphasis on various types of bridges. SHM is implemented on various structures over the last two decades, yet, there are some issues such as considerable cost, field implementation time and excessive labor needs for the instrumentation of sensors, cable wiring work and possible interruptions during implementation. These issues make it only viable when major investments for SHM are warranted for decision making. For other cases, there needs to be a practical and effective solution, which computer-vision based framework can be a viable alternative. Computer vision based SHM has been explored over the last decade. Unlike most of the vision-based structural identification studies and practices, which focus either on structural input (vehicle location) estimation or on structural output (structural displacement and strain responses) estimation, the proposed framework combines the vision-based structural input and the structural output from non-contact sensors to overcome the limitations given above. First, this study develops a series of computer vision-based displacement measurement methods for structural response (structural output) monitoring which can be applied to different infrastructures such as grandstands, stadiums, towers, footbridges, small/medium span concrete bridges, railway bridges, and long span bridges, and under different loading cases such as human crowd, pedestrians, wind, vehicle, etc. Structural behavior, modal properties, load carrying capacities, structural serviceability and performance are investigated using vision-based methods and validated by comparing with conventional SHM approaches. In this study, some of the most famous landmark structures such as long span bridges are utilized as case studies. This study also investigated the serviceability status of structures by using computer

vision-based methods. Subsequently, issues and considerations for computer vision-based measurement in field application are discussed and recommendations are provided for better results. This study also proposes a robust vision-based method for displacement measurement using spatio-temporal context learning and Taylor approximation to overcome the difficulties of vision-based monitoring under adverse environmental factors such as fog and illumination change. In addition, it is shown that the external load distribution on structures (structural input) can be estimated by using visual tracking, and afterward load rating of a bridge can be determined by using the load distribution factors extracted from computer vision-based methods. By combining the structural input and output results, the unit influence line (UIL) of structures are extracted during daily traffic just using cameras from which the external loads can be estimated by using just cameras and extracted UIL. Finally, the condition assessment at global structural level can be achieved using the structural input and output, both obtained from computer vision approaches, would give a normalized response irrespective of the type and/or load configurations of the vehicles or human loads.

To my parents and my wife

ACKNOWLEDGMENTS

I would first sincerely express my deepest appreciation and gratitude to my advisor and my mentor, Dr. F. Necati Çatbaş. His support and guidance contributed greatly to the success of this work. What I learned from him during my Ph.D. period is not only the special knowledge about structural engineering, but also spirits of seriousness, honesty, modesty, and confidence as a dependent researcher and an educator.

I would like to thank my committee members, Dr. Kevin R. Mackie, Dr. Georgios Apostolakis, and Dr. Ulas Bagci for serving in my committee and for providing me with excellent course instructions and guidance.

I would like to express appreciation to two of my collaborators and mentors: Dr. Eugene O'Brien from Ireland and Dr. Su Taylor from UK for their help and feedback during my Ph.D.

I would like to thank all the research team members of Civil Infrastructure Technologies for Resilience and Safety (CITRS), and specially thank Dr. Ozan Celik, Dr. Selcuk Bas, Dr. Ninel Alver, Dr. Recep Birgul, Kailea Moseley, Wesley Shattenkirk and Analiese Majetich.

I also would like to acknowledge the financial support for this research that is provided by NSF Division of Civil, Mechanical and Manufacturing Innovation [grant number 1463493].

I would like to thank my parents, my sister and my brother for their support during my Ph.D. life.

Last but most certainly not the least, I would like to express my deepest love to my wife, Ru, for her understanding and support to my whole graduate study. Without her encouragement and patience, I cannot overcome the difficulties in finishing my Ph.D.

TABLE OF CONTENTS

LIST OF FIGURES	xv
LIST OF TABLES	xxiv
CHAPTER ONE: INTRODUCTION	1
1.1 Background	1
1.2 Organization of Thesis	2
CHAPTER TWO: REVIEW OF COMPUTER VISION-BASED STRUCTURAL HEALTH MONITORING	4
2.1 Introduction	4
2.2 Projective Geometry Applied in CV-SHM	8
2.3 Computer Vision-based Structural Health Monitoring at Local Level (CV-SHM-LL)	13
2.3.1 Crack Detection of Concrete Structures	13
2.3.2 Crack Detection of Concrete Structures	18
2.3.3 Spalling Detection of Concrete Structures	20
2.3.4 Delamination Detection	21
2.3.5 Crack Detection of Steel Structures	23
2.3.6 Crack Propagation Monitoring	24
2.3.7 Rust Detection of Steel Structures	25
2.3.8 Bolt Loose Detection of Steel Structures	27
2.4 Computer Vision-based Structural Health Monitoring at Global Level (CV-SHM-GL) ...	28

2.4.1 CV-based Structural Response (Output) Monitoring	28
2.4.2 CV-based Load (Input) Estimation.....	45
2.4.3 SHM Using CV-based Input-Output Information	48
2.5 Challenge and Concerns in Real-life Practices of CV-SHM.....	48
2.5.1 Challenges.....	49
2.5.2 Concerns	51
2.6 Summary	52
CHAPTER THREE: MARKER FREE MONITORING OF THE GRANDSTAND STRUCTURES AND MODAL IDENTIFICATION USING COMPUTER VISION METHODS	54
3.1 Introduction.....	54
3.2 Methodology and System Development.....	57
3.2.1 General Procedure for Vision-based Displacement Measurement System	57
3.2.2 Camera Calibration	57
3.2.3 Feature extraction.....	57
3.2.4 Visual tracking using optical flow	60
3.2.5 Displacement calculation	66
3.3 Laboratory verification	67
3.3.1 Experimental Setup.....	67
3.3.2 Comparative Study of Displacement Measurement Using Different Features.....	69

3.3.3 Multi-point Displacement Measurement Using Multi-camera	71
3.3.4 Structural Identification	74
3.4 Field Application	79
3.4.1 Experimental Setup	79
3.4.2 Analysis and Results	81
3.5 Summary	83
CHAPTER FOUR: A NON-TARGET STRUCTURAL DISPLACEMENT MEASUREMENT METHOD USING ADVANCED FEATURE MATCHING STRATEGY	86
4.1 Introduction	86
4.2 Methodology	90
4.2.1 General Procedure for Vision-based Displacement Measurement Method Using Feature Matching	90
4.2.2 Image Data Acquisition	91
4.2.3 Feature Detection and Description	91
4.2.4 Feature Matching and Outlier Removal	93
4.2.5 Structural Displacement Calculation	95
4.3 Laboratory Verification	96
4.3.1 Experimental Setup	96
4.3.2 Comparative Study of Displacement Measurement Using Different Methods	97
4.4 Field Application	102

4.4.1 Experimental Setup.....	102
4.4.2 Analysis and Results	103
4.5 Summary	105
CHAPTER FIVE: STRUCTURAL DISPLACEMENT MONITORING USING DEEP LEARNING-BASED FULL FIELD OPTICAL FLOW METHODS.....	107
5.1 Introduction.....	107
5.2 Methodology and System Development.....	111
5.2.1 General Procedure for Vision-Based Displacement Measurement System.....	111
5.2.2 Image Data Acquisition and Considerations.....	113
5.2.3 Full Field Optical Flow Estimation Using Deep Learning Methods	117
5.2.4 Camera Motion Subtraction.....	122
5.2.5 Structural Displacement Calculation	124
5.3 Laboratory Verification	131
5.3.1 Experimental Setup.....	131
5.3.2 Comparative Study of Displacement Measurement Using Different Methods	132
5.3.3 Comparison of Computation Times of Vision-Based Methods Using Different Full Field Optical Flow Algorithms	140
5.4 Field Application	141
5.4.1 Experimental Setup.....	141
5.4.2 Analysis and results	142

5.4.3 Recommendations for Practice in Field Application	146
5.5 Summary	147
CHAPTER SIX: A ROBUST VISION-BASED METHOD FOR DISPLACEMENT MEASUREMENT UNDER ADVERSE ENVIRONMENTAL FACTORS USING SPATIO- TEMPORAL CONTEXT LEARNING AND TAYLOR APPROXIMATION.....	149
6.1 Introduction.....	149
6.2 Methodology	153
6.2.1 General Procedure for The Proposed Displacement Measurement System	153
6.2.2 Visual Tracking Using Spatio-Temporal Context (STC) Learning.....	154
6.2.3 Subpixel Level Estimation Using Taylor Approximation	160
6.3 Laboratory Verification	163
6.3.1 Experimental Setup.....	163
6.3.2 Results Analysis and Comparative Study of Case 1 (Ideal Case)	167
6.3.3 Results Analysis and Comparative Study of Case 2 (Illumination Change)	174
6.3.4 Results Analysis and Comparative Study of Case 3 (Fog Interference).....	177
6.4 Summary	181
CHAPTER SEVEN: DISPLACEMENT MONITORING OF LONG SPAN BRIDGES USING VISION-BASED METHODS	185
7.1 Introduction.....	185
7.2 Methodology	185

7.3 Experiment on the First Bosphorus Bridge.....	186
7.4 Experiment on the Second Bosphorus Bridge	188
7.5 Experiments on the Osman Gazi Bridge.....	190
7.5.1 The First Experiment	190
7.5.2 The Second Experiment.....	193
7.6 Summary	197
CHAPTER EIGHT: INVESTIGATION OF VIBRATION SERVICEABILITY OF A FOOTBRIDGE USING COMPUTER VISION-BASED METHODS.....	198
8.1 Introduction.....	198
8.2 Vibration Limit for Serviceability Assessment in Current Standards and Codes.....	201
8.3 Vision-based Displacement/Velocity Measurement Using Feature Matching	204
8.4 Experimental Verification and Field Application.....	209
8.4.1 Experimental setup.....	209
8.4.2 Result analysis	212
8.5 Summary	225
CHAPTER NINE: A PRACTICAL APPROACH FOR THE ESTIMATION OF BRIDGE DISTRIBUTION FACTOR AND LOAD RATING FACTOR	228
9.1 Introduction.....	228
9.2 Methodology	229
9.3 Application on a Real Bridge.....	231

9.3.1 General Features of the Bridge	231
9.3.2 Experimental Setup	232
9.3.3 Result analysis	236
9.4 Summary	246
CHAPTER TEN: A COMPLETELY NON-CONTACT RECOGNITION SYSTEM FOR	
BRIDGE UNIT INFLUENCE LINE USING PORTABLE CAMERAS AND COMPUTER	
VISION	248
10.1 Introduction	248
10.2 Methodology	252
10.2.1 UIL	252
10.2.2 Vision-based Structural Input Estimation: Vehicle Location	255
10.2.3 Vision-based Structural Output Estimation: Displacement Responses	259
10.3 System Configuration	262
10.4 Laboratory Demonstration	263
10.4.1 Experimental Setup	263
10.4.2 Result Analysis	267
10.5 Field Application	270
10.5.1 Experimental Setup	270
10.5.2 Result Analysis	272
10.6 Summary	274

CHAPTER ELEVEN: SUMMARY AND CONCLUSIONS	277
REFERENCES	281

LIST OF FIGURES

Figure 1 The projection from real world object to image	9
Figure 2 Examples of spalling and crack	20
Figure 3 General procedure of CV-based displacement measurement	29
Figure 4 Sliding the template across another image to find the best match	29
Figure 5 Visualized SIFT descriptor: image processed with the codes released by (VLFeat 2018)	34
Figure 6 Feature matching procedure	35
Figure 7 Examples of feature points: a) Shi-Tomasi corners; b) SURF	59
Figure 8 Motion in image sequence: a) temporal change; b) spatial change	60
Figure 9 Aperture problem	62
Figure 10 a) Visual tracking results; b) Illustration of bidirectional error	64
Figure 11 Grandstands in laboratory and stadium	68
Figure 12 Experimental setup	69
Figure 13 Comparison of displacement data from vision using different features and potentiometer at P2	70
Figure 14 Comparison of displacement data from vision using different cameras and potentiometer at P3	73
Figure 15 Displacement time histories from computer vision	75
Figure 16 Acceleration time histories from accelerometers	75
Figure 17 CMIF calculated by displacement time histories from vision method	75
Figure 18 CMIF calculated by acceleration time histories from accelerometers	76
Figure 19 The first three modal shapes of the front beam from vision method	77

Figure 20 The first three modal shapes of the front beam from accelerometers	78
Figure 21 The first three modal shapes of the full structure from accelerometers	78
Figure 22 Grandstand from a real structure	80
Figure 23 Time history of the chorus part of “Zombie Nation” song.....	80
Figure 24 Comparison of displacement data from vision and potentiometer	81
Figure 25 Acceleration time histories from the accelerometer	82
Figure 26 Comparison of operational frequencies under human jumping: a) vision; b) accelerometer	82
Figure 27 Periodogram power spectral density estimate of time history of the chorus part of Zombie Nation	82
Figure 28 Flowchart for proposed vision-based displacement measurement method using feature matching.....	91
Figure 29 An example of SIFT feature points (a) and descriptors (b).....	93
Figure 30 An example of feature matching and outlier removal: (a) SIFT feature matching; (b) SIFT outlier removal; (c) SIFT-VGG feature matching; (d) SIFT-VGG outlier removal.	95
Figure 31 Experimental setup	97
Figure 32 Comparison of displacement time histories from displacement sensor and vision-based methods	98
Figure 33 Correlation matrix of time displacement histories	99
Figure 34 Distributions of measurement error for different methods Experimental setup for a railway bridge	101
Figure 35 Experimental setup for a railway bridge	103
Figure 36 Displacement time history obtained from the proposed method.....	104

Figure 37 Flowchart for proposed full field structural displacement measurement method	112
Figure 38 Image projection using the planar homography matrix and four point correspondences	113
Figure 39 Image data acquisition and processing	116
Figure 40 Optical flow estimations using different methods.....	122
Figure 41 Structural displacement rectification using camera motion subtraction.....	123
Figure 42 Structural displacement at a discrete point using kernels.....	124
Figure 43 Two different strategies to process image sequences and get displacement time histories	128
Figure 44 Displacement results obtained from the two different image sequence processing strategies and the displacement sensor	130
Figure 45 Experimental setup of the grandstand monitoring.....	132
Figure 46 Comparison of displacement time histories from displacement sensor and vision-based methods	133
Figure 47 Statistical analysis of time spent on image collection for each frame.....	134
Figure 48 Resampled displacement time histories using displacement sensor and vision-based methods	135
Figure 49 Distributions of measurement error for different methods.....	137
Figure 50 Comparison of displacement data in the frequency domain	139
Figure 51 Experimental setup for a footbridge	142
Figure 52 Displacement time histories obtained from the proposed methods. FlowNet2+G org and FlowNet2+M org represent the original displacement data obtained using FlowNet2 with Gaussian	

kernel and Mean kernel, respectively. While FlowNet2+G w/ cam. mot. subtr and FlowNet2+M w/ cam. mot. subtr represent those with camera motion subtraction.....	143
Figure 53 Acceleration time history obtained from accelerometer	144
Figure 54 Operational frequencies of bridge under human jumping excitation: (a) vision-based system; (b) accelerometer signal.....	144
Figure 55 Graphical model of spatial relationship between the target and its surrounding context	155
Figure 56 Sketches of motion estimation using STC tracking and Taylor approximation.....	162
Figure 57 Flowchart of STC-based subpixel tracking using Taylor approximation	162
Figure 58 Experimental setup	164
Figure 59 Illumination change. (a) 34 lux, (b) 18 lux.....	166
Figure 60 Fog simulation	166
Figure 61 Vertical displacement time histories of P1 in pixel units using non-subpixel, image upsampling, and Taylor approximation techniques	167
Figure 62 A zoomed in section at the beginning of the vertical displacement time histories of P1	168
Figure 63 Horizontal displacement time histories of P1 in pixel units using non-subpixel, image upsampling, and Taylor approximation techniques	169
Figure 64 A zoomed in section of the horizontal displacement time histories of P1 in pixel units	169
Figure 65 Case 1 (ideal condition): displacement time histories of P1 obtained from different methods	171
Figure 66 Correlation matrix of time displacement histories of Case 1 (ideal condition).....	172

Figure 67 Displacement comparison in the frequency domain of Case 1	173
Figure 68 Case 2 (illumination change): (a) displacement time histories of P1 obtained from different methods and (b) average image intensity time history.....	174
Figure 69 Correlation matrix of time displacement histories of Case 2 (illumination change)..	176
Figure 70 Displacement comparison in the frequency domain of Case 2 (illumination change)	177
Figure 71 Poor matches when using feature-based methods (fog interference).....	178
Figure 72 Case 3 (fog interference): displacement time histories of P1 from different methods	179
Figure 73 Zoomed in section of the horizontal displacement time histories of P1.....	179
Figure 74 Correlation matrix of time displacement histories of Case 3 (fog interference)	180
Figure 75 Displacement comparison in frequency domain of Case 3 (fog interference)	181
Figure 76 Experimental setup of the First Bosphorus Bridge.....	186
Figure 77 Measurement point and static point.....	187
Figure 78 Displacement results of the mid-span of the First Bosphorus Bridge	187
Figure 79 Experimental setup of the Second Bosphorus Bridge	189
Figure 80 Displacement results of the mid-span of the Second Bosphorus Bridge.....	190
Figure 81 Experimental setup of the first experiment of the Osman Gazi Bridge: (a) image from camera, (b) camera location in map; (c) wood chessboard, (d) accelerometers in vertical and horizontal direction, (e) computer for data acquisition, (f) camera setup.....	191
Figure 82 Results of the first experiment on the Osman Gazi Bridge in time domain: (a) horizontal displacement from vision-based method, (b) vertical displacement from vision-based method, (c) horizontal acceleration from accelerometer, (d) vertical acceleration from accelerometer.....	192
Figure 83 Results of the first experiment on the Osman Gazi Bridge in frequency domain: (a) FFT of horizontal displacement from vision-based method, (b) FFT of vertical displacement from	

vision-based method, (c) FFT of horizontal acceleration from accelerometer, (d) FFT of vertical acceleration from accelerometer	192
Figure 84 Experimental setup of the second experiment of the Osman Gazi Bridge.....	194
Figure 85 Results of the second experiment on the Osman Gazi Bridge in time domain: (a) horizontal displacement from vision-based method, (b) vertical displacement from vision-based method, (c) horizontal acceleration from accelerometer, (d) vertical acceleration from accelerometer	195
Figure 86 Vehicles on the bridge at time A and B.....	195
Figure 87 Results of the first experiment on the Osman Gazi Bridge in frequency domain: (a) FFT of horizontal displacement from vision-based method, (b) FFT of vertical displacement from vision-based method, (c) FFT of horizontal acceleration from accelerometer, (d) FFT of vertical acceleration from accelerometer	196
Figure 88 Acceleration limits for footbridge vibration serviceability in different standards and codes	203
Figure 89 Procedure for proposed vision-based displacement/velocity measurement method using feature matching.....	204
Figure 90 Feature extraction and outlier removal.....	207
Figure 91 Feature matching for displacement and velocity.....	207
Figure 92 Matched pairs of an image sequences using different feature matching strategy	208
Figure 93 Experimental setup	210
Figure 94 Comparison of displacement results from two different feature matching strategies of Case 5: (a) comparison in time domain, and (b) comparison in frequency domain	213

Figure 95 Comparison of velocity results from two different feature matching strategies of Case 5: (a) comparison in time domain, and (b) comparison in frequency domain	215
Figure 96 Acceleration (raw data) directly calculated from the displacement data by using vision-based method with no frame update of Case 5: (a) in time domain, and (b) in frequency domain	216
Figure 97 Acceleration (raw data) directly calculated from the velocity data by using vision-based method with frame update of Case 5: (a) in time domain, and (b) in frequency domain	217
Figure 98 Acceleration data collected by accelerometer of Case 5: (a) in time domain, and (b) in frequency domain.....	218
Figure 99 Comparison of filtered acceleration obtained from vision-based methods with the data from accelerometer of Case 5: (a) in time domain, and (b) in frequency domain	220
Figure 100 Frequency spectra of the acceleration time histories obtained from vision-based method and accelerometer of all cases.....	225
Figure 101 Schematic diagram of the proposed method for distribution factor estimation	230
Figure 102 The prestressed concrete highway bridge in this study	231
Figure 103 Sensor instrumentation plan	232
Figure 104 Sensors and cameras used in the test: (a) displacement sensor, (b) strain gauge, (c) camera for traffic monitoring, and (d) camera for displacement measurement.....	233
Figure 105 Sensors installed on the bridge	233
Figure 106 Overview of the instrumentation	234
Figure 107 Truck loading on the bridge	235
Figure 108 Loading plan of the static test.....	235

Figure 109 Displacement results of the midspan of Case 1 (T1L1, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5	237
Figure 110 Strain results of the 1/4 span of Case 1 (T1L1, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5	238
Figure 111 Displacement results of the midspan of Case 2 (T1L2, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5	239
Figure 112 Strain results of the 1/4 span of Case 2 (T1L2, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5	239
Figure 113 Distribution factor calculated from the static test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (e) two same trucks in both lanes	241
Figure 114 Load rating factor calculated from the static test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (e) two same trucks in both lanes	242
Figure 115 Displacement results of the midspan of Case 3 (T1L1-35, dynamic): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5	243
Figure 116 Displacement results of the midspan of Case 4 (T1L2-35, dynamic): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5	243
Figure 117 Distribution factor calculated from the dynamic test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (e) two same trucks in both lanes	244
Figure 118 Load rating factor calculated from the dynamic test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (e) two same trucks in both lanes	245
Figure 119 UIL decomposition	253
Figure 120 General procedure of vehicle tracking	255

Figure 121 Vehicle tracking example: (a) Status of vehicle at the beginning of a truss bridge; (b) Status of vehicle at the end of a truss bridge	257
Figure 122 In-plane transformation using Homography matrix	259
Figure 123 Digital correlation-based template matching: (a) Normalized cross-correlation coefficient of grayscale image; Normalized cross-correlation coefficient of Canny edge map.	261
Figure 124 System configuration.....	263
Figure 125 Experimental setup in laboratory.....	264
Figure 126 Image from fisheye camera	265
Figure 127 Camera calibration using a black white chessboard.....	266
Figure 128 Rectified image after camera calibration.....	266
Figure 129 Video time synchronization using normalized cross-correlation based pattern matching of audio signals	267
Figure 130 Vehicle tracking in the rectified images from fish camera.....	268
Figure 131 Displacement comparison between the proposed vision-based method and the potentiometer	269
Figure 132 Extracted UIL using the proposed system.....	270
Figure 133 Experimental setup of a footbridge	271
Figure 134 Vehicle tracking in the planar transformed images: (top) tracking when the vehicle starts from the left end of the footbridge; (middle) tracking when the vehicle is at the midspan; (bottom) tracking when the vehicle arrives the right end	272
Figure 135 Displacement of the midspan under different loads	273
Figure 136 Extracted UIL of the midspan of the footbridge using the proposed system	274

LIST OF TABLES

Table 1 Summation of different feature points	59
Table 2 Optical flow methods.....	65
Table 3 Comparative study of modal parameters	77
Table 4 Comparative study of operational modal frequencies from vision and accelerometers ..	83
Table 5 Measurement accuracy and resolution ($\pm 2\sigma$) analysis: unit (mm).....	101
Table 6 Comparison of computation times of vision-based methods using different feature matching methods	102
Table 7 Pros and cons of the two different strategies to process image sequence.....	130
Table 8 Normalized root mean square error (NRMSE) and normalized cross-correlation (NCC) of the fit between the vision-based displacements and the benchmark.....	136
Table 9 Measurement accuracy and resolution ($\pm 2\sigma$) analysis	137
Table 10 Frequencies of human jumping load extracted from displacement data: unit (Hz).....	139
Table 11 Comparative study of operational modal frequencies from vision and accelerometer signals	145
Table 12 Time consumption of processing one image using different STC-based methods.....	170
Table 13 Detailed acceleration limits for footbridge vibration serviceability in different standards and codes.....	203
Table 14 Experimental cases	211
Table 15 Correlation matrix of time acceleration histories	221
Table 16 RMS and peak value of the acceleration data.....	222
Table 17 Serviceability assessment of different cases.....	223
Table 18 Experimental cases	234

Table 19 Distribution factor calculation in AASHTO	238
Table 20 Distribution factor calculated from the static test cases	241
Table 21 Load rating factor calculated from the static test cases	242
Table 22 Distribution factor calculated from the dynamic test cases	244
Table 23 Load rating factor calculated from the dynamic test cases	245
Table 24 Applicable scenarios and restraints for the proposed computer vision-based structural output-only monitoring methods.....	280
Table 25 Recommendations for different structure types and different computer vision-based structural output-only monitoring methods	280

CHAPTER ONE: INTRODUCTION

1.1 Background

Structural health monitoring (SHM) has been researched and implemented in the field of civil engineering and is still receiving more attention from both researchers and engineers (Catbas and Aktan 2002; Gul and Catbas 2008, 2011a; b). Significant progress has been made in structural health monitoring and performance evaluation of structures as well as structural identification, damage detection, model updating, structural reliability, condition assessment, decision making, structure management and maintenance (Aktan et al. 2000; Ghosn et al. 2016; Gul et al. 2014; Gul and Catbas 2009), as a result of the advances in sensing technologies and data processing techniques. SHM is playing an essential role in the diagnosis and prognosis of performance and safety of civil infrastructures. However, certain challenges have been presented and discussed in the development of current SHM such as challenges in field-work for sensor instrumentation, cable wiring, data acquisition, power and transmission arrangement for wired and wireless sensing. These challenges can make certain SHM applications inconvenient, time consuming, and expensive. Additional considerations that need to be resolved are: 1) the inspection and monitoring of existing structures may experience service interruptions such as bridges may require traffic closure, 2) it may be difficult to obtain access for instrumentation work for some large or sensitive infrastructures, and 3) some structures may warrant just intermittent monitoring with portable systems instead of permanent and continuous monitoring systems. As a result, the development of effective, convenient and inexpensive monitoring tools for cases outlined above and also for large populations of civil infrastructure is becoming more important.

Recently, the combination of camera technology and computer vision algorithms has led to great advancements in the field of SHM (Catbas et al. 2012b; Ye et al. 2016a; Yeum and Dyke 2015; Zaurin et al. 2015; Zaurin and Catbas 2010a; b). Consequently, computer vision-based structural health monitoring and identification research and implementation are gaining increasing attention in the community of infrastructure/structural engineering. This dissertation conducts a detailed investigation of computer vision concepts and methods for SHM and structural identification.

1.2 Organization of Thesis

This chapter, Chapter 1, introduces the subject matters to be discussed and a description of the research approaches, methods, objectives and applications. Chapter 2 delves into literature to discuss the advances in computer vision based SHM at local and global level of structural evaluation. This chapter presents the challenges and concerns in current research and practices of computer vision-based application in civil infrastructures. Chapter 3 proposes a vision-based displacement measurement framework which combines optical flow and different types of feature points without using manual markers. The proposed framework is verified on a grandstand model in laboratory and a real-life stadium. In addition, structural dynamic properties are identified using the proposed framework during the verification, and validated in a comparative fashion with conventional sensors. Chapter 4 proposes a novel marker free displacement measurement method by using advanced feature matching strategy. A new feature descriptor is implemented to original feature detector. The improvement of the new method is verified by a series experiments on a two-span bridge model in laboratory and a real-life railway bridge. Chapter 5 proposes a practical vision-based displacement measurement method by using a pretrained deep learning based optical

flow model. This method makes it much easier to use for the researchers or engineers in civil engineering who may not have enough knowledge or experiences of digital image processing or computer vision. The proposed method is verified by the experiments on a large-scale grandstand structure in laboratory and a real-life footbridge under human jumping loads. Considerations and recommendations are also discussed and provided in detail for accurate results. Chapter 6 proposes a robust vision-based displacement measurement method to overcome the difficulties and limitations for the traditional vision-based displacement measurement method in application cases under adverse environmental factors such as illumination change and fog. Chapter 7 presents the field applications of the vision-based displacement measurement methods on three large, landmark structures (three long span bridges in Turkey). The findings as well as the challenges and issues encountered during the field applications for such large structures are presented and discussed. Chapter 8 focuses on the application of the proposed vision-based vibration monitoring to determine vibration serviceability assessment of a footbridge. Chapter 9 proposes a practical approach for the estimation of bridge distribution factor and load rating factors by using computer vision-based methods. Chapter 10 presents a non-contact approach which can assess the load distribution on structures. By combining the structural input (load) and structural output (responses) obtained by cameras and computer vision, a conceptual index, unit influence line (UIL), is identified under daily traffic. In addition, the loads on the structures are estimated by using UIL. Finally, Chapter 11 presents a brief summary, conclusions and recommendations of this doctoral dissertation.

CHAPTER TWO: REVIEW OF COMPUTER VISION-BASED STRUCTURAL HEALTH MONITORING

2.1 Introduction

Structural health monitoring is the process of tracking the operational status, assessing the condition, and detecting the damage of various types of structures (Catbas et al. 2008a; Gul and Catbas 2009, 2011a). Material deterioration or structural change can induce the service interruption or can potentially pose safety risk for structures. Damage at the local level may be considered as the changes in the effective material properties and condition due to causes such as crack, spalling, corrosion, delamination and void. For example, qualified engineers and inspectors implement hammer sounding and/or chain drag, and visual inspection for concrete bridge deck evaluations, yet these methods require substantial field labor, experience, and operational interruptions. Based on visual inspections, National Bridge Inventory (NBI) defined the condition rating categories to evaluate three primary components of a bridge: deck, superstructure, and substructure (Item 58, 59, 60). The condition rating categories of NBI are divided into 10, from 0 to 9, and inspectors rate the bridge condition subjectively based on their experiences in accordance with the descriptions provided by US Department of Transportation (USDOT) and Federal Highway Administration (FHWA) (USDOT and FHWA 2000).

In the new manual for bridge element inspection released by American Association of State Highway and Transportation Officials (AASHTO 2015), the condition of each element is standardized by recording quantities of all defects detected from a field inspection. In the manual,

categorization of each element depending with the extent of defect type is defined. Seven types of defects are defined for prestressed concrete deck, such as delamination/spall/patched area, exposed rebar, exposed prestressing, cracking, efflorescence/rust staining, abrasion/wear, damage. Each defect is recorded with the size and volume, and rated condition state including “Good, Fair, Poor and Severe” separately based on the criteria provided by the manual. In addition, non-destructive evaluation (NDE) techniques such as impact echo, ultrasound electrical resistivity, thermography and ground-penetrating radar have been developed to inspect and monitor aging and deteriorating structures effectively in place of visual and sounding inspection methods. The NDE process is a localized condition evaluation and it can be described as structural health monitoring at local level (SHM-LL) if data are collected over time and condition is tracked. The critical deterioration mechanisms and severity levels need to be well-established before being monitored so that SHM-LL can properly address the problems that have implications on local as well as global condition. Even though a complete local scan throughout a structure can be done for the purpose of effective structural maintenance and management, it is still necessary to understand how local damage or condition changes affect the performance of the whole structure (Catbas and Aktan 2002).

The global damage can be described as structural behavior change that can distinctly influence the input-output (loads-responses) of the critical regions/elements of a structure. The process of addressing the global damage and condition assessment problems can be described as structural health monitoring at global level (SHM-GL). The relation between SHM-LL and SHM-GL is bidirectional: (a) the process of understanding the input-output structural behavior which is one of the tasks of SHM-GL can benefit from the condition assessment from SHM-LL. For example, load rating is a very important task of SHM-GL, while the condition factor in general load rating

equation is affected by the local condition rating (USDOT and FHWA 2000) according to the manual for bridge evaluation (AASHTO 2018). (b) the global condition evaluation and damage detection from SHM-GL can assist the SHM-LL to understand how localized condition and damage affect the complete system.

SHM-LL and SHM-GL play significant roles in the assessment, management and maintenance of structures to assist the assets' owners for decision making. According to the 2017 ASCE Infrastructure Report Card (ASCE 2017), the grade point average (GPA) of America's civil infrastructures is D+ on an A through F grading scale and the cost for the improvement is around 4.59 trillion USD. Consequently, it may become an overburden on the shoulders of the State Departments of Transportation (DOTs), which have already been stretched to limits in terms of budget and manpower in several states. At current stage, there are still a lot of limitations when implementing the conventional SHM-LL and SHM-GL approaches. The efficiency, cost, time, labor force are big issues for inspection and maintenance of the infrastructures. For SHM-LL, visual inspection is still the most preferred method by engineers to inspect structures. Although the nondestructive testing methods such as impact echo, ultrasound, electrical resistivity, and ground-penetrating radar make great progress in the effective inspection, the costs are too high to be promoted widely. For SHM-GL, sensors are always necessary to be installed on the critical regions of structures to acquire the external loads and structural responses. The challenges in fieldwork for sensor instrumentation, cable wiring, data acquisition, and power and transmission arrangement for wired and wireless sensing make certain SHM-GL applications inconvenient, time-consuming, and expensive. Traffic closure and access problem of sensor instrumentation on large scale structures are among other problems. In addition, some structures may warrant just

intermittent monitoring with portable systems instead of permanent and continuous monitoring systems. It is necessary to develop effective, convenient, and inexpensive SHM tools to better support the owner's decision making for structural condition assessment at local and global level.

In recent years, with the development of imaging devices having low cost and high quality, and with the remarkable progress of computer vision (CV) techniques, vision-based SHM has been gathering increasing attention in the SHM community. In addition, the corresponding techniques can be easily implemented into SHM to dig effective structural condition information and provide benefit to the conventional practice of SHM. The tasks of CV as image classification, detection, image segmentation, optical flow, and visual tracking are very similar in both SHM-LL and SHM-GL. SHM-LL indices for local condition assessment such as crack, spalling, corrosion, delamination and void can be extracted from visual images of the structure's surface and infrared images reflecting the inside of structural elements with computer techniques such as image classification, detection, image segmentation. SHM-GL indices for global condition assessment such as structural responses including displacement, vibration, modal parameters, cable force, curvature and profile can be obtained by optical flow and visual tracking, and the external loads including vehicle distribution on bridge structures and human occupancy in building structures can be obtained by objected detection. The advantages of CV-based structural health monitoring at local and global level (CV-SHM-LL and CV-SHM-GL) are these methods enable long distance, non-contact, low cost and automated inspection. CV-based approach can not only give detailed distribution map of damages/structural changes, but also quantify the extent of the structural conditions with detailed information. It is also compatible with the current AASHTO (American Association of State Highway and Transportation Officials) codes and NBI (National Bridge

Inventory) condition rating codes. In addition, CV-based approaches can supply denser spatial information compared with the conventional SHM which relies on installing sensors on discrete nodes of structures. Although it is not possible for CV-SHM to cover every aspect of the concerns of SHM, CV-SHM (LL and GL) provides a promising complementary of the current SHM practice to improve the present grade of the infrastructures all over the world.

This chapter dedicates to make a comprehensive review of computer vision-based structural health monitoring at local and global level. The paper is structured as follows: Section 2.2 introduces the projective geometry implemented in CV-SHM; Section 2.3 introduces the methods, algorithms, and applications of CV-SHM at local level; Section 2.4 introduces those at global level; Section 2.5 presents the challenges and concerns in real-life practices of CV-SHM; Section 2.6 concludes the literature review.

2.2 Projective Geometry Applied in CV-SHM

In both CV-SHM-LL and CV-SHM-GL, the object in real world is projected to an image via camera and lens. By analysing the changes in the images such as a motion or abnormal phenomena, those happen in the real world can be estimated. In AASHTO code, for structural element inspection, requirements such as the width/length of the crack and the size of the spalling area are necessary to assess the element condition. In CV-SHM-GL applications, requirements such as vibration and deflection are necessary to obtain the modal properties and load bearing capacities of bridges. While in the process of converting from images to the real world, it is necessary to figure out the relation between the camera and the real world: “what is the pose of the camera and

lens?” “how long does one pixel represent in the real world in each direction?”. To answer these questions, the projective geometry calculation needs to be implemented and camera calibration is essential to reverse the camera projection. Fig. 1 is a diagram of pinhole camera model which shows the projective transform from three-dimensional (3D) world to two-dimensional (2D) image. The extrinsic and intrinsic matrices need to be estimated during the camera calibration.

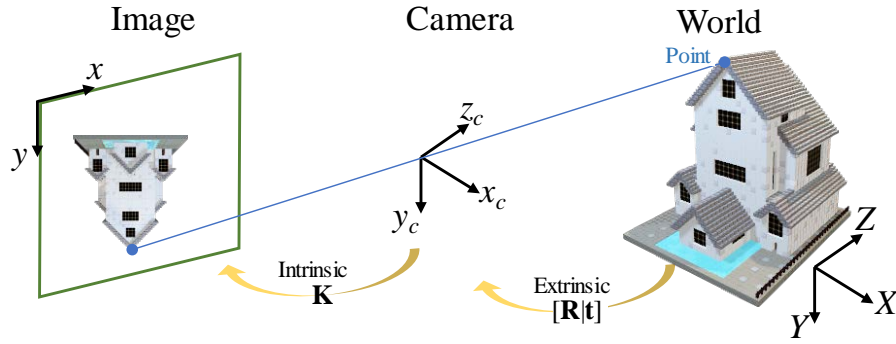


Figure 1 The projection from real world object to image

The projection transform from world coordinates to the image coordinates through camera coordinates can be expressed by the formula below:

$$s\mathbf{x} = \mathbf{K}[\mathbf{R} | \mathbf{t}]\mathbf{X} \quad (1)$$

and it can be expanded as

$$s \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2)$$

where s is the scale factor, $\mathbf{x} = (x, y, 1)^T$ are image coordinates, $\mathbf{X} = (X, Y, Z, 1)^T$, are world coordinates, and \mathbf{K} is the camera intrinsic parameters which represents the projective

transformation from the 3D world to the 2D image plane. \mathbf{R} and \mathbf{t} are camera extrinsic parameters which represent the rigid rotation and translation from the 3D real world coordinates to the 3D camera coordinates. In the intrinsic matrix, f_x and f_y are the focal lengths of the lens in horizontal and vertical directions, c_x and c_y are offsets of the optical axis in horizontal and vertical directions, and γ is the skew factor of the lens. In the extrinsic matrix, r_{ij} ($i, j = 1, 2, 3$) and t_i ($i=1, 2, 3$) are the elements of \mathbf{R} and \mathbf{t} , respectively. From Eqs. (1) and (2), it is indicated that the camera intrinsic parameters are relevant to the camera and lens, and the camera extrinsic parameters are relevant to the relative position between the camera-lens and real objects, i.e., camera pose. Once the camera is calibrated with a specific lens, as soon as the focal lens doesn't change, the intrinsic parameters don't change. However, the extrinsic parameters should be calibrated in different application scenarios since the pose changes. Zhang's practical calibration approach (Zhang 2002) is commonly used, which utilizes images of a black and white chessboard taken in different camera pose. Commercial software such as MATLAB Vision Toolbox, NI Vision, Halcon and open source library such as OpenCV supply user friendly interface to implement camera calibration procedure.

In the application of CV-SHM, no matter LL or GL, the implementation of projective transform for camera calibration is always necessary. For CV-SHM-LL application, (Karaaslan et al. 2019) implemented CV to estimate the camera pose of a headset and determine the length/width/area of detected cracks on structures. Then they assessed the structural condition as "Good, Fair, Poor or Severe" according to AASHTO codes. (Adhikari et al. 2014) implemented projective geometry to calculate the crack densities of concrete bridges. (Ellenberg et al. 2016) implemented it to rectify the cameras on an Unmanned Aerial Vehicle (UAV) to estimate the size of the delamination in a bridge deck. (Yang et al. 2015) implemented the projective transform to two cameras to detect the

crack distribution, calculate the crack width and monitor the propagation of cracks in a concrete pier. Also, in CV-SHM-GL applications, especially when there's a need for 3D displacement monitoring, two cameras need to be implemented and the pose estimations of the two cameras have to be performed respectively in order to build the connection between two cameras. (Shan et al. 2018a; b, 2019) implemented the projective transform to calibrate two cameras and estimate the camera pose to monitor the 3D displacements of a three-story structure under seismic loading and a concrete beam in a destructive test. (Chang and Ji 2007) implemented projective geometry to monitor the 3D vibration of a bridge deck in wind tunnel test and a frame structure in a shaking table test. References (Kim et al. 2006; Lages Martins et al. 2015; Park et al. 2015b) also presented the similar implementations. When a wide-angle camera is used, the projective geometry can also rectify the image due to radial distortion. (Xu et al. 2018) employed a GoPro camera with wide angle to monitor the vibration of a cable-stayed bridge and applied the similar procedure presented above to rectify distorted images. (Dong et al. 2019a) implemented the projective transform to rectify distorted images and estimate the location of vehicles on the bridge deck.

In practical CV-SHM applications, the projective transform is simplified by scale ratio or homography transform. The scale ratio, SR , can be expressed as:

$$SR = \frac{L}{l} = \frac{D}{f} \quad (3)$$

where L is the length of the object in real world, l is the length with the unit of pixel in image, D is the distance from the camera to the object, f is the focal length. Compared to the projective transformation in Eqs. (1) and (2), the scale ratio is much simpler and easier to calculate. In practical applications, the scale ratio is always an option. References (Jahanshahi and Masri 2012;

Li et al. 2014; Zhong et al. 2018c) employed the scale ratio for the crack width and length estimation. References (Celik et al. 2018a, 2019a; Dong et al. 2015; Feng and Feng 2017; Ye et al. 2016f, 2013, 2015, 2016c, b, d; a) applied the scale ratio to calculate the displacement of monitored structures. When the axis of camera-lens is perpendicular to the measurement plane, the calculation of scale ratio can be obtained from Eq. (3). If the axis of camera-lens is not perpendicular to the measurement plane, Eq. (3) has to be modified. References (Dong et al. 2019b; Feng et al. 2015a) presented the detailed modifications. Another option is to use homography matrix if the motion/change of the structure is limited in one plane. The projection from the real world plane to the image plane is expressed by the homography transform:

$$\mathbf{X} = s\mathbf{H}\mathbf{x} \quad (4)$$

Eq. (4) is a degraded version of Eq. (1). Eq. (4) can be formed by

$$\begin{Bmatrix} X \\ Y \\ 1 \end{Bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{Bmatrix} x \\ y \\ 1 \end{Bmatrix} \quad (5)$$

where \mathbf{X} is degraded to $(X, Y, 1)^T$. In this formulation, \mathbf{H} is the 3×3 homography matrix and h_i ($i = 1, 2, \dots, 9$) is the element of \mathbf{H} . The homography matrix \mathbf{H} has 9 unknowns and only 8 of them are independent. Here, at least four point correspondences are needed to calculate \mathbf{H} . The detailed parameter estimation procedure is presented in (Hartley and Zisserman 2003). References (Dong et al. 2019c; Wu et al. 2014; Xu et al. 2018) applied homography matrix method to measure the in-plane structural displacement. References (Dong et al. 2019a; Khuc and Catbas 2018; Zaurin and Catbas 2010a) applied homography matrix method to estimate the location of the vehicles on bridge deck.

2.3 Computer Vision-based Structural Health Monitoring at Local Level (CV-SHM-LL)

According to AASHTO codes (AASHTO 2015), the structural condition is assessed with the inspection of items such as crack, spalling, delamination, corrosion, etc. Quantities are necessary to categorize the conditions. The conventional procedure done by inspectors is first to check the structures visually and then to manually mark the location on the drawings of structures (Adhikari et al. 2014). This procedure is time-consuming and need lots of labor forces. Photos or videos may just be regarded as references or back-up for documentary. In CV-SHM-LL, the images are the main materials for analysis and assessment, and the whole procedure is towards to be automated. The final assessment is performed according to the condition map extracted from images.

This section reviews the applications of CV-SHM-LL in different structures types including crack, spalling, and delamination detection in concrete structures, crack and delamination detection in pavement structures, and crack detection, crack propagation monitoring, rust detection and bolt loose detection in steel structures.

2.3.1 Crack Detection of Concrete Structures

Generally, the crack detection methods can be classified into two categories (Zhu et al. 2011). The first category is the patch-based method. The basic procedure is to recognize whether cracks exist in a patch. The patch can be a sliding window crossing the whole images to do an exhaustive search with predefined stride or can be the subregion which is segmented from the original image. Within the patch, pattern recognition, template matching or classifier can be implemented to recognize

whether there are cracks. In this process, machine learning, deep learning or matching/recognition with manual features can be applied. The final crack distribution can be obtained by putting all the crack results with the patches together. The patch-based method may only produce the possible range of cracks depending on the size of the patch. The connectivity, detailed shape and angle cannot be obtained (Mei and Gül 2019). Further process such as local edge detection, morphological process or visual segmentation are necessary to refine and segment the cracks from background.

The second category is the pixel-based method. The whole image is processed and cracks are segmented from the background. At the end, a detailed cracks shape and distribution map is obtained. Reference (Ni et al. 2019) introduced crack detection results by using different methods including: 1) GoogLeNet which is one of the patch based crack detection methods; 2) Crack Delineation Network (CDN) which is a convolutional neural network (CNN) based crack segmentation method (pixel based); 3) Otsu's segment using threshold to binarize the image to get the morphology (pixel based); and 4) Canny edge detector using edge detection to get crack profile (pixel based). This section reviews the two categories of crack detection methods for concrete structures respectively.

2.3.1.1 Patch-based Crack Detection of Concrete Structures

(Liu et al. 2003) combined the image intensities, edges and Support Vector Machine (SVM) to train a crack classifier and then categorized small image patches into three classes, "crack, non-crack and intermediate". The classifier slides on the whole image to predict the possible crack

regions. (Abdel-Qader et al. 2006) introduced a PCA (Principal Component Analysis)-based crack cluster method for small image patches, and this method can achieve unsupervised bridge crack detection. (Li et al. 2017) extracted and segmented cracks from image patches by using edge detection and modified active contour and the crack width was also calculated. To eliminate the noises caused by inconsistent intensity and possible shadows, they applied SVM and greedy search strategy on the whole tested image. (Silva and Lucena 2018) retrained the VGG16 neural network and developed a crack classifier, however, this classifier can only determine whether there're cracks in images or not. It cannot show the crack shape. (Cha et al. 2017a) trained CNN based crack classifier for small image patch (256×256 pixels) and slid the classifier on the tested images to predict whether there were cracks inside the overlapped patches. The final crack detection map was extracted by only keeping the patches with cracks. (Yokoyama and Matsumoto 2017) did the similar work and found a phenomenon that when there is stain, the detection rate decreases. (Chen and Jahanshahi 2018) developed a CNN-based crack classifier and detected the possible crack area by sliding a small patch on images to classify cracks. They maintained the spatiotemporal coherence of cracks in videos using data fusion and decreased the false positive rate by using Naïve Bayes decision making.

(Prasanna et al. 2016) utilized line segment detector, spatially tuned multi-feature and machine learning classifiers such as Random Forest, SVM and Adaboost to get the local crack map in image patches. Then by stitching all the patches together, they got the global crack map. (Jahanshahi and Masri 2012) built a crack classier by using the features extracted from 3D scene reconstructed segmentation map. The crack segmentation parameters could be automatically adjusted based on depth parameters obtained from multi-view geometry. This crack detection method also recognizes

cracks within small patches. (Karaaslan et al. 2019) retrained the VGG (a CNN framework developed by Visual Geometry Group in University of Oxford) weights in SSD architecture to recognize crack regions with a bounding box and used the SegNet model to segment the cracks inside the bounding box. (Kim and Cho 2018) trained a CNN-based crack classifier for small image patches and by sliding the crack classifier on images, they acquired the crack detection results and probability map. (Dorafshan et al. 2018) retrained the AlexNet and by transfer learning, they developed an CNN-based crack classifier. Then they divided an image into small patches and used the trained classifier to infer whether there were cracks inside the small patches. Also, by doing edge detection inside the patches, they obtained the crack shapes. (Jang et al. 2019) retrained the GooLeNet to a crack classifier for standard visual images to recognize the cracks in small patches and used another CNN to train the images from infrared thermal (IRT) cameras to minimize the false detection.

2.3.1.2 Pixel-based Crack Detection of Concrete Structures

Edge detection is a very popular technique for crack detection at pixel level and to the best of the authors' knowledge, it might be the first technique to be implemented to perform crack detection. Edge detection is a manual processing method for crack detection. (Abdel-Qader et al. 2003) compared four different edge detection algorithms for crack detection including fast Haar transform (FHT), fast Fourier transform, Sobel, and Canny detector, and found that FHT gives more reliable results than the other three. (Yu et al. 2007) applied Sobel detector to segment cracks and used neighbor region linking method to determine real cracks from edges. (Li et al. 2013, 2014) combined C-V model with the edge detection and filtering to segment the cracks and also to estimate the width of the cracks.

Besides edge detection, filtering, morphology, segmentation with threshold is also one of manually processing approaches. (Iyer and Sinha 2006) implemented mathematical morphology and curvature evaluation to detect the cracks in buried sewers. (Fujita and Hamamoto 2011) used median filter, image subtraction and multi-scale line filter to separate cracks from background and noises. Then they used probabilistic relaxation and a locally adaptive thresholding to detect cracks with coarse and fine steps. (Sinha and Fieguth 2006) proposed a crack detection approach using statistical filter and defined global cracks among filtered segments using cleaning and linking. (Yamaguchi and Hashimoto 2010) extracted cracks from large size concrete using percolation-based methods and (Zhu et al. 2011) implemented the percolation-based methods to retrieve the crack properties such as width, length, and orientation. (Nayyeri et al. 2019) extracted cracks by fusing the local structure and global texture distributions. (Adhikari et al. 2014) segmented cracks using skeletonization and applied artificial neural network (ANN) based data fitting to get the crack depth and width. (Liu et al. 2016b) extracted the crack profile and calculated the crack width property by using 3D scene reconstruction. The approaches mentioned above for pixel-based crack detection rely on manual feature extraction and segmentation. The application scenario might limit the wide application of them.

While with large data base for cracks, deep learning-based methods could be another good option for pixel level crack detection. References (Dung and Anh 2019; Ye et al. 2019) trained fully convolutional networks (FCN) for semantic segmentation to extract cracks from images and to give detailed crack map with shape and distribution. (Ni et al. 2019) combined the GoogLeNet for classification and Crack delineation network (CDN) to achieve the pixel level crack detection. The main advantage of deep learning-based crack detection at pixel level is that it provides an end-to-end framework. The users just need to feed the images into the framework and then the crack map with detailed information will be produced.

2.3.2 Crack Detection of Concrete Structures

The crack detection in pavement structures is similar to the concrete, which is also divided into patch-based and pixel-based, while the dataset and challenges may be different. Crack detection in pavement structures is introduced in this section.

2.3.2.1 Patch-based Crack Detection of Pavement Structures

(Gavilán et al. 2011) developed an SVM classifier for different road types and regarded the non-crack features. The result image patches then were processed by a seed-based approach which combined with multiple directional non-minimum suppression with a symmetry check to extract cracks. (Huang and Xu 2006) divided pavement image into patch of 8×8 pixels and classified whether there were cracks inside using the information of border pixels. The crack shape inside the patch was formed by crack clustering. The total crack map of a pavement image was obtained by putting all the crack results of patches together. (Mathavan et al. 2015) developed a crack cluster for small image patches using an unsupervised learning technique called self-organizing map. (Hu et al. 2010) extracted texture features and shape descriptors from pavement surface and trained an SVM classifier to recognize cracks in small image patches of the original image. Inside the patch, segmentation, fake-crack eliminating, and crack-measuring were operated to extract the crack shape. By combining all the patches, they obtained the full crack map. (Shi et al. 2016) developed a crack classifier using random structure forest and SVM to recognize cracks in image patches. Then the final crack map was extracted after doing erosion and the dilation in each patch. (Zhang

et al. 2016b) trained a CNN-based classifier to recognize cracks in small patches. (Gopalakrishnan et al. 2017) retrained the VGG-16 network with the FHWA/LTPP database and developed a pavement crack classifier for patches. (Nhat-Duc et al. 2018) compared the CNN based pavement crack classification method in patches with the edge-based methods and found that CNN-based classifier performed better than edge-based methods for cracks. (Tong et al. 2018) designed two CNN-based crack patch classifiers, one is to recognize whether there's crack inside a patch and the other one is to recognize the range of the crack length within 1 cm.

2.3.2.2 Pixel-based Crack Detection of Pavement Structures

(Cheng et al. 2003) determined the real time thresholding from image intensities by using sample space reduction and interpolation approach and extracted the cracks by using image thresholding-based segmentation. (Zou et al. 2012) developed the CrackTree to perform crack detection in pixel level. The CrackTree is based on tensor voting for crack probability map and crack seed sampled from probability map is used to identify desirable cracks. (Chambon and Moliard 2011) developed a pixel-based crack detection approach by combining multi-scale extraction and a Markovian segmentation. (Amhaz et al. 2016) proposed a crack detection approach based on minimal path selection and with the skeleton refinement the artifacts could be eliminated. Also, after post processing, the width of cracks could be obtained. (Fei et al. 2019) developed a CNN-based pixel level crack detection approach for 3D asphalt pavement images. References (Sajedi and Liang 2019; Yang et al. 2018a) trained the FCN using the semantic segmentation methods to extract crack map at pixel level. (Mei and Gül 2019) developed a conditional Wasserstein generative adversarial network to extract pavement cracks in pixel level. The images were taken from a GoPro camera

installed on a vehicle running with normal speed. (Yang et al. 2019) trained a feature pyramid and hierarchical boosting network (FPHBN) to detect cracks in pixel level. They found that FPHBN outperforms the semantic segmentation methods in five datasets.

2.3.3 Spalling Detection of Concrete Structures

Spalls are flakes of a material that are broken off of a large solid body or surface. In reinforced concrete (RC) structures, spalling is a severe problem which can let the steel bars lose the concrete cover; accordingly the steel bars may be easily corroded. In addition, spalling is regarded as an important indicator of significant damage to structural elements formed during an earthquake (German et al. 2012). Figure 2 shows some examples of spalling compared with cracks. It can be seen that spalling is larger than cracks generally. Mechanical spalling occurs at high stress contact areas.

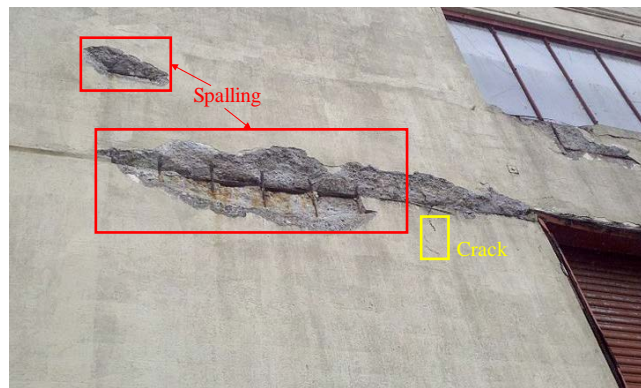


Figure 2 Examples of spalling and crack

(Gao and Mosalam 2018) retrained the VGG net by using transfer learning to develop a spalling classifier to recognize spalling in concrete structures. (Dawood et al. 2017) developed a spalling

detection approach by using hybrid algorithm including image smoothing, thresholding, histogram equalization, Gaussian blur, color transformer, smart filter, image scaling etc. Also, the most important point is that they proposed a method to predict the depth and severity of spalling using regression analysis. This makes their approach become a decision-making tool. (German et al. 2012) proposed a spalling recognition approach using a local entropy-based thresholding algorithm and by combining a novel global adaptive thresholding algorithm with template matching and morphological operations, they could measure the depth and length of spalling in concrete columns. (Karaaslan et al. 2019) first retrained the SSD (Single Shot MultiBox Detector) to detect spalling in an image and then proposed an attention guided segmentation network to segment spalling from concrete columns and walls. The guided segmentation with human aided operation doesn't need full image search and improve the accuracy.

2.3.4 Delamination Detection

Delamination is one of the subsurface deteriorations of concrete structures and it could further evolve to become cracks and spalling. This kind of structural damage eventually reduce the structural load carrying capacities. It is essential to detect the possible delamination areas and retrofit the structures. Current practice to detect delamination is generally using chain drag or hammer, however, this kind of approaches might may further damage to the structures during the inspection (Hiasa et al. 2016a). Ground-Penetrating Radar (GPR) technique is one of the Non-Destructive Testing (NDT) methods to detect the delamination, however, the speed is slow. The Infrared Thermography (IRT) has been developed to detect existing subsurface deteriorations including delaminations and voids in concrete (Hiasa et al. 2016a). When there is delamination inside the concrete, the temperature is different with the sound area (Hiasa et al. 2017a). With this feature, by scanning the surface of the concrete, the delamination can be detected. The IRT camera

can also be installed on a vehicle with a normal moving speed to achieve faster inspection speed compared to other NDT methods (Hiasa et al. 2016b, 2017d). (Hiasa et al. 2017d) gave an example of concrete scanning using vehicle carried IRT cameras. The detection performance relies on temperature gradients (Watase et al. 2015), which means it is quite important to select the scanning time range in a day (Hiasa et al. 2017d). (Matsumoto et al. 2013) presented the time zone when the inspection can be executed.

(Watase et al. 2015) investigated the favourable time windows for IRT for concrete delamination evaluation by using plates with different thickness and delamination with different depth. The different thickness of plate and the depth of delamination could influence the time window. (Hiasa et al. 2018) explored the time window for good inspection of IRT by using experimental and numerical methods and found that optimal conditions for IRT implementation on concrete bridge decks was night time application under the clear sky condition. (Hiasa et al. 2017b) investigated the effect and correlation of delamination size and shape for using IRT through finite element modeling (FEM) and found that the delamination depth information could be estimated by incorporating IRT with FEM. To segment delamination from IRT images, a proper temperature threshold is necessary as IRT images can also be processed by using the similar techniques for visual images taken by standard cameras. (Hiasa et al. 2017c) investigated the temperature threshold using FEM and found that the temperature threshold of delaminated areas of concrete slab with the depths of 1.27 cm and 2.54 cm defined by FEM simulation could give better prediction performance than directly judging from IRT images with naked eye. (Omar et al. 2018) used *k*-means clustering to segment the mosaicked thermogram of entire bridge deck and identified the objective threshold separately. Based on the different thresholds, the detection of delamination performed at higher accuracy. (Hiasa et al. 2017d) discussed the considerations and issues in the

application of IRT for concrete scanning at normal driving speeds, such as thermal contrast, time window, camera specification, distance, and utilization speed and gave detailed recommendations. They also implemented a High Definition (HD) camera along with the IRT cameras to scan concrete to get visual images. The visual images from HD camera could assist IRT to discard false positive prediction of delamination. Instead of using a vehicle to do concrete scanning, References (Ellenberg et al. 2016; Omar and Nehdi 2017) installed the IRT cameras on unmanned aerial vehicle (UAV) to scan concrete bridges. This definitely makes the IRT-based delamination detection become more flexible.

2.3.5 Crack Detection of Steel Structures

The phenomena of cracks on a metallic surface, especially in steel structures are different from those in concrete or pavement structures. There are few studies about crack detection in steel structures. Current approaches used in crack detection are similar to those used in concrete crack detection. (Yeum and Dyke 2015) detected the bolts on the steel structures by training a classifier which combined histogram of gradient (HOG), Haar like feature window and boosting. Then they detected the cracks around the bolts by using edge-based methods. The approach they proposed can be categorized as a patch-based detection. (Chen et al. 2017) also proposed a patch-based approach to detect cracks on the metallic surface. They trained a classifier using local binary patterns and SVM. By sliding the classifier on the image, they extracted the possible crack regions.

2.3.6 Crack Propagation Monitoring

When subjected to loading and a certain stress level is reached within a material, cracks form and propagate due to increase in load. It is essential to monitor the pattern of crack propagation and take actions to prevent further damage. Using computer vision to monitor the crack propagations is a very promising approach. (Bhattacharjee and Deb 2016) detected and classified the cracked zones by using the displacement field obtained from digital image correlation (DIC). During load test, the crack changes can also be monitored by using their approaches. (Yin et al. 2014) detected and monitored the crack propagation by using keypoint/feature-based tracking and they found that displacements at keypoint/feature points were well correlated with cracks appeared. (Yang et al. 2015) investigated the thin crack propagation of an RC pier using the displacement changes at manual marked grid tracked from stereo cameras. (Yang et al. 2018c) monitored the crack generation, propagation, and distribution of an RC column during load test by estimating the full field surface displacement using optical flow. The proposed method is consistent with the results obtained from that manually marked on the concrete surface. (Lee et al. 2019) investigated the development of crack width of an RC specimen during tensile test by using cameras and manual marker to measure the displacement and strain change. (Kong and Li 2018a) monitored the crack propagation during a fatigue loading test by tracking the displacement of keypoint/features extracted around the crack opening. (Yu et al. 2019) monitored the crack propagation of wood structures by using DIC to measure the crack mouth opening displacements.

2.3.7 Rust Detection of Steel Structures

For steel structures, one of the major expenses is the anti-corrosion maintenance tasks (Liao and Lee 2016). According to the instructions of AASHTO inspection manual (AASHTO 2015), the rust on the surface of structures' coating reflect the extent of steel corrosion. (Shen et al. 2018) listed some examples of rust on the surfaces of steel structures.

Computing the area percentage of rust is one of the conventional ways of assessing the coating quality of steel structures. American Society for Testing and Materials (ASTM) proposed a zero-to-ten scale to rate the corrosion performance (Chen et al. 2012a) and gave suggestions to do maintenance. The rust percentage is a crucial indicator for the contractors or owners of structures to repair paintings, thus, it is quite important to develop accurate and reliable inspection approaches to support the owners' decision making (Lee et al. 2005). Detecting and quantifying the rust is a promising approach and can provide faster and more reliable inspection results compared with the visual inspection (Lee 2005). One of the challenges of rust detection is non-uniform illumination. Most of the studies focus on overcoming the problem. (Chen and Chang 2002) proposed the neuro-fuzzy recognition approach for rust detection and solved the non-uniform illumination problem. In this approach, the threshold value for image binarization and segmentation is generated from a pre-trained neural network.

(Lee et al. 2005) compared four rust detection methods including Neuro-Fuzzy Recognition Approach, Illumination-Based Segmentation and K-Means Algorithm, K-Means Algorithm, Simplified K-Means Algorithm. They found that Simplified K-Means Algorithm had the best performance and took the shortest time to process an image. (Lee et al. 2006b) proposed a rust

detection approach using color information instead of grayscale image processing. They also used multivariate discriminant functions to build the statistical model to perform rust detection. (Chen et al. 2009) investigated 14 color spaces and found the best color configuration a^*b^* (a^* is the color component of red minus green and b^* is that of green minus blue) has a moderate ability to filter light and eliminate non-uniform illumination problem. They proposed an adaptive ellipse approach with the identified color configuration to detect rust areas in images. (Chen et al. 2010) implemented the Bidimensional Empirical Mode Decomposition (BEMD) and morphology-based rust detection approach to eliminate the non-uniform illumination problem including shades and highlights. References (Chen et al. 2011; Shen et al. 2013) applied the Fourier transform to filter the images and detected the rust by using the color and texture information. This approach can be adopted to different background color and eliminate the non-uniform illumination in some extent. (Chen et al. 2012a) combined the Fourier transform with SVM to detect the rust area and the comparison with Simplified K-Means Algorithm showed better performance. (Kim et al. 2014) proposed a rust detection approach based on color space transformation and decision tree. Also they implemented this approach to a robotic system to achieve rapid and automated defect inspection. (Liao and Lee 2016) proposed three rust detection approaches: (a) cluster method with K-Means in hue component of HSI (Hue-Saturation-Intensity) images; (b) double-center double-radius algorithm in RGB space; and (c) double-center double-radius algorithm in HSI space. They also implemented least-square SVM to achieve rust radius prediction. (Shen et al. 2018) proposed an ANN-based rust intensity recognition approach which extracts more clusters and can better reflect the rust intensity and severity.

2.3.8 Bolt Loose Detection of Steel Structures

Bolt is a critical component in the joint connections of steel structures and bolt loose may cause repetitive loads and vibrations of the connection parts. It may weaken the structural integrity of the bolted steel joints (Kong and Li 2018b) and reduce the capacity of the whole structure. The AASHTO inspection manual (AASHTO 2015) considered bolt loose as one of the inspection requirements to do condition assessment. Current practice of bolt loose inspection mainly relies on human visual inspection. This section reviews the recent development in vision-based automated bolt loose detection. (Kong and Li 2018b) proposed a non-contact bolt loose detection method by tracking motion of the keypoints/feature points around the bolts. (Park et al. 2015a) monitored the bolt loose by tracking the bolt angle motion using Hough line alignment. Ramana et al. (Ramana et al. 2019) combined the Viola–Jones algorithm and SVM classifier to recognize bolt loose status. (Zhang et al. 2019b) retrained the Fast R-CNN object detector with bolt loose images to detect loose bolts in steel structures. (Wang et al. 2019) trained a CNN based classifier to recognize the bolt loose status and implemented Hough transform to estimate the rotation angle. (Sun et al. 2019) proposed a bolt loose detection approach using binocular vision which combines edge detection, segmentation, CNN prediction and 3D feature points matching. Their method can achieve bolt loose detection on a running train.

2.4 Computer Vision-based Structural Health Monitoring at Global Level (CV-SHM-GL)

2.4.1 CV-based Structural Response (Output) Monitoring

In this section, the CV-based structural response (output) monitoring refers to CV-based structural displacement measurement. By recording the structures under external loads with a still camera, the trajectories of selected regions can be obtained by tracking the motion of them. This is the basic idea of CV-based displacement measurement. This section reviews core techniques implemented in the CV-based displacement measurement and their applications in SHM.

2.4.1.1 General Procedure of CV-based Displacement Measurement

Figure 3 shows a general procedure of CV-based displacement measurement. Usually there are five steps to extract displacement information from digital image sequences or videos. At the very beginning, the projective geometry relation between the camera and real world has to be determined through camera calibration. This step is introduced and discussed in Section 2. In next step, the region of interest (ROI) is selected. ROI can be the manual markers or targets installed on the surface of structures or the parts of the structures with distinct surface features or textures. Then features are extracted from the selected ROIs for visual tracking. In the fourth step, visual tracking algorithm which is selected according to the measurement requirement and convenience is implemented to track the motion of the selection ROI. Finally, the displacement of the selected ROI can be calculated based on the visual tracking results and camera calibration information.

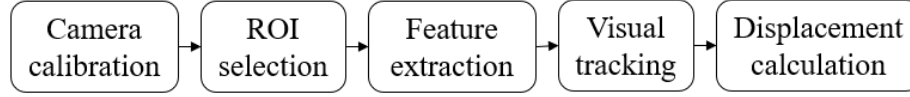


Figure 3 General procedure of CV-based displacement measurement

2.4.1.2 Core step review: visual tracking

In the general procedure of CV-based displacement measurement, the visual tracking is the most important step and the tracking performance directly affects the measurement accuracy. Selection of a visual tracking algorithm in the meantime also needs the suitable feature extraction. This section reviews the visual tracking algorithm along with the feature used during tracking.

(1) **Template matching.** As shown in Figure 4, the basic idea of template matching is sliding the selected ROI as a template across another image to search the best matching by calculating the similarity between the template and overlapped region of the image during sliding.

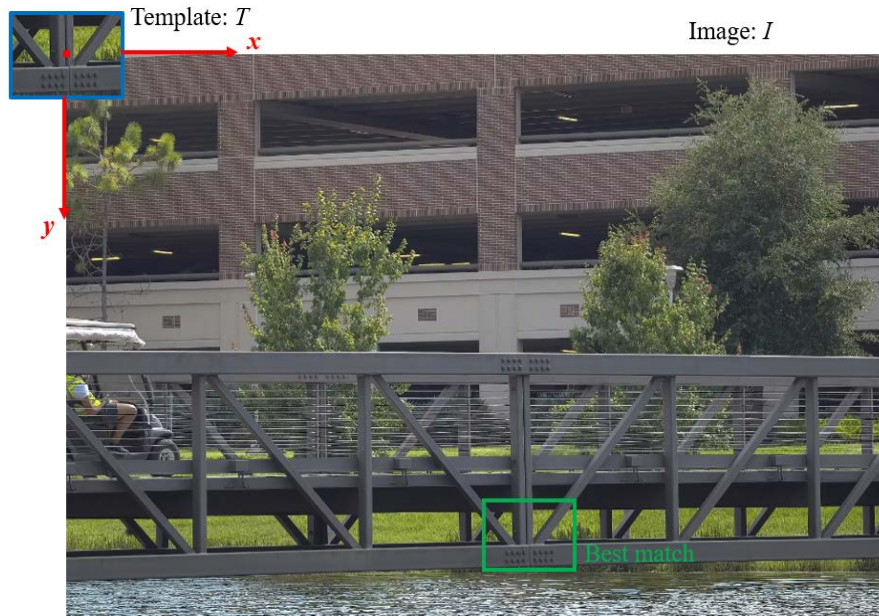


Figure 4 Sliding the template across another image to find the best match

Several ways to calculate similarities are listed with formulas as following (Bradski and Kaehler 2008):

a) Square difference matching method:

$$\alpha_{sq_diff}(x, y) = \sum_{x', y'} [T(x', y') - I(x + x', y + y')]^2 \quad (6)$$

b) Normalized square difference matching method:

$$\alpha_{nsq_diff}(x, y) = \frac{\sum_{x', y'} [T(x', y') - I(x + x', y + y')]^2}{Z(x, y)} \quad (7)$$

$$Z(x, y) = \sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2} \quad (8)$$

c) Cross correlation matching methods:

$$\alpha_{ccorr}(x, y) = \sum_{x', y'} (T(x', y') \cdot I(x + x', y + y')) \quad (9)$$

d) Normalized cross correlation matching methods:

$$\alpha_{nccorr}(x, y) = \frac{\sum_{x', y'} (T(x', y') \cdot I(x + x', y + y'))}{Z(x, y)} \quad (10)$$

e) Correlation coefficient matching methods

$$\alpha_{ccoeff}(x, y) = \sum_{x', y'} (T'(x', y') \cdot I'(x + x', y + y')) \quad (11)$$

$$T'(x', y') = T(x', y') - 1/(w \cdot h) \sum_{x'', y''} T(x'', y'') \quad (12)$$

$$I'(x + x', y + y') = I(x + x', y + y') - 1/(w \cdot h) \sum_{x'', y''} I(x + x'', y + y'') \quad (13)$$

f) Normalized correlation coefficient matching methods

$$\alpha_{nccoeff}(x, y) = \frac{\sum_{x', y'} (T'(x', y') \cdot I'(x + x', y + y'))}{Z'(x, y)} \quad (14)$$

$$Z'(x, y) = \sqrt{\sum_{x', y'} T'(x', y')^2 \cdot \sum_{x', y'} I'(x + x', y + y')^2} \quad (15)$$

g) Zero mean normalized square difference matching methods

$$\alpha_{znsq_diff}(x, y) = \frac{\sum_{x', y'} [T'(x', y') - I'(x + x', y + y')]^2}{Z'(x, y)} \quad (16)$$

In above equations, T is the grayscale image intensity of template, I is the grayscale image intensity of the image searching region, (x, y) , (x', y') and (x'', y'') represent the location coordinate in image searching regions and template. The perfect match will be small for the square difference matching methods and bad matches will be large. On the contrary, the perfect match will be large for the cross correlation method and bad matches will be small or 0. However, for the correlation coefficient methods, the perfect match will be 1 and a perfect mismatch will be -1 , and a value of 0 simply means that there is no correlation. The normalized methods can help reduce the effects of light/illumination change between the template and the image.

In CV-based displacement measurement, the normalized correlation coefficient matching method is the most popular one and there are numerous applications of the method. (Ye et al. 2013, 2015, 2016b; a) implemented the normalized correlation coefficient matching method to monitor the structural displacement of two long span bridges (Tsing Ma Bridge and Stonecutters Bridge), a steel arch bridge (Hangzhou Changyun Bridge) and several scaled bridges and structures in laboratory. (Brownjohn et al. 2017; Xu et al. 2016, 2018) implemented it to measure the

displacement of the Humber Bridge and a cable stayed foot bridge. (Busca et al. 2014) implemented it to measure the displacements of a railway bridge under train loads. (Pan et al. 2016; Tian and Pan 2016) implemented zero mean normalized method to measure the deflections of two railway bridges. References (Feng and Feng 2018; Xu and Brownjohn 2018; Ye et al. 2016c) reviewed the recent practical applications of CV-based displacement measurement using template matching.

The correlation coefficient matching methods (or normalized) need huge computation expenses during the template matching since there are lots of correlation/convolution operations. According to convolution theorem, the correlation/convolution in spatial domain becomes multiplication operation frequency domain. It saves the computation time once the correlation methods are calculated in frequency domain. (Feng et al. 2015a) implemented cross correlation in frequency domain for template matching and measured the displacement of a railway bridge. Generally, the template matching can achieve displacement measurement in pixel level and if the requirement of the practical application need higher accuracy, optimized searching process are necessary to achieve subpixel level. (Feng et al. 2015a) developed an subpixel level template matching based on upsampled cross correlation method. (Zhang et al. 2016a) applied two subpixel processing techniques including the modified Taylor approximation refinement and subpixel localization refinement to after the normal template matching using cross correlation in frequency domain. References (Berenstein et al. 1987; Foroosh et al. 2002; Liang et al. 2015; Pan et al. 2006) introduced grayscale interpolation, correlation coefficient curve fitting, phase correlation interpolation, space gradient and finite element for subpixel processing techniques. In general, subpixel is an estimation and it is a good complementary of pixel level image registration or

displacement measurement. However, it cannot go beyond the physical mechanism of sensors of digital cameras: the integral pixel level is the more reliable fraction of the matching estimation. In practical application, due the system and environmental noises, the precision of subpixel level estimation might be reduced.

Digital image correlation (DIC) also follows equations presented in this part and has become a well-established technique in experimental mechanics and material test area (Chen et al. 2018b; Schreier et al. 2009; Zhong et al. 2019a; b, 2017, 2018a; b, 2019c; Zhong and Quan 2017, 2018b). The difference between DIC and template matching is that DIC first divides the whole image into multiple sub regions with grids and uses template matching to calculate the displacement of the sub region. Then by using shape functions, the displacement in the whole image, in other word, full field displacement can be obtained. Full field strain estimation is also one of the tasks of DIC. (Pan et al. 2009) made a detailed survey of DIC and its application in displacement and strain application. (Spencer et al. 2019) reviewed the applications of DIC in static measurement.

(2) **Feature matching.** Instead of using the ROI to represent the tracking target, feature points (also called keypoints) extracted from the ROI can also represent the tracking target. The feature points are generally not a point, but a small region, e.g. a subimage with 5×5 pixels or 7×7 pixels, which is a low sample of target with distinct features such as intensities, gradients in different directions, textures, etc. Feature matching-based image registration can achieve more accurate performance than template matching because general template matching is usually considered as rigid matching, i.e. matching scenarios are translation and rotation. In general, when talking about a type of feature point, it refers to the feature detector and feature descriptor. Feature detector is a

subregion of an image selected by certain algorithms such as Scale Invariant Feature Transform (SIFT) (Lowe 2004), while feature descriptor is the matrix or a vector to describe the subregion. Figure 5 shows the SIFT detector and also visualizes the SIFT descriptors on the original image.

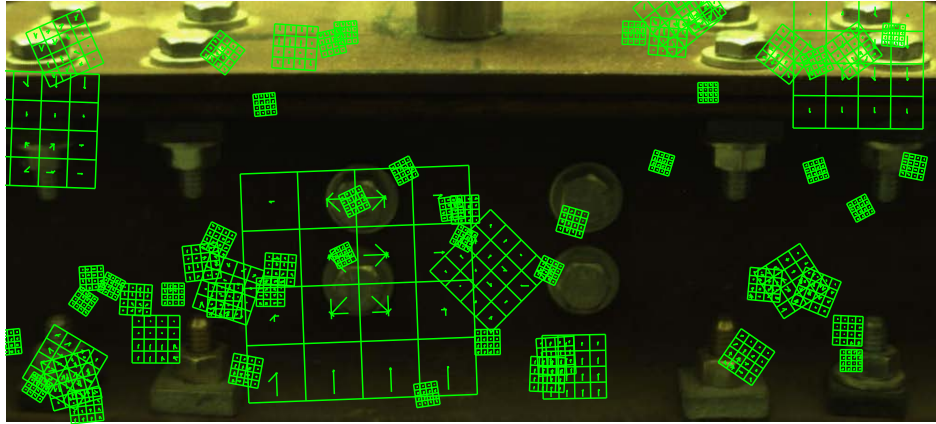


Figure 5 Visualized SIFT descriptor: image processed with the codes released by (VLFeat 2018)

The feature matching process is also a visual tracking process. Figure 6 shows the procedure of feature matching. First, the feature points in two different images are extracted, then the similarity of descriptors of feature points in different images are calculated by Euclidean or Hamming distances and based on the extent of similarity, feature points are matched. The higher the similarity, the smaller the distance is. The matching process can be executed by Brute-Force matching which traverses all the feature points but takes a long time (OpenCV 2019a) or K-Nearest Neighbor method and threshold can be used to reduce the computation expenses (Lowe 2004). The selection of Euclidean or Hamming distance is based on the type of the descriptor vector. The SIFT and SURF (Speeded-Up Robust Features) (Bay et al. 2008) descriptors are float type and Euclidean distance is needed, while the AKAZE (Accelerated-KAZE), FREAK (Fast Retina Keypoint), BRISK (binary robust invariant scalable keypoints), BRIEF (Binary Robust Independent

Elementary Features), ORB (Oriented FAST and Rotated BRIEF) descriptors are binary and Hamming distance is needed (OpenCV 2019b). Finally, the wrong matches are removed in this step. In general, the homography transform based random sample consensus (RANSAC) or least median square (LMS) are used in outlier removal.

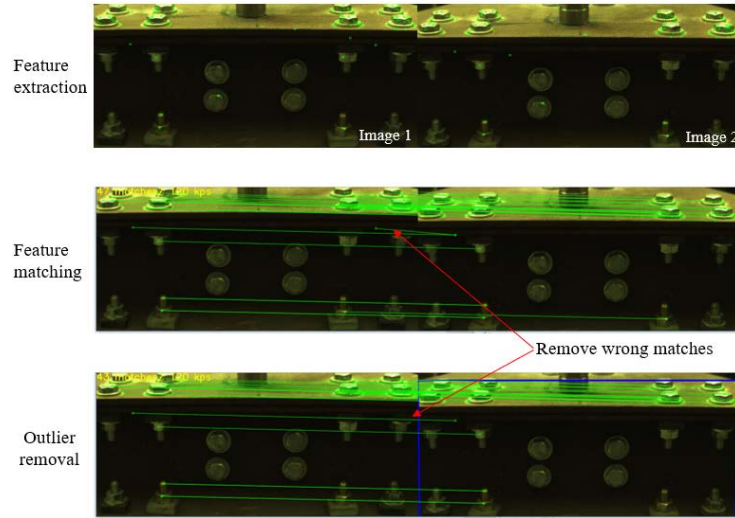


Figure 6 Feature matching procedure

The location change can be obtained from the average value of the change of each matched feature point or can be extracted from the estimated homography matrix. In this procedure, the matching achieves subpixel level. (Khuc and Catbas 2017) extracted Harris corner as feature point and by using FREAK as the descriptor to achieve subpixel level matching, they monitored the vibration of a stadium under human loads. (Khuc and Catbas 2016) also employed SIFT feature detector and descriptor to do feature matching and monitored the displacement of a pre-stressed transit guideway under train loads. Instead of using SIFT descriptors, (Dong and Catbas 2019) implemented the SIFT detector and VGG descriptor to do feature matching and this improved the measurement precision by about 24%. (Hu et al. 2017) implemented ORB detector and descriptor

feature matching to monitor the displacement of a viaduct.

The application scenarios of feature matching-based tracking are usually the cases when there're distinct textures on the surface of the structure and no manual marker is necessary. The advantages of feature points are scale and rotation invariants and stability under light change, while there is big uncertainty in feature extraction. Depending on the light condition, the threshold of feature extraction needs to be adjusted manually so that a certain number of feature points can be extracted. Up to now, there're still few discussions about the number to use in feature matching-based displacement measurement. It is still an open problem to decide the number of feature points and how to adjust thresholds for feature extraction. In addition, currently the application of feature matching-based displacement measurement is still limited to close range monitoring. It is still a challenge to do long distance monitoring by using feature matching.

(3) **Full field dense optical flow.** Optical flow is the velocity estimation of a video or an image sequence and it represents the motion between two images. In the mathematical language, it is a vector which points from the initial location to the end location of the motion. It could be caused by camera motion or the motion of the objects inside the field of view. To calculate the optical flow for two images, two basic assumptions are made (Szeliski 2011): (a) brightness constancy which means the pixel intensities of an object in an image do not change between consecutive frames; (b) temporal regularity which means the between-frame time is short enough to consider the motion change between images using differentials (used to derive the central equation below) and it assumes a small motion between two consecutive images (Dong et al. 2019b). With the first assumption the basic optical flow formula is

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (17)$$

where I is the image intensity, x and y are coordinates in the image and t is time. Expanding Eq. 17 with Taylor's Formula, and discarding the high order items with the second assumption, Eq. 17 can be expressed as:

$$I_x u + I_y v + I_t = 0 \quad (18)$$

where I_x and I_y are image gradients and I_t is the gradient at time t , u and v are the elements of optical flow vector in x and y directions. There are two unknowns in one equation, thus Eq. 18 is underdetermined. More constraints are necessary to solve the equation. Based on different assumptions, classic optical flow algorithms such as Lucas-Kanade algorithm (Lucas and Kanade 1981), Horn-Schunck algorithm (Lucas and Kanade 1981), Black and Anandan algorithm (Black and Anandan 1996), CLNL algorithm (Sun et al. 2014) and phased-based optical flow algorithm (Gautama and Van Hulle 2002) were proposed to estimate the optical flow at each pixel of the image, i.e., full field/dense optical flow. Due to the small motion assumption, the classic optical flow has a big advantage in small displacement estimation, while when estimating the optical flow with large displacement, pyramid architecture is implemented to reduce the estimation error (Bouguet 1999). (Khaloo and Lattanzi 2017) implemented four classical optical flow to monitor the displacement of a frame structure and compared their pros and cons. With the development of deep learning-based computer vision, optical flow estimation using deep learning has also made a great progress. References (Dosovitskiy et al. 2015; Ilg et al. 2017) proposed FlowNet and FlowNet2 which are CNN-based optical flow algorithm and they can estimate optical flow with large displacement. (Dong et al. 2019c) implemented FlowNet2 to monitor the displacement of a grandstand and a footbridge.

(4) **Sparse optical flow at feature points.** The full field dense optical flow methods calculate the flow vector at each pixel and need large computation expense, therefore it's difficult to achieve real-time online monitoring. Also, the classic full field optical flow methods face the challenge of handling the estimation at boundaries. Instead of calculating the full field optical flow, (Shi and Tomasi 1994) proposed to extract the feature points from image first and then calculate the optical flow at the feature points by using Lucas-Kanade algorithm. They made an assumption that at the feature point with a neighborhood of several pixels it has the same motion. It makes the Eq. (18) become overdetermined and can be solved with the least square methods. This algorithm simplifies the procedure of calculation, reduces the computation expense and eliminate the poor estimation at boundaries for classical optical algorithms. By calculating the optical flow vectors at the feature points at two different images, it achieves feature point tracking and in other words, feature matching just like the effects presented in (2). This kind of optical flow calculation is also called sparse optical flow. (Dong et al. 2019b) conducted forward estimation and backward estimation to calculate the bidirectional error and by setting a threshold, they could remove the erroneous optical flow estimation. In practical applications, the homography transform-based RANSAC and LMS can also be implemented to remove the outliers. (Yoon et al. 2016) combined the Lucas-Kanade optical flow and Shi-Tomasi corner (Shi and Tomasi 1994) to track the motion of a slender steel frame, and they implemented RANSAC to remove the outliers. (Dong et al. 2019b) combined Lucas-Kanade optical flow with SURF, SIFT and Shi-Tomasi corner separately to measure the displacement of a grandstand and a stadium under human loads. (Lydon et al. 2018, 2019) combined Lucas-Kanade optical flow with SURF and SIFT to measure the displacement of several bridges including a truss bridge, a prestressed concrete and a cable-stayed steel footbridge. (Hoskere et al. 2019) implemented Lucas-Kanade optical flow and Shi-Tomasi corner to process

the video captured from a UAV and measured the displacement of fiducial markers installed on a suspension bridge.

(5) **Geometry matching.** The geometry matching generally achieve target tracking by using the geometry features including edge, curve, square, circle or angle. Manual markers are necessary. The advantages of geometry matching are: (a) the scale change of tracking targets is easy to match; (b) the target can still be tracked even there's partial occlusion; (c) geometry matching shows robust properties to non-uniform illumination and partial edge blur, (d) it shows low computational complexity, and this reduces the computation time. The basic idea of geometry matching is to determine the geometrical position of the selected targets by edge detection or Hough transform. By comparing the geometrical positions of targets in different images, the matching is achieved. References (Chen et al. 2015a; Ho et al. 2012; Lee et al. 2017, 2006a; Lu et al. 2017; Tian et al. 2019) developed CV-based displacement measurement methods based on geometric matching. By tracking the scale change of a circle target, (Lu et al. 2017) estimated the structural displacement out of plane, while the limitation is that the axis of the camera and lens has to be perpendicular to the structural plane. (Xu and Brownjohn 2018) made a survey of targets with different geometric shapes and the corresponding processing approaches.

(6) **Color-based tracking.** The color of the surface of the structural element or manual targets is another good feature to track. The general color-based tracking algorithms include color-based template matching (National Instruments 2016) and histogram based mean-shift tracking (Ye et al. 2016a). The advantage of color-based tracking is that the target can be easily identified and tracked when the color of the target has a large difference with the background, still the performance of

color-based tracking is vulnerable to environmental light, image blur and noises. (Ye et al. 2016a) compared the performances of the color-based template matching with normalized correlation coefficient-based template matching in structural displacement measurement and observed that color-based template matching was not suitable for outdoor monitoring and long-distance monitoring.

(7) **Particle image velocimetry (PIV).** The initial application PIV is for the velocity monitoring of fluid (Kitagawa et al. 2005). By setting particles in the fluid field and capture the image of the motion of particles, the full field fluid can be tracked (Grant 1997). One of the characteristics of PIV is that it is good at tracking large displacements. However, the application of PIV is very limited in structural displacement measurements. (Hosseini et al. 2014) monitored the full field displacement and strain of steel and RC beams using PIV in laboratory. (Tian et al. 2018) implemented PIV to monitor the displacement of the human jumping in vertical direction. Although they stated the selection of PIV was because PIV could achieve more accurate measurement than others, they didn't provide any comparison analysis with the other methods or proof in that paper. To the best knowledge of the authors, there're very few comparison analyses reported in existing publications.

(8) **Deep learning-based object tracking methods.** With the wide application of deep learning in computer vision, object tracking has made great progress. The accuracy of deep learning-based object tracking has outperformed the traditional algorithms and seems to be going mainstream in object tracking area (Kristan et al. 2018). By building deep convolutional neural network and training the network with manually labeled dataset, the optimized neural network can achieve very

good tracking performance even with the cases of scale change, illumination change, different view, occlusion and background clutter. Currently, examples of deep learning-based visual tracking algorithms are Siamese FC (Wang et al. 2018b), Siamese Mask (Wang et al. 2018a), Siamese RPN++ (Li et al. 2018a), MFT (Bai et al. 2018), UPDT (Bhat et al. 2018). Although the deep learning-based object tracking algorithms have made great progress in accuracy and robustness, they require large volume of datasets and time to train their networks and the trained data also need to be labeled manually. In addition, even in the test stage, the processing still costs lots of time. Most of algorithms cannot supply real-time tracking. Also, the tracking algorithm predict the target location at pixel level, which makes the deep learning-based tracking algorithms not suitable for the structural displacement measurement with higher requirements, e.g., sub pixel level. In literature, the applications of deep learning-based structural displacement are very rare. Most of the applications still implement traditional tracking algorithms. Still, in the near future, with the development of fast graphic processing unit (GPU) processing speed and more efficient networks, the deep learning-based object tracking may make a big hit for the structural displacement measurement.

There're many visual tracking algorithms developed by the community of computer vision, even though not all of them are suitable for structural displacement measurement. The selection of visual tracking should consider the practical application scenarios, measurement accuracy and whether real time monitoring is required or not.

2.4.1.3 SHM Application: Using CV-based Displacement

Displacement obtained from CV-based methods can be further processed to become critical indicators for structural performance evaluation and health condition assessment. This part covers most of the categories of applications of CV-based displacement measurement in SHM. The applications are reviewed as following:

(1) **Structural behavior analysis.** (Ye et al. 2013) analyzed the range of vertical deflections of a long span suspension bridge under operational loads using the CV-based displacement measurement and also extracted the unit influence line of a cable stayed bridge by using CV-based displacement measurement during load test. (Tian and Pan 2016) measured the deflection profile of a multi-span railway bridge by using multi-point tracking during load test. (Guo and Zhu 2016) investigated the behavior of the sound barriers of a viaduct when a train was crossing by using the CV-based displacement measurement.

(2) **Load capacity analysis.** The displacement measured during load test under designated loads reflects the structural load carrying capacity. Comparing the displacement obtained from CV-based methods with the designated displacement limit is a non-contact approach for structural load carrying capacity. With the measured displacement, the load imposed on the structure can also be derived. (Ojio et al. 2016) combined the CV-based displacement measurement with the vehicle axle location on the deck to estimate the axle weight. (Lee et al. 2006a) did bridge load carrying capacity evaluation by using dynamic displacement records obtained from CV-based methods.

(3) **Dynamic analysis.** Structural dynamic displacement time histories can be processed to analyze the structural dynamic properties. (Xu et al. 2018) monitored the dynamic displacements of a cable-stayed footbridge under human crowds using CV-based methods and analyzed the instantaneous frequency and amplitude change when a large human crowd crossed the bridge. (Celik et al. 2019a) conducted human comfort analysis of a stadium using CV-based displacement records. (Dong et al. 2019b) monitored the displacement of a stadium under human jumping loads during a football game and analyzed the relation between the structural responses and the song played when human crowds were jumping on the structure.

(4) **Modal identification.** Multi-point displacement time histories can also be processed using operational modal analysis algorithms to extract structural modal parameters such as frequency, damping ratio and mode shape. References (Dong et al. 2018; Feng and Feng 2016, 2017; Yoon et al. 2016) monitored the displacements of multiple points on structures using one camera and extracted the modal parameters using the displacement histories. (Dong et al. 2019b) synchronized two cameras to measure multiple points of a grandstand and extracted the modal parameters from the displacement time histories. (Hoskere et al. 2019) roved a camera with a UAV to measure the displacements of multiple points in part of a whole bridge and after scanning all the bridge, they extracted the mode shapes of the whole bridge by connecting the mode shape from each bridge part. References (Bharadwaj et al. 2019; Poozesh et al. 2017; Srivastava and Baqersad 2019) applied 3D DIC and multiple cameras to monitor the full field displacement of wind turbine structures and extracted the operational deflection mode shapes. (Ji and Chang 2008a) estimated the mode shape of a stayed cable using stereo vision techniques. (Chen et al. 2015b, 2018a) implemented phase-based optical flow, Fourier transform and motion magnification to extract the

instantaneous mode shape of a beam in lab and a truss bridge. (Yang et al. 2017a; b) applied blind source separation, phase based optical flow, motion magnification and edge detection to analyze the video of a small steel frame structure and extract the operational mode shape. (Fioriti et al. 2018) extracted the mode shape of several ancient constructions by using motion magnification. (Tian et al. 2019) applied CV-based displacement measurement method to measure the displacement of a beam during impact test and extracted the structural frequencies, scaling factor and mode shape.

(5) **Modal updating.** Finite element model (FEM) need to be updated with real monitoring data. Modal parameters can be used to update the FEM. (Feng and Feng 2016) first obtained the modal parameters from a CV-based measurement and then used these modal parameters to update the FEM.

(6) **Damage detection.** The changes of global characteristics of structures such as boundary condition and stiffness can induce the change of structural modal parameters. With the modal information extracted from CV-based displacement records, damages at global level can be detected. (Felipe-Sesé and Díaz 2018) investigated the damage detection of a car bonnet using 3D DIC full field displacement measurement and modal analysis. (Feng and Feng 2017) monitored the displacement of multiple points on a small beam and extracted the mode shape change before and after the damage of the beam. By comparing the damage indicators calculated using mode shape elements, they detected the damage of the beam. (Cha et al. 2017b) filtered the displacement measurement obtained from phased-based optical flow of a structure and extracted the stiffness and damping coefficient values to detect damages on boundaries of cantilever beams in laboratory.

Currently most of studies using CV-based methods to do damage detection at global level are for small structures and limited in laboratory.

(7) **Cable force monitoring.** The tension force in cable is related to its vibration frequencies and deformations. With CV-based displacement measurement, the vibration frequencies or cable deformation can be obtained. Then the cable force can be estimated. References (Feng et al. 2017; Kim et al. 2013; Zhao et al. 2017) estimated the cable force by using the vibration frequencies obtained from CV-based displacement records. (Ye et al. 2016b) measured the deformation of a cable segment and estimated the cable force by using the relation between force and material deformation.

2.4.2 CV-based Load (Input) Estimation

The CV-based techniques are not only used to monitor the structural responses (output), but also to monitor the external loads (input). This section reviews the CV-based human load and vehicle load estimation respectively.

2.4.2.1 Human Load Estimation

Human load is the main live load on structures such as stadia and footbridges, and the accurate human load estimation imposed on structures would be very helpful to investigate the human-induced vibration problems such as vibration comfort, vibration serviceability, human-structure interaction, response spectra and structural design. By tracking the motion of human behavior such

as jumping and bobbing, the human loads imposed on structures might be estimated by using Newton's second law. (Mazzoleni and Zappa 2012) implemented DIC to track the motion of a single person and a small group of people respectively to get the load time histories of human jumping and bobbing. (Zheng et al. 2016) trained a tracking-by-detection system to monitor the motion of a small group of people to estimate the human crowd loads. The very interesting part is that their tracking system can be updated with new detected subjects during the tracking process. This makes the load estimation more accurate. (Li et al. 2018b) estimated the dynamic load factors of human crowd jumping by using 3D motion tracking system to track the markers on human bodies. (Celik et al. 2018a) implemented dense optical flow to track the motion of a small group of people on the grandstand in laboratory and a large human crowd jumping during a football game. Comparison study with load cells, accelerometers and FEM showed promising performance of the proposed approach. (Tian et al. 2018) implemented PIV to track the human jumping and obtained the acceleration of human body. Utilizing the acceleration of human jumping as input and the acceleration of structures collected from accelerometers as output, they performed the impact test of a footbridge and extracted the frequency, modal scaling and mode shape.

2.4.2.2 Vehicle load estimation

Vehicle load is the most common external excitations to bridges and the estimation of vehicle load is significant to the bridge structural behavior analysis. By using only CV-based methods, it is hard to estimate the exact axle weights. Still, with the size of the detected vehicle, the approximate load range can be estimated. Sometimes, with the known vehicle weight, the vehicle distribution on

structures is essential to analyze structural behavior. (Catbas et al. 2012b; Zaurin et al. 2015; Zaurin and Catbas 2010b; a) implemented background subtraction to estimate the load distribution of a vehicle on the bridge deck and combining with the strain data from strain gauges and the known vehicle weight, the unit influence line (UIL) of bridges could be extracted. A UIL-based damage detection approach was proposed and verified in laboratory, and load rating based on UIL was validated in a real life bridge. (Khuc and Catbas 2018) implemented AdaBoost technique and Cascade classifier using histograms of oriented gradients features to train and detect vehicle types. With known vehicle weight, the unit influence surfaces (UIS) were extracted and a UIS-based damage detection approach was proposed and verified in laboratory.

(Zhang et al. 2019a) retrained the Faster RCNN to detect and track the spatial information of vehicles on bridge deck, including vehicle location, axle and length. (Pan et al. 2018) trained a vehicle detector by using HOG feature and RF classifier to monitor the vehicle distribution and speed on a bridge with noise barrier. With the obtained vehicle information, they investigated the effect of vehicle-induced aerodynamic load to the noise barriers. (Dan et al. 2019) estimated the moving vehicle load on pavement by combining the weigh-in-motion (WIM) system and visual tracking. (Chen et al. 2016) combined the WIM system, multiple cameras and visual tracking to estimate the vehicle load distribution in long span bridges. (Daize et al. 2018) investigated the relation between the detected vehicle length and the weight from WIM data and proposed to predict vehicle load based on the vehicle length. (Chen et al. 2014a) estimated the traffic load spectra of a long span bridge by detecting vehicles from traffic videos.

2.4.3 SHM Using CV-based Input-Output Information

As reviewed in 2.4.1 and 2.4.2, with CV-based methods both the structural input and output can be obtained. Still, the studies of SHM using CV-based input-output information are very limited. (Ojio et al. 2016) calibrated the relation between the size of a vehicle and the distance from the camera and used the size-distance relation to recognize the vehicle location in one lane. Based on the displacement of a girder measured by camera and the location information recognized by size-distance relation, they extracted the UIL of a girder and created a contactless bridge WIM system with UIL. (Dong et al. 2019a) used one camera to extract the location of a moving vehicle on bridge deck by visual tracking and another camera to measure the bridge displacement using CV-based methods. By synchronizing the input-output signal with audio-based pattern recognition, they extracted the UIL of a two-span bridge and the UIL of a footbridge. With the UIL, they predicted the human loads on the footbridge. Although there're very few SHM applications on CV-based input-output, it is still very promising to see more development in this area.

2.5 Challenge and Concerns in Real-life Practices of CV-SHM

With the benefit of computer vision techniques, numerous difficult problems faced in the past have been solved in the area of SHM at local and global level. While CV-based approaches cannot cover every respect of SHM tasks and within current research and application, there are still many challenges and concerns to better use CV-based approaches and make our infrastructures safe, reliable and resilient.

2.5.1 Challenges

2.5.1.1 Influence Factors to Vision Systems

When using CV-based techniques to solve problems in SHM, such as local damage detection and displacement measurement, influence factors such as hardware (electric noise, camera self-heating and lens distortion), environment (illumination change, partial occlusion, ground vibration, wind, rain, fog), limitations of computer vision algorithms may have adverse effect on the performance of the vision system. (Dong et al. 2019e) proposed a robust vision-based method for displacement measurement under adverse environmental factors such as illumination change and fog interference using spatio-temporal context learning and Taylor approximation. References (Feng and Feng 2018; Ye et al. 2016c) made detailed survey for the error source analysis of vision based systems.

2.5.1.2 Long Term Monitoring

At present, research and application of CV-based approaches focus on short-term inspection or monitoring. The stability and reliability of CV-SHM in long time span are still a big issue. Problems in long term monitoring such as management of big data, data interpretation, process of abnormal data, and the maintenance and calibration of CV-SHM systems are still challenges.

2.5.1.3 Subtle and Fast Motion Detection from Images

Structures such as short span concrete bridges and small or medium height buildings have large stiffness so the deformations of them are small and the vibration frequencies are very high. With

the limitations of imaging sensors in resolution and exposure time, the resolution and sampling rate of CV-based system cannot outperform the conventional sensors such as strain gauges, displacement transducers or accelerometers. Even with motion magnification, the subtle motion components in videos can be extracted, while the quantification of absolute deformation is still a problem.

2.5.1.4 Quantification of Local Damage

Although with the development of deep learning based local damage detection, the quantification of a certain local damage is still a challenge, such as the estimation of depth of crack/spalling. The researches of quantification of local damages in literature are very limited as reviewed in Section 2.3.

2.5.1.5 Simplification of Deep Learning for Local Damage/Change Detection

At present, the research of using deep learning to do local damage detection requires large datasets to train neural networks. Even just for one type of local damage detection, e.g., crack, a certain dataset may need to be created separately. For different application scenarios, crack detection of different bridges, buildings or tunnels may need to retrain the networks. The procedure of deep learning-based local damage detection needs to be simplified. How to transfer the human experiences for local damage detection to neural networks is still a big challenge.

2.5.1.6 Selection of Markers in CV-based Displacement Measurement

In literatures, manual markers can indeed improve the measurement accuracy while it takes some

efforts to install manual markers on structures. In the cases when there's no access to the structure to install manual markers, algorithms without using markers can achieve displacement measurement. The performance comparison between CV-based methods with and without manual markers for displacement measurement in literature didn't make big difference when a marker was not necessary in some cases. The balance of trade-off between marker and no marker is still a problem in real-life practices for displacement measurement.

2.5.2 Concerns

2.5.2.1 Local and Global Condition Assessment Using CV-based Methods

Up to now, local and global SHM using CV-based techniques achieve high accuracy in local damage detection or displacement measurement, while how to utilize the monitoring results to assess the structural condition according to current standards or codes is still a big concern. Whether or not CV-SHM can really assist the asset owners to do decision making and make better management and maintenance of structures might be one of the criteria for the success of CV-SHM.

2.5.2.2 Connection between Computer Vision-based SHM at Local and Global Level

As mentioned in the Introduction part, understanding how local damage or condition changes affect the performance of the whole structure is very important. It raises the question of how to build the connection between computer vision-based SHM at local and global level. The bidirectional relation between CV-SHM-LL and CV-SHM-GL can benefit the fully condition

assessment and the study of the integrity of different elements of a structure.

2.5.2.3 Industry Certificate, Standards and Codes for CV-SHM

Although CV-SHM has made a great progress and gained increasing attention in the community of SHM and states departments of transportation or other funding agencies have supported the CV-SHM projects every year, the industry certificates, standards and codes for CV-SHM are necessary. Still, up to now, relative standards or codes are very rare.

2.6 Summary

This chapter introduces the current state of art in research and state of practice for computer vision-based structural health monitoring at local level (e.g. element, crack, delamination) and global level (e.g. structural deflection, vibration). Concepts, approaches and experiences of computer vision techniques used to solve SHM problems at local and global level in practice are presented along with representative studies. Computer vision-based approaches have the advantages such as non-contact measurements, data collection from a long distance, rapid and low cost application as well as reduced labor needs with minimum interference or intrusion to the daily operation of structures. CV-SHM can monitor numerous structural conditions and collect data towards damage indices or parameters such as crack, spalling, delamination, displacement, acceleration, modal parameters, load factors, while it cannot fulfill every aspect of SHM tasks. It should be indicated that there are a number of adverse influence factors can easily affect the accuracy of the computer vision applications if possible precautions are not taken during the image data collection and data

analysis. Such challenges and concerns do still exist and these must be taken into account in the future research and applications. CV-SHM is still very promising in fulfillment of the structural condition assessment and CV-SHM can be regarded as a good complement to conventional SHM.

CHAPTER THREE: MARKER FREE MONITORING OF THE GRANDSTAND STRUCTURES AND MODAL IDENTIFICATION USING COMPUTER VISION METHODS

Previously published as Dong, C. Z., Celik, O., and Catbas, F. N. (2019b). “Marker free monitoring of the grandstand structures and modal identification using computer vision methods.” *Structural Health Monitoring*, 18(5–6), 1491–1509.

3.1 Introduction

Recently, the combination of camera technology and computer vision algorithms has led to great advancements in the field of SHM (Catbas et al. 2012b; Yeum and Dyke 2015; Zaurin et al. 2015; Zaurin and Catbas 2010a; b). In past studies (Ye et al. 2016c), the use of computer vision to measure structural displacement was discussed extensively since displacement is a critical indicator of a structure’s performance. Current vision-based displacement measurement methods face many obstacles in field applications because they typically use manual markers, which must be attached to the surface of a selected structure (Feng et al. 2015a). Additionally, most researchers have used a digital image correlation (DIC) algorithm to do vision-based structural displacement measurement along with the manual markers (Henke et al. 2015; Jáuregui et al. 2003; Kohut et al. 2013; Lee et al. 2006a; Lee and Shinozuka 2006; Olaszek 1999; Park et al. 2010; Ye et al. 2016f, b, a; c; Yoneyama and Kitagawa 2007). Digital image correlation has many disadvantages because it is easily affected by changes in illumination, slight occlusions, blurring due to motion, target shape deformation, scale change and rotation. These disadvantages are critical issues for field

applications and thus have limited the popularization of vision-based monitoring methods. Using manual markers is a way to improve the robustness of the digital image correlation methods but it does not solve the access requirement problem. Feng et al.(Feng et al. 2015b) introduced an orientation-code matching (OCM) based displacement measurement method and compared the performance of different target types including target panel (actually QR codes), feature, rivet and LED.

Khuc and Catbas proposed a new vision-based displacement measurement method that did not require installation of manual markers and instead used robust features extracted from the image as virtual makers (Khuc and Catbas 2016, 2017). The displacement measurement was achieved using feature matching between the consecutive images. However, in their work, they were limited to a single point dynamic displacement measurement and did not focus on multi-point measurement. Other studies were able to achieve multi-point displacement monitoring but only through multiple manual markers(Fukuda et al. 2010; Ho et al. 2012; Ye et al. 2015).Yoon et al.(Yoon et al. 2016) introduced a target-free approach for vision-based structural system identification using Kanade–Lucas–Tomasi (KLT) tracking algorithm and Shi-Tomasi corners. This work could accommodate multi-point displacement measurement of a six-story building model in the laboratory, however, it did not provide verification with conventional displacement sensors. In addition, the vibration of the structure was taken perpendicular to the line of sight of the camera, which would be a limitation in field applications and for the measurement of several common structures, such as high-rise buildings, grandstands and bridges. Celik et al.(Celik et al. 2018a) applied the sparse optical flow and dense optical flow algorithms to estimate the load time histories of lively individuals and crowds. Literature (Chen et al. 2015b; Yang et al. 2017a; b,

2018b) shows the full-field vibration mode identification approaches using phase-based video motion magnification and edge detection. The instantaneous mode shapes extracted from videos are vulnerable to the background clutter. Although displacement measurement results compared with laser vibrometer on a cantilever beam exist in the literature (Chen et al. 2015b), such work mainly focus on the modal information and its application of structures in laboratory. Displacement measurement is still an important indicator especially for real-life applications. It is useful and important to obtain multi-point displacement records and analyze them for comprehensive assessment of structural performance.

In this chapter, a general computer vision-based structural dynamic monitoring framework that utilizes marker-free techniques is proposed and demonstrated. Physical markers used for target localization are replaced with virtual markers (feature points) that are extracted from video frames by robust feature detection algorithms. These virtual markers represent textures or other unique surface characteristics of the structure. The virtual markers can be selected and plugged into the framework according to the best application for each scenario, which makes the whole framework more adaptive. The extracted virtual markers are combined with optical flow to achieve general dynamic displacement monitoring. In addition, the proposed method can synchronously monitor multi-point dynamic displacement responses in real time. In this framework, a mechanism for synchronization of multi-camera and conventional sensors is also designed.

The proposed method is first verified on a grandstand in the laboratory by identifying modes of the structure through multi-point displacement records, and then comparing the vision-based results with traditional displacement sensors (i.e., potentiometer and accelerometers). Subsequently,

a second experiment is conducted in a football stadium during a game to validate the feasibility of field application and identification of operational modes under human crowd motion. It is shown that the computer vision-based method provides good results and it is also applicable for SHM.

3.2 Methodology and System Development

3.2.1 General Procedure for Vision-based Displacement Measurement System

The flowchart of the proposed vision-based displacement measurement method is illustrated in Figure 3 and the general procedure is presented in Section 2.4.1.1. Here they are not repeated. Critical steps related to this chapter will be discussed in detail in the following.

3.2.2 Camera Calibration

In this chapter, the term “camera calibration” mainly refers to the scale ratio calculation. Before the scale ratio calculation, the camera should be setup to ensure that the captured image contains the necessary objects and has high quality resolution. The calculation of scale ratio follows Section 2.2.

3.2.3 Feature extraction

Features are the unique characteristics of a structure used to distinguish measurement regions for visual tracking. Extracting features that are good and robust is the foundation of visual tracking and helps assure a high level of accuracy of the displacement measurement. As mentioned

previously, these features can be low-level features like image intensity, color, and geometrical shape. However, these features are easily occluded by adverse factors such as illumination change and partial occlusion, which are challenges for visual tracking. Generally, manual markers are employed to distinguish measurement points from their surroundings. However, manual markers introduce new problems because they require additional work and also require access to the structure being measured. In this situation, higher level features, such as corners and other feature points, are the better options because they are more distinct than edges. A main advantage of the feature points is that key points permit matching even in the presence of clutter (occlusion) and large scale and orientation changes (Szeliski 2011). Consequently, feature points are the preferred features in the proposed method. Feature points, also called key points, are obtained from gradient operations, moments, or other mathematical steps on image matrices. In the proposed method, the necessity for physical markers is eliminated by extracting useful and strong features within the scene. With contributions from the field of computer vision, there are many kinds of feature points that can be selected. A feature point should be selected based on the difficulty of the problem, application scenarios, requirements of on-line or off-line monitoring, and environmental factors. The most prevalent feature points are the Harris corner, Shi-Tomasi corner, SIFT, SURF, FAST, BRIEF and ORB. Brief characteristics of each feature are listed in Table 1 and further information can be found in reference (Alahi et al. 2012; Bay et al. 2008; Calonder et al. 2010; Harris and Stephens 1988; Lowe 2004; Rosten and Drummond 2006; Rublee and Bradski 2011; Shi and Tomasi 1994). Figure 7 displays two examples of feature points from the same image. Feature points remove the need for manual markers and can be regarded as virtual markers, which are distinct from the surroundings in the image and perform the functions that the manual markers did.

Feature extraction is a critical step for achieving successful marker-free vision-based displacement measurements.

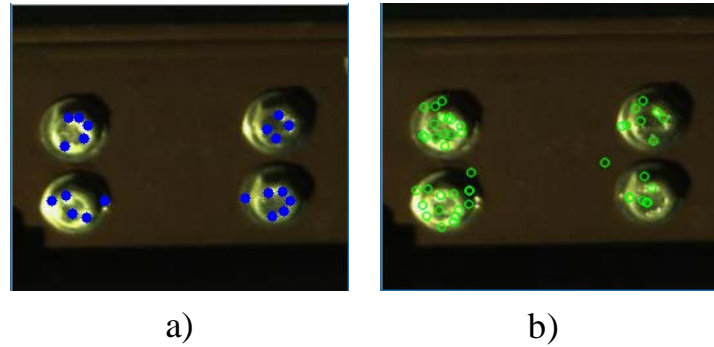


Figure 7 Examples of feature points: a) Shi-Tomasi corners; b) SURF

Table 1 Summation of different feature points

No.	Feature points	Characteristics
1	Harris Corner	The most basic feature, not scale invariant
2	Shi-Tomasi Corner	Modified from Harris corner, better than Harris corner
3	SIFT	Scale-invariant features, insufficient speed
4	SURF	Faster version of SIFT
5	FAST	Faster than SURF and can be used in real-time
6	BRIEF	Less memory, faster matching, and higher recognition rate than SIFT
7	ORB	ORB is much faster than SURF and SIFT, and ORB descriptor works better than SURF

3.2.4 Visual tracking using optical flow

Figure 8 shows the motion issues that can be developed in an image sequence (video), which is a function of both space (x, y) and time t . In this study, optical flow is used to solve the visual tracking task. Optical flow refers to the estimation of a vector field of local displacement for a sequence of images. It is the pattern of object motion between two consecutive frames in a sequence. This motion can be due to the movement of the objects or the camera. As illustrated in Figure 8b, in the $(t-1)^{\text{th}}$ frame of the sequence, the ball is at location A , and in the t^{th} frame, the ball is at location B . The vector pointing from the initial location to the current location is the optical flow. If the ball's location in t^{th} frame back is put back into the $(t-1)^{\text{th}}$ frame, location B' , then this optical flow vector can be represented by s . To calculate the optical flow for two images, two basic assumptions are needed:

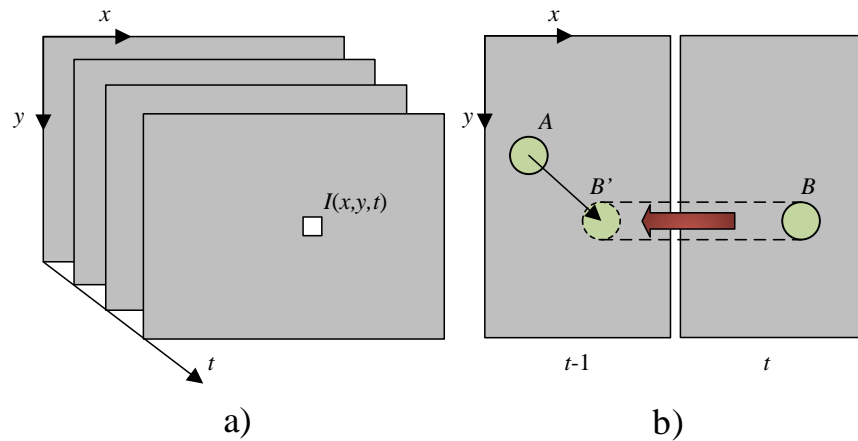


Figure 8 Motion in image sequence: a) temporal change; b) spatial change

1) Brightness constancy: the pixel intensities of an object in an image do not change between consecutive frames;

2) Temporal regularity: the between-frame time is short enough to consider the motion change between images using differentials (used to derive the central equation below), which assume a small motion between two consecutive images;

In many cases, these assumptions may not hold, but for small motions and short time steps between images, it is a good model. Consider a pixel $I(x, y, t)$ in the first frame. It moves by distance (dx, dy) in next frame taken after a period of time dt . Since the pixels are the same and their intensity does not change, the following equation can be applied:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (19)$$

Then assuming I is a differentiable function, by expanding the first term using the Taylor series, removing higher order terms and dividing by dt , the following equation can be obtained:

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \quad (20)$$

or

$$I_x u + I_y v + I_t = 0 \quad (21)$$

where $I_x = \frac{\partial I}{\partial x}$, $I_y = \frac{\partial I}{\partial y}$, $I_t = \frac{\partial I}{\partial t}$, $u = \frac{dx}{dt}$, $v = \frac{dy}{dt}$.

Eq. 21 is called the Optical Flow equation. I_x and I_y are image gradients and I_t is the gradient at a time t . A unique solution cannot be obtained from this equation with two unknown variables, (u, v) . It is known as the aperture problem (Szeliski 2011) and is shown in Figure 9. Aperture is a

patch which defines a region of view. Patches with gradients in at least two different orientations are easy to localize.

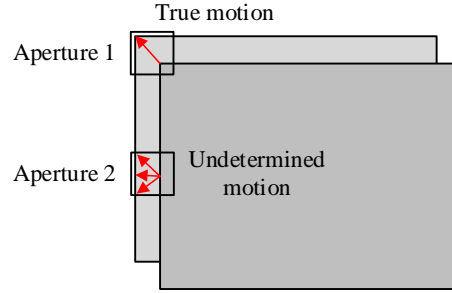


Figure 9 Aperture problem

As is outlined in Figure 9, the motion of Aperture 2 cannot be determined due to the inadequate amount of boundary condition. However, as in Aperture 1, by enforcing some spatial consistency, it is possible to obtain solutions. There are many methods to solve this problem, one of which was developed by (Lucas and Kanade 1981). This method operates under the assumption that neighboring pixels have similar motion. With this assumption, it is possible to stack many of these equations into one system, as in Eqs. 22, for a neighborhood of n pixels.

$$\begin{bmatrix} I_x(p_1) \\ I_x(p_2) \\ \dots \\ I_x(p_n) \end{bmatrix} u + \begin{bmatrix} I_y(p_1) \\ I_y(p_2) \\ \dots \\ I_y(p_n) \end{bmatrix} v + \begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \dots \\ I_t(p_n) \end{bmatrix} = 0 \quad (22)$$

or in the format below:

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \dots & \dots \\ I_x(p_n) & I_y(p_n) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \dots \\ I_t(p_n) \end{bmatrix} \quad (23)$$

Eqs. 23 are over-determined since they have more equations than unknowns; they can be solved using the least square methods:

$$E(u, v) = \int_{x,y} (I_x u + I_y v + I_t)^2 dx dy \quad (24)$$

From the equation above Eq. 12 is obtained:

$$\frac{\partial E(u, v)}{\partial u} = \frac{\partial E(u, v)}{\partial v} = 0 \quad (25)$$

Eqs. 26 to 28 are derived:

$$2(I_x u + I_y v + I_t)I_x = 0 \quad (26)$$

$$2(I_x u + I_y v + I_t)I_y = 0 \quad (27)$$

$$\begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_y I_x & \sum I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix} \quad (28)$$

Eqs. 28 can be represented as a structural tensor representation as in Eq. 29:

$$\begin{bmatrix} T_{xx} & T_{xy} \\ T_{xy} & T_{yy} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} T_{xt} \\ T_{yt} \end{bmatrix} \quad (29)$$

When the matrix on the left, $\begin{bmatrix} T_{xx} & T_{xy} \\ T_{xy} & T_{yy} \end{bmatrix}$, is invertible, Eqs. 29 can be solved, and the solutions are:

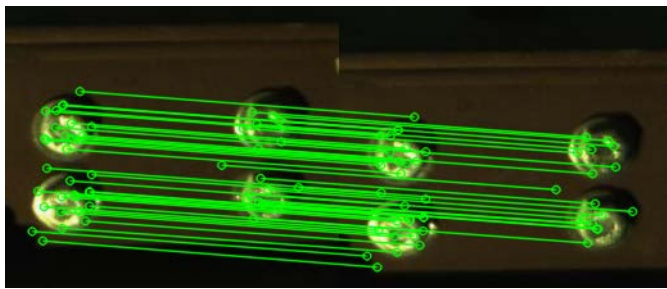
$$u = \frac{T_{yt}T_{xy} - T_{xt}T_{yy}}{T_{xx}T_{yy} - T_{xy}^2} \quad (30)$$

$$v = \frac{T_{yt}T_{xy} - T_{xt}T_{yy}}{T_{xx}T_{yy} - T_{xy}^2} \quad (31)$$

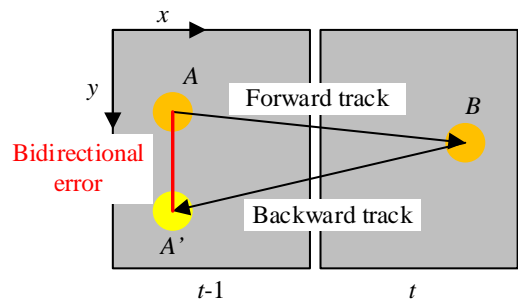
When the matrix on the left of Eqs. 28 is compared with a feature point detector of the feature extraction in reference (Harris and Stephens 1988), it is seen that the matrix is invertible, and a feature point is also found at the same point. It means that at the feature points' location, Eqs. 28

and Eqs. 29 can be solved, making feature points easy points to track. With feature points, the assumption made in Lucas-Kandade method is not a drawback but a good thing.

Feature points are first extracted in the current frame and then the optical flow vectors are calculated to track the locations of the feature points in the next frame. For the optical flow vector (u, v) , the new location of the feature point (x, y) in next frame is $(x+u, y+v)$. Figure 10a shows the visual tracking results of feature points in two consecutive frames. To improve the accuracy of tracking and eliminate the outliers of tracked feature points in the next frame, bidirectional error detection is conducted, as shown in Figure 10b. When going forward, the optical flow vector is calculated from the $(t-1)^{\text{th}}$ frame to t^{th} frame, and the new location of A in the $(t-1)^{\text{th}}$ frame is B in the t^{th} frame. However, when going backward, the optical flow vector is calculated from the t^{th} frame to the $(t-1)^{\text{th}}$ frame, and the location of B in the t^{th} frame might not exactly be A but A' in $(t-1)^{\text{th}}$ frame. The difference between A and A' is called the bidirectional error. A threshold is needed to eliminate the error in the tracking when the bidirectional error is too high.



a)



b)

Figure 10 a) Visual tracking results; b) Illustration of bidirectional error

The optical flow calculation method used in this research, Lucas-Kanade method, is a sparse optical flow method that calculates the motion at the feature points. While Lucas-Kanade method can also be applied to calculate dense optical flow (calculating optical flow vector at each pixel) if the calculate window slides all over the image. There are other algorithms, such as Horn-Schunck method (Horn and Schunck 1981), Farneback method (Farneback 2003), Block match method (Liu and Zaccarin 1993) and Phase based optical flow (Fleet and Jepson 1990). Comparisons of listed examples are listed in Table 2.

Table 2 Optical flow methods

No.	Methods	Characteristics
1	Lucas-Kanade	Sparse/dense flow, local smooth assumption, can use pyramid estimation for large motion estimation, fast and easy to implement
2	Horn-Schunck	Dense flow, global smooth assumption, cannot use pyramid estimation for large motion estimation, sensitive to noise, slow
3	Farneback	Dense flow, use polynomial expansion, consider local information, can use pyramid estimation for large motion estimation, slow
4	Block match	Dense flow, only integer displacements, only use local information, can use pyramid estimation for large motion estimation, two-frame time derivatives are inaccurate, slow
5	Phase based optical flow	Dense flow, use phase information, more robust to smooth shading and lighting variations, unstable in the vicinity of phase singularities, slow

Even though with the optical flow on every pixel, the displacement of selected regions can be obtained, the noise will be added into the final results due to unremarkable flow in these regions. In this study, the focus is on motion tracking of selected measurement regions – specifically robust feature points - not the whole image. Therefore, the Lucas-Kanade algorithm is more suitable for our tracking task than dense optical flow methods. To negate the drawbacks of the optical flow method due to the two basic assumptions stated above, the steps below need to be followed:

- 1) For the first assumption, select features that are resistant to illumination change, since during practical application of structural displacement measurement, changes in lighting can occur.
- 2) For the second assumption, use Lukas-Kanade with Pyramid to solve problems due to large motion.

3.2.5 Displacement calculation

Since this study focuses on multi-point displacement measurement, each of the measurement points is represented by a region of interest (ROI), which is a subset selected from the whole image as mentioned above. In this part, the procedure to calculate the displacement of each ROI is introduced and take one ROI for example. With feature extraction and visual tracking, the displacement decrement for each feature point from the selected ROI in the $(t-1)^{\text{th}}$ frame to the one in t^{th} frame is obtained, $(u_{t-1 \rightarrow t}^i, v_{t-1 \rightarrow t}^i)$, where i is the number of feature points. Then the displacement decrement of each selected measurement point from the $(t-1)^{\text{th}}$ frame to t^{th} frame is calculated by averaging the decrements for each feature point:

$$\begin{bmatrix} u_{t-1 \rightarrow t} \\ v_{t-1 \rightarrow t} \end{bmatrix} = \begin{bmatrix} \frac{1}{n_t} \sum_{i=1}^{n_t} u_{t-1 \rightarrow t}^i \\ \frac{1}{n_t} \sum_{i=1}^{n_t} v_{t-1 \rightarrow t}^i \end{bmatrix} \quad (32)$$

where n_t is the total number of the tracked feature points. Then the displacement in pixel coordinates in the t^{th} frame from the initial frame is:

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^t u_{i-1 \rightarrow i} \\ \sum_{i=1}^t v_{i-1 \rightarrow i} \end{bmatrix} \quad (33)$$

Here t is equal or greater than 2. When t is equal to 1, u_1 is defined as 0 since displacement at the first frame is regarded as 0.

Using the scale ratio, SR , calculated in the section of camera calibration, the actual displacements of the selected measurement point at the t^{th} frame are

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = SR \begin{bmatrix} u_t \\ v_t \end{bmatrix} \quad (34)$$

where X_t and Y_t are the horizontal and vertical displacements, respectively. With Eq. (34), the horizontal and vertical displacements of one ROI at time t , i.e. one measure point, is obtained. For multi-point displacement measurement, the same procedure will be executed on all of the selected ROIs, which represent all the measurement points to get the displacements of all the measurement points.

3.3 Laboratory verification

3.3.1 Experimental Setup

This section focuses on verifying the feasibility and performance of the proposed displacement measurement method. The first experiment used for verification is a grandstand, depicted in Figure 11. The structure is modeled after those found in a football stadium. Throughout a football game,

spectators will stand up, jump, and dance, producing lots of vibration. These same scenarios are simulated on the grandstand scale model by having people jump on the grandstand. In this study, only the displacement of the front beam is measured. Figure 12 shows the experimental setup. Here five ROIs are selected as the measure point for the proposed method, i.e. P1-5. At each measure point a conventional displacement sensor (i.e., potentiometer) and an accelerometer are installed. The cameras here are MindVision- MV-GE131gc-t with a maximum frame rate of 60 Hz, a resolution of 1280 pixel \times 960 pixel and the zoom lens has a focal length of 5~100 mm. The cameras are connected to the same data acquisition system as the potentiometer. The sampling rate for potentiometer is 100 Hz and it is down-sampled to be comparable to the camera. The laboratory verification is divided into three experiments, which are introduced in detail below.

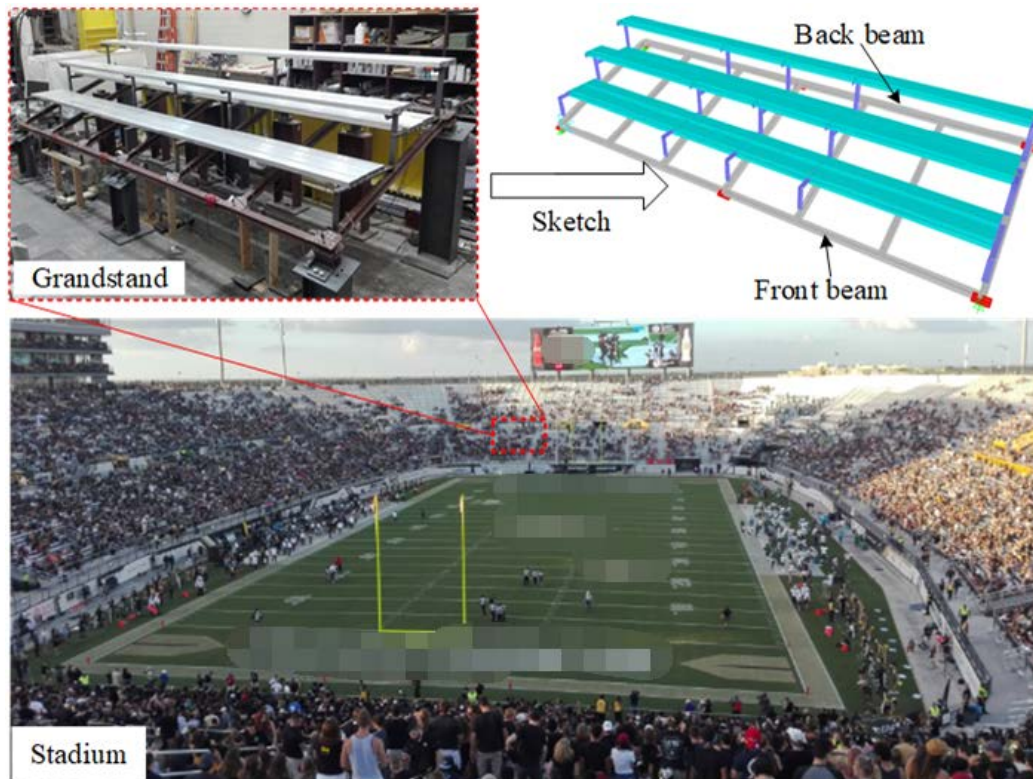


Figure 11 Grandstands in laboratory and stadium

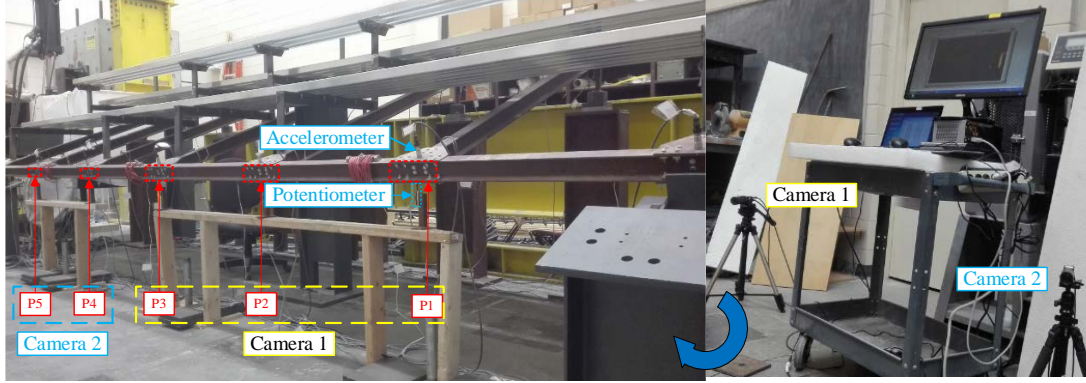


Figure 12 Experimental setup

3.3.2 Comparative Study of Displacement Measurement Using Different Features

In the first experiment, P2 on the front beam of the grandstand was selected as the displacement measurement point, as illustrated in Figure 12. During the experiment, one person stood on the grandstand and jumped, as one camera recorded the motion of P2 with 30 frames per second and a potentiometer measured the displacement of P2 at a sample rate of 30Hz simultaneously. Shi-Tomasi corners and SURF features were chosen as the virtual markers respectively. The displacement is obtained using the proposed, aforementioned procedure for vision-based displacement measurement system. Figure 13 reveals that the displacement records from the vision method match those from the potentiometer well and that they accurately depict the movements of the test structure: first, the person came onto the grandstand, causing an increase in displacement (0s~5s), then walked to P2, producing fluctuations in the displacement (5s~7s), stood for two seconds (7s~9s), and then began to jump (9s~17s), which produced a continuous up and down pattern. The person then stopped, resumed jumping for two seconds (17s~19s); then finally got down from the grandstand, returning the displacement to zero. From this experiment, it can be

seen that in this lab case, both Shi-Tomasi corners and SURF features give the same measurement results compared with the conventional displacement sensor (potentiometer).

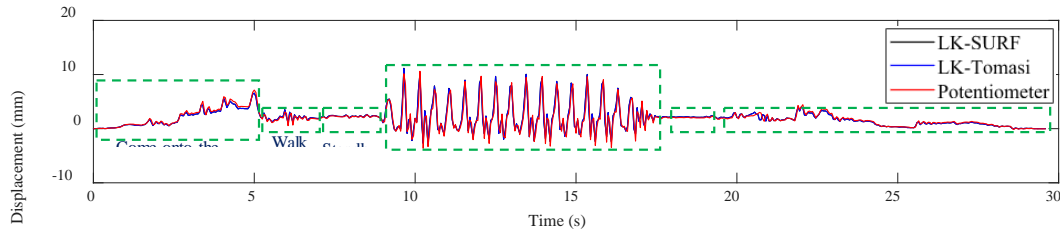


Figure 13 Comparison of displacement data from vision using different features and potentiometer at P2

Although both the proposed method and Khuc and Catbas's previous methods (Khuc and Catbas 2016, 2017) include feature extraction, the way how these features are used to obtain the displacement is different in this proposed work. In Khuc and Catbas' work (Khuc and Catbas 2016, 2017), after feature extraction of both of the two consecutive images, feature matching are performed between these two images using the minimum Euclidean distance (for SIFT) or shortest Hamming distance (for FREAK) of the feature points' descriptor vectors. Feature matching is a critical step to determine the location changing of the selected regions in the two consecutive images for displacement measurement. While in the proposed method, after feature extraction, the optical flow vectors at the locations of feature points are obtained with bidirectional calculation (forward and backward). Although both approaches can finish the tracking task and build feature matches in two consecutive images, and both need to extract feature points from two consecutive images, the bidirectional calculation of optical flow is a bidirectional prediction and check, which can discard bad feature matches. Khuc and Catbas' methods employed trimmed mean algorithm (for SIFT) and planar geometric transformation (for FREAK) respectively to discard bad feature

matches. As is stated in literature (Khuc and Catbas 2016) since the percentage in the trimmed mean algorithm is a parameter, it must be adjusted corresponding to a particular monitoring condition. Instead of using a different distance to carry out feature matching and different operations to discard bad matches, the proposed method gives a general and convenient framework, which just needs to plug in different feature points into bidirectional optical flow calculation to give good feature matches. Planar geometric transformation can be regarded as an additional step in the proposed method to further discard bad matches and improve the final results. (Yoon et al. 2016) used the MLESAC modeling fitting method to remove the displacements that are not consistent with the dominant geometric transformation between two consecutive frames, but did not use bidirectional optical flow calculation. Up to now, there is no research study on how much the planar geometric transformation can improve the performance in discarding bad matches after bidirectional optical flow calculation.

In this experiment, Shi-Tomasi corner and SIFT are employed respectively to be combined with bidirectional optical flow calculation to obtain displacement records. No additional steps are taken to further discard bad matches. The feasibility of the proposed framework is validated by comparing the results from two approaches and the actual displacement measurements using potentiometer. In future studies, the comparative evaluation of different feature points will be performed to further investigate their superiority and inferiority.

3.3.3 Multi-point Displacement Measurement Using Multi-camera

In order to carry out structural identification using computer vision, multiple cameras are needed to measure all the required responses since one camera cannot measure multiple points due to the

distance limitation in the laboratory and the limitation of the camera lens. In the second experiment, the feasibility of using multi-camera to measure displacement synchronously was demonstrated and verified. Two cameras and one potentiometer were used to measure the same point (P3) at the same time. Since each camera can include multiple measurement points with different depth of field, the scale ratio of the ROI at the measurement point was calculated separately. Both of the camera axes were perpendicular to the motion plane of the measure points so that the scale ratios were calculated. Figure 14 shows that the displacement records from both cameras matched well with those from the potentiometer measurements. Both of the displacement records basically gave the same structural response for the following loading procedure: first, the grandstand was still (0s~4s), then the person came onto the grandstand, causing an increase in displacement (4s~6s), then walked to P2, producing fluctuations in the displacement (6s~8s), stood for two seconds (8s~10s), and then began to jump (10s~16s), which produced a continuous up and down pattern. The person then stopped, resumed jumping for a second (16s~17s); then finally got down from the grandstand, the grandstand displacement came back to zero. These results from these two cameras indicate that the proposed multi-camera displacement measurement method can work synchronously and accurately. The correlation coefficient between the displacement time histories from Camera1 and potentiometer is 98.76%, while the correlation coefficient between the displacement time histories from Camera 2 and potentiometer is 97.93%. Since no person occupied the grandstand within the first four seconds, it can be assumed that the grandstand was stationary during that time. The displacement time histories in the first four seconds from both cameras are selected for statistical analysis. The standard deviation is chosen to represent the minimum small change of the proposed method (Khuc and Catbas 2017; Xu et al. 2018). The minimum small change of displacement using the proposed method and the cameras is 0.0154 mm. Although it is

not better than that of potentiometer, which is 0.0064 mm, using the proposed method and both cameras is still quite good. Comparing this value to the maximum displacement range (around 18mm), the accuracy is roughly $0.0154/18=0.0856\%$.

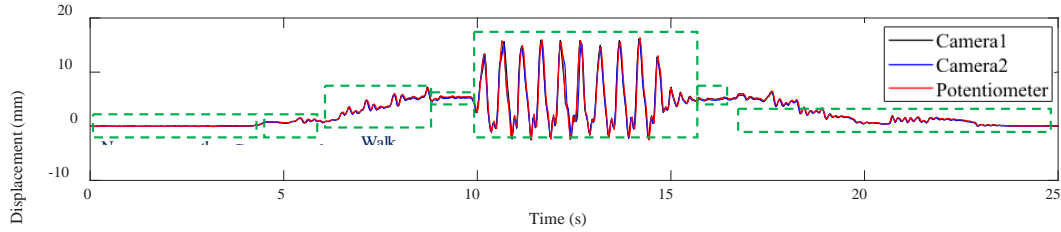


Figure 14 Comparison of displacement data from vision using different cameras and potentiometer at P3

During the experiments, all the cameras and the potentiometers were synchronized using a NI Multifunction I/O Device with the model number of USB-6343. The USB-6343 offers analog I/O, digital I/O, and four 32-bit counters/timers for PWM, encoder, frequency, event counting, and more. Onboard NI-STC3 timing and synchronization technology delivers advanced timing functionality, including independent analog and digital timing engines and retriggerable measurement tasks (National Instruments 2017). The cameras used here have the external trigger function. The cameras were triggered to capture images by the edge-triggered generated by the digital I/O of USB-6343, meanwhile responses of the conventional sensors including potentiometer and accelerometers were acquired by analog I/O of USB-6343. These tasks were synchronized with one internal timer of USB-6343 to make sure multiple cameras and sensors work synchronously.

3.3.4 Structural Identification

In this section, the displacement time histories obtained from the proposed method are used for structural identification. As shown in Figure 12, two cameras were employed to measure the displacement of the front beam of the grandstand. P1~P3 were measured by Camera 1, and P4 and P5 were measured by Camera 2. At the same locations (P1~P5), five accelerometers were installed to measure the accelerations from the structural vibrations. The experiment was performed under impact excitation. Both cameras and accelerometers recorded the free vibration attenuation process of the grandstand synchronously after the impact pulse. The frame rate of the cameras was 60 frames per second and the sample rate of the data acquisition system attached to the accelerometers was 100 Hz.

Figure 15 and Figure 16 illustrate the displacements from the proposed method and the accelerations from the accelerometers, respectively. The Complex Mode Indicator Functions (CMIF) (Catbas et al. 2004; Wang et al. 2016b; a) were calculated from the time histories from the vision method and accelerometers as shown in Figure 17 and Figure 18. In these two figures, the first (blue line) and second (red line) singular value curves of CIMFs are shown. In this experiment, the first singular value curve (blue line) is much clearer and is used to pick the peaks of the modes. Three different modes were successfully identified with pick-peak method from the CMIF plots.

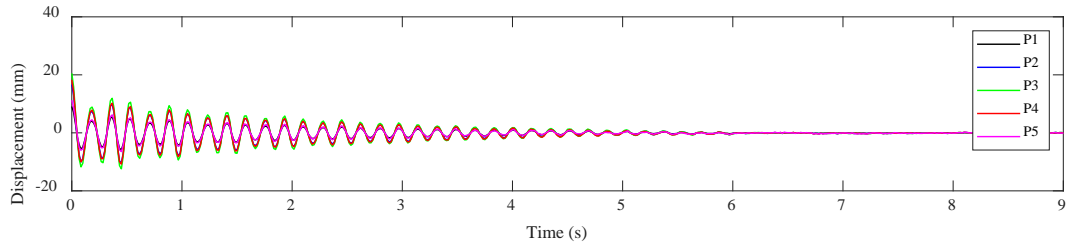


Figure 15 Displacement time histories from computer vision

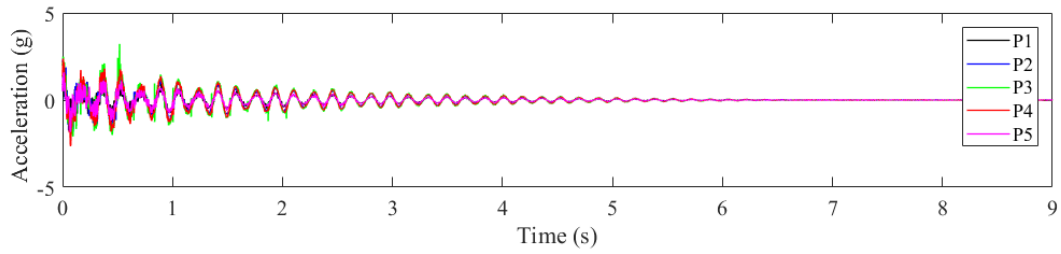


Figure 16 Acceleration time histories from accelerometers

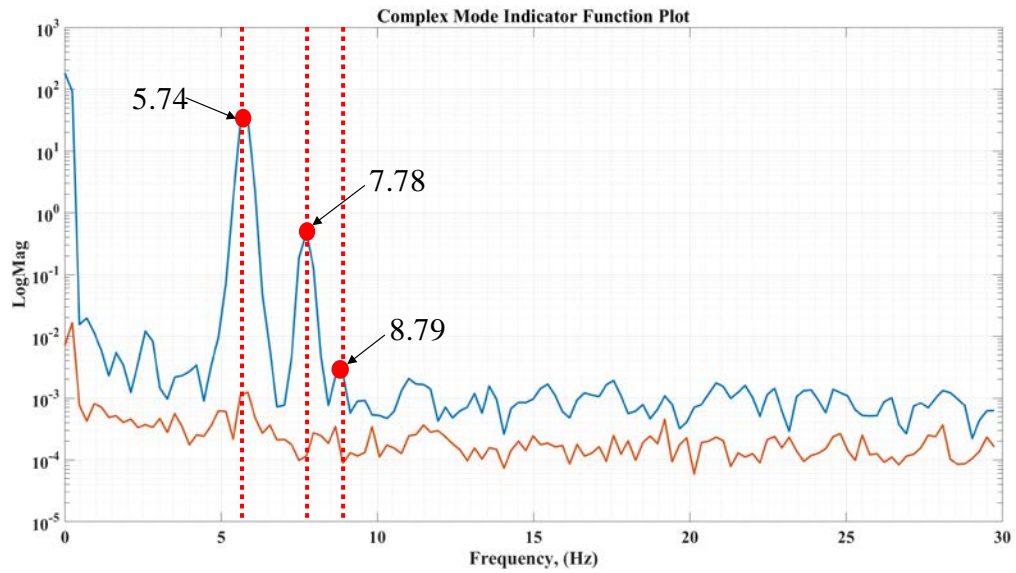


Figure 17 CMIF calculated by displacement time histories from vision method

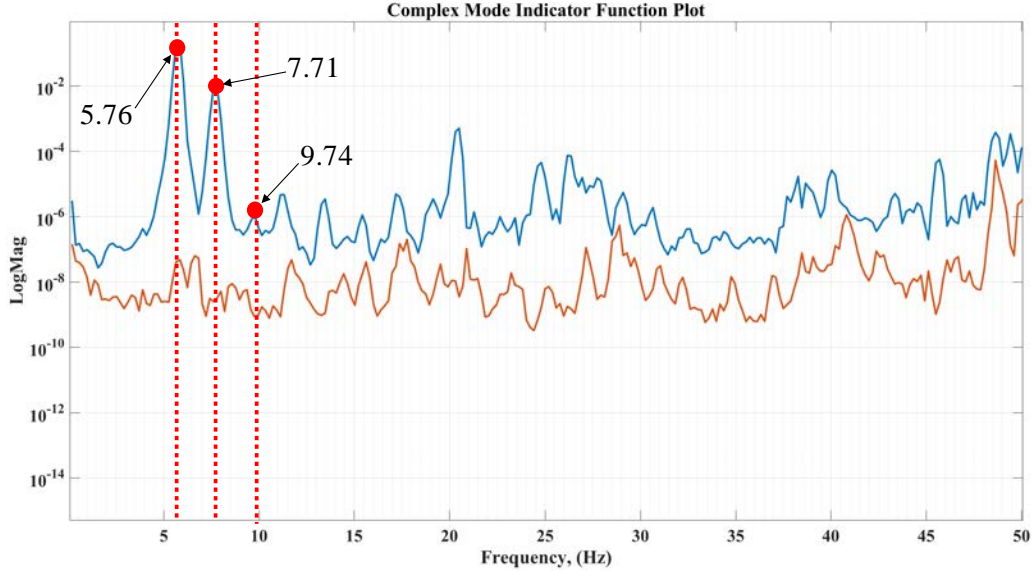


Figure 18 CMIF calculated by acceleration time histories from accelerometers

Table 3 shows the comparison of modal parameters obtained from vision and accelerations. The first three frequencies obtained from the vision method are 5.74 Hz, 7.78 Hz and 8.79 Hz, and those obtained from accelerations are 5.76 Hz, 7.71 Hz and 9.74 Hz. From the comparisons one can see that the differences between the modal frequencies are small (the first and second are less than 1% and the third is less than 10%). The difference between the damping ratios of the first order are also small, just 2.9%, but the differences between the second (264.9%) and third are larger (19.62%). The first three damping ratios obtained from the vision method are 1.36%, 2.08%, and 1.68%, and those obtained from the acceleration data are 1.40%, 0.57% and 2.09%. Both percent error of damping ratios and also difference of damping ratios are given in percentages to provide a more complete perspective of the differences. The large differences of damping ratio might come from the different type of data since one is displacement and another is acceleration. Usually in vibration displacement signal is relatively weaker than acceleration data and easy to be

contaminated by noise. That might be a reason to explain the large difference between the damping ratio obtained from the vision data and acceleration. However, it should be noted that damping is one of the most challenging modal parameters, and further studies need to be conducted especially when computer vision data are utilized.

Table 3 Comparative study of modal parameters

Mode	f (Hz) Vision	f (Hz) Acc	Diff. (%)	Damp ratio (%) Vision	Damp ratio (%) Acc.	Error % and Difference (%)
1	5.74	5.76	0.35	1.36	1.40	2.9, (0.04)
2	7.78	7.71	0.91	2.08	0.57	264.9, (1.51)
3	8.79	9.74	9.75	1.68	2.09	19.62, (0.41)

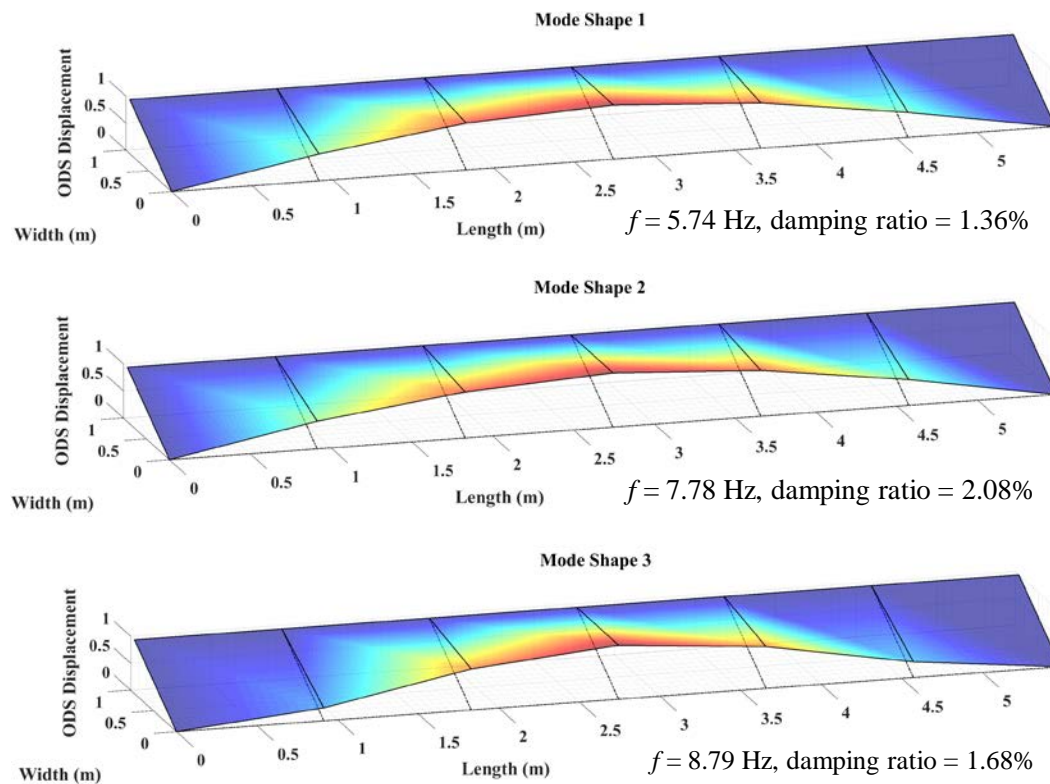


Figure 19 The first three modal shapes of the front beam from vision method

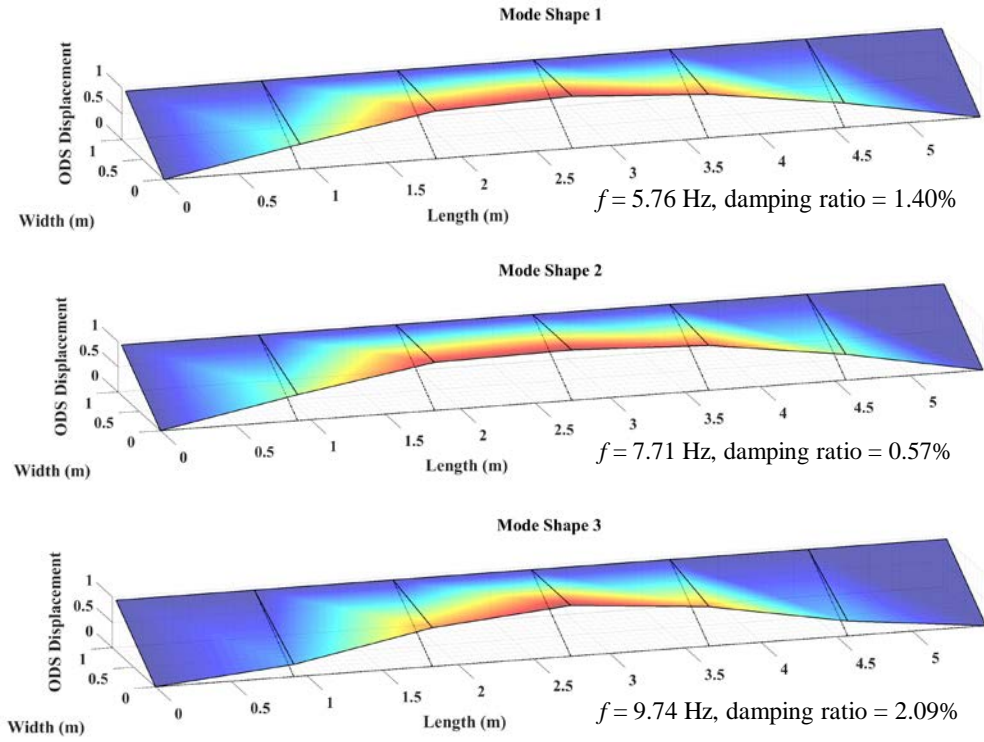


Figure 20 The first three modal shapes of the front beam from accelerometers

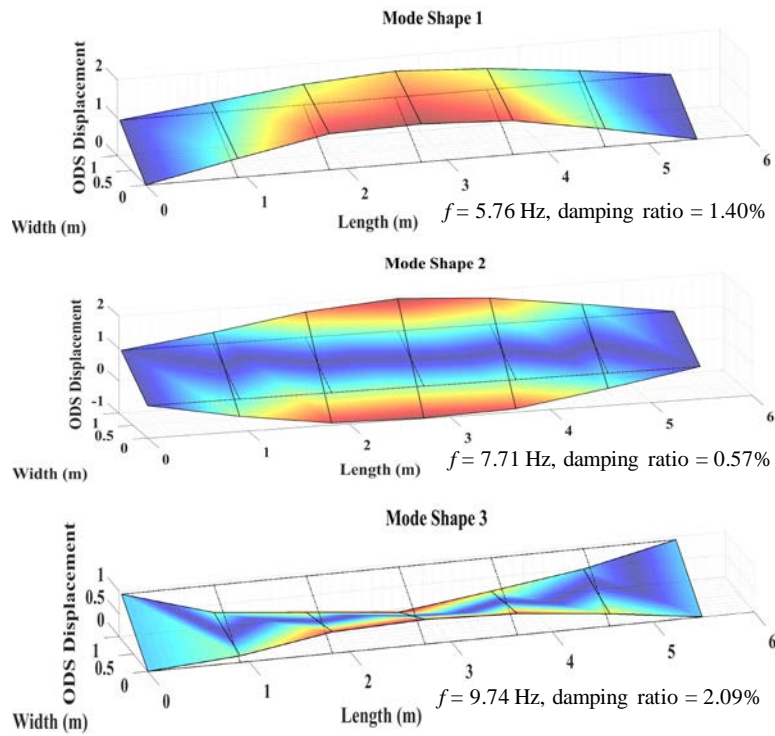


Figure 21 The first three modal shapes of the full structure from accelerometers

Figure 19 and Figure 20 show the first three modal shapes obtained from the vision data and acceleration data, respectively. Here the modal shapes are operating deflection shapes (ODS). The first three modal shapes look similar as bending modes of the front beam, which is because the modes are extracted only using the information of the front beam of the grandstand. Without the data from the back beam, the deflections of the back beam are set as zero. When using additional measurements, DOFs (Degrees of freedom) are added from the back beam, the validity of the identified modal vectors can be better observed visually. As shown in Figure 21, the first mode shape indicates a bending mode and the second and third mode shapes indicate two different torsion modes. But these mode shapes are consistent with those shown in Figure 19 and Figure 20, which are a subset with reduced DOFs. The experimental results indicate that it is feasible to use the proposed vision-based method for modal identification for extracting the natural frequency and modal shape. More studies are to be conducted to explore the accuracy of damping ratios using computer vision methods.

3.4 Field Application

3.4.1 Experimental Setup

In this section, a field application that was performed in a football stadium (as shown in Figure 11) is discussed and used to verify the feasibility of the proposed displacement measurement method and structural identification. As shown in Figure 22, a beam under the grandstand was selected for monitoring. At a predetermined measurement point, a potentiometer and an accelerometer were

installed, and a camera was positioned 10.21 meters away (33.5 ft) from the structure with an upward angle of 13°.

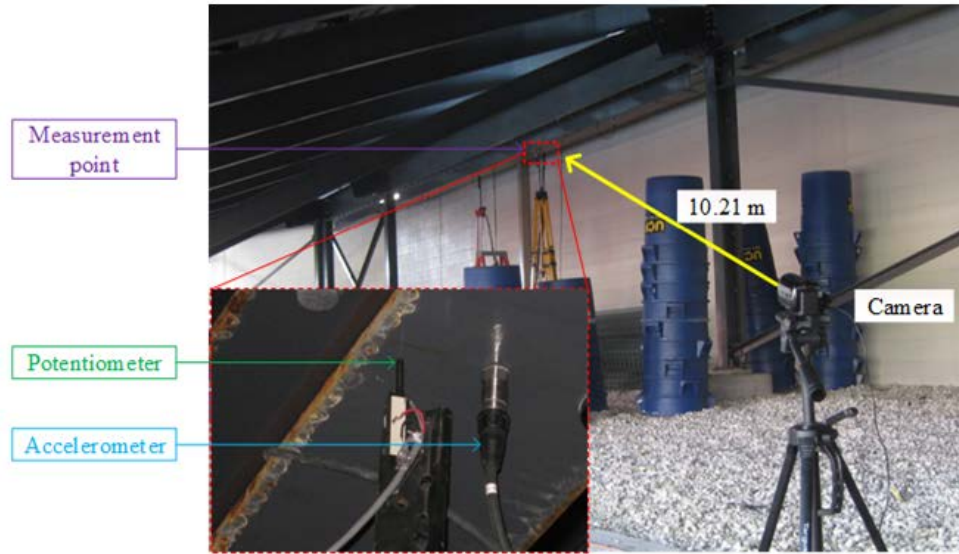


Figure 22 Grandstand from a real structure

The sensors and camera recorded the structural vibrations synchronously during periods of intense crowd motion throughout the football game, such as during a kick-off or touchdown. The sample rate of the potentiometer and the accelerometers was 60 Hz and the frame rate of the camera was 30 Hz. In this experiment, structural vibrations, which were produced by crowd jumping during a touchdown scored by a football team, were recorded. The song that the crowd jumping synchronized was the chorus of the “Zombie Nation” song. The time history of the chorus is illustrated in Figure 23.

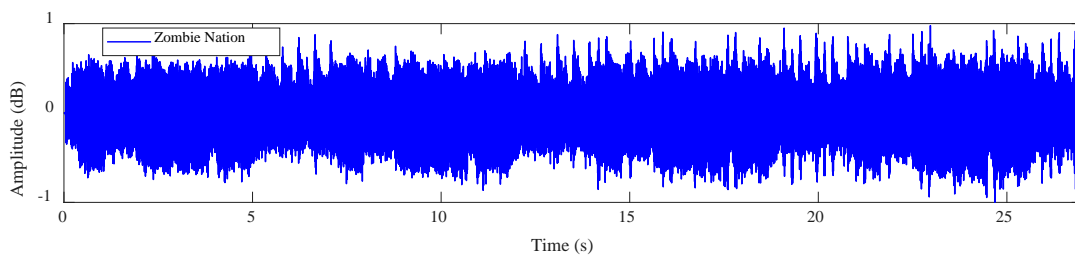


Figure 23 Time history of the chorus part of “Zombie Nation” song

3.4.2 Analysis and Results

Figure 24 illustrates the displacement time histories obtained from the proposed vision method and the potentiometer and Figure 25 illustrates the acceleration records obtained from the accelerometers. The results from computer vision method are consistent with those from the potentiometer measurements, which indicates that the proposed method is feasible for displacement measurement in field applications. The comparison of operational modal identification between vision and accelerometer-based data also gives satisfactory results as shown in Figure 26 and Table 4. It can be seen from this data that the computer vision and accelerometer results for the first three operational modal frequencies under a human jumping load are almost the same. Figure 27 illustrates the periodogram power spectral density estimate of the time history for the chorus part of the “Zombie Nation” song. Figure 26, Figure 27 and Table 4 indicate that the operational modal frequency results are closely spaced with the modal frequencies extracted from the chorus part of the “Zombie Nation” song. Even though there is a slight difference between the results from Figure 26 and Figure 27, it can still be concluded that both results obtained from the vision-based method and from accelerometers provide reliable operational modal frequencies.

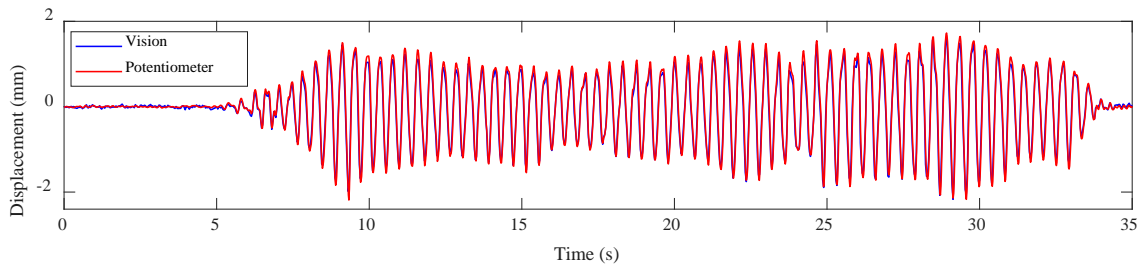


Figure 24 Comparison of displacement data from vision and potentiometer

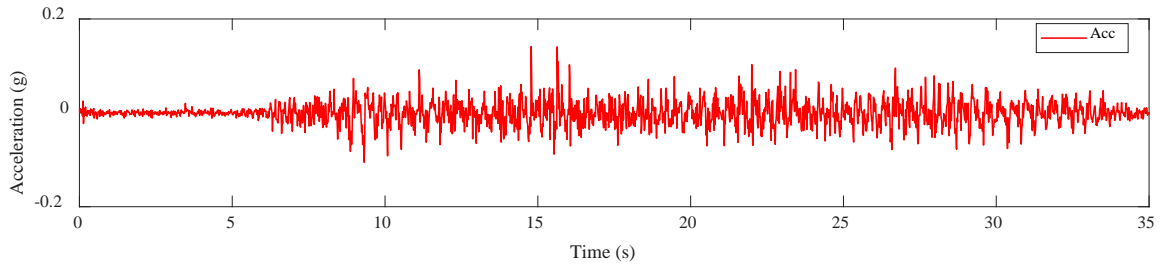


Figure 25 Acceleration time histories from the accelerometer

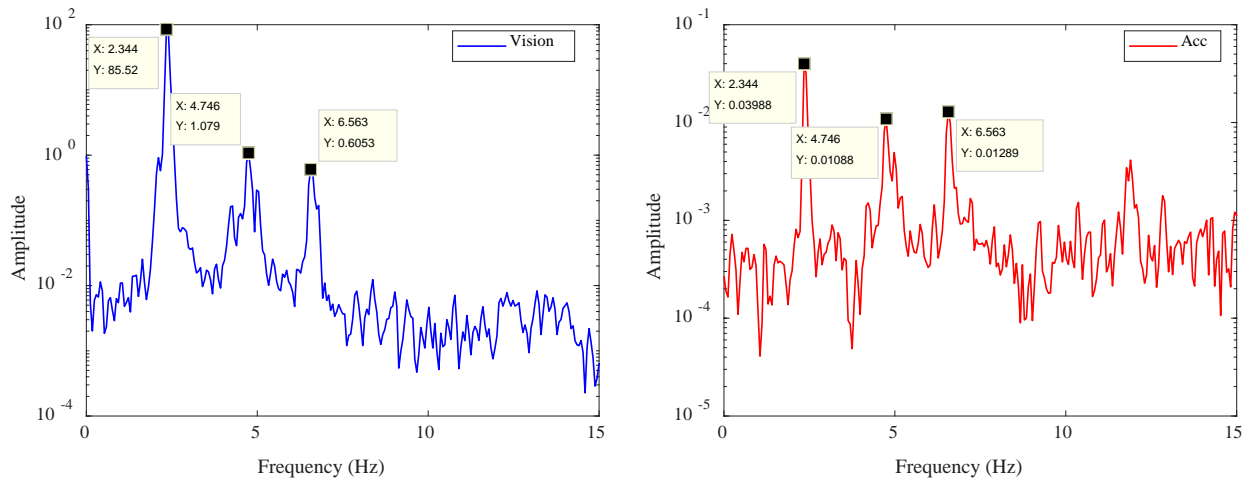


Figure 26 Comparison of operational frequencies under human jumping: a) vision; b) accelerometer

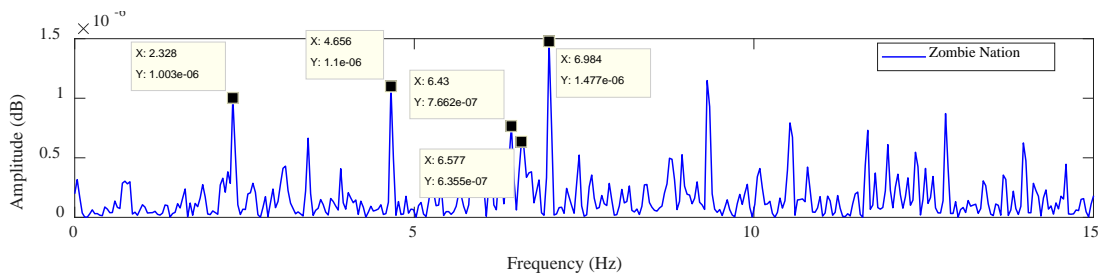


Figure 27 Periodogram power spectral density estimate of time history of the chorus part of Zombie Nation

Table 4 Comparative study of operational modal frequencies from vision and accelerometers

Mode	f (Hz) Vision	f (Hz) Acc	Difference between Vision and Acc. (%)	f : corresponding range in Zombie Nation
1	2.344	2.344	0	2.328
2	4.746	4.746	0	4.656
3	6.563	6.563	0	6.430, <u>6.577</u> , 6.984

3.5 Summary

In this chapter, a computer vision based structural dynamic monitoring method using marker-free and multi-camera synchronization techniques is proposed. An adaptive framework for the proposed method is introduced in detail, including camera calibration, feature extraction, visual tracking and displacement calculation. Instead of manual markers, this study extracts various kinds of feature points as virtual makers and combines them with optical flow based tracking algorithms that can be utilized for different application scenarios. The performance and accuracy of the proposed method are validated by comparing the displacement results with those obtained from a potentiometer sensor on a large grandstand structure in the laboratory. A modal identification for the structure is conducted using multi-point displacement records obtained from the proposed method, a multi-camera setup, impact testing, and the CMIF method. The computer vision-based results are checked by comparing them with the results from the conventional accelerometers. The proposed method is also validated by comparing with conventional displacement and acceleration sensors from an experiment in a stadium. The comparative operational modal analysis of the real

stadium structure using conventional sensors and cameras provided accurate results under human-induced loading during football games. It is feasible to use the proposed method for structural health monitoring and modal identification particularly for natural frequencies and mode shapes.

The proposed method and application to civil structures especially to bridges have major positive impacts for structural safety and assessment. There are sensor based structural health monitoring (SHM) technologies that are available yet permanent and continuous SHM may be economically feasible for the major (e.g. landmark) or structurally critical structures. In the meantime, a large population of routine structures (such as highway bridges) will still rely on visual inspection by the inspectors/engineers. Cost-effective and easy to use camera based technologies can support field inspections of such structures while also eliminating time-taking field installation of sensors. Besides the data obtained using the proposed framework can be further processed to conduct structural damage detection, model updating, condition assessment, etc., which are also currently carried out using conventional sensors. As a result, cost and time could be saved for the SHM of ordinary bridges. It is anticipated that such SHM would support decision-making and planning for structure management and maintenance. Future work will focus on widening the field of application to not only grandstand structures, but possibly to bridges, buildings, tunnels, dams and power transmission towers, as well as focusing on developing more robust algorithms for scenarios with harsh environmental conditions. Although it is very promising to use the vision based method as stated above, there are still some limitations, such as the errors induced by the camera shaking, illumination change, image processing time and other adverse environmental factors. Also, from a practical point of view, putting a camera underneath a grandstand and going through a sophisticated algorithm to reveal displacement might be cumbersome than attaching

accelerometers. This can be resolved by putting all the processing procedures in a package and just give the displacement at the end. Future work will focus on how to eliminate errors caused by the adverse environmental factors and make it more convenient to use.

CHAPTER FOUR: A NON-TARGET STRUCTURAL DISPLACEMENT MEASUREMENT METHOD USING ADVANCED FEATURE MATCHING STRATEGY

Previously published as Dong, C. Z., and Catbas, F. N. (2019). “A non-target structural displacement measurement method using advanced feature matching strategy.” *Advances in Structural Engineering*, 22(16), 3461–3472.

4.1 Introduction

Structural Health Monitoring (SHM) has been widely explored and implemented by the community of structural engineering for the last twenty years (Catbas et al. 2008b; Ni et al. 2011, 2010). There are a number of centers of excellence or research groups around the world that have contributed to the development and implementation of the SHM methods, technologies and implementations. Research group led by Prof. Ko at Hong Kong Polytechnic was one of the leading groups with particular real life contributions such as long span bridges and tall buildings (Ko et al. 2002; Ko and Ni 2005; Ni et al. 2003, 2008, 2009; Xu et al. 1997). SHM focuses on the condition assessment of different type of structures for better decision making (Gul and Catbas 2011b). To achieve this goal, various advanced sensing technologies are employed with interdisciplinary knowledge to support SHM with convenient monitoring ways and reliable data acquisition. Interdisciplinary research is needed for the development of sensing techniques for various types of measurements (Duan et al. 2016; Ye et al. 2017). Displacement is one of the most

commonly used indicators for structural assessment (Chen et al. 2014b, 2015c, 2019); and various displacement measurement methods using different technologies have been proposed and implemented. Some examples are the dial gauges, which use basic mechanical principles, linear variable differential transformer (LVDT), magnetostrictive displacement sensor (MDS), which utilize the electromagnetic properties, the liquid-level meter using liquid pressure principle, and the potentiometer using the change of electric resistor. Such displacement measurement technologies require some type of contact at the measurement location. These sensors are preferred and used, also to be compared with new methods and technologies. However, the contact type displacement methods need static reference to fix the measurement base and this makes it very challenging if impossible for large structures such as high-rise buildings and large bridges. Laser Doppler Vibrometer (LDV) (Chen et al. 2018b) as a noncontact type method is very expensive and not suitable for multi-point measurement of large structures. Global Position System (GPS) utilizes remote sensing and modern communication techniques to achieve displacement measurement, but its accuracy can easily be affected by electromagnetic interference (Ye et al. 2015). The measurement resolution is in centimeter level and currently it is suitable for the flexible structures like long span bridges (Ye et al. 2013) and towers (Ni et al. 2009; Xia et al. 2014; Yi et al. 2013). While there are other studies which convert accelerometer and strain gauge data to the displacement by leveraging mechanical relationships between them (Ye et al. 2015). This can achieve indirect displacement measurement but the error accumulated during the mathematical operations and approximations become significant. In addition, these methods may need special attachment and operational requirements such as traffic closure if used in bridge structures (Ye et al. 2015). Those limit the practical applications.

In recent years, with the development of computer vision and the progress of image data acquisition devices, vision-based methods are gathering increasing attention in the field of SHM (Catbas et al. 2017, 2012b; Celik et al. 2018a; Ye et al. 2016c; Zaurin et al. 2015; Zaurin and Catbas 2010a; b). Vision-based displacement measurement methods become viable alternatives for the current practice of structural displacement measurement in SHM (Feng and Feng 2016, 2017, 2018). Various tracking algorithms are applied to extract displacement from videos and unmanned aerial vehicles (UAVs) are also employed in field a field application (Yoon et al. 2017, 2018). There are limitations also for current practice of vision-based displacement measurement. Digital image correlation (DIC) and DIC-based template matching are the most popular methods (Zhong et al. 2017, 2018b; Zhong and Quan 2017, 2018a), while it has to create manual speckles on the surface of measurement structures to achieve good accuracy of DIC (Ye et al. 2016a). Manual targets are necessary to improve the performance of DIC based template matching (Brownjohn et al. 2017). Also, DIC and DIC-based template matching are easily affected by the viewpoint and illumination change (Ye et al. 2016f) which means that appropriate camera alignment and light sources are crucial. Feature points are extracted from images and keep invariant properties in different situations, which can be regarded as virtual markers to achieve non-target measurement (Khuc and Catbas 2017). In general, there are two different approaches to utilize the feature points to do displacement measurement: 1) combining the sparse optical flow with feature points (Dong et al. 2019b; Yoon et al. 2016); and 2) doing feature matching with the similarity calculations of feature points in different images (Khuc and Catbas 2017). Although both of the two different approaches can utilize the non-target properties of feature points, combining optical flow and feature points only employs the feature points without an important component: feature descriptor. In general, when talking about a type of feature point, it refers to

the feature detector and feature descriptor. Feature detector is a sub region of an image selected by certain algorithms such as SIFT (Scale Invariant Feature Transform) detector (Lowe 2004). Feature detector is also called feature point or keypoint, which provides visual information of image characteristics. Feature descriptor is generally represented by a vector and used to depict the feature point mathematically. Just using optical flow with feature points does not take full advantage of the feature point and the advantage of its descriptors. And also, the optical flow-based method can be easily affected by illumination change since the first assumption in optical flow calculation is that illuminations in consecutive frames are constant. Some advantages such as illumination invariant properties of the feature points cannot be reflected when combining with optical flow. While using feature matching-based methods, the appropriate properties of the feature points can be better utilized. The matching procedure is first calculating the distance between the descriptors' vector and giving matching candidate. Then outlier removing methods are employed to get rid of the wrong matches. During this procedure, all of the four components including the feature point, its descriptor, matching calculation and outlier removal play important roles.

In this chapter, descriptor is given more attention and explored to improve the displacement measurement accuracy with comparisons to the existing feature matching based methods (Khuc and Catbas 2016, 2017). Instead of using the hand-crafted feature descriptor such as SIFT descriptor and SURF (Speeded-Up Robust Features) descriptor (Bay et al. 2008), a descriptor generated by a learning processing is implemented and combined with SIFT feature point to do feature matching. The proposed feature matching strategy is put into a general procedure for

vision-based displacement measurement. The feasibility, accuracy and performance are verified through a comparative study with a laboratory experiment and then a field application.

4.2 Methodology

4.2.1 General Procedure for Vision-based Displacement Measurement Method Using Feature Matching

Figure 28 illustrates the flowchart for proposed vision-based displacement measurement method using feature matching. In the first step, the camera is calibrated to obtain the relationship between the image coordinates and the real-world coordinates. In other words, it should be figured out how many physical units (e.g., millimeter) in the real world represent one pixel in the image plane. In the second step, image data either color or grayscale image sequence is acquired with an appropriate device. In the third step, image feature points, also called keypoints or corners which are salient parts of the images are extracted and determined by descriptors. In the fourth step, the feature points in the current image and the initial image are matched with certain similarity criteria and suppression algorithms are employed to remove the outliers (i.e., wrong matches). In the last step, the displacements at feature points are obtained by converting the displacement in pixels to the displacement in physical units. By taking the average of the displacements at the feature points in the region of interest (ROI), the structural displacements expected to measure are obtained. In this chapter, for the convenience of the application, scale factor method is employed. Some critical issues which can affect the measurement accuracy are introduced in detail below.

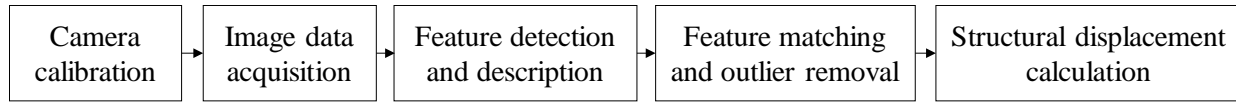


Figure 28 Flowchart for proposed vision-based displacement measurement method using feature matching

4.2.2 Image Data Acquisition

Image data acquisition means to take the image sequence of the measurement region with appropriate device which usually includes cameras, lenses and computers. Depending on the application scenarios, one camera or multiple cameras can be used to measure the displacement of single point or multiple points. It should be noted that if only one camera is used for multi-point displacement measurement, the accuracy is reduced because of the resolution reduction of each measurement region. Also, there is the trade-off between the image sampling rate and image resolution due to the limitations of the data transmission and storage. Depending on the monitored structures, the image sampling rate and resolution have to be selected properly. In practical applications, the camera may be affected by the ground motion or wind loads. To eliminate the camera motion effects, one of the solutions is to subtract motion of the stationary part in the image. This is also necessary to be considered in image data acquisition. More details and image data acquisition strategies are presented in the authors' previous work (Dong et al. 2019c).

4.2.3 Feature Detection and Description

Image features represent the characteristics of the image, which make it be distinguished from the surroundings. Features can be the grayscale intensity, color, geometrical shape and edges.

However, these features can be affected by illumination, scale, partial occlusion and viewpoint, etc. Feature points, also called feature detectors, are the sub regions of the image which show robust properties with influencing factors and keep invariant. The most popular feature points include Harris corner (Harris and Stephens 1988), Shi-Tomasi corner (Shi and Tomasi 1994), SIFT (Scale Invariant Feature Transform) (Lowe 2004), SURF (Speeded-Up Robust Features) (Bay et al. 2008) and ORB (Oriented FAST and rotated BRIEF) (Rublee and Bradski 2011), etc. In order to use these features to do feature matching in different images with mathematical tools, e.g., distance, the feature points need mathematical representation, not just the visual region of the images. In general, descriptors are applied to represent the feature points. Feature descriptor combines the location, intensity, gradient and orientation etc. of the region where the feature point stands, and it is usually denoted by a vector. SIFT, SURF, ORB, FREAK (Fast Retina Keypoint) (Alahi et al. 2012) and BRISK (Binary Robust Invariant Scalable Keypoints) (Leutenegger et al. 2011) are common feature descriptors used in computer vision.

Figure 29 gives an example of SIFT feature points and descriptors. Here the SIFT descriptors are visualized on the image to give a picture with the uniqueness of different feature points using VLFeat packages (VLFeat 2018). Currently the features mentioned above are implemented for vision-based displacement measurement and give good results (Catbas et al. 2018; Dong et al. 2019b; d; Khuc and Catbas 2016, 2017). The feature descriptors are hand-crafted and without learning process. In this study, a learning local feature descriptor, also called VGG feature descriptor, which is proposed by Visual Geometry Group (VGG) of University of Oxford using convex optimization (Simonyan et al. 2014), is employed. As presented by (Simonyan et al. 2014), VGG feature descriptor shows better performance for the general benchmarks. In this study, the

SIFT feature point and VGG descriptor (SIFT-VGG) are combined together as a strategy for vision-based displacement measurement. The results from SIFT-VGG are compared with SIFT, SURF and the combination of SURF feature point and VGG descriptor (SURF-VGG).

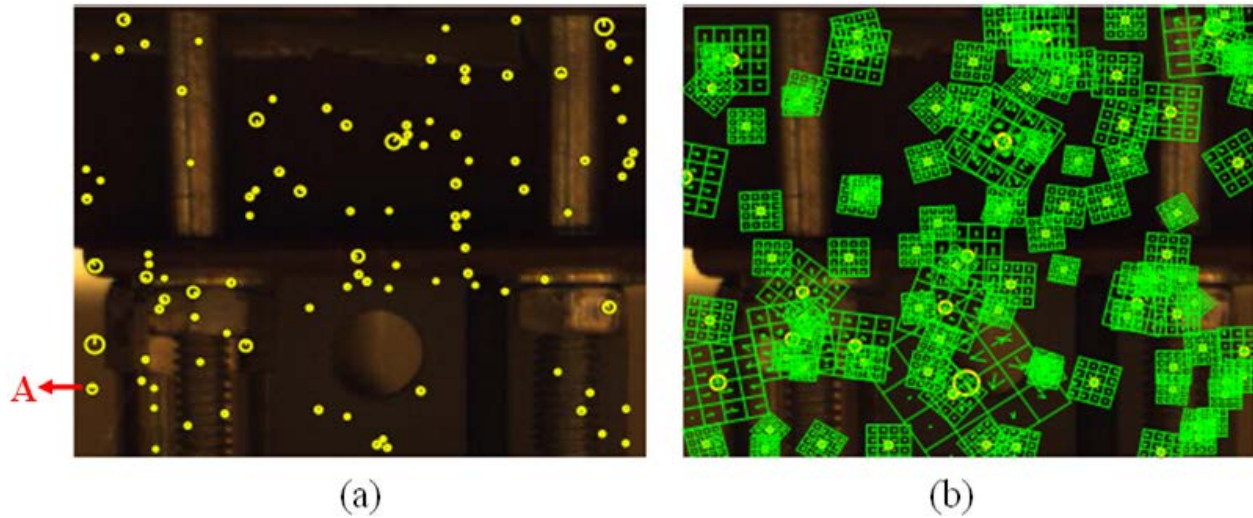


Figure 29 An example of SIFT feature points (a) and descriptors (b)

4.2.4 Feature Matching and Outlier Removal

To match the feature points in different images, the similarities between different feature points are measured. In general, to measure the similarities, the distance between the feature descriptors are calculated. By default, Euclidean distance is employed, while the feature descriptor is a binary vector, then Hamming distance is selected. For example, ORB is a binary descriptor. To complete the feature matching, there are two approaches: 1) Brute-Force matcher and 2) k nearest neighbors (KNN) matcher. Here k is a constant and always selected as 2, which is also utilized as FLANN (Fast Library for Approximate Nearest Neighbors). Brute-Force matcher calculates all the distance and just picks the closest one and does the matching. While FLANN matcher holds two closest

distances and in the next step just keeps the one that satisfies the condition that one of the distance is smaller than a threshold times the other. In Lowe's paper (Lowe 2004), this threshold is 0.65. There are still wrong matches (outliers) using just distance and to obtain better feature matching results, RANSAC (RANDOM SAMPLE Consensus) method or LME (Least Median Estimation) method is applied. RANSAC can handle practically any ratio of outliers but need a threshold to distinguish inliers from outliers, while LME does not need any threshold but it works correctly only when there are more than 50% of inliers (Simonyan et al. 2014). Khuc and Catbas implemented RANSAC method to remove the outliers and shows good structural displacement measurement results (Khuc and Catbas 2017). In this study, FLANN and RANSAC are implemented for feature matching and outlier removal. Figure 30 gives an example of feature matching using SIFT and SIFT-VGG and outlier removal using RANSAC method. The wrong matches in Figure 30a are successfully removed in Figure 30b for the original SIFT. The same also shows in Figure 30c and Figure 30d for SIFT-VGG. The difference is that when using SIFT and RANSAC, the right matches in Figure 30a are also removed in Figure 30b. While they are kept when using SIFT-VGG. In this view, it can be seen that SIFT-VGG shows better performance than the original SIFT. The blue bounding boxes in Figure 30b and Figure 30d are generated when applying perspective transformation with RANSAC (OpenCV 2019c).

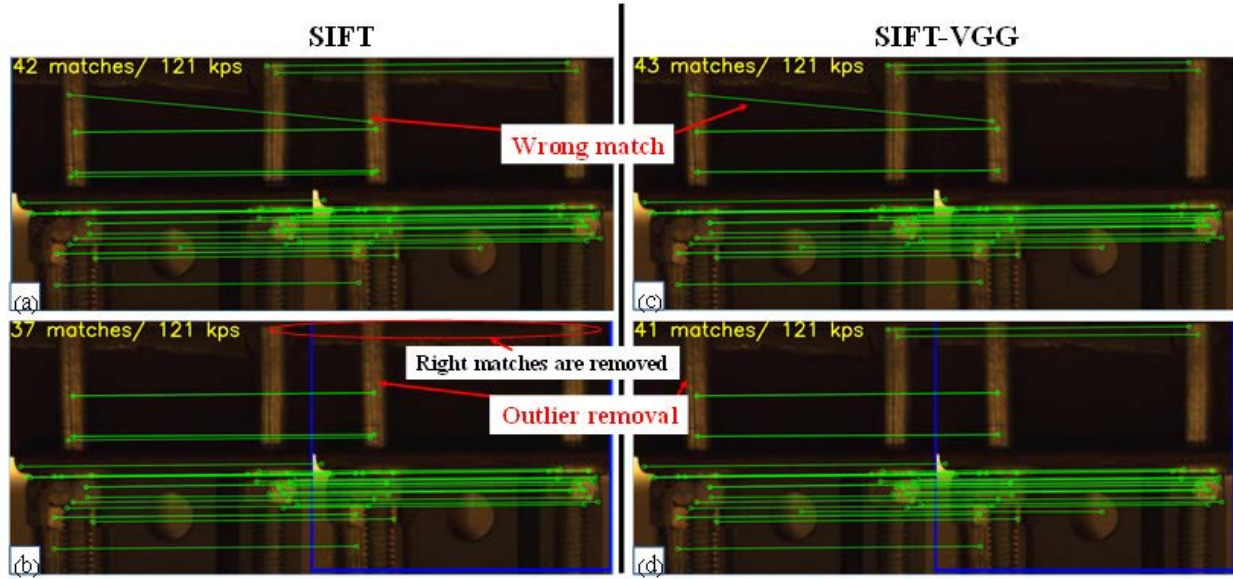


Figure 30 An example of feature matching and outlier removal: (a) SIFT feature matching; (b) SIFT outlier removal; (c) SIFT-VGG feature matching; (d) SIFT-VGG outlier removal.

4.2.5 Structural Displacement Calculation

Based on the feature matching results, the motion of the feature points can be calculated by the location change of the feature points in image coordinates. By taking the average of the displacements of all the feature points that matched, the displacement of the measurement region in image coordinate is obtained. Then with the camera calibration results obtained in the first step, the displacement in real world with physical unit is eventually estimated. Since this approach takes the average value to calculate the displacement, the feature points which have no motion might be included and finally make the displacement smaller than the real value. For example, the feature point, A, in Fig. 2a, is part of the background and supposed to be stationary. If this feature point is included in the measurement region, then it will affect the measurement result when taking the average value. For making more accurate measurements, the measurement region should only include the motion part as possible when using feature matching-based methods.

4.3 Laboratory Verification

4.3.1 Experimental Setup

As shown in Figure 31, a two-span bridge model is selected to be the experimental object, which is constructed in the UCF Civil Infrastructure Technologies for Resilience and Safety (CITRS) Experimental Design and Monitoring (EDM) laboratory. The bridge is a scaled down model of a mid-size real-life structure in terms of structural responses and toy trucks with variable weights are used to model moving loads (Dong et al. 2019d). The two-span bridge is modified from the original UCF four-span bridge which was introduced in literature (Khuc and Catbas 2018). It consists of two 300 cm main continuous spans with a 3.18 mm steel sheet at 120 cm wide, which makes the deck 600 cm long by 120 cm wide. In order to verify the feasibility and performance of the proposed method, an experiment was conducted on this bridge. An industrial camera (MindVision- MV-GE131gc-t) with a zoom lens (focal length: 5~100 mm) is set up in front of the bridge to record images at a measuring point (P1) during the moving truck load trials. The camera has a maximum frame rate of 60 Hz, a resolution of 1280 pixel \times 960 pixel. In addition, a conventional displacement sensor, potentiometer, is mounted under the deck to measure the displacement of P1 and is assumed as the ground truth. The model of the potentiometer is BEI 9615 with an active electrical travel distance of 35.6 mm and linearity over active electrical travel of $\pm 2\%$. More detail specifications can be found from reference (BEI 2019). Both of the camera and potentiometer are synchronized by using a NI Multifunction I/O Device with the model number of USB-6343. Detailed specifications and implementations can be found in

reference(Dong et al. 2019b). During the experiments, the truck moves from one side of the bridge to the other while the potentiometer and the camera record the motion of P1 (midspan of the left span) synchronously. The sampling rate for potentiometer is 100 Hz and it is down-sampled to be comparable to the camera. The acquired image sequence is analyzed using the proposed methods. The displacement obtained from image sequences is compared with that of the displacement sensor.

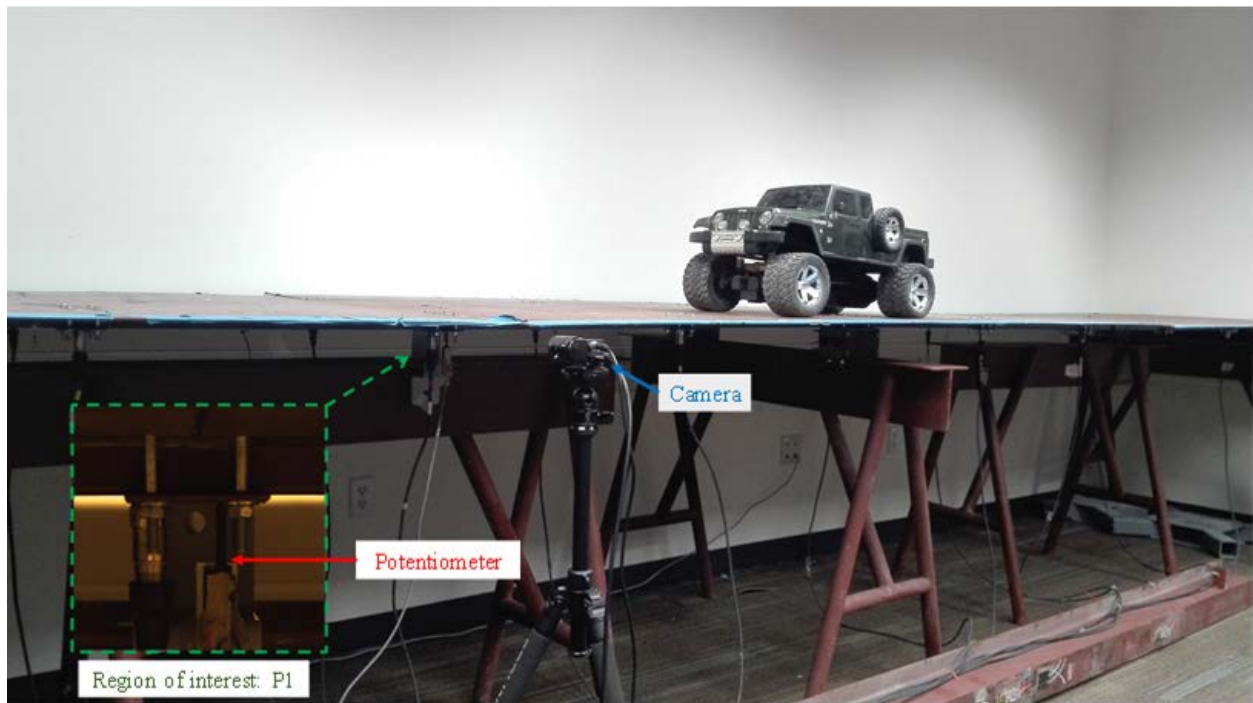


Figure 31 Experimental setup

4.3.2 Comparative Study of Displacement Measurement Using Different Methods

In this experiment, four vision-based displacement measurement methods using feature matching methods including SIFT, SURF, SIFT-VGG and SURF-VGG and one conventional displacement sensor, i.e., potentiometer (PM) are used to obtain the displacement time histories of P1, when the

toy truck passes over the bridge model. Figure 32 shows the displacement measurement results using different methods. Figure 32 depicts the whole loading process and displacement responses at the measurement point P1. In the beginning, the truck stays at the left end of the first span, then it moves to the right and approaches P1. In the meantime, the displacement of P1 (the downward direction is positive) gradually increases to a maximum when the truck is passing at P1. Then the truck begins to move away from P1 and keep heading to the right, while the displacement of P1 gradually decreases. When the truck moves to the right span, the displacement begins to be negative (i.e., upward displacement) due to loading on the other span of the two-span bridge. As it approaches the right end of the right span, the absolute value of the displacement at P1 first increases and achieves a maximum and then decreases. When the toy truck arrives at the right end of the bridge, the displacement of P1 becomes stable but does not go back to zero. This is because the rear axle is still left on the bridge.

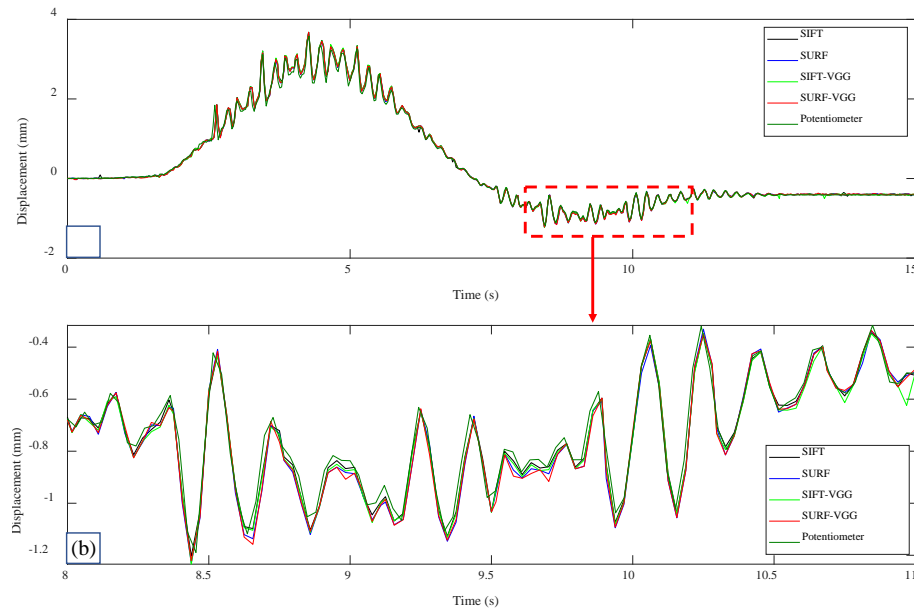


Figure 32 Comparison of displacement time histories from displacement sensor and vision-based methods

By comparing the displacement time histories, it is easy to see that the result obtained from the proposed method (i.e., SIFT-VGG) is quite consistent with those obtained from the potentiometer (PM, regarded ground truth) and other three vision-based methods (SIFT, SURF and SURF-VGG). Figure 33 illustrates the correlation matrix of these time displacement histories. The figures on the diagonal of the correlation matrix are the histograms of the displacement time histories whereas the others are data plots and linear fits between the displacement time histories from the two methods. The correlation matrix is symmetric, and the last row and the last column give the correlation coefficients between the displacement data obtained from the vision-based methods and the potentiometer. The correlation coefficients of SIFT, SURF, SIFT-VGG and SURF-VGG with ground truth, i.e., potentiometer are all 0.99. From the correlation matrix, it can be seen that the correlation effect between the results from all vision-based methods (SIFT, SIFT-VGG, SURF, SURF-VGG) and potentiometer data are worse when the excitation amplitudes are higher.

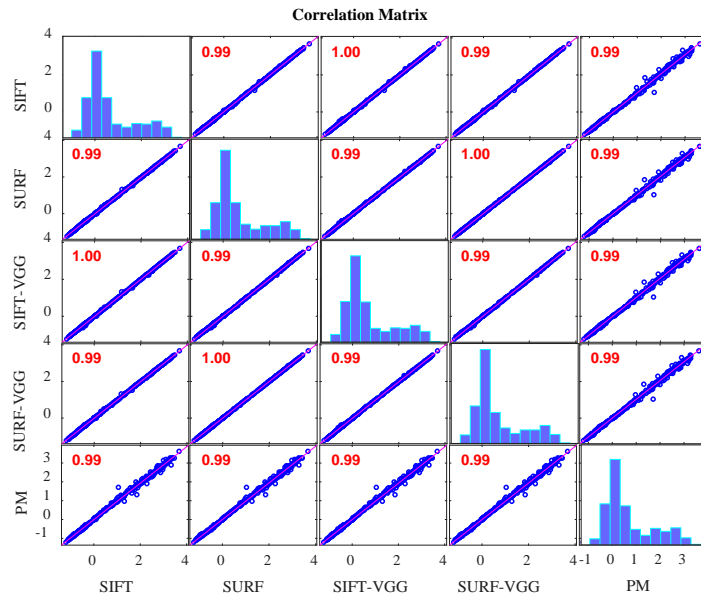


Figure 33 Correlation matrix of time displacement histories

The displacement time histories when no loads are present on the structure are extracted as the measurement error distribution and are used to estimate the measurement accuracy. Here “no loads” means there’s no vehicle passing through on the bridge and the bridge specimen is subjected to minor vibration sources in the laboratory environment. Here the bridge specimen is assumed static even with this minor vibration sources. accuracy evaluation method is carried out by exploring the statistical characteristics of the measurement of the assumed static status. The measurement obtained in this status are considered as noise level. The index of the measurement accuracy is defined by ± 2 standard deviations (Khuc and Catbas 2017), which corresponds to a level of confidence of 95%. The smaller the value of 2σ , the smaller the error and the better the measurement accuracy. Figure 34 shows the distributions of measurement error from different methods. Table 1 gives the measurement resolution ($\pm 2\sigma$) analysis. From Figure 34 and Table 5, it is seen that the accuracy of the proposed methods using SIFT-VGG is ± 0.0087 mm which is the smallest one among the four vision-based methods, and closest one to the ground truth (displacement sensor) which is ± 0.0026 mm. As can be seen, there is still a difference between SIFT-VGG and potentiometer which is a contact measurement method. The measurement accuracy of SIFT-VGG can be considered acceptable as it improved the original SIFT method about 24%. For the SURF feature, it doesn’t give better performance than the SIFT feature. Even replacing the descriptor with VGG, it doesn’t improve the measurement accuracy. Further, from Figure 34, it can be seen that the proposed methods using SIFT-VGG give much better stability in displacement measurement.

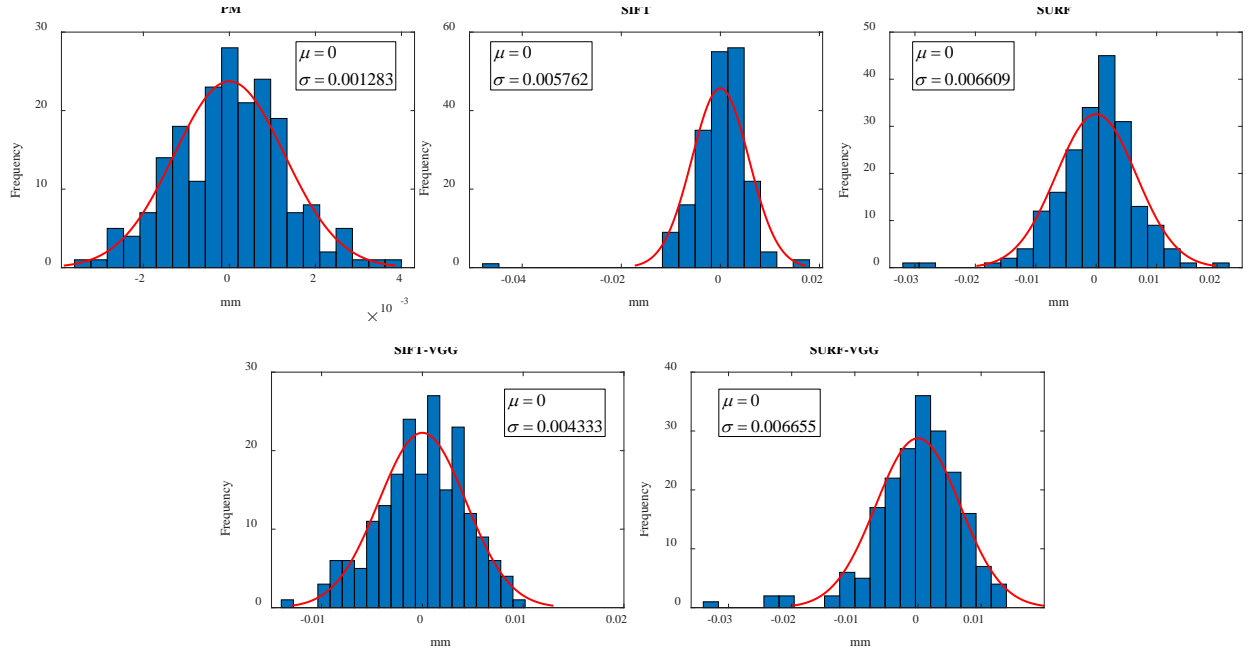


Figure 34 Distributions of measurement error for different methods Experimental setup for a railway bridge

Table 5 Measurement accuracy and resolution ($\pm 2\sigma$) analysis: unit (mm)

Method	Potentiometer	SIFT	SURF	SIFT-SURF	SURF-VGG
$\pm 2\sigma$	± 0.0026	± 0.0115	± 0.0132	± 0.0087	± 0.0133

Table 6 gives the comparison of computation times of an image pair matching by using different feature matching methods. From this table, it is indicated that SIFT is much slower than SURF which is already known from the literature (Bay et al. 2008) and it takes even more computation time to replace the SIFT descriptor with VGG descriptor for SIFT feature matching: 77% more computation time. Considering that it improves accuracy by 24%, the 77% more computation time may be acceptable for post processing. Table 6 also presents the processing speed in frame per second for different feature matching methods. Although speed of the proposed method (SIFT-

VGG) is 16.83 FPS, it may be viable choice for post processing when there is no requirement for real-time monitoring.

Table 6 Comparison of computation times of vision-based methods using different feature matching methods

Method	SIFT	SURF	SIFT-SURF	SURF-VGG
Matching time of an image pair (second)	0.0335	0.0242	0.0594	0.0491
Frame per second (FPS)	29.83	41.37	16.83	20.39

4.4 Field Application

4.4.1 Experimental Setup

In this section, a field application performed on a railway bridge (Reverend Kenneth C. Crossman Bridge) in Orlando is presented to verify the feasibility of the proposed displacement measurement method. The railway bridge is located at 1792 Orlando Ave, Maitland, Florida, named by Reverend Kenneth C. Crossman who became knowns as the “Bridge Builder”. The structure (Figure 35) is a two-span steel bridge and the displacement measurement location is the two thirds of the first span. The measurement region is manually selected and denoted as P1. A stationary region, P0 in the background was manually selected as the static reference and used to eliminate the camera motion caused by ground vibration and wind effects. The distance from the camera location to the measurement region is around 33 meters. A portable camera (Z Camera E1) with a resolution of 4K (3840×2160 pixels) and a frame rate of 30 FPS and an Olympus zoom lens with a focal length

of 75-300mm was used to collect images. During the experiment, the camera captured the bridge response when a train crossed the bridge.



Figure 35 Experimental setup for a railway bridge

4.4.2 Analysis and Results

Figure 36 illustrates the displacement time history of P1 obtained from the proposed method (using SIFT-VGG feature matching) when a train was crossing the railway bridge. From Figure 36, it can be seen that the original displacement result is heavily affected by the camera motion. The displacement of the stationary region, P0, is regarded as the camera motion, and after the camera motion subtraction from the original displacement, the revised response of the bridge gives the more reasonable displacement results. The maximum deflection of the bridge is around 11 mm. Such displacement monitoring can be carried out intermittently to track changes in structural behavior. In addition, displacement under heavy and long freight trains can be carried out very

efficiently without any interruption to train operations and also without any access needs for improved safety. In this experiment, a stationary region is selected to eliminate the camera motion effect, while in other field applications it might not be able to find a stationary region. Possible solutions are: 1) the stationary object can be included by enlarging the field of view of camera, but it would decrease the resolution of the measurement target with a fixed camera sensor; 2) the components of camera motion can be filtered out by adding accelerometers on the camera; but this also induce another problem that is when the frequency of camera motion is close to the structural motion, it doesn't work. Up to now, it is still an open question to eliminate the camera motion effectively.

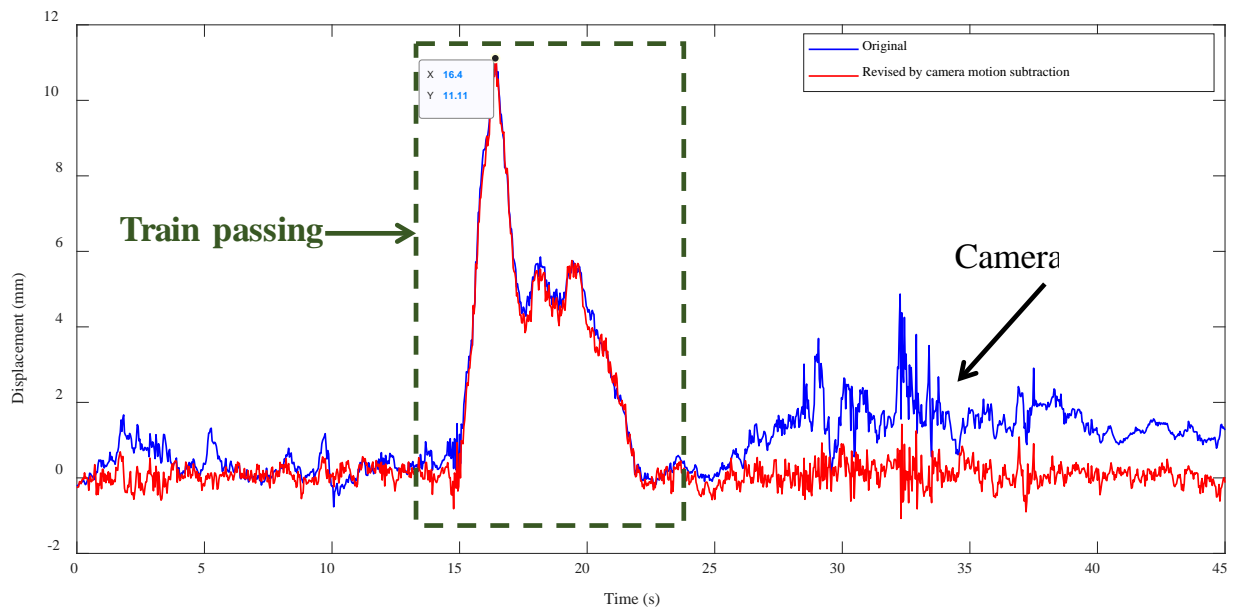


Figure 36 Displacement time history obtained from the proposed method

4.5 Summary

To achieve non-target displacement monitoring for civil structures without manual targets and to overcome the limitations of common vision-based methods as well as to improve the performance and accuracy of current vision-based methods using feature matching, a novel method using advanced feature matching combining SIFT feature points and VGG descriptors is proposed. The feasibility of the proposed method is verified through a comparative study with a laboratory experiment on a two-span bridge model and a field application on a railway bridge. The main findings and conclusions are as follows:

- (1) A general procedure for vision-based displacement measurement using feature matching is presented and this approach provides a standard reference for future users.
- (2) An advanced feature matching strategy by combining SIFT feature points and VGG descriptors are proposed and this integrated strategy improves the measurement accuracy of the original SIFT method by 24%.
- (3) Critical issues in vision-based methods using feature matching are discussed and recommendations such as camera calibration, image data acquisition and ROI selection for reducing measurement errors are provided.
- (4) The camera motion issue is discussed in the context of field application. Camera motion subtraction is proposed to address the errors induced by camera motion.

Although the proposed method (SIFT-VGG) requires more computation time and may not be able to handle the real-time displacement monitoring task, considering the practical engineering measurement experience, it may be a viable option for post processing for non-target structural displacement measurement especially if real-time monitoring is not a requirement. In the future, the authors will focus on improving the processing speed, and will investigate the effectiveness of long-distance monitoring and the robustness of method when there is illumination change, as well as explore solutions for removing camera motion when there is no stationary reference in the background.

CHAPTER FIVE: STRUCTURAL DISPLACEMENT MONITORING USING DEEP LEARNING-BASED FULL FIELD OPTICAL FLOW METHODS

Previously published as Dong, C. Z., Celik, O., Catbas, F. N., O'Brien, E. J., and Taylor, S. (2019c). "Structural displacement monitoring using deep learning-based full field optical flow methods." *Structure and Infrastructure Engineering*, 16(1), 51–71.

5.1 Introduction

Displacement is a critical indicator for structural performance evaluation and health condition assessment of infrastructure. Static and dynamic characteristics of structures such as bridge load capacity (Lee et al. 2006a; Ojio et al. 2016), bridge deflections (Moreu et al. 2016) and deformation profiles (Xu et al. 2018), load distribution (Fuchs et al. 2004), load input information , unit influence line (UIL) and unit influence surface (UIS) (Khuc and Catbas 2018), modal frequency and shape (Chen et al. 2014b, 2015c, 2019; Dong et al. 2018, 2019b; Yang et al. 2017b; Yoon et al. 2016) can be extracted from displacement data. Currently, displacement measurement is still a difficult task in conventional structural health monitoring (SHM) (Catbas et al. 2018).

Ye et al. (Ye et al. 2015) summarized the current displacement measurement methods in the field of SHM, including 1) contact type: linear variable differential transformers (LVDT), double integration of recorded acceleration data, displacement derivation from the strain-deflection

relationship; and 2) non-contact type: global positioning systems (GPS) and integration of data from Laser Doppler Vibrometers (LDV). Non-contact type displacement methods do not need a reference level, and do not need the access to the measured structures and can save on road closures, which are key advantages of this approach. Ye et al also indicated the limitation of conventional non-contact type methods: 1) GPS has low accuracy and sampling rate; and 2) LDV has high cost. A total station is also a non-contact displacement measurement tool. However, the use of total station is unsuitable for bridge monitoring (da Silva et al. 2018) due to difficulties of the installations of control points and continuous automatic monitoring at high frequency. In order to achieve desired continuous measurement, additional equipment has to be added to the original total station (Ehrhart and Lienhart 2016; Omidalizarandi et al. 2018). Extracting displacement measurements from image sequences has become a popular research topic in various applications of civil engineering (Feng and Feng 2015, 2016; Feng et al. 2015b; Pan et al. 2009) since the manufacture of advanced cameras improved and computer vision techniques progressed. The advantages of non-contact, long-distance, high precision, low cost and less time-consuming measurement capabilities has caused vision-based displacement methods to get increasing attention from the community of structural health monitoring with the potential of becoming an alternative to the conventional displacement measurement methods in SHM as well as to infrastructure inspections (Chen et al. 2018a; Khuc and Catbas 2017; Luo and Feng 2018; O’Byrne et al. 2015; Wu et al. 2014; Xu and Brownjohn 2018; Ye et al. 2016c).

In general, current vision-based displacement measurement methods are divided into five categories: 1) image correlation based template matching (Feng and Feng 2016; Lava et al. 2009; Pan et al. 2016; Sutton et al. 2008; Ye et al. 2016b), 2) color based template matching (Ye et al.

2016a), 3) key point matching (Khuc and Catbas 2016, 2017; Lydon et al. 2019), 4) Lucas-Kanade optical flow estimation at key points (Celik et al. 2018a; Dong et al. 2019b; Lydon et al. 2018; Yoon et al. 2016, 2018), and 5) full field optical flow estimation (Celik et al. 2018a; Chen et al. 2015b; Khaloo and Lattanzi 2017).

Image correlation-based template matching is the most popular (Chen et al. 2018b; Zhong et al. 2017, 2018b; Zhong and Quan 2017, 2018a). However, it is sensitive to changes in shading, illumination and background condition, especially when used in field applications (Xu and Brownjohn 2018). To improve the measurement performance, manual light sources or targets are designed to be fixed on the structures and then tracked. Ye and Dong (Dong et al. 2015, 2018; Ye et al. 2016a; b; f) installed light emitting diode (LED) and QR (quick response) codes on structures to improve the texture contrast of the visual tracking area and tried to eliminate the influence of illumination changes. (Tian and Pan 2016) combined the use of LED targets and a coupled bandpass optical filter to mitigate the ambient light interference. Color based template matching is not robust to color change and the application is limited to the close-range displacement measurements. For long distances, the color condition of the measurement area could easily be affected by the light and shading, which makes it hard to get the right measurement results. To improve the measurement performance, artificial targets with specific colors could be utilized. Key point matching is a non-target method which calculates the displacement by averaging location change of the robust key points extracted from images. The method relies on calculating the similarities of the descriptors of key points in consecutive images based on statistical distance. Once similar key point pairs are recognized, the locations are confirmed to be the continuation of the former motion. Generally, the key points may have robust properties such as being invariant

to shading, illumination change, and scale. The most popular key points are Harris corner (Harris and Stephens 1988), Shi-Tamasi corner (Shi and Tomasi 1994), Scale-Invariant Feature Transform (Lowe 2004) (SIFT) feature points and Speeded-Up Robust Features (SURF) (Bay et al. 2008). The performance of key point matching methods is highly dependent on the saliency of the texture of the measurement surface. The number of key point extractions from a measurement area of an image is not easy to decide and it has been an open question as to how many key points should be extracted for displacement measurement to achieve the best performance. Lucas-Kanade optical flow is a sparse flow calculation algorithm and is usually combined with key points to do visual tracking. This displacement measurement methodology involves similar limitations as the key point matching does. Besides, the “small motion” assumption of optical flow restricts its application for large structural deflections (Dong et al. 2019b) although a pyramid method is used to refine the displacement evaluation in large displacement cases.

Full field optical flow can calculate the displacement vector of each pixel of images and give the displacement information of the entire structure. Classical full field optical flow estimation algorithms (Sun et al. 2010) originated from the core work of Horn and Schunck (Horn and Schunck 1981). These algorithms are derived from variational methods which are based on the gradient change in images and need filters to smooth the motion in images. They are adversely affected by illumination change and give inaccurate flow estimation at the motion boundaries. A phase-based optical flow algorithm is another method implemented into some structural displacement measurement problems (Chen et al. 2015b; Yang et al. 2017a, 2018b), but the applications are limited to those cases without background clutter. Parameters in these algorithms have to be adjusted to accommodate the differences in applications and it is too complicated for

practical use. Also, full field optical flow calculation is a heavy task that needs longer computation time, which makes it unsuitable for continuous structural displacement measurement, especially for real time monitoring. Detailed comparisons can be found in (Dong et al. 2019b).

In this study, a novel structural displacement measurement method using deep learning based full field optical flow is proposed. A general procedure for vision-based displacement system is presented and the planar homography matrix is applied for camera calibration. By implementing a pre-trained deep neural network for optical flow calculation, i.e., FlowNet2, the full field optical flow is obtained, and the displacement of the measurement region is calculated by using a mean kernel or Gaussian kernel. The proposed method does not need manual target and can be operated with less human participation than the key point matching, Lucas-Kanade optical flow with key points and classical full field optical methods. Image collection strategies, tracking strategies in image sequences, non-uniform image sampling and camera motion problems are also discussed in this paper. Useful strategies are identified to address the problems that could occur in practical application. Laboratory experiments on a grandstand structure and field experiments on a footbridge are conducted to verify the feasibility of the proposed method.

5.2 Methodology and System Development

5.2.1 General Procedure for Vision-Based Displacement Measurement System

Figure 37 illustrates the flowchart for the proposed full field structural displacement measurement method. In the first step, the camera is calibrated to obtain the relationship between the image

coordinates and the real-world coordinates, i.e., to find how many physical units (e.g., millimetres) in the real world represent one pixel in the image. In the second step, image data from the structure in question are collected and are transferred to the next step for real time or post processing. In the third step, optical flow algorithms are implemented to do visual tracking and calculate the full field structural motion, thereby obtaining the motion vector at each pixel of the image. In the fourth step, the false structural motion induced by the camera vibration is mitigated by subtracting the motion of the static parts in the image. In the last step, the full field structural displacement is obtained by converting the displacement in pixels to the displacement in physical units. In the flowchart in Figure 37, the three steps: camera calibration, full field optical flow estimation and mitigation of camera vibration are crucial as they directly affect the measurement accuracy. In this chapter, the planar homography matrix method presented in Section 2.2 is implemented for camera calibration. The four point correspondences (a-A, b-B, c-C and d-D) marked in Figure 38 are used to show the projection from real world plane to image plane. The other steps will be introduced in detail below.

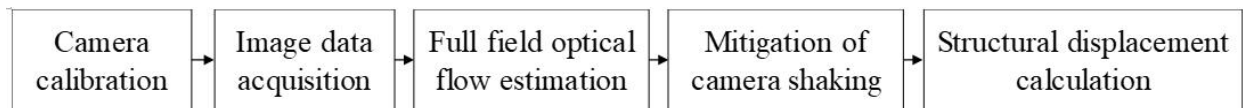


Figure 37 Flowchart for proposed full field structural displacement measurement method

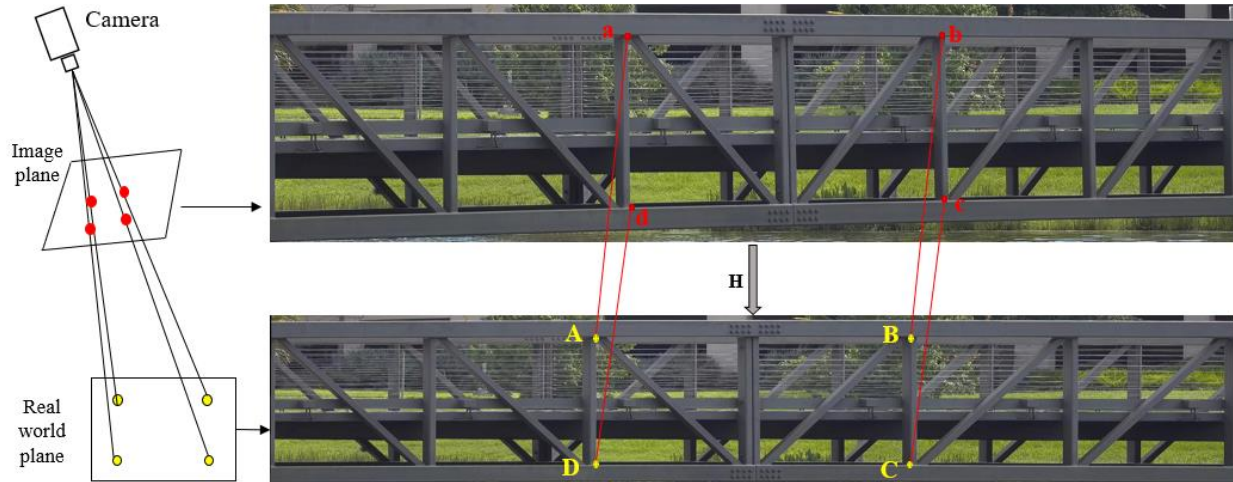


Figure 38 Image projection using the planar homography matrix and four point correspondences

5.2.2 Image Data Acquisition and Considerations

Unlike data from conventional sensors such as displacement sensors, accelerometers, strain gauges or tiltmeters, which provide one dimensional data (i.e., temporal data), image data is two dimensional and contains temporal and spatial information. This increases the demand to sample larger amounts of data and results in a reduction in the sampling rate of image data acquisition systems (i.e., cameras and image grabbers) compared to the conventional sensors. For image data acquisition systems, the sampling rate is referred to as frame rate, expressed in frames per second (FPS). When used within the context of single point or full field displacement time history from a vision-based system, the sampling rate is different from its frame rate. The frame rate is related to the camera exposure time, time trigger, etc. Usually frame rate is a critical factor to be considered when doing image data acquisition in vision-based displacement measurement and it always influences the selection of image acquisition methods and devices. Depending on the monitoring or measurement requirements (whether real time monitoring or not), generally there are two ways

to do image data acquisition as shown in Figure 39.

(1) If there is no requirement to carry out real time displacement monitoring, the image data acquisition systems grab images continuously (also called video recording) and then processes the image data afterwards (post processing). The image data acquisition system can be a portable camera, or a digital camera which can be connected to a computer through different types of interface such as GigE, USB2/3, Camera Link, FireWire (IEEE 1394/IIDC DCAM Standard), or an analog camera which needs an image grabber card to be connected with a computer, or even a smart phone. The normal portable cameras and smart phones usually have internal clock and the frame rate can be set as a fixed number such as 30 FPS, 60 FPS or 120 FPS. The images or videos are stored in the on-board storage card. In practical applications, the frame rate is not always fixed. For example, when the frame rate is set to 60 FPS, practically it might be less than that. For instance, within the scope of this study, a Canon portable camera was tested, capturing video at a frame rate of 60 FPS at a resolution of 1920×1080 pixels. However, analysis showed that the real frame rate was 59.94 FPS on average. The frame rate reduction might be associated with some frames being delayed or dropped as a result of longer exposure time or unsuccessful triggering. As for the off-the-shelf portable cameras, the real and the pre-set frame rates are not distinctly different than each other. This feature makes them a convenient option for the monitoring of structures with low frequency dynamic characteristics. However, it is hard to find information on image timestamps and dropped images. On the other hand, computer controlled analogue or digital cameras can deliver this information accurately during the exposure, since the image data acquisition procedure is programmed into a software package. When using these kinds of cameras, the exposure time can be auto-adjusted to acquire images with good quality. Nevertheless, image

data acquisition is still a non-uniform sampling process. In general, the frame rate is calculated as $1/\Delta t$, where Δt is ideally the time interval for uniform sampling. In reality, the time interval between consecutive images varies every time, as Δt_i ($i = 1, 2, \dots, k$) and the average frame rate is calculated as the ratio of the total number of frames and the total acquisition time. To partially remedy this problem, a triggering function (i.e., an edge signal) that controls the exposure can be sent to the Input/Output (I/O) interface of the camera satisfying the nominal uniform sampling (Dong et al. 2019b). It should be noted that the trigger frequency has to be less than the camera's maximum frame rate and a high trigger frequency may lead to an increased probability of frame drop.

(2) If there is a requirement to do real time displacement monitoring, then the selection has to be among computer-controlled cameras. The data sampling process is divided into three steps: a) image grabbing; b) image transmission to the computer; c) image processing. The computation time spent in each step (Δt_{ig} , Δt_{it} , Δt_{ip} ,) may also vary to produce different numbers of samples because the exposure time, transmission time, and motion tracking time interval (during image processing) at each step may differ. For instance, as described earlier, a triggering function is applied to make the image grabbing time intervals, Δt_{ig} , equal and uniform, while the transmission time, Δt_{it} , and image processing time, Δt_{ip} , may vary. For these reasons, the data sampling rate is decreased and becomes less than that of (1). Also, it is a non-uniform sampling process. To make this data sampling process uniform, a waiting time, Δt_{ipw} , can be added into the image processing time. This waiting time forces the total sampling time to be fixed and, as a result, the real sampling rate is decreased.

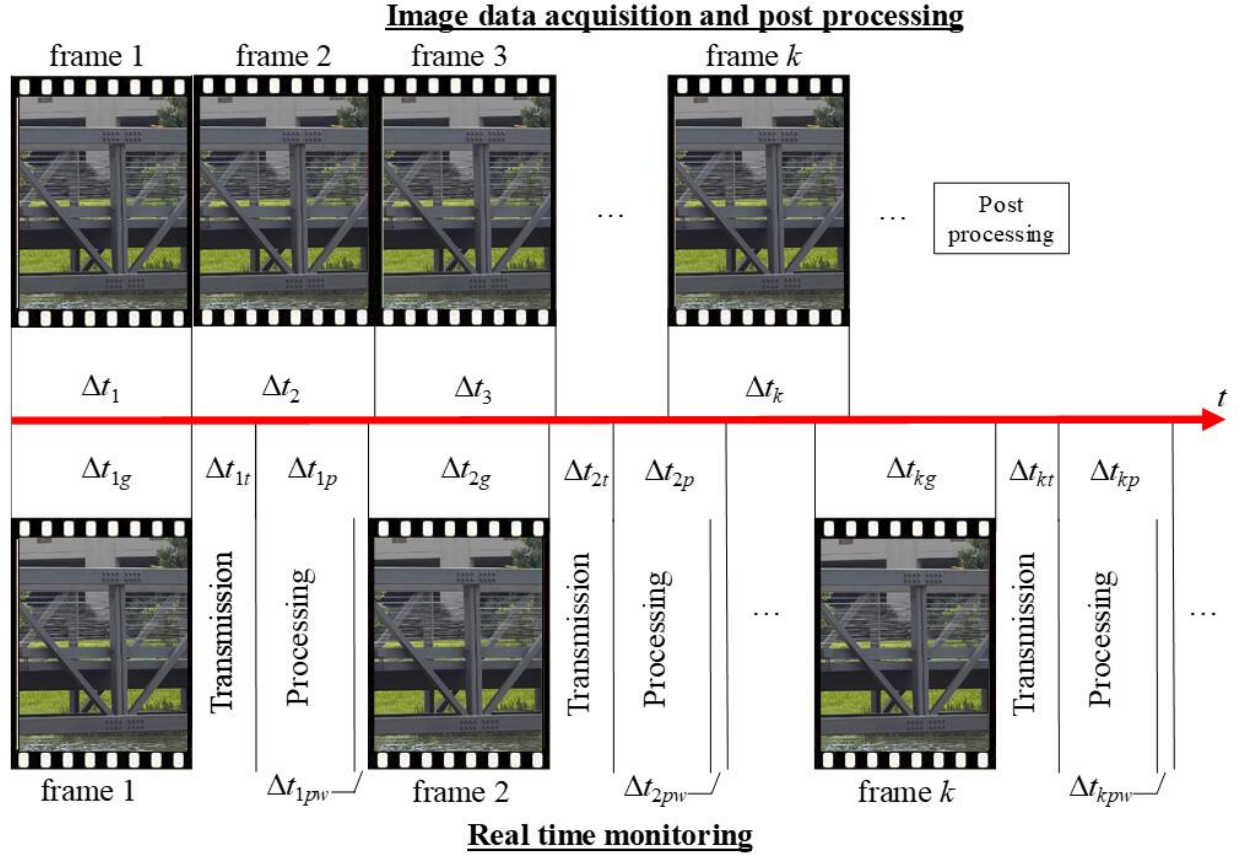


Figure 39 Image data acquisition and processing

It is essential to know whether or not the application will require real time monitoring as it impacts the selection of motion tracking algorithms. In this study, full field displacement is estimated by computationally demanding optical flow algorithms which means that some of them cannot afford real time monitoring. That is why the selection of the optimal optical flow algorithm is a crucial step. Besides, in the cases when multiple cameras are needed, an array of cameras with time synchronization can be designated to satisfy the measurement requirement (Wilburn et al. 2005).

5.2.3 Full Field Optical Flow Estimation Using Deep Learning Methods

Optical flow is the distribution of motion velocity vectors in image data. The motion can be an image sequence captured by a single camera or two images captured by two different cameras. The optical flow is usually represented by a 2D vector, i.e., the horizontal and vertical components, along with the two different directions. To estimate the full field structural displacement, optical flow estimation of the image sequences containing the motion of structures, is a good option. In general, there are two different optical flow estimation subsets: 1) local optical flow estimation methods, which calculate the flow vector on the selected pixels, blobs, or key points (e.g., Lucas-Kanade algorithm (Lucas and Kanade 1981)); and 2) global optical flow estimation methods, which calculate the flow vector at each pixel of the image (e.g., Horn-Schunck algorithm (Horn and Schunck 1981)). Global optical flow estimation methods are ideal choices for full field structural displacement measurement, while utilizing pyramid, window and smoothing techniques. The local optical flow estimation methods can also be used to estimate the flow vector at each pixel. (Bouguet 1999) improved the original Lucas-Kanade method by implementing pyramid, feature tracker and interpolation to get the optical flow at each pixel. (Sun et al. 2014) analyzed the current practices in optical flow estimation quantitatively and most of the optical flow methods were developed using the formulation structured by Horn-Schunck. These methods are called classical methods. Based on the classical optical flow methods, Sun et al. implemented non-local smoothing techniques to develop a new method named Classic+NL (Classic with non-local). The performance of Classic+NL was validated by comparing with the classical optical flow methods on the popular optical flow datasets and showed better estimation results. (Khaloo and Lattanzi

2017) implemented the Lucas-Kanade (LK) method, Horn-Schunk (HS) method, Black and Anandan (BA) method, and Classic+non local smoothing (Classic+NL) methods, which are investigated in Sun's work (Sun et al. 2014), and developed pixel-wise structural motion tracking methods and verified them on two shaking table tests.

Even though these classical optical flow methods are programmed as built-in functions into current computer vision libraries such as MATLAB or OpenCV and successfully implemented for structural displacement monitoring, there are lots of parameters in the functions that need to be adjusted based on experience. Limitations of classical optical flow methods such as the small displacement assumption, brightness consistency, motion boundary problems are still the main sources of errors. For structural engineers without enough experience in the computer vision field, it is difficult to use such methods. Further, some of the classical methods are too slow to satisfy the requirement of real time monitoring.

Instead of the aforementioned classical methods in computer vision, deep learning has been a very popular tool to help address the challenges in the field of computer vision in recent years (Goodfellow et al. 2017; Lecun et al. 2015). With pre-trained deep neural networks, the optical flow can be easily estimated (Dosovitskiy et al. 2015; Ilg et al. 2017) by processing the image sequences through the pipeline of the forward propagation of the networks, without adjusting too many parameters as in classical optical flow methods. Since the dataset can be augmented by adding artifact noise, illumination change and other interference factors, the deep learning-based optical flow methods can perform better than the classical methods (Ilg et al. 2017). With GPU acceleration, the deep learning-based optical flow method can do real-time monitoring. Also, deep

learning based methods such as FlowNet (Dosovitskiy et al. 2015) and FlowNet2 (Ilg et al. 2017) are good at large displacement estimation, which is one of the drawbacks of the classical methods. In this study, a deep learning-based optical flow method, i.e., FlowNet2, is implemented to achieve full field structural displacement monitoring.

FlowNet2 is based on the work of FlowNet that was firstly proposed by Dosovitskiy et al. (Dosovitskiy et al. 2015). The study represented a paradigm shift in optical flow estimation by allowing the use of a simple Convolutional Neural Network (CNN) architecture to directly learn the concept of optical flow from the dataset. In FlowNet, Dosovitskiy et al proposed two CNN architectures: FlowNet-S and FlowNet-C. In FlowNet-S, Dosovitskiy et al first stacked two input images together and fed them through a generic network with 9 convolutional layers, allowing the network to decide itself how to process the image pair to extract the motion information. The first layer has a CNN kernel size of 7×7 , the second and third layers have kernel sizes of 5×5 , and the fourth to ninth have kernel sizes of 3×3 . The dimensions of each layer are conv1 ($354 \times 512 \times 6$), conv2 ($192 \times 256 \times 64$), conv3 ($96 \times 128 \times 128$), conv3_1 ($48 \times 64 \times 256$), conv4 ($24 \times 32 \times 512$), conv4_1 ($24 \times 32 \times 512$), conv5 ($12 \times 16 \times 512$), conv5_1 ($12 \times 16 \times 512$) and conv6 ($6 \times 8 \times 1024$). Finally, they added a refinement operation of the coarse feature maps to the high-resolution prediction and then provided the optical flow prediction. The detailed CNN architecture can be found in (Dosovitskiy et al. 2015). In FlowNet-C, instead of directly stacking two images, they first fed the two images to three convolutional layers separately and then combined them together with an explicit correlation layer. After another 6 convolutional layers, a refinement operation was added and then the optical flow prediction output. The training dataset FlowNet used is their homemade FlyingChair dataset, which simulates the motions and illumination change.

The pre-trained FlowNet performs well on the current optical flow dataset, especially for the case of large displacement. However, FlowNet still cannot compete with variational methods in small displacement and real-world data. (Ilg et al. 2017) proposed FlowNet2 based on FlowNet, where multiple FlowNet-S and FlowNet-C networks were stacked and a sub-network specializing in small motions to improve the accuracy and the speed of the original FlowNet was integrated. They trained the new CNN architecture on the FlyingChair and FlyingThings3D datasets. Illumination change, background clutter and other noises were added to the training data set as data augmentation to simulate real scenarios. Combining different training dataset and orders, FlowNet2 finally gave nine different CNN architectures for optical flow prediction. These CNN architectures are suitable for various application requirements such as small displacement, large displacement, good accuracy and fast speed as well as being capable of dealing with illumination change and background clutter. FlowNet2 performs well on small displacement and real-world data and is fast enough for real-time motion estimation. It should be noted that FlowNet2 will not always be the strong choice for full field optical flow estimation, and with the development of the computer vision techniques, more advanced optical flow algorithms will come out and be the alternatives for the purpose given in this study.

Figure 40 shows the optical flow estimations of a beam motion in two images using six different methods, namely, Horn-Schunk (HS), Lucas-Kanade with pyramid and sparse to dense interpolation (LKPyrSD), Farneback, BA, Classic+NL and FlowNet2. The reason the specific methods are chosen for comparison is that they are implemented and validated for structural displacement measurement in literature aforementioned. The beam has a downward deflection from Frame 1 to Frame 2 and since there are sensors installed on it, the motion in the images are

not just those of the beam but also those of the cables hanging from the sensors. In the flow field color coding, the color indicates the motion direction and the distance away from the center indicates the motion amplitude. In this case, the beam moves down from Frame 1 to Frame 2 so it should be colored yellow in the full field optical flow map according to the flow field color coding. In the optical flow estimation results, HS provides the worst estimate and is not robust in the presence of image noise. The beam motion is interfered with motions in other directions and the background causes an excessive amount of incorrect motion estimation. LKPyrSD gives poor results for motion boundaries and the motions on boundaries are blurred. The Farneback method cannot give accurate results for the whole beam but only those parts with salient textures. BA and Classic-NL perform better than these three but still give unsatisfactory results on boundaries. FlowNet2 gives the best results for beam motion, especially at boundaries. It even gives more detail about the motion of the cable. On this basis, the authors have selected FlowNet2 for full field optical flow estimation. The results using HS, LKPyrSD, BA and Classic-NL are similar to those by (Khaloo and Lattanzi 2017) and Classic-NL gives the best prediction among these four classical optical flow methods.

In the comparative studies of the experimental verification section, the authors will compare the results from FlowNet2 with Classic-NL for structural displacement monitoring. Instead of performing image re-projection using the planar homography matrix to mitigate the distortion caused by projection first and then estimating optical flow just as (Khaloo and Lattanzi 2017) did in their work, in this study the authors directly estimate the optical flow and shift the original points to the new location. Then the authors implement the planar homography matrix to project the location in the image to the real world. The consideration is that if the planar homography matrix

is applied to the image first, the re-projection might break the pixel structures and this makes optical flow estimation inaccurate.

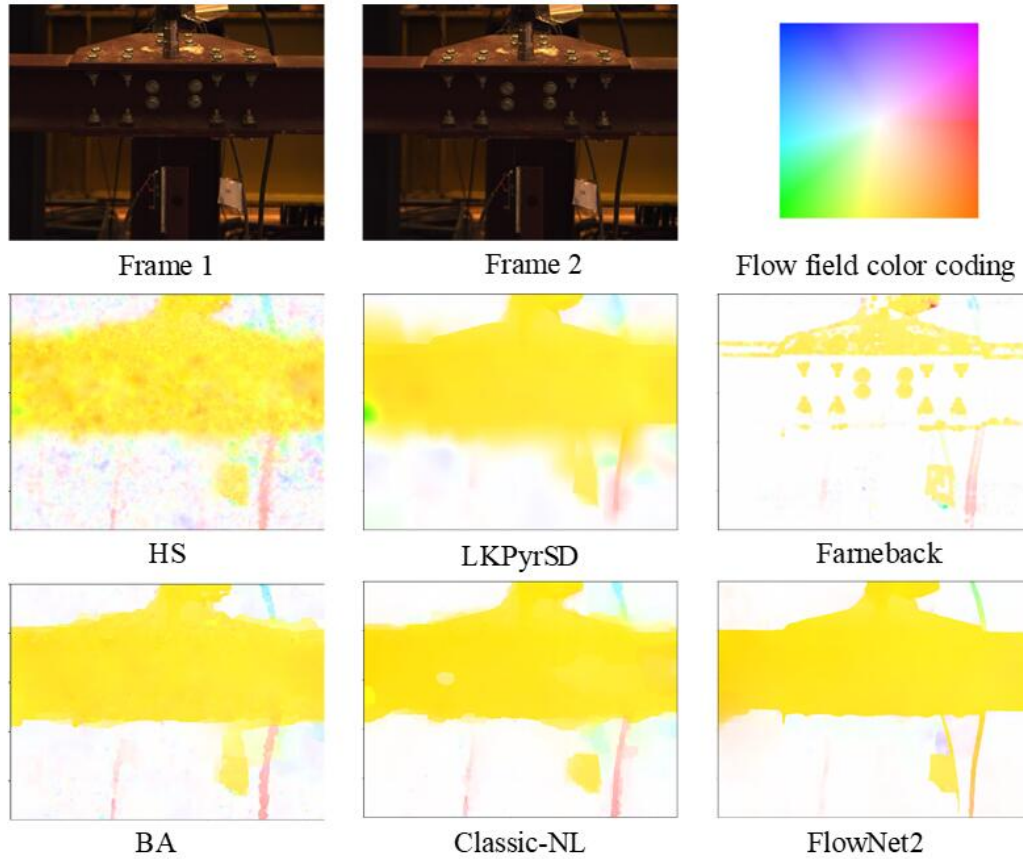


Figure 40 Optical flow estimations using different methods

5.2.4 Camera Motion Subtraction

When using vision-based methods to estimate structural displacement, especially in field application, camera motion is always a big issue which can induce displacement measurement errors. Camera motion may be caused by the ground vibration or wind. It will be mixed into the structural motion and has to be removed to rectify displacement measurement. There are two main

approaches to mitigate the camera motion: 1) filtering out the displacement components related to the frequencies of the camera motion; 2) directly subtracting the motion of a static object/scene in the video from the total motion. The first approach is suitable for the case when the frequency of the camera motion is not close to the structural motion with a trade-off that accelerometers are necessary to be installed on the camera to identify the frequencies of camera motion. The second approach is suitable for the case when there are static areas (objects assumed to be static) in the field of view of the camera (Feng and Feng 2017). In this study, the second approach is implemented to eliminate the displacement errors induced by the camera motion. As shown in Figure 41, the areas A, B, C and D can be assumed to be static. When a camera affected by ambient motion is utilized to measure the structural displacement of the footbridge, i.e., the displacement of M, the rectified displacement is M subtracted by the average displacement of A, B, C and D.

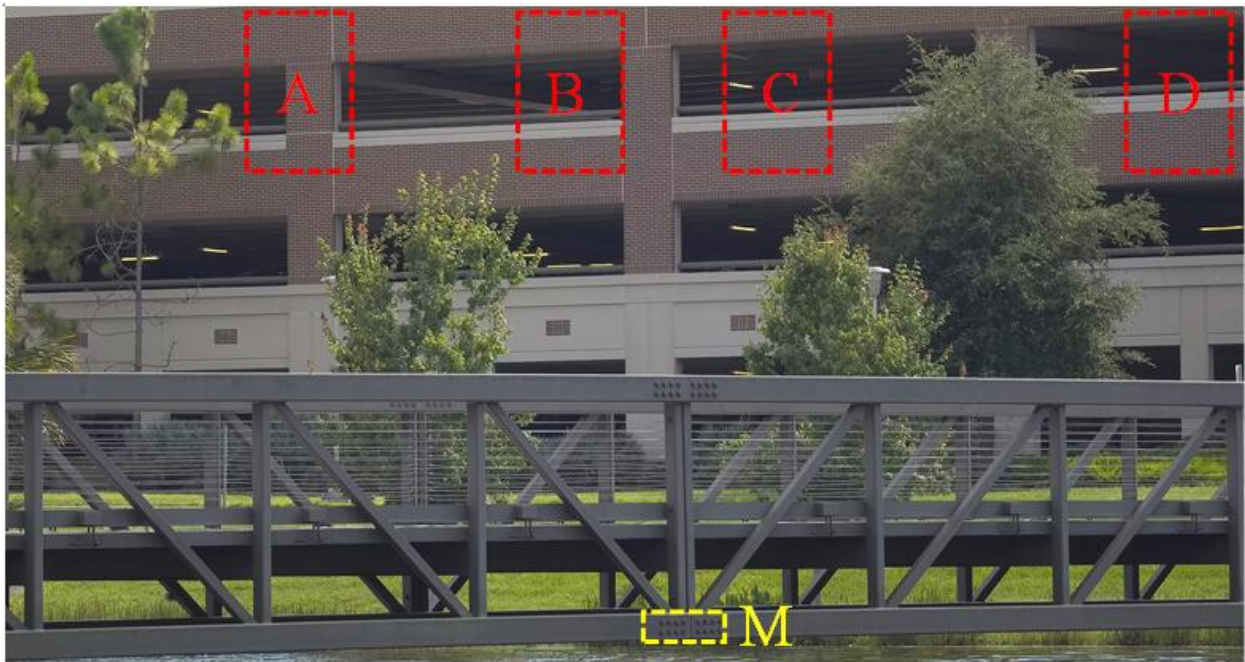


Figure 41 Structural displacement rectification using camera motion subtraction

5.2.5 Structural Displacement Calculation

Once the rectified optical flow is obtained, the planar homography matrix is applied to convert the displacement in pixels to their actual physical counterparts. By using full field optical flow and the planar homography matrix, the full field structural displacement is obtained. Theoretically, the displacement of any point of the structure can be acquired by taking the value of the full field structural displacement map. Conversely, in conventional SHM, displacement sensors are installed to measure the structural displacements at discrete points. As shown in Figure 42, a displacement sensor is installed to measure the structural displacement of the beam, and the result obtained here is the displacement at a discrete point. The experimental setup and loading condition are outlined in Section 5.3 and Figure 45.

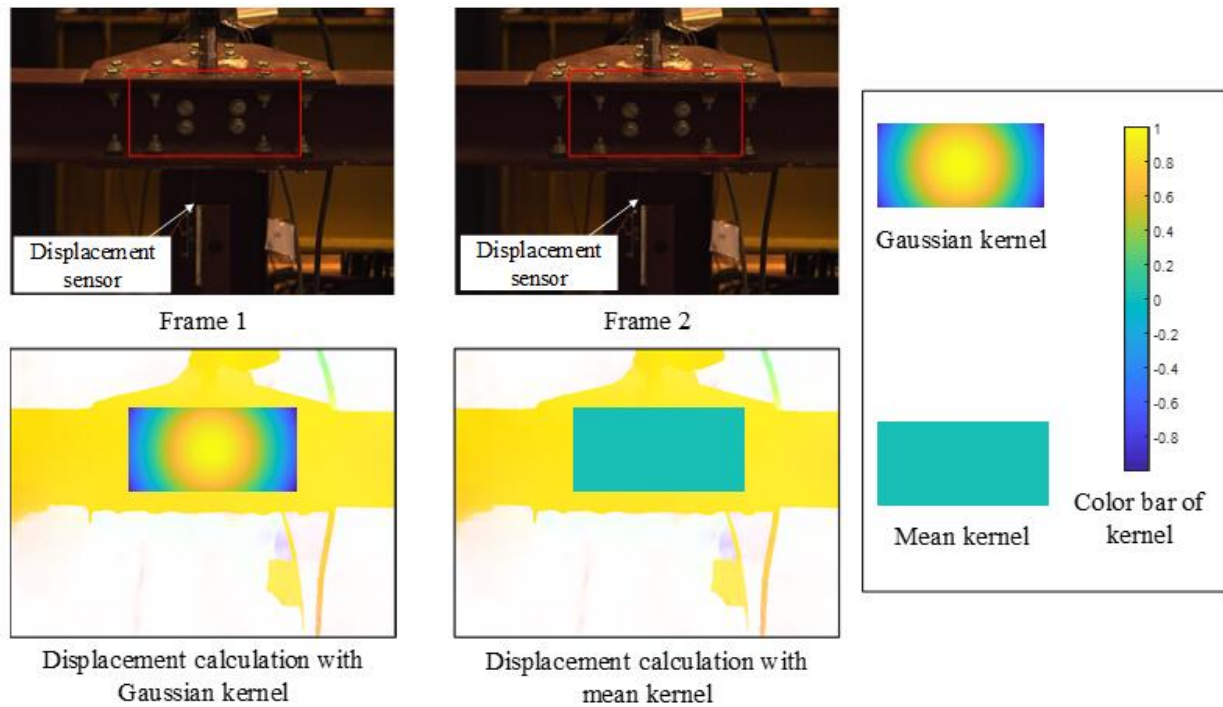


Figure 42 Structural displacement at a discrete point using kernels

The structural displacement at a discrete point can be obtained by using the displacement in the area close to the single measurement point. For example, the area marked by the red box in the two frames of Figure 42 is used to estimate the structural displacement of the discrete point that the installed displacement sensor measures. Two methods can be applied to estimate the structural displacement: 1) displacement calculation with Gaussian kernel; and 2) displacement calculation with mean kernel. The Gaussian kernel applied in this study is represented by \mathbf{G} , described in Eq. 35:

$$\mathbf{G}(m,n) = \frac{\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{\left(m-\frac{h_c}{2}\right)^2 + \left(n-\frac{w_c}{2}\right)^2}{2\sigma^2}}}{\sum_{m=1}^h \sum_{n=1}^w \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{\left(m-\frac{h_c}{2}\right)^2 + \left(n-\frac{w_c}{2}\right)^2}{2\sigma^2}}} \quad (35)$$

$$h_c = \left\lfloor \frac{h+1}{2} \right\rfloor \quad (36)$$

$$w_c = \left\lfloor \frac{w+1}{2} \right\rfloor \quad (37)$$

$$\sigma = \frac{h+w}{2} \quad (38)$$

where, m and n are the row column number of the Gaussian kernel, respectively, and h and w are its height and width. The symbol, $\left\lfloor \right\rfloor$ indicates the floor function that takes as input a real number, x , and gives as output the greatest integer less than or equal to x .

The mean kernel is represented by \mathbf{M} , given in Eq. (39):

$$\mathbf{M}(m,n) = \frac{1}{hw} \quad (39)$$

The Gaussian kernel used in this study actually models the focus of attention that is motivated by the biological visual system which concentrates on certain image regions requiring detailed analysis (Zhang et al. 2013). The closer to the focus center, the greater the weight is set. This Gaussian kernel implements the concept of attention guide tracking. Also, since most of the classical optical flow methods do not perform well on the motion boundaries, the Gaussian kernel can decrease the weight when calculating the weighted average displacement. When applying camera motion subtraction, if the displacement of the assumed static areas on the background are also calculated using a Gaussian kernel, the error then can be reduced by giving lesser weight to the parts away from the focus center of the static areas. This is an indirect way of suppressing the outliers, especially those close to the motion boundaries but far away from the focus center. When manually selecting the assumed static area, the assumption is more accurate as the kernel is placed closer to the center. In addition, the mean kernel is a well-known strategy which is applied in key points-based tracking using Lucas-Kanade optical flow (Dong et al. 2019b), key point matching using Fast Library for Approximate Nearest Neighbors (FLANN) and Kanade-Lucas-Tomasi based template matching (Yoon et al. 2016). Correlation based template matching also uses the mean kernel method to find the best location (Dong et al. 2018).

The displacement of a discrete point is estimated by combining the full field displacement of the selected region with either Gaussian kernel, \mathbf{G} , or mean kernel, \mathbf{M} :

$$d_{Gi} = \sum_1^m \sum_1^n (\mathbf{X}_i - \mathbf{X}_0) \circ \mathbf{G} \quad (40)$$

or

$$d_{Mi} = \sum_1^m \sum_1^n (\mathbf{X}_i - \mathbf{X}_0) \circ \mathbf{M} \quad (41)$$

where \circ is the element-wise product operator, d_{Gi} and d_{Mi} are the displacements estimated using Gaussian kernel and mean kernel, \mathbf{X}_0 is the original coordinate of the pixel-wise location in the real world, \mathbf{X}_i is the current (the i^{th} frame) coordinate of the pixel-wise location in the real world and $\mathbf{X}_i - \mathbf{X}_0$ is the displacement vector.

The displacement time history is obtained by calculating the optical flow between the current image (frame k as shown in Figure 43) and the original image (frame 1 in Figure 43, top row). In this strategy, there is neither frame nor tracking location update and the displacement at every single time point is independent of the others. Another strategy which is indicated in the bottom row of Figure 43 calculates the optical flow between two consecutive frames with updating the frame each time. In the displacement time history, the displacement at every single time point is dependent on its previous neighbors. The final displacement time history is the cumulative operation of the incrementals obtained by calculating the displacement of two consecutive frames.

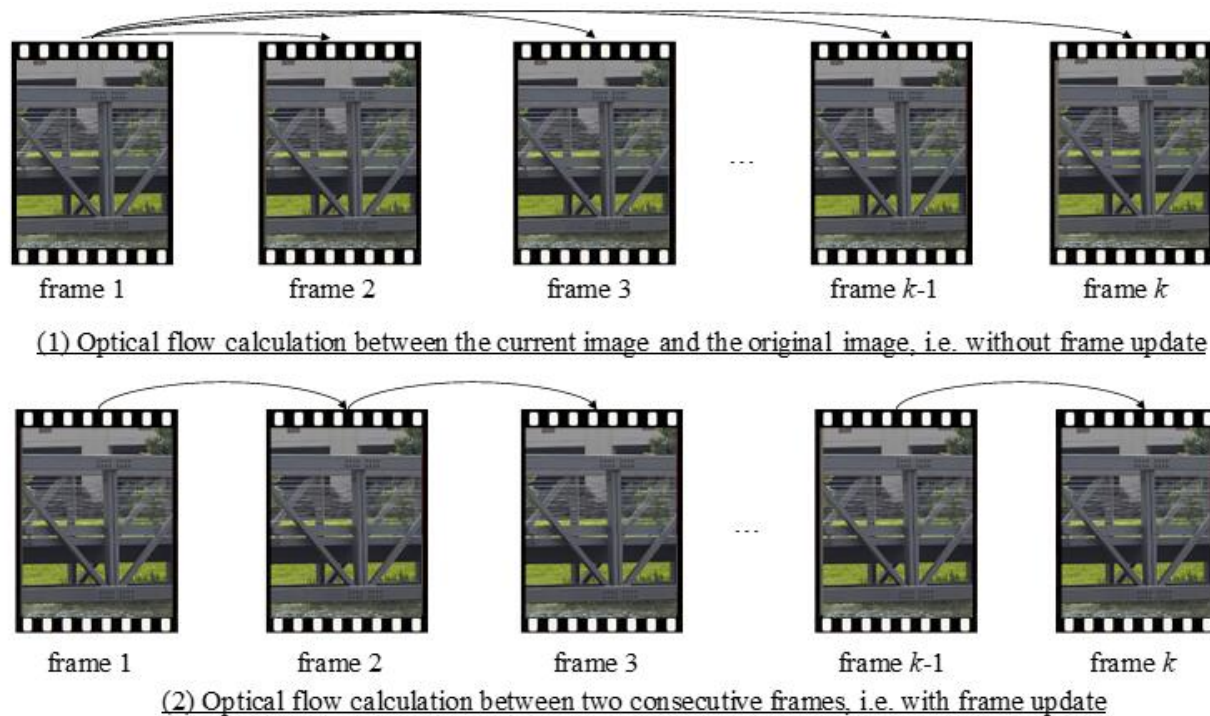


Figure 43 Two different strategies to process image sequences and get displacement time histories

With practical experiences and observations of conducting vision-based displacement measurement, general suggestions are summarized as the pros and cons of the two different strategies, which are listed in Table 7. The main reason that the first strategy is recommended is that when the second strategy (with frame update) is applied, errors tend to accumulate. (Stiros 2008) analyzed the accumulated errors in velocities and displacement deduced from accelerographs using numeric integration, which provided a possible way to eliminate the errors in the second strategy. Stiros indicated that the errors depend on the characteristic errors of accelerometers such as the sensitivity/accuracy of the measurements described by standard deviation, duration of the record and instabilities in the sampling rate. Also, the peaks in the accelerograms contribute to the errors during numeric integration. In this study, when using the second strategy, the process of calculating displacements from consecutive frames is very similar

to the process of calculating velocity from accelerations using numeric integration. It may be beneficial for the elimination of errors in the second strategy from the theoretical analysis of Stiros' work. While the formulas summarized by Stiros are limited to the analysis of linear movements and rotations are ignored; at the end, Stiros stated that if baseline corrections are taken into consideration for the formulas of errors, the numeric integration errors may be reduced. In the study, the first strategy to take the first frame as the baseline actually applies the way of baseline correction to some degree and it is more practical than using numerical integration with Stiros' theoretical analysis.

Figure 44 shows the displacement results obtained from the two different image sequence processing strategies and the ground truth (displacement sensor) from the same experiment introduced in Section 5.3. Due to the accumulation of errors, the displacement result obtained using this strategy with frame update deviates from the ground truth, while the result obtained using the strategy without frame update is consistent with it. While not updating brings inconvenience and possible errors to structural displacement measurement, these can be overcome by controlling the image quality and using pyramid methods for optical flow estimation or visual tracking. In the practice of conventional object tracking task, the second strategy (with frame update) is more popular and practical when processing the tracking problems with scale/view changes and illumination changes (OpenCV 2019d; Zhang et al. 2013). However, in vision-based displacement measurement, the first strategy (without frame update) is preferred, regardless of approach: digital image correlation based template matching, feature point matching, or optical flow (Dong et al. 2019b).

Table 7 Pros and cons of the two different strategies to process image sequence

Strategie s	Without frame update: Optical flow calculation between the current image and the original image	With frame update: Optical flow calculation between two consecutive frames
1	The displacement at every instant in the time history depends on the tracking between the current image and the original image and is independent of the others. The error in current time instant will not be accumulated.	Target scale changes, deformation, illumination changes and other changes of image quality can be updated in current tracking task and adjusted tracking scenarios give high chances of accurate tracking.
5	Without updating the frames, the target scale changes, deformation, illumination changes may affect tracking performance. Classical optical flow may fail to estimate large displacements in non-consecutive frames since there is a small motion assumption in classical optical flow methods. The measurement target may be out of view in the original frame.	The displacement at every instant in the time history is dependent on its previous neighbour. When calculating displacement time history, error in the current instant will be accumulated afterwards. This cumulative effect may cause a drift through the time history and a gradual loss of accuracy.

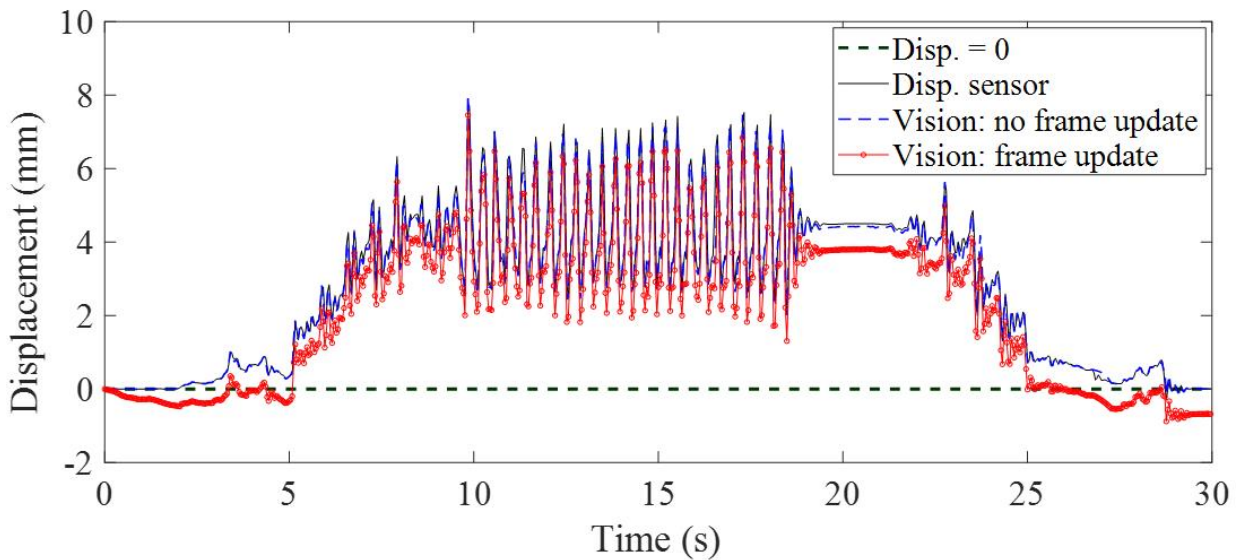


Figure 44 Displacement results obtained from the two different image sequence processing strategies and the displacement sensor

5.3 Laboratory Verification

5.3.1 Experimental Setup

In this section, an experiment on a model grandstand in the Structures Laboratory of the University of Central Florida is designated to verify the feasibility and performance of the proposed displacement methods. The grandstand, shown in Figure 45, is a scaled model of part of a real American football stadium. Detailed information can be found in previous papers (Celik et al. 2018a; Dong et al. 2019b). One region of interest (ROI), P1, is selected as the measurement point for the proposed method. At this point, a conventional displacement sensor (potentiometer) is installed to measure the displacement for comparison and is used as the ground truth. The cameras are MindVision-MV-GE131gc-t with a maximum frame rate of 60 Hz, a resolution of 1280×960 pixel and a zoom lens with a focal length of 5~100 mm. The cameras are connected to the same acquisition system as the displacement sensor. Unlike previous work(Dong et al. 2019b), in this study no trigger module is applied to enforce uniform sampling in the image data acquisition. The average frame rate of the camera is around 29 FPS (frame per second). The sampling rate for the displacement sensor is 100 Hz and it is down-sampled for comparison with the camera data. During the experiment, one person stands on the grandstand and jumps as the camera and potentiometer record the structural motion at P1. The acquired image sequence is analysed using the proposed methods. The displacement obtained from image sequences is compared with that of the displacement sensor.

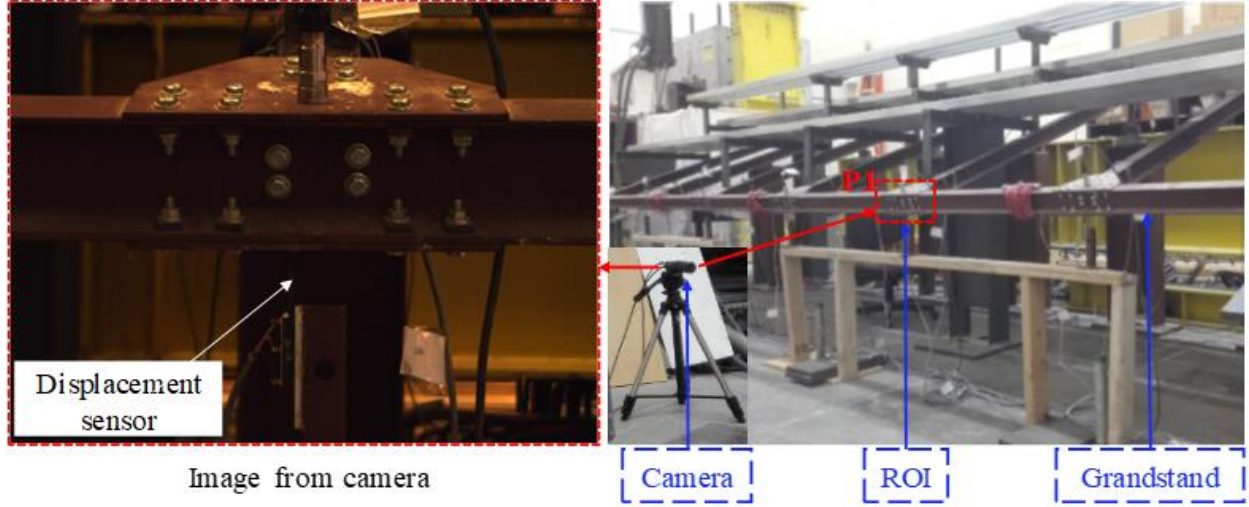


Figure 45 Experimental setup of the grandstand monitoring

5.3.2 Comparative Study of Displacement Measurement Using Different Methods

In this chapter, FlowNet2 is implemented and is verified through comparison with Classic+NL and the displacement sensor. Here, to obtain the displacement, both mean kernel and Gaussian kernel are applied to the full field optical flow results estimated by Classic+NL and FlowNet2. Figure 46 illustrates the comparison of displacement time histories of P1 using displacement sensor (Disp. Sensor) and vision-based methods, i.e., Classic+NL full field optical flow with Mean kernel (C+NL+M), Classic+NL full field optical flow with Gaussian kernel (C+NL+G), FlowNet2 full field optical flow with Mean kernel (FlowNet2+M), and FlowNet2 full field optical flow with Gaussian kernel (FlowNet2+G). The synchronization of different data sources is done manually. Different segments of the time history plot correspond to different events happening on the grandstand under human load. Firstly, the subject climbs up the grandstand causing an increase in displacement (0s~5s); then walks to P1 causing fluctuations and an increase in the displacement (5s~9s); then begins to jump (9s~19s), which produces a continuous up-and-down pattern; then

briefly stops; resumes jumping for two more seconds (19s~21s) and finally climbs down from the grandstand, allowing the displacement to return to zero. The figure indicates that the results obtained from all the vision-based methods are consistent with the benchmark.

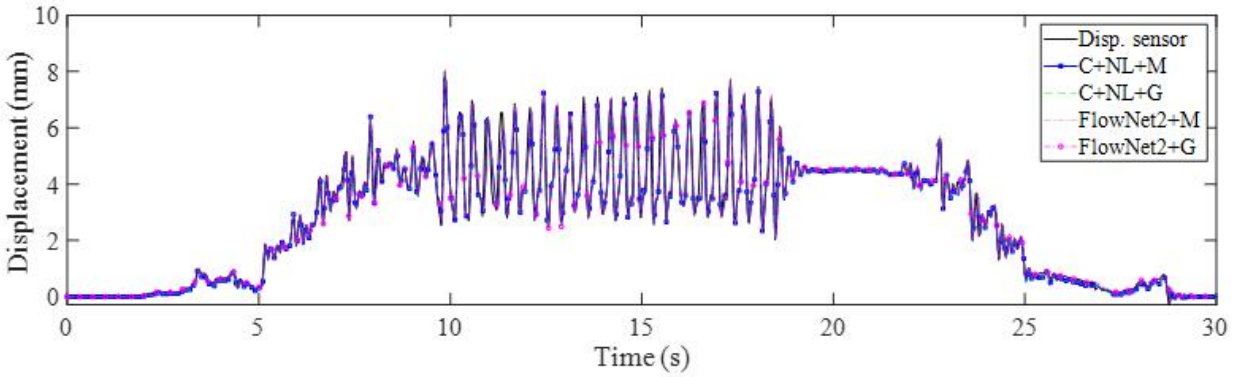


Figure 46 Comparison of displacement time histories from displacement sensor and vision-based methods

Before comparing the vision-based methods and displacement sensor quantitatively, the displacement time histories from the vision-based methods have to be pre-processed. As mentioned, the image sampling is non-uniform in this experiment. Figure 47a shows the time spent on the image collection for each frame, Δt and Figure 11b gives the histogram and normal distribution fit for the same variable. The mean time interval for a frame is $\mu = 0.0337$ s which gives an average camera frame rate of 29.7 FPS. This frame rate cannot be directly approximated to 30 FPS, because the standard deviation of the time interval, σ is 0.012 s, which is significant. To use the mean frame rate would cause misalignment problems for displacement time histories. The interval of mean ± 2 standard deviations, $[\mu - 2\sigma, \mu + 2\sigma]$, is, [0.010s, 0.057s] at a level of confidence of 95%. It should be noted that the uncertainties of image process measurement are the

selection of visual tracking algorithms/optical flow algorithms, region of interest and average methods (mean kernel or weighted kernels) etc. These are also the sources of the uncertainties.

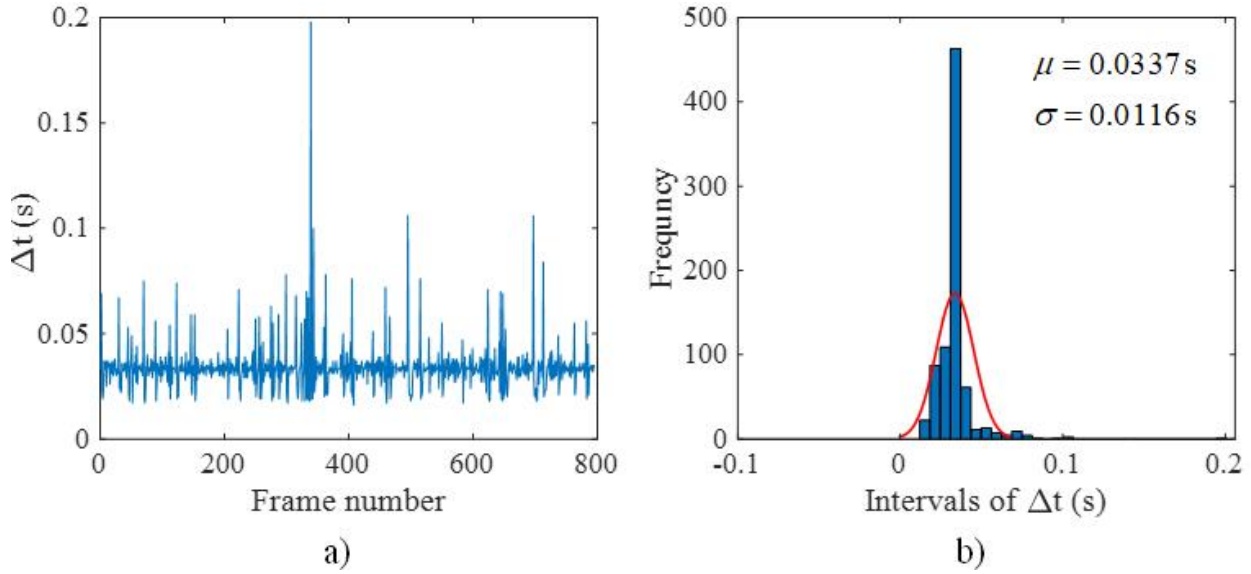


Figure 47 Statistical analysis of time spent on image collection for each frame

The non-uniformly sampled displacement time histories obtained from the vision-based methods are first resampled at 25 Hz using a cubic spline interpolation, while the uniformly sampled displacement time history obtained from the displacement sensor is directly down-sampled to 25 Hz. Cross correlation (Oppenheim et al. 1996) is applied to synchronize the resampled displacement time histories obtained from the vision based methods and the displacement sensor. Figure 48 shows the resampled displacement time histories of P1 using all methods. After resampling, the consistency between the displacement time histories obtained from vision-based methods and displacement sensor are still very good and do not change, compared to Figure 46. Normalized root mean square error (NRMSE) is applied to evaluate the goodness of fit between the signals, and normalized cross-correlation (NCC) is calculated to evaluate the similarities between them:

$$\text{FIT}_{\text{NRMSE}} = 1 - \frac{\|d_v(i) - d_s(i)\|}{\|d_s(i) - \mu_{d_s}\|} \quad (42)$$

$$\text{NCC} = \frac{|\sum (d_v(i) - \mu_{d_v}) \times (d_s(i) - \mu_{d_s})|}{\sqrt{\sum (d_v(i) - \mu_{d_v})^2} \sqrt{\sum (d_s(i) - \mu_{d_s})^2}} \quad (43)$$

where $d_v(i)$ and $d_s(i)$ are the displacements from the vision-based methods and displacement sensor, respectively, and μ_{d_v} and μ_{d_s} are their mean values. The greater the values of $\text{FIT}_{\text{NRMSE}}$ and NCC, the better the fit. It can be seen in Table 8 that the proposed methods, FlowNet2+M and FlowNet2+G, perform a little bit better than the alternatives, C+NL+M implemented in literature (Khaloo and Lattanzi 2017). For example, the normalized mean square error for the former two is 0.8758 which is slightly better than that of the latter (0.8727). Similarly, the NCC of the former at 0.9923 is slightly better than that of the latter at 0.9921.

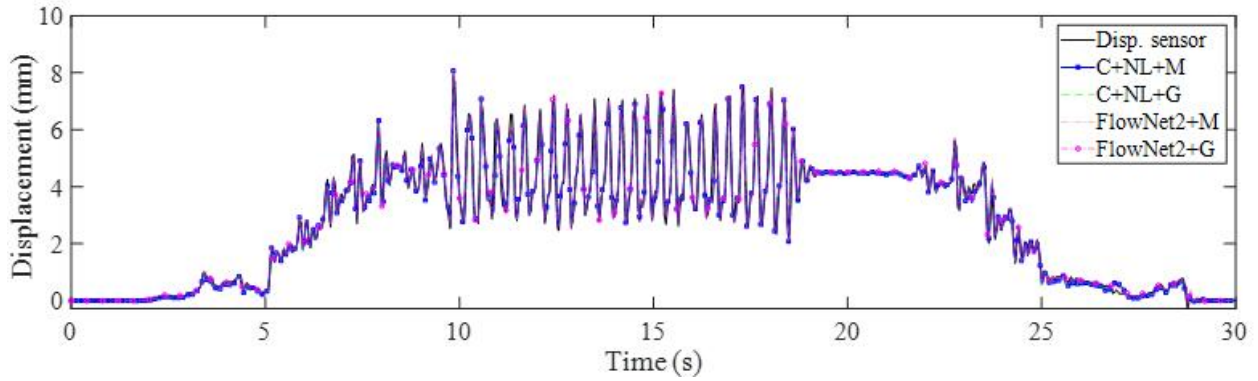


Figure 48 Resampled displacement time histories using displacement sensor and vision-based methods

Table 8 Normalized root mean square error (NRMSE) and normalized cross-correlation (NCC) of the fit between the vision-based displacements and the benchmark

Method	FIT_{NRMSE}	NCC
C+NL+M	0.8727	0.9921
C+NL+G	0.8726	0.9921
FlowNet2+M	0.8758	0.9923
FlowNet2+G	0.8758	0.9923

The displacement time histories when no loads are present on the structure are extracted as the measurement error distribution and are used to estimate the measurement accuracy and resolution. The index of the measurement accuracy and resolution is defined by ± 2 standard deviations, which corresponds to a level of confidence of 95% (Khuc and Catbas 2017; Xu et al. 2018). The smaller the standard deviation, the smaller the error and the better the measurement accuracy and resolution. Figure 49 shows the distributions of measurement error from different methods. Table 9 gives the measurement resolution ($\pm 2\sigma$) analysis. From Figure 49 and Table 9, it is indicated that the accuracy and resolution of the proposed methods using FlowNet2 is ± 0.0029 mm which is very close to the ground truth (displacement sensor) which is ± 0.0021 mm. The vision-based method using Classic+NL, has a resolution of around 0.0240 mm, which is almost 10 times that of the proposed methods. Further, from Figure 49, it can be seen that the proposed methods using FlowNet2 give much better stability than Classic+NL in displacement measurement, especially from the sixth subplot “Distribution comparison”.

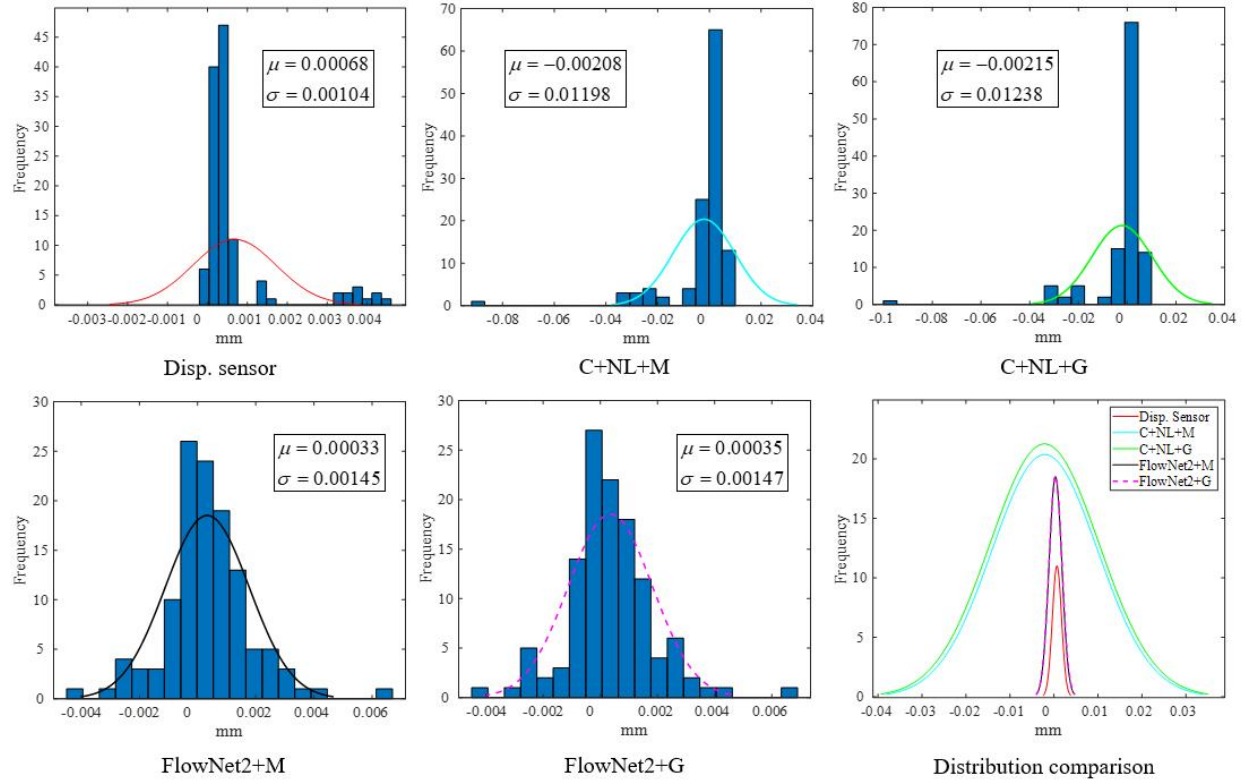


Figure 49 Distributions of measurement error for different methods

Table 9 Measurement accuracy and resolution ($\pm 2\sigma$) analysis

Method	Disp. sensor	C+N+M	C+N+G	FlowNet2+M	FlowNet2+G
$\pm 2\sigma$	± 0.0021	± 0.0240	± 0.0248	± 0.0029	± 0.0029

Since the displacements obtained from the vision-based methods are non-uniformly sampled, they cannot be directly processed by general Fast Fourier Transform (FFT) or Power Spectral Density (PSD). Therefore, in this study, two ways are applied to extract the frequency information from the non-uniformly sampled displacement data: 1) Lomb-Scargle Periodogram for the non-uniform sampled data (Lomb 1976); and 2) Power Spectral Density Using FFT for uniformly sampled data from cubic spline data interpolation. Figure 50 shows the comparison of displacement data in the

frequency domain. The modal testing of the same structure was conducted in previous work and the first natural frequency is 5.74 Hz (Celik et al. 2018a; b; Dong et al. 2017, 2019b). From this Figure 50, it is suggested that all the methods give the operational modes for human jumping (frequency around 2.86 Hz). However, Lomb-Scargle method does not give unique peaks (peak at 2.825 Hz and peak at 2.881 Hz) and the peaks have a clear shift from those obtained from the displacement sensor which is regarded as ground truth. This is a distortion in dynamics which may be induced by non-uniform sampling. The peaks obtained from the cubic spline data interpolation of the original non-uniform data is very close to the ground truth. By picking the peak of the PSD curve, the frequencies of human jumping load are extracted as shown in Table 10. From this table, it is indicated that using cubic spline data interpolation seems better than directly using the Lomb-Scargle method. Using cubic spline interpolation data gives the exact human jumping mode frequency as the displacement sensor. In this view, using cubic spline interpolation may correct the distortion that may be induced by non-uniform sampling. It might be because the sampling rate is low and, in this case, cubic spline interpolation works.

Besides, in Figure 50 there is a discrepancy between spline and displacement sensor results at higher frequencies, which is expected due to the resampling step. This observation reflects the limitation of the vision-based methods: inherently vision-based methods give a lower frequency range than conventional displacement measurement for hardware and software processing reasons, so that over the entire range, especially the high frequency range, vision-based methods are not as sensitive as their conventional counterparts.

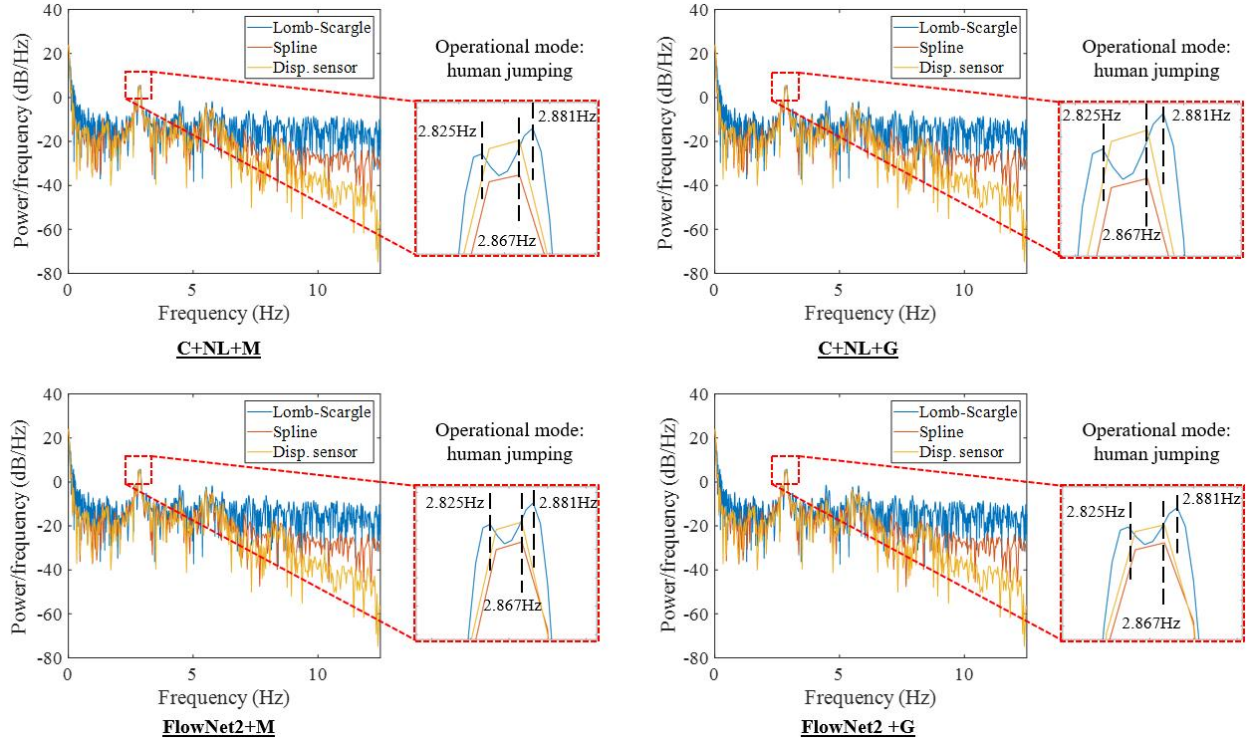


Figure 50 Comparison of displacement data in the frequency domain

Table 10 Frequencies of human jumping load extracted from displacement data: unit (Hz)

Method	Disp. sensor	C+N+M	C+N+G	FlowNet2+M	FlowNet2+G
Direct PSD	2.867	--	--	--	--
Spline+PSD	--	2.867	2.867	2.867	2.867
Lomb-Scargle	--	2.881	2.881	2.881	2.881

From the accuracy and resolution analyses, as shown in Table 8 and Table 9, using mean kernel seems to be a little better than the Gaussian kernel, which was not expected from the very beginning. The authors believe this to be because the region using the kernel did not contain too many motion boundaries and the advantage of Gaussian kernel was not manifest. As a result, at this time, it remains an open question whether it is better to use mean kernel or Gaussian kernel.

5.3.3 Comparison of Computation Times of Vision-Based Methods Using Different Full Field

Optical Flow Algorithms

When comparing the image processing time, it takes 1.7 seconds to calculate the full field optical flow of two images with a resolution of 1280×960 pixel using FlowNet2. The computation time is accelerated by a Graphics Processing Unit (GPU) on a Linux system (Ubuntu 18.04) with the AMD Ryzen 5 2600X CPU, 16Gb RAM, and the NVIDIA GeForce GTX 1080 Graphics Card. It takes about 1600 seconds for the same operation on the same system using Classic+NL. The Classic+NL used in this study is the same as that of Khaloo and Lattanzi and does not implement GPU acceleration. During this experiment, 1159 images were collected, and it took about 32.8 minutes using FlowNet2 to calculate the full field optical flow of the image sequence whereas it took about 21.5 days for Classic+NL. The processing tests were conducted on the same computer and computation times were directly extracted from the internal clock when running optical flow codes. This is a limited test, but it is a practical and simpler way to compare the speeds of different algorithms. To date, it is unknown whether Classic+NL can use GPU acceleration. Perhaps in the future, the current Classic+NL can be extended to a GPU version and the processing speed accelerated. At this time, the proposed method of implementing FlowNet2 gives a much higher processing speed and at the same time, provides better accuracy. As a deep learning-based optical flow estimation algorithm, FlowNet2 can take advantage of GPU acceleration technology, which makes the highly time-consuming optical flow estimation task much faster. As stated in the literature (Ilg et al. 2017), FlowNet2 and its sub-networks can achieve 8 to 140 FPS real time optical flow estimation on the Middlebury data set with an NVIDIA GeForce GTX 1080 Graphics

Card. This means that FlowNet2 and its sub-networks can be easily implemented to do real time full field displacement measurement. From this experiment, it can be clearly seen that the vision-based method using Classic+NL algorithm is much slower than using FlowNet2.

5.4 Field Application

5.4.1 Experimental Setup

In this section, a field application is presented that was performed on a footbridge on campus to verify the feasibility of the proposed displacement measurement methods. The structure (Figure 51) is a three-span (7.31m + 39.01m + 7.3m) truss bridge with a width of 3.65m. A portable camera (Z Camera E1) with a resolution of 1980×1080 pixels, a frame rate of 60 FPS and an Olympus zoom lens with a focal length of 75-300mm was used to collect images during the experiment. The measurement location is at midspan and T1 (see figure) was selected as the region for the vision-based measurement. The distance from the camera to the measurement region was about 52 meters. An accelerometer was installed at midspan to measure the vibration. The sampling rate of the accelerometer was 200 Hz. Point T0 in the background was selected as the static reference and used to eliminate the camera motion caused by ground vibration and wind effects. During the experiment, two persons jumped at the bridge midspan and both the camera and the accelerometer recorded the vibration of the bridge.

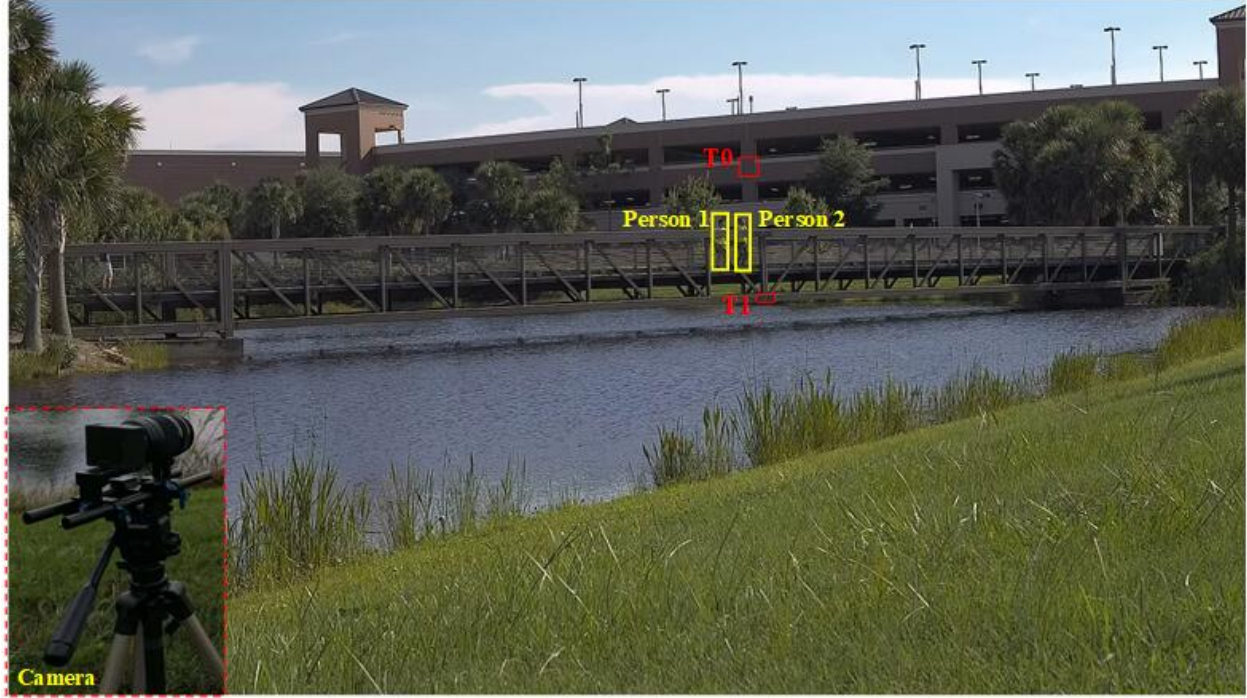


Figure 51 Experimental setup for a footbridge

5.4.2 Analysis and results

Figure 52 illustrates the displacement time histories obtained from the proposed methods. Here both Mean kernel and Gaussian kernel are used to calculate the displacement at midspan. In this experiment, using Mean kernel and Gaussian kernel give almost identical displacement results. The red curve (-) and the cyan dashed curve with circle (-o-) show the original displacement at T1 without doing camera motion subtraction, while the blue curve (-) and the magenta dashed curve with asterisks (-*-) show the displacement at T1 with camera motion subtraction. Further, the black dashed curve is added to show the zero line. In this figure FlowNet2+G org and FlowNet2+M org represent the original displacement data obtained using FlowNet2 with Gaussian Kernel and mean kernel, respectively. While FlowNet2+G w/ cam. mot. subtr and FlowNet2+M w/ cam. mot. subtr

represent the ones with camera motion subtraction. Please note that FlowNet2+G org and FlowNet2+M org as well as FlowNet2+G org and FlowNet2+M org are almost identical in Figure 16. After camera motion subtraction, the structural vibration varies up and down by around ± 2 mm. The camera motion subtraction shifts the displacement downwards by about 1.5 mm, especially in the first 32 seconds, and the range is reduced by about 38%. Clearly the camera motions caused by ground vibration and wind have a substantial influence on the displacement measurement and it is necessary to correct for the camera motions. Technological advances in cameras might provide vibration reduction features, which can also be considered for camera shake correction in the future.

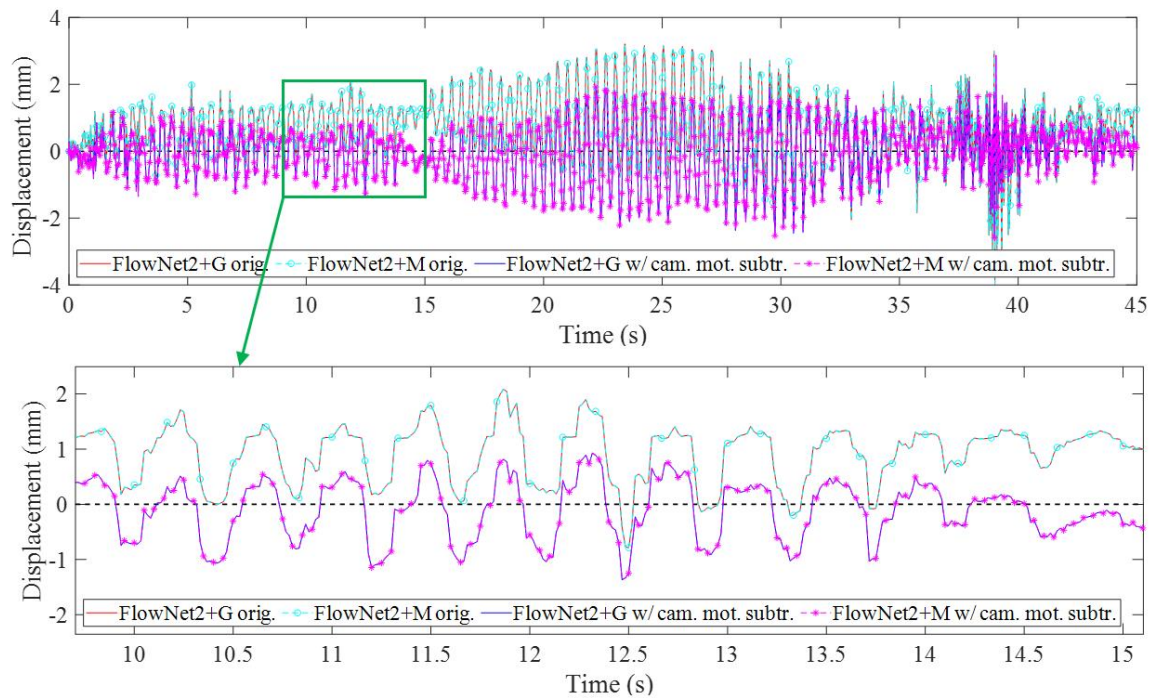


Figure 52 Displacement time histories obtained from the proposed methods. FlowNet2+G org and FlowNet2+M org represent the original displacement data obtained using FlowNet2 with Gaussian kernel and Mean kernel, respectively. While FlowNet2+G w/ cam. mot. subtr and FlowNet2+M w/ cam. mot. subtr represent those with camera motion subtraction

Figure 53 shows the acceleration time history obtained from the accelerometer installed at T1. A comparison was conducted between the displacement from the proposed methods and the

measured acceleration in the frequency domain. The portable camera used in this experiment provides a uniform sampling rate so that there is no need to process the displacement data using non-uniform frequency analysis methods. By directly applying an FFT to the displacement and acceleration data, the frequency spectra are obtained and are shown in Figure 54. By using a peak-picking method, the operational modal frequencies of the footbridge under human loads are extracted and summarized in Table 11.

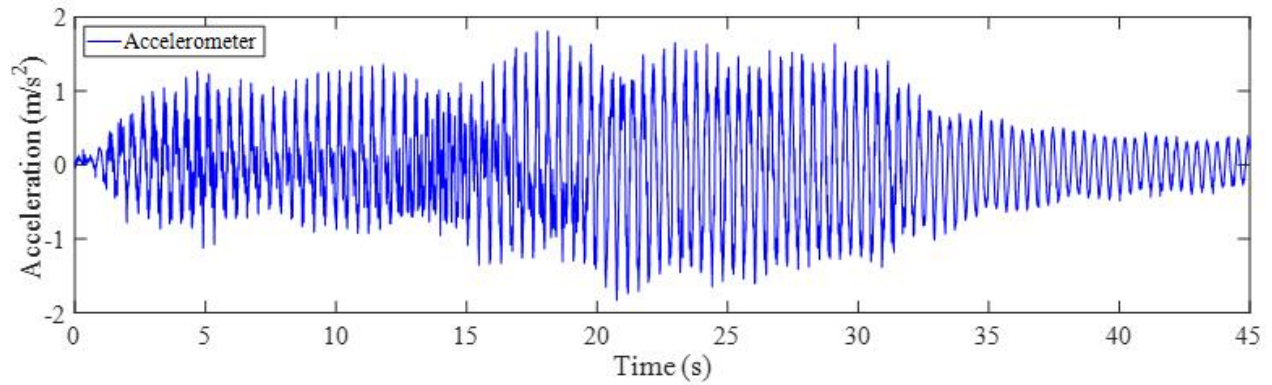


Figure 53 Acceleration time history obtained from accelerometer

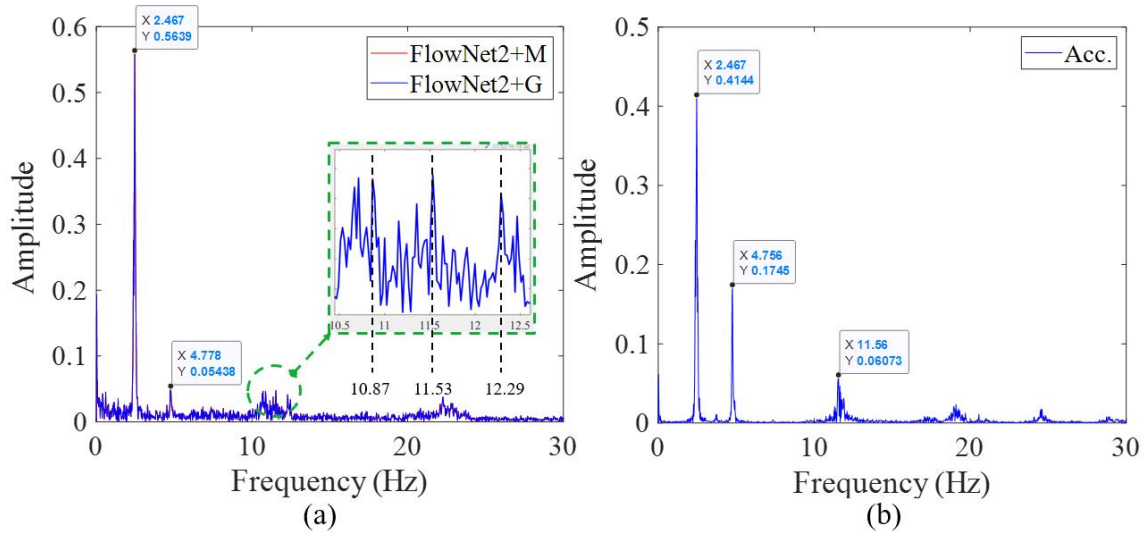


Figure 54 Operational frequencies of bridge under human jumping excitation: (a) vision-based system; (b) accelerometer signal

Table 11 Comparative study of operational modal frequencies from vision and accelerometer signals

Operational Mode	$f(\text{Hz})$: Vision	$f(\text{Hz})$: Acc.	Difference between Vision and Acc
1	2.467	2.467	0.00%
2	4.778	4.756	0.46%
3	10.87, <u>11.53</u> , 12.29	11.56	5.97%, <u>0.26%</u> , 6.31%

In this study, the peak frequencies are simply those observed from displacement and acceleration data. There was no investigation to determine whether they are caused by human jumping or are natural modal frequencies. The first three operational modes are listed and for the modes extracted from the proposed methods, they are very close to those from the accelerometer. However, around the third mode (i.e., 11.53 Hz), there are two more pseudo-modes (10.87 Hz and 12.29 Hz) which make it hard to pick the right operational mode. Compared with the third natural frequency measured by the accelerometer (i.e., 11.56 Hz), it is assumed that 11.53 Hz in the vision-based signal, is most likely to be the third operational mode. The pseudo-modes may come from the camera motion caused by wind or ground motion. Even though camera motion subtraction is applied in this case, there may still be vibration which cannot be removed completely. From Table 11, it can be seen that the difference between the modal frequencies obtained from the proposed methods and the accelerometer are all less than 0.5%. This gives a lot of confidence that the proposed method is accurate and capable of use in field application.

It should be noted that the camera motion effects are not completely eliminated as shown in Figure 53 and Figure 54, even camera motion subtraction is applied. Camera motion subtraction is partially efficient in getting rid of the tripod/camera vibration effects. The pseudo modes can be

removed by filtering out the vibration frequencies of the tripod/camera set-up. While they are not measured in this study and it is one of the limitations.

5.4.3 Recommendations for Practice in Field Application

In field application, the measurement environment is different to that in a laboratory, and the following is recommended:

(1) Camera motion: The influence of camera motion needs to be minimized. The effect of wind and ground vibration should be reduced by careful selection of the camera location. It is useful to include a stationary object in the field of view to facilitate camera motion subtraction. Putting a triaxial accelerometer on the camera and filtering out the camera motion effects are not inconvenient in field application.

(2) Background clutter: Background clutter (e.g. due to leaves moving in the wind) should be avoided because, when calculating full field optical flow, the motions in the background may cause difficulties. It tends to reduce the accuracy of the flow prediction of the measurement.

(3) Region selection of target: The target in the image should be sufficiently large and it should be ensured that no other moving object is inside the region that is not part of the target.

(3) Kernel selection: For simplicity and convenience, the mean kernel can be selected since there is very little difference in the use of mean kernel and Gaussian kernel in this experiment.

(4) Camera calibration: Drawings of the structure should be used, or dimensions measured to facilitate image calibration.

5.5 Summary

To achieve non-contact displacement monitoring for civil structures with less user involvement and to overcome the limitations of common vision-based methods, a novel full field structural displacement measurement method using deep learning-based optical flow, is proposed. The feasibility of the proposed method is verified through a comparative study of a series of laboratory experiments and a field application. The main conclusions are as follows:

(1) A procedure for vision-based displacement measurement is presented and provides a standard reference for future users.

(2) A deep learning-based full field optical flow algorithm, FlowNet2, is implemented in the proposed approach. It decreases the requirement for human involvement in the operation and gives more accurate measurement results with less computation time.

(3) Issues in vision-based methods for real-time monitoring and post processing are explored and strategies for the use of portable cameras, industrial cameras, triggers and time control are presented. The non-uniform sampling problems are discussed, and camera trigger, spline interpolation, Lomb-Scargle method are recommended to solve the problems.

(4) Strategies for displacement calculation in common vision-based methods are discussed, specifically the issue of whether to calculate the motion between consecutive images or between the current image and the initial one. To reduce drift caused by an accumulation of errors in calculating the differences between consecutive images, the authors recommend the latter approach.

(5) The camera motion issue is discussed in the context of field application. Camera motion subtraction is proposed to address the errors induced by camera motion.

In the future, more work will be done to process the non-uniform sampled image data and explore the application of kernels in calculating displacements at discrete structural points. Additional study will be focused on the investigation of bridge deflection profile, full field structural modal analysis and distribution factor calculation using the proposed method. Also, how shading and illumination affect the proposed method will be evaluated.

CHAPTER SIX: A ROBUST VISION-BASED METHOD FOR DISPLACEMENT MEASUREMENT UNDER ADVERSE ENVIRONMENTAL FACTORS USING SPATIO-TEMPORAL CONTEXT LEARNING AND TAYLOR APPROXIMATION

Previously published as Dong, C. Z., Celik, O., Catbas, F. N., OBrien, E., and Taylor, S. (2019e). “A Robust Vision-Based Method for Displacement Measurement under Adverse Environmental Factors Using Spatio-Temporal Context Learning and Taylor Approximation.” *Sensors*, 19(14), 3197.

6.1 Introduction

Computer vision-based displacement measurement using cameras has attracted increasing attention in the community of structural health monitoring (SHM) because of its characteristics as a non-contact, long-distance, multi-point, high-precision, time-saving, and cost-effective sensing technique (Brownjohn et al. 2017; Celik et al. 2018a; Chen et al. 2018b; Dong et al. 2019b; Feng et al. 2015a; b; Feng and Feng 2016, 2017, 2018; Fukuda et al. 2013, 2010; Khuc and Catbas 2016, 2017; Lydon et al. 2016; Schumacher and Shariati 2013; Xu and Brownjohn 2018; Yoon et al. 2016). Structural displacement is a critical indicator for evaluating performance and identifying and determining the effects of damage/change under external loads (Dong et al. 2019c; Dong and Catbas 2019). For instance, during the regular operation of a structure, displacement can be monitored to ensure that it stays within a specified tolerance and safety range (Ye et al. 2015). Once the displacement time histories from the monitored structures are extracted using vision-

based methods, traditional structural health monitoring and behavior analysis can easily be conducted (Chen et al. 2015b, 2019). Vision-based displacement measurement methods are also applied for bridge load testing to evaluate the bridge load carrying capacity (Lee et al. 2006a) and have even been used for contactless bridge weigh-in-motion (Ojio et al. 2016). Combining the multi-point displacement response with structural input data extracted from vehicle tracking, structural identification can be carried out using traditional structural indicators such as the unit influence line (UIL) and unit influence surface (UIS) (Catbas et al. 2018; Dong et al. 2019a; Khuc and Catbas 2018). Without the need for the deployment of conventional sensor networks, operational modal analysis can be performed using vision-based displacement measurement methods, which may provide multi-point synchronization and, therefore, a much denser spatial resolution than is practical with conventional sensors (Chen et al. 2015a; Dong et al. 2018; Feng and Feng 2016; Fioriti et al. 2018; Ji and Chang 2008b; a; Poozesh et al. 2017). Full field motion estimation and instantaneous mode shapes can even be obtained with high spatial and temporal resolution (Chen et al. 2015b; Yang et al. 2017b; a, 2018b). Reference (Fioriti et al. 2018) introduced an application based on motion magnification for modal identification of an on-the-field, full-scale, large historic masonry bridge by using videos taken from a common smartphone device. Modal properties and other indices derived from vision-based displacement time histories can be turned into sensitive indicators for structural damage detection and model updating (Cha et al. 2017b; Feng and Feng 2015; Yang and Nagarajaiah 2015). There are also numerous studies related to the estimation of stay cable forces that use vision-based displacement measurement (Feng et al. 2017; Ye et al. 2016b). In addition to structural response monitoring, the external loading information can be predicted. (Celik et al. 2018a) estimated the load time histories of individuals and crowds with the displacement time histories obtained using computer vision-based

measurements. These successful research applications make computer vision-based displacement methods a very promising complementary tool to conventional structural health monitoring practices, particularly for bridges.

The majority of applications and experiments in the literature are conducted in an ideal measurement environment so that an adequate measurement performance and accuracy is ensured. In addition, when these experiments are performed for the purposes of new method verification or comparison, the measurement time span is generally short and the adverse factors which can influence the measurement accuracy and stability are mostly avoided. For a general proof of concept, it makes sense to conduct such studies. However, when vision-based systems are intended for long-term deployment, either as a standalone or to complement a conventional SHM system, some unfavorable contingencies may affect the measurement quality. Even in the short term, the accuracy and stability of a vision-based system can be affected adversely. In a review of the current literature, (Feng and Feng 2018) summarized the possible measurement error sources in vision-based methods, including: (1) camera motion; (2) coordinate conversion; (3) hardware limitations; and (4) environmental sources. (Brownjohn et al. 2017) investigated the challenges in field application of a commercial vision-based system resulting from camera instability, the nature of the target (artificial or structural feature), and illumination. (Ye et al. 2016c) reviewed the state-of-the-art systematic errors, assessment, and reduction, including: (1) target size and texture; (2) camera alignment; (3) motion blur; and (4) the ratio between target size and full view. (Xu and Brownjohn 2018) reviewed subpixel techniques used in vision-based displacement measurement methods. (Ma et al. 2012) studied the measurement error in the digital image correlation method caused by the self-heating of digital cameras. (Ye et al. 2016f) conducted a series of shaking table

experiments in the laboratory to examine the influence of environmental factors which affect the accuracy and stability of vision-based systems. The targets used in the experiments were QR (quick response) codes and the textures of the QR codes showed rich sparkle patterns. It is suggested that measurement results are adversely affected by illumination and vapor. Subsequently, (Dong et al. 2015; Ye et al. 2016d) investigated the possibility of improving the accuracy and the stability of vision-based systems by mitigating the adverse effects of vapor. They used light emitting diodes (LEDs) and infrared emitting diodes as the measurement target and the experimental results showed that these emitting diodes can mitigate the adverse effects of vapor. However, installing these kinds of targets on the structure can be difficult, perhaps requiring wiring and a main power supply, which may not be feasible for a bridge.

These problems may decrease the accuracy of the measurement results and affect the evaluation of structural performance and health conditions when using vision-based monitoring systems over long-term time spans. In the literature, there are lots of studies on the analysis of sources of error, but only a few (Ye et al. 2016c; f) seek to improve system performance under adverse influencing factors. Therefore, it is essential to develop a robust vision-based displacement measurement method for long-term structural monitoring, which can handle some of these adverse factors.

While one study cannot address all issues related to computer vision-based monitoring, this paper focused on the mitigation of environmental factors such as illumination change and fog interference, and improvement of the measurement sensitivity at the subpixel level. A robust vision-based displacement measurement method was developed, leveraging the advantages of high-resolution imaging and computer-vision techniques to mitigate the interferences induced by

illumination change and fog, and was adapted for long-term bridge monitoring. The proposed method utilizes the spatio-temporal context (STC) learning algorithm to track measurement objects in image sequences and obtain the locations. The STC algorithm (Zhang et al. 2013) builds the spatio-temporal relationships between the measurement target and its local context based on a Bayesian framework, which models the statistical correlation between the low-level features (i.e., image intensity and position) from the measurement target and its surrounding regions. The tracking problem is solved by computing a confidence map and obtaining the best target location by maximizing an object location likelihood function. Combining this with the Taylor approximation (Chan et al. 2010), the accuracy of the proposed method achieves a subpixel level without sacrificing processing speed. The objectives of this study were: (1) developing a new vision-based displacement method using spatio-temporal context learning; (2) achieving a subpixel level estimation based on a Taylor approximation for the new vision-based method; and (3) verifying the feasibility, stability, and robustness of the proposed method via comparison with the current vision-based methods and conventional displacement sensor (Linear Potentiometer, LP) by conducting a series of experiments under two adverse environmental factors (illumination change and fog) on a two-span three-lane model bridge in the laboratory.

6.2 Methodology

6.2.1 General Procedure for The Proposed Displacement Measurement System

The general procedure of the proposed method in this chapter is the same with the one presented in Section 2.4.1.1. And the scale ratio is selected as camera calibration approach. In this chapter,

the authors implemented the spatio-temporal context (STC) learning method to conduct the visual tracking. The horizontal and vertical displacements in pixels— $x_t - x_0$ and $y_t - y_0$, respectively—are found by subtracting the coordinates of the initial target position (x_0, y_0) from the current target position (x_t, y_t) . When pattern matching methods such as DIC are used in this step, the displacements in pixels are integer values. And with STC, it also turns out integer value pixel results. One way to increase the sensitivity and to improve the measurement accuracy is by applying subpixel techniques. For instance, Feng et al. (Feng et al. 2015a) implement upsampled cross correlation in the local region to obtain the displacement at the subpixel level. In this chapter, the authors utilized the Taylor approximation method to achieve the subpixel level without upsampling the images and without sacrificing the image processing speed. Finally, with the scale ratio, SR , and the displacement in pixels, the actual displacement at time t of the physical unit is obtained: $(x_t - x_0) * SR$, horizontally, and $(y_t - y_0) * SR$, vertically. The visual tracking method and subpixel estimation used in this paper are introduced in detail in the next sections.

6.2.2 Visual Tracking Using Spatio-Temporal Context (STC) Learning

The spatio-temporal relationship among the local scenes containing the target in consecutive frames can be used to model the statistical correlation between the low-level features, such as image intensity and position, extracted from the target and its local context (Zhang et al. 2013). As illustrated in the footbridge example of Figure 55, the yellow (smaller) box is the target to be tracked and the red (larger) box is the local context.

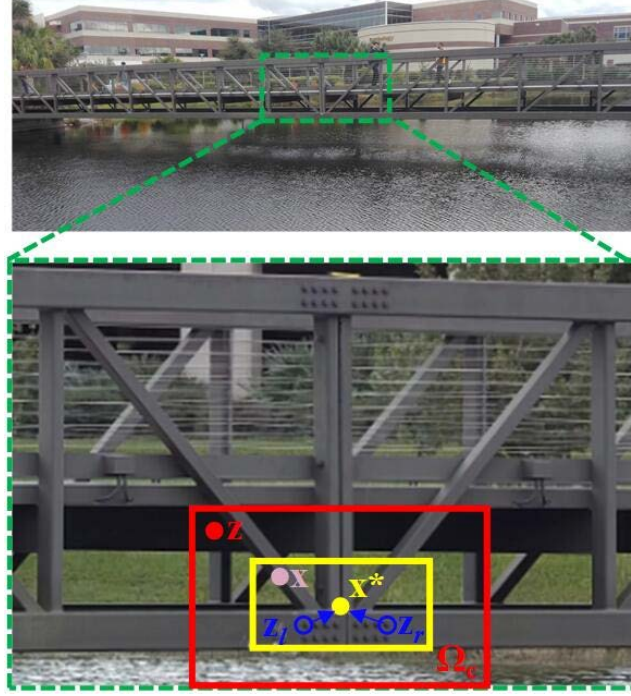


Figure 55 Graphical model of spatial relationship between the target and its surrounding context

The visual tracking task can be obtained by maximizing an object location likelihood function $c(\mathbf{x})$ as (Zhang et al. 2013):

$$\begin{aligned} c(\mathbf{x}) &= P(\mathbf{x} | o) = \sum_{c(\mathbf{z}) \in X^c} P(\mathbf{x}, c(\mathbf{z}) | o) \\ &= \sum_{c(\mathbf{z}) \in X^c} P(\mathbf{x} | c(\mathbf{z}), o) P(c(\mathbf{z}) | o) \end{aligned} \quad (44)$$

where \mathbf{x} is the target location which can be represented with the coordinates defined above, (x, y) and o denotes the target present in the scene. The context feature set, X^c , is defined as:

$$X^c = \left\{ c(\mathbf{z}) = (I(\mathbf{z}), \mathbf{z}) \mid \mathbf{z} \in \Omega_c(\mathbf{x}^*) \right\} \quad (45)$$

where $I(\mathbf{z})$ denotes the image intensity at location \mathbf{z} and $\Omega_c(\mathbf{x}^*)$ is the neighborhood of location \mathbf{x}^* .

The conditional probability $P(\mathbf{x} | c(\mathbf{z}), o)$ in Eq. 44 models the spatial relationship between the object location and its context information. It can help to resolve ambiguities when the image

measurements allow different interpretations, which are introduced in the following parts. $P(\mathbf{c}(\mathbf{z})|o)$ is a context prior probability which models the appearance in the local context. The conditional probability $P(\mathbf{x}|\mathbf{c}(\mathbf{z}), o)$ in Eq. 44 is defined as:

$$P(\mathbf{x}|\mathbf{c}(\mathbf{z}), o) = h^{sc}(\mathbf{x} - \mathbf{z}) \quad (46)$$

where $h^{sc}(\mathbf{x} - \mathbf{z})$ is the spatial context model function, which is only of the direction and the relative distance between the target location \mathbf{x} and its local context location \mathbf{z} , which means this function contains the spatial relationship between the target and its spatial context. Eq. 46 defines the spatial context model. It is worth noting that Eq. 46 is not a radially symmetric function which means that $h^{sc}(\mathbf{x} - \mathbf{z})$ is not equal to $h^{sc}(|\mathbf{x} - \mathbf{z}|)$. It considers different spatial relationships between the target and its local context, which facilitates the solving of the ambiguities when similar objects appear in close proximity. As shown in Figure 55, when a visual tracking method tries to track a bolt based only on the appearance denoted by \mathbf{z}_l , it might be distracted to the right one denoted by \mathbf{z}_r , because both bolts and the local surroundings have a similar appearance. This would cause ambiguities and consequently decrease the tracking accuracy, especially when the target moves fast, and the search region is not small. With the non-radially symmetric characteristics of $h^{sc}(\mathbf{x} - \mathbf{z})$, the ambiguities can be resolved.

In Eq. 44, $P(\mathbf{c}(\mathbf{z})|o)$ can be calculated according to the target location that has been initialized manually in the first frame. It is modeled by:

$$P(\mathbf{c}(\mathbf{z})|o) = I(\mathbf{z})w_{\sigma}(\mathbf{z} - \mathbf{x}^*) \quad (47)$$

where $w_{\sigma}(\cdot)$ is a weighted function defined by:

$$w_{\sigma}(\mathbf{z}) = ae^{-\frac{|\mathbf{z}|^2}{\sigma^2}} \quad (48)$$

In Eq. 48, a is the normalization constant which restricts $P(\mathbf{c}(\mathbf{z})|o)$ to be in the range from 0 to 1 and σ is a scale parameter. Eq. 47 ensures that, the closer the context location \mathbf{z} is to the current tracking target location \mathbf{x}^* , the more important it is to predict the location and a greater weight is set. The confidence map of an object location is modeled as:

$$c(\mathbf{x}) = P(\mathbf{x} | o) = be^{-\left|\frac{\mathbf{x}-\mathbf{x}^*}{\alpha}\right|^\beta} \quad (49)$$

where b is a normalization constant, α is a scale parameter, and β is a shape parameter. According to the literature (Zhang et al. 2013), robust results can be obtained when $\beta = 1$. Based on the context prior model in Eq. 47 and the confidence map function in Eq. 49, the objective is to learn the spatial context model, i.e., Eq. 44. Combining Eqs, 44, 46, 47 and 49, gives:

$$\begin{aligned} c(\mathbf{x}) &= P(\mathbf{x} | o) = be^{-\left|\frac{\mathbf{x}-\mathbf{x}^*}{\alpha}\right|^\beta} \\ &= \sum_{\mathbf{z} \in \Omega_c(\mathbf{x}^*)} h^{sc}(\mathbf{x}-\mathbf{z}) I(\mathbf{z}) w_\sigma(\mathbf{z}-\mathbf{x}^*) \\ &= h^{sc}(\mathbf{x}) \otimes I(\mathbf{x}) w_\sigma(\mathbf{x}-\mathbf{x}^*) \end{aligned} \quad (50)$$

where \otimes denotes the convolution operator. The fast Fourier transform (FFT), Equation (9) transforms the function to the frequency domain:

$$\mathcal{F}\left(be^{-\left|\frac{\mathbf{x}-\mathbf{x}^*}{\alpha}\right|^\beta}\right) = \mathcal{F}\left(h^{sc}(\mathbf{x})\right) \odot \mathcal{F}\left(I(\mathbf{x}) w_\sigma(\mathbf{x}-\mathbf{x}^*)\right) \quad (51)$$

where \mathcal{F} denotes the FFT function and \odot is the element-wise product. Then, the spatial context model is:

$$h^{sc}(\mathbf{x}) = \mathcal{F}^{-1} \left(\frac{\mathcal{F} \left(be^{-\frac{|\mathbf{x}-\mathbf{x}^*|^\beta}{\alpha}} \right)}{\mathcal{F} \left(I(\mathbf{x}) w_\sigma(\mathbf{x}-\mathbf{x}^*) \right)} \right) \quad (52)$$

where \mathcal{F}^{-1} denotes the inverse FFT function. Then in the whole image sequence, the spatio-temporal context model of the $(t + 1)$ th frame, H_{t+1}^{stc} , can be updated using the spatio-temporal context model, H_t^{stc} , and the spatial context model, h_t^{sc} , of the t th frame. It is formulated as:

$$H_{t+1}^{stc} = (1 - \rho) H_t^{stc} + \rho h_t^{sc} \quad (53)$$

where ρ is a learning parameter and t denotes the t th frame. It should be noted that in the first frame, i.e., when t is equal to 1, the spatio-temporal context model H_t^{stc} is equal to the spatial context model, h_t^{sc} .

Finally, the target location \mathbf{x}_{t+1}^* in the $(t + 1)$ th frame is determined by maximizing the new confidence map:

$$\mathbf{x}_{t+1}^* = \arg \max_{\mathbf{x} \in \Omega_c(\mathbf{x}_t^*)} c_{t+1}(\mathbf{x}) \quad (54)$$

Deduced from Eq. 51, the new confidence map $c_{t+1}(\mathbf{x})$ is represented as:

$$c_{t+1}(\mathbf{x}) = \mathcal{F}^{-1} \left(\mathcal{F} \left(H_{t+1}^{stc}(\mathbf{x}) \right) \odot \mathcal{F} \left(I_{t+1}(\mathbf{x}) w_{\sigma_t}(\mathbf{x}-\mathbf{x}^*) \right) \right) \quad (55)$$

As the scale of the target may tend to change over time, the scale parameter σ in the weight function w_σ in Eq. 48 is updated by:

$$\left\{ \begin{array}{l} s'_t = \sqrt{\frac{c_t(x_t^*)}{c_{t-1}(x_{t-1}^*)}} \\ \bar{s}_t = \frac{1}{n} \sum_{i=1}^n s'_{t-i} \\ s_{t+1} = (1-\lambda)s_t + \lambda\bar{s}_t \\ \sigma_{t+1} = s_t\sigma_t \end{array} \right. \quad (56)$$

where s'_t is the estimated scale among two consecutive frames. The estimated target scale s_{t+1} is obtained through filtering, in which \bar{s}_t is the average of the estimated scale from n consecutive frames to avoid oversensitive adaptation and to reduce noise, and $\lambda > 0$ is a fixed parameter. More details about the derivation can be found in the literature (Zhang et al. 2013). In general, the scale updating should be considered, but in this study, only in-plane motion is considered for two-dimensional displacement measurement, so that scale updating is neglected. If this method is used to do three-dimensional displacement measurement, which means there is out-plane motion, scale updating has to be considered.

To obtain robust tracking results, the reference gives rules of thumb regarding the selection of the parameters used in STC tracking: $\alpha = 2.25$, $\beta = 1$, $\rho = 0.075$, $s_1 = 1$, $\lambda = 0.25$, and $n = 5$. Additionally, for Equation (12), with the deductions,

$$\begin{aligned} H_{t+1}^{stc} &= (1-\rho)H_t^{stc} + \rho h_t^{sc}; \\ \int H_{t+1}^{stc} e^{-j\omega t} dt &= \int (1-\rho)H_t^{stc} e^{-j\omega t} dt + \int \rho h_t^{sc} e^{-j\omega t} dt; \\ \text{LHS} &= \int H_{t+1}^{stc} e^{-j\omega t} dt \stackrel{t \rightarrow t-1}{=} \int H_t^{stc} e^{-j\omega(t-1)} d(t-1) \\ &= e^{j\omega} \int H_t^{stc} e^{-j\omega t} dt = e^{j\omega} H_w^{stc}; \\ \text{RHS} &= \int (1-\rho)H_t^{stc} e^{-j\omega t} dt + \int \rho h_t^{sc} e^{-j\omega t} dt \\ &= (1-\rho) \int H_t^{stc} e^{-j\omega t} dt + \rho \int h_t^{sc} e^{-j\omega t} dt \\ &= (1-\rho)H_w^{stc} + \rho h_w^{sc}; \end{aligned}$$

LHS = RHS;

$$H_w^{stc} = \frac{\rho}{e^{jw} - (1 - \rho)} h_w^{sc}$$

a temporal filtering procedure can be easily obtained in the frequency domain, which is:

$$H_w^{stc} = \frac{\rho}{e^{jw} - (1 - \rho)} h_w^{sc} \quad (57)$$

where:

$$H_w^{stc} = \int H_t^{stc} e^{-jw t} dt \quad (58)$$

is the temporal Fourier transform of H_t^{stc} and similar to h_w^{sc} . The temporal filter can be represented by:

$$F_w = \frac{\rho}{e^{jw} - (1 - \rho)} \quad (59)$$

which is a low-pass filter (Oppenheim et al. 1996). With this low-pass filter, the spatio-temporal context model is able to filter out image noise caused by appearance variations and this leads to more stable measurement results. The properties of the spatio-temporal model contribute to the resolution of the adverse effects induced by environmental factors such as illumination change and fog.

6.2.3 Subpixel Level Estimation Using Taylor Approximation

With Eq. 54, the targets can easily be tracked in the image sequence, but the change of locations can only be obtained with integer pixel values. To achieve subpixel level motion, the Taylor approximation method is applied to solve the optical flow estimation. Assuming there are two consecutive images, $f(x, y)$ and $g(x, y)$, with a shift $(\Delta x, \Delta y)$, the following estimation applies:

$$\begin{aligned}
g(x, y) &= f(x + \Delta x, y + \Delta y) \\
&\approx f(x, y) + \Delta x \frac{\partial}{\partial x} f(x, y) + \Delta y \frac{\partial}{\partial y} f(x, y)
\end{aligned} \tag{60}$$

which is the first order Taylor series approximation. The shift in the image can be calculated by minimizing the sum of squared errors (SSEs):

$$\arg \min \Phi(\Delta x, \Delta y) \tag{61}$$

where:

$$\Phi(x, y) = \sum_{x,y} \left[g(x, y) - f(x, y) - \Delta x \frac{\partial}{\partial x} f(x, y) - \Delta y \frac{\partial}{\partial y} f(x, y) \right]^2 \tag{62}$$

Using the ordinary least squares (OLS) method to solve Eq. 61, the optimal Δx and Δy can be determined by setting the partial derivatives of Eq. 62 to zero, i.e.,

$$\begin{cases} \frac{\partial \Phi}{\partial \Delta x} = 0 \\ \frac{\partial \Phi}{\partial \Delta y} = 0 \end{cases} \tag{63}$$

Combining Eqs. 62 and 63 gives the system of linear equations:

$$\begin{bmatrix} \sum_{x,y} \left(\frac{\partial f}{\partial x} \right)^2 & \sum_{x,y} \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \\ \sum_{x,y} \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} & \sum_{x,y} \left(\frac{\partial f}{\partial y} \right)^2 \end{bmatrix} \begin{Bmatrix} \Delta x \\ \Delta y \end{Bmatrix} = \begin{Bmatrix} \sum_{x,y} (g - f) \frac{\partial f}{\partial x} \\ \sum_{x,y} (g - f) \frac{\partial f}{\partial y} \end{Bmatrix} \tag{64}$$

The optimal shift, $(\Delta x, \Delta y)$, is obtained by solving Equation (23). It should be noted that to make the Taylor approximation valid, the assumptions $|\Delta x| \ll 1$ and $|\Delta y| \ll 1$ have to be satisfied. When small motion is estimated, i.e., motion less than one pixel, the assumption holds. The procedure simplified from optical flow estimation is called Taylor approximation here and it will be utilized to solve the subpixel level motion estimation (Chan et al. 2010).

Figure 56 and Figure 57 illustrate the proposed motion estimation at the subpixel level. At first, the spatio-temporal context (STC) tracking method is employed to determine the integer pixel displacements, $(\bar{\Delta x}, \bar{\Delta y})$.

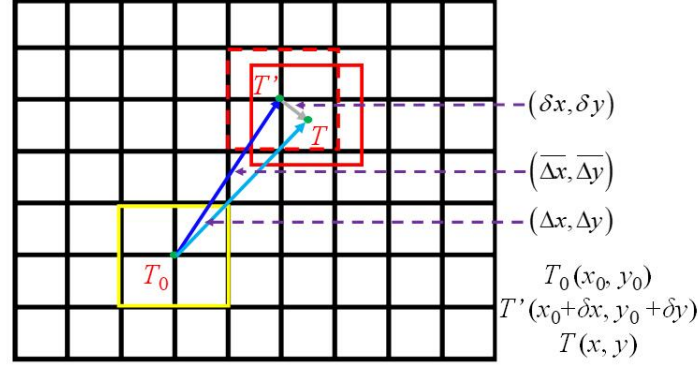


Figure 56 Sketches of motion estimation using STC tracking and Taylor approximation

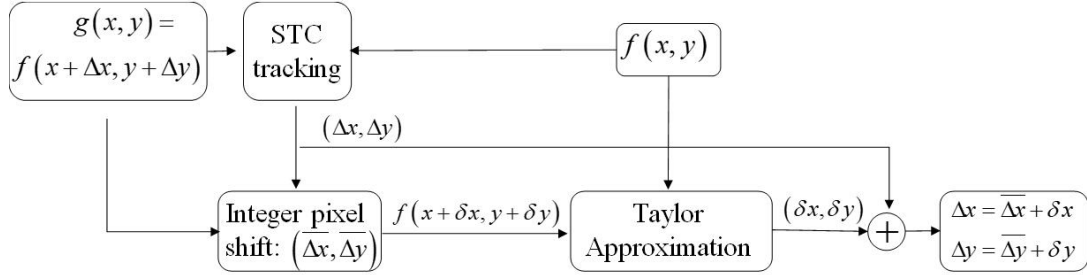


Figure 57 Flowchart of STC-based subpixel tracking using Taylor approximation

In Figure 56, the yellow, solid line box represents the original target location in the initial frame and the red, dashed line box represents the target recognized in the current frame using STC tracking, which has an accuracy at the subpixel level. Here the centers of the targets are used to represent their locations, i.e., T_0 and T' . Assuming the real target in the current frame is the red, solid line box at location T , the true displacements are $(\Delta x, \Delta y)$. Then, the displacements $(\bar{\Delta x}, \bar{\Delta y})$ are the integer estimations of the true displacements, $(\Delta x, \Delta y)$. The differences between $(\bar{\Delta x}, \bar{\Delta y})$ and $(\Delta x, \Delta y)$ are $(\delta x, \delta y)$, from T' to T , where $|\delta x| < 1$ and $|\delta y| < 1$, and the assumption of

using the Taylor approximation is satisfied with the conditions of $|\delta x| < 1$ and $|\delta y| < 1$. Secondly, the Taylor approximation is employed to estimate the displacement between the target tracked by STC (red, dashed line box) and the real target (in red, solid line box), i.e., $(\delta x, \delta y)$. Finally, the total displacements are:

$$\begin{cases} \Delta x = \overline{\Delta x} + \delta x \\ \Delta y = \overline{\Delta y} + \delta y \end{cases} \quad (65)$$

According to the literature (Chan et al. 2010), the Taylor approximation gives an error bound of less than 0.0125 pixels, without any image upsampling and the error is much smaller than that of the normal template matching methods using image upsampling. The feasibility and performance of the proposed method for structural displacement will be verified through laboratory experiments in the next sections.

6.3 Laboratory Verification

6.3.1 Experimental Setup

Figure 58 shows the two-span bridge model constructed at the University of Central Florida's (UCF) Civil Infrastructure Technologies for Resilience and Safety (CITRS) Experimental Design and Monitoring (EDM) laboratory. The bridge was a scaled down model of a mid-sized real-life structure and a toy truck with variable weights was used to model moving loads. The bridge consisted of two 300 cm main continuous spans, which were rebuilt from UCF's original four-span bridge (Khuc and Catbas 2016). The bridge deck, which included a 3.18 mm steel sheet, was

120 cm wide and 600 cm long. The steel deck was supported by two 25 mm \times 25 mm \times 3.2 mm steel girders separated 0.61 m from each other. To provide a connection between girders and the deck, a set of four M6 bolts and 3.18 mm thick plates were utilized. The total length of the toy truck was 68.6 cm and the distance among the two axles was 45.72 cm. The width and the height of the toy truck were 35.6 cm and 33.1 cm, respectively. The total weight of the truck was 13.48 kg with a front axle weight of 1.9 kg and a rear axle weight of 11.58 kg. The average moving speed of the toy truck was manually controlled to achieve 0.4 m/s, while the actual speed in the experiment for three cases were 0.36 m/s, 0.38 m/s, and 0.33 m/s, respectively.

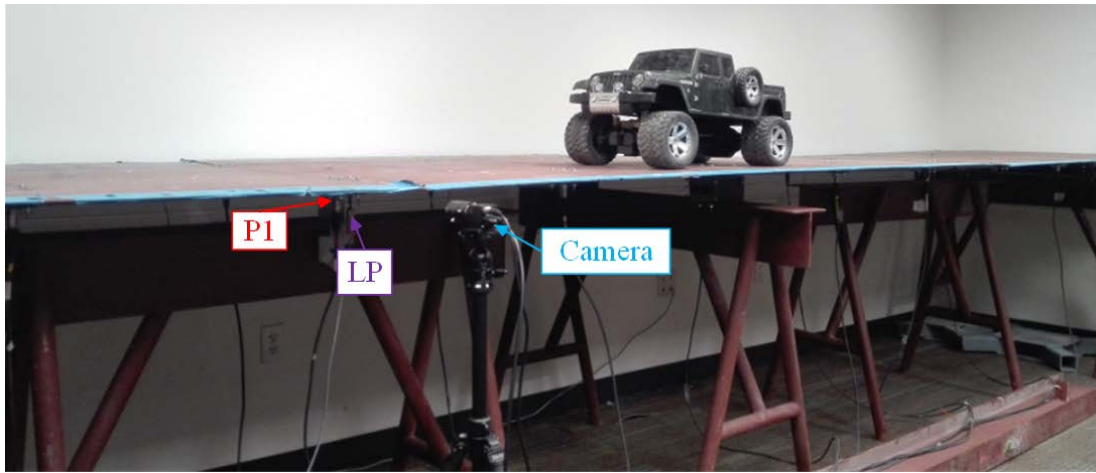


Figure 58 Experimental setup

An industrial camera was set-up in front of the bridge to record images at a measuring point (P1) during the moving load trials. A linear potentiometer (LP) was mounted under the deck to measure the displacement of the P1 and was assumed as the ground truth. The model number was BEI Duncan 9615. Detailed information about the sensor can be found in Reference (BEI 2019). During the experiments, the truck moved from one side of the bridge to the other while the LP and the camera recorded the motion of the P1 (midspan of the left span) synchronously. The resolution of the camera was 1280 \times 960, with a maximum frame rate of 60 FPS (frames per second). Here, the frame rate was set to 30 FPS. The focal length of the lens was within a zoom range of 6~60 mm.

The sampling rate of the data acquisition system for the LP was 300 Hz, which was then downsampled to 30 Hz during post processing. Three experimental cases were specified to achieve the objectives of this chapter:

An industrial camera was set-up in front of the bridge to record images at a measuring point (P1) during the moving load trials. A linear potentiometer (LP) was mounted under the deck to measure the displacement of the P1 and was assumed as the ground truth. The model number was BEI Duncan 9615. Detailed information about the sensor can be found in Reference (BEI 2019). During the experiments, the truck moved from one side of the bridge to the other while the LP and the camera recorded the motion of the P1 (midspan of the left span) synchronously. The resolution of the camera was 1280×960 , with a maximum frame rate of 60 FPS (frames per second). Here, the frame rate was set to 30 FPS. The focal length of the lens was within a zoom range of 6~60 mm. The sampling rate of the data acquisition system for the LP was 300 Hz, which was then downsampled to 30 Hz during post processing. Three experimental cases were specified to achieve the objectives of this paper:

Case 1: The truck moves on the bridge in ideal conditions and no adverse factors are imposed in the measuring environment. A light meter (Dr. Meter LX1010B Digital Illuminance, London, England) is used to measure the illumination change. The illumination is 34 lux and the relative humidity is 49% at the displacement measurement location under the ideal conditions;

Case 2: The truck moves on the bridge while the illumination of the laboratory is changed several times by switching a manual controller. Normally, 9 light panels in the lab are on and the

illumination at the measurement location is 34 lux. By turning off the 3 light panels, which are close to the measurement target, the illumination drops to 18 lux. As shown in Figure 59, the left image is lighter, which was taken when the illumination was 34 lux, while the right image is darker which was taken when the illumination was 18 lux;

Case 3: A humidifier (Honeywell HUL520B Mistmate Cool Mist Humidifier, Seattle, Washington) is placed between the camera and measuring targets (Figure 60). The humidifier produces a mist at the maximum status to simulate natural fog which decreases the visibility in the camera's field of view. Normally, the temperature is 24 °C and the relative humidity is 49%. While in the center of the mist, the temperature is 20.3 °C and the relative humidity is 95%.

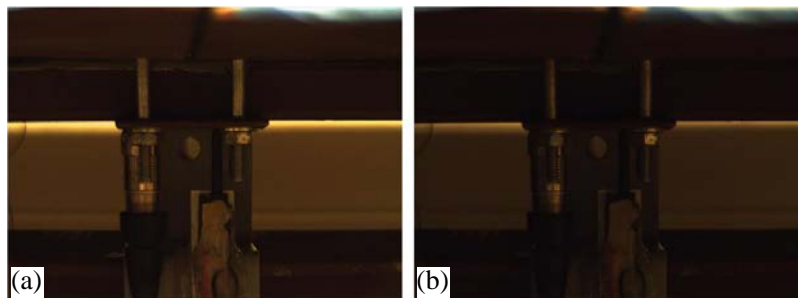


Figure 59 Illumination change. (a) 34 lux, (b) 18 lux

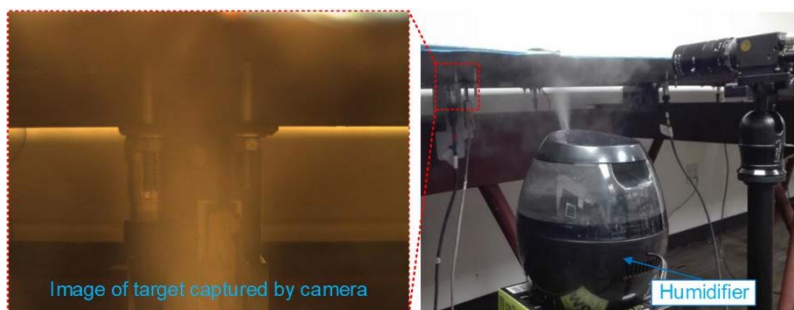


Figure 60 Fog simulation

6.3.2 Results Analysis and Comparative Study of Case 1 (Ideal Case)

The objectives for Case 1 were: to evaluate the performance of the subpixel estimation method presented in this paper and verify the feasibility and performance of the proposed method (i.e., STC tracking plus Taylor approximation, STC-Taylor App) by comparing the conventional LP data with the current vision-based displacement methods, e.g., Lucas–Kanade optical flow with SURF features (LK-SURF), key point matching with Fast Library for Approximate Nearest Neighbors and SURF features (FLANN-SURF), and digital image correlation (DIC). Figure 61 illustrates the vertical displacement time histories of P1 in pixel units induced by the vehicle loading when the toy truck traveled along the two-span bridge.

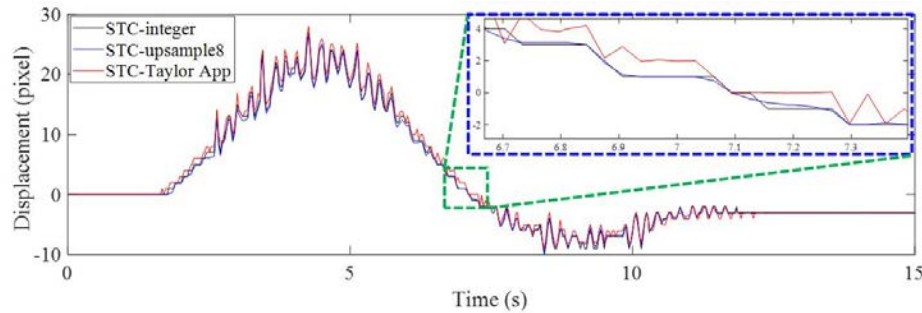


Figure 61 Vertical displacement time histories of P1 in pixel units using non-subpixel, image upsampling, and Taylor approximation techniques

The vision-based methods used here included STC tracking with non-subpixel (STC-integer), STC tracking with image upsampling (STC-upsample8), and STC tracking with Taylor approximation (STC-Taylor App). In Figure 61, by zooming in on the green, dashed line box area of the displacement time histories, it is clear that the results of the STC-integer approach were rounded

to integer values, i.e., (4, 3, 3, 3, 2, 1, 1, 1, 1, 1, 0, 0, (−1), (−1) pixels, ...). The image upsampling technique means that each image recorded during the experiment was upsampled 8 times in the horizontal and vertical directions using bicubic interpolation. Then, the minimum resolution was $1/8 = 0.125$ pixel. The result of using image upsampling was a much smoother curve and more subpixel-level displacement records. However, it still cannot provide more details about small motions (Dong et al. 2019b), especially at the very beginning and at the end. When there are no apparent loads on the structure, there is still very small structural motion induced by random environmental loads, such as wind, nearby machine operations, ambient ground vibration, etc. As illustrated in Figure 62, during the first several seconds before the toy truck moves, the displacements measured by both STC-integer and STC-upsample8 were exactly zero, which might not be true. Even though the structure was not loaded, it can still vibrate under random environmental loading. The STC-Taylor App indicated the small motions of the structure caused by random environmental loadings. This indicates that the proposed method which combines STC tracking and Taylor approximation has a higher sensitivity, resolution, and accuracy.

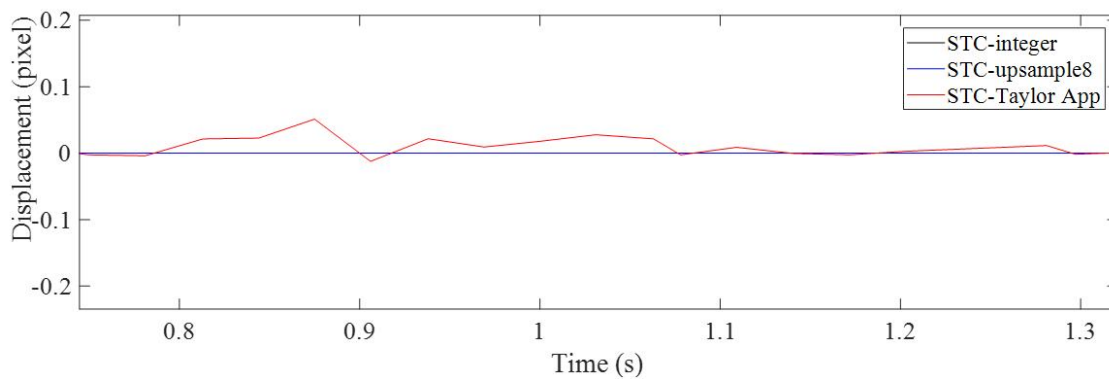


Figure 62 A zoomed in section at the beginning of the vertical displacement time histories of P1

Figure 63 verifies the previous findings. In this figure, the horizontal time displacement histories showed the bridge motion in the longitudinal direction induced by the moving truck impact. The motion was very small, around 1 pixel. The result from the proposed method (STC-Taylor App) provided very detailed information about the vehicle impact, while the results of the STC-integer and STC-upsample8 were almost zero except for one or two points, which means the bridge was stationary in the longitudinal direction. Figure 64 depicts a zoomed in area of the green, dashed line box in Figure 63.

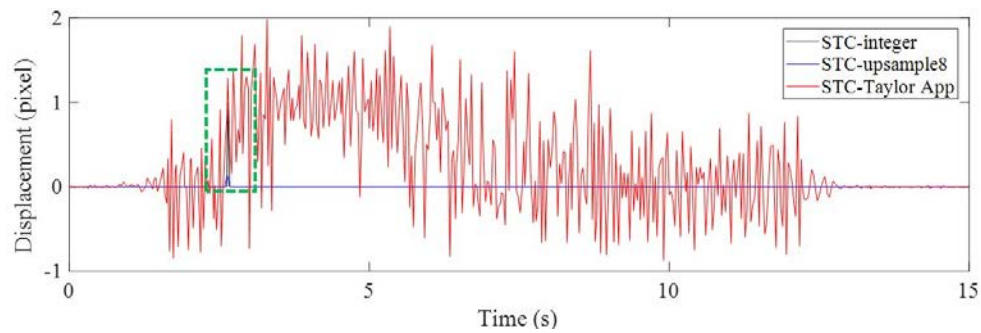


Figure 63 Horizontal displacement time histories of P1 in pixel units using non-subpixel, image upsampling, and Taylor approximation techniques

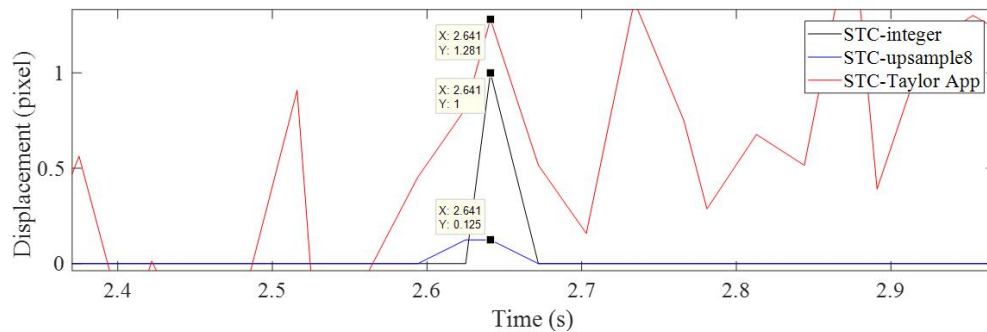


Figure 64 A zoomed in section of the horizontal displacement time histories of P1 in pixel units

In the zoomed in Figure 64, the displacements of the non-zero points measured by STC-integer, STC-upsample8, and STC-Taylor App were 1, 0.125, and 1.281 pixels. For STC-integer and STC-

upsample8, 1 and 0.125 were their minimum measurement resolutions, and statistically, these points were outliers which should be removed from the displacement time histories. In addition, the image processing speed of the proposed method was much faster than using image upsampling. Table 12 shows the elapsed processing time of one image using three different STC-based methods. The program environment was MATLAB running on a computer with the CPU of i7, 8 processors, and 16 Gb RAM. The original image had a resolution of 1280×960 . It took 0.0481 s to process one image to obtain the displacement at the integer pixel level (STC-integer). However, when conducting subpixel level estimation using image upsampling, it took 2.4895 s, which was about $(2.4895 - 0.0481)/0.0481 = 50.76$ times that of the STC-integer. It took only 0.0495 s to conduct this and provided even better subpixel results when using STC-Taylor App. The proposed method was about 50 times faster than using image upsampling techniques.

Table 12 Time consumption of processing one image using different STC-based methods

Methods	STC-Integer	STC-Upsample8	STC-Taylor App
Time (s)	0.0481	2.4895	0.0495

Overall, it is suggested that the proposed method using STC tracking and Taylor approximation can provide displacement measurements at the subpixel level with high sensitivity, resolution, accuracy, and faster speed.

The next step is to convert the displacement in pixel units to physical units (e.g., millimeter) and verify the feasibility and performance of the displacement measurement by comparing the vision-based methods with the conventional displacement sensor. As illustrated in Figure 65, four vision-

based displacement measurement methods (i.e., LK-SURF, FLANN-SURF, DIC, and STC-Taylor App) and one conventional displacement sensor (i.e., LP) were used to obtain the displacement time histories of P1 when the toy truck passed over the bridge. At the very beginning, the toy truck stood at the left end of the first span, then moved to the right and approached the measurement point P1. In the meantime, the displacement of P1 (the downward direction was positive) gradually increased to a maximum when the truck was located at P1. Then the toy truck began to drive off P1 and kept heading to the right, while the displacement of P1 gradually decreased. When the toy truck moved to the right span, the displacement began to be negative (i.e., upward displacement) due to the loading on the other span of the two-span bridge. As it approached the right end of the right span, the absolute value of the displacement at P1 first increased and achieved a maximum and then decreased. When the toy truck arrived at the right end of the bridge, the displacement of P1 became stable but did not go back to zero. This was because the rear axle still rested on the bridge.

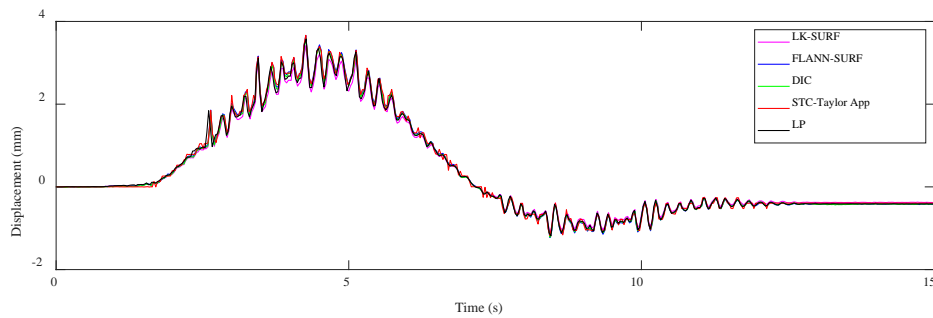


Figure 65 Case 1 (ideal condition): displacement time histories of P1 obtained from different methods

By comparing the displacement time histories, it is easy to see that the result obtained from the proposed method (i.e., STC-Taylor App) was quite consistent with those obtained from the LP and other three vision-based methods.

Figure 66 illustrates the correlation matrix of these time displacement histories for Case 1. The figures on the diagonal of the correlation matrix are the histograms of the displacement time histories, whereas the others are data plots and linear fits among the displacement time histories from the two methods. The correlation matrix is symmetric, and the last row and the last column give the correlation coefficients between the displacement data obtained from the vision-based methods and the conventional displacement sensor, i.e., LP. The correlation coefficients of the LK-SURF, FLANN-SURF, and DIC with the ground truth (i.e., LP) were all 0.99, while the correlation coefficient between the proposed method (i.e., STC-Taylor App) and LP was 0.98, which is also quite good. The performance of the vision-based displacement measurement methods can also be obtained from the similarity of the histograms of each method comparing with the one of LP. Here, from the diagonal element of the correlation matrix, it is indicated that the histograms of these time displacement histories were highly consistent with each other.

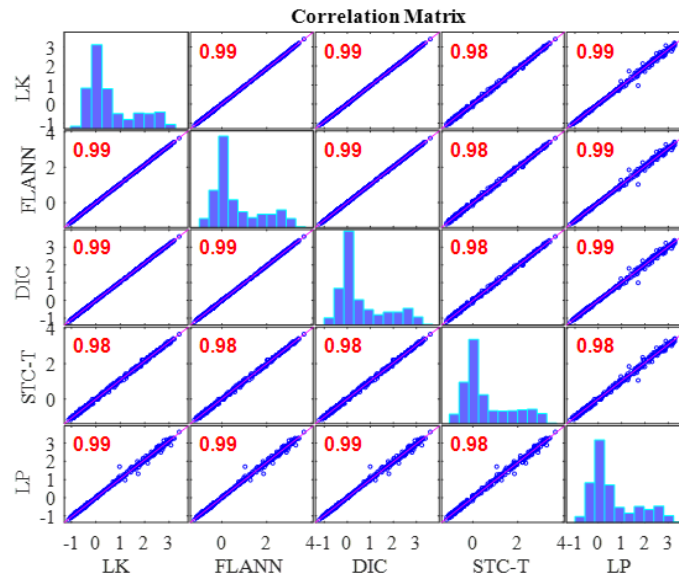


Figure 66 Correlation matrix of time displacement histories of Case 1 (ideal condition)

Figure 67 shows the displacement comparison in the frequency domain of Case 1. The first set of peaks from the different methods which gave a frequency value of 0.083 Hz was not the structural mode. The frequency was related to the vehicle loading procedure on the two-span bridge, which produced a sinusoidal shape displacement time history. It seems to be a response cycle and causes the pseudo vibration mode. The frequency value here was related to the speed of the loading. Since the speed of the truck was manually controlled and in each case the speed was different, the frequencies of the first set of peaks were different. This can be observed from Figure 70 and Figure 75 in the following sections. The second set of peaks around 5 Hz (actually 4.87 Hz) was the structural operational vibration frequency corresponding to the higher modes in the displacement time histories in Figure 66. This operational frequency was related to the roughness of the deck and speed of the loading truck. From this figure it can be seen that the structural operational vibration frequencies obtained from different methods were quite consistent.

In this case, under ideal experimental conditions and no adverse factors added to the experiment, the robustness and advantages of the proposed method (STC-Taylor App) did not reveal itself. In the next two cases, the robustness and advantages of the proposed method will be verified.

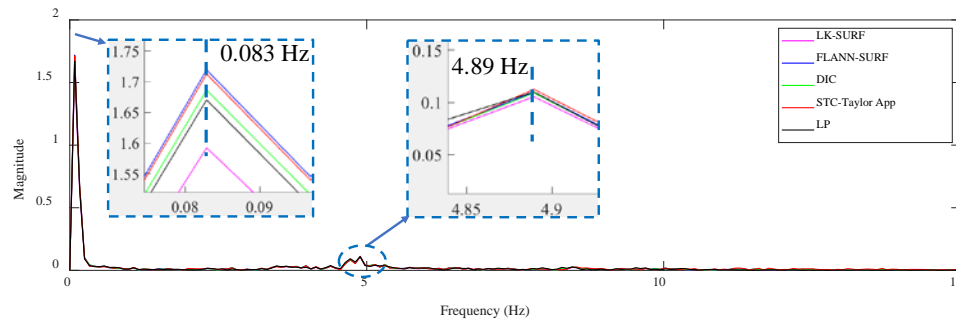


Figure 67 Displacement comparison in the frequency domain of Case 1

6.3.3 Results Analysis and Comparative Study of Case 2 (Illumination Change)

This case was designated to verify the robustness of the proposed vision-based displacement method under the adverse environmental condition: Illumination change. For vision-based methods, illumination is a serious problem when conducting field applications since the image quality is easily affected by illumination change. Consequently, the visual tracking performance and the displacement measurement accuracy are affected by the poor quality in the formation of images. In this experiment, the environmental illumination was determined by the fluorescent light in the lab. By turning the light switches in the laboratory on and off, the image quality changed, as shown in Figure 59. The time histories of P1 obtained from different vision-based methods and LP under environmental illumination change are illustrated in Figure 68a.

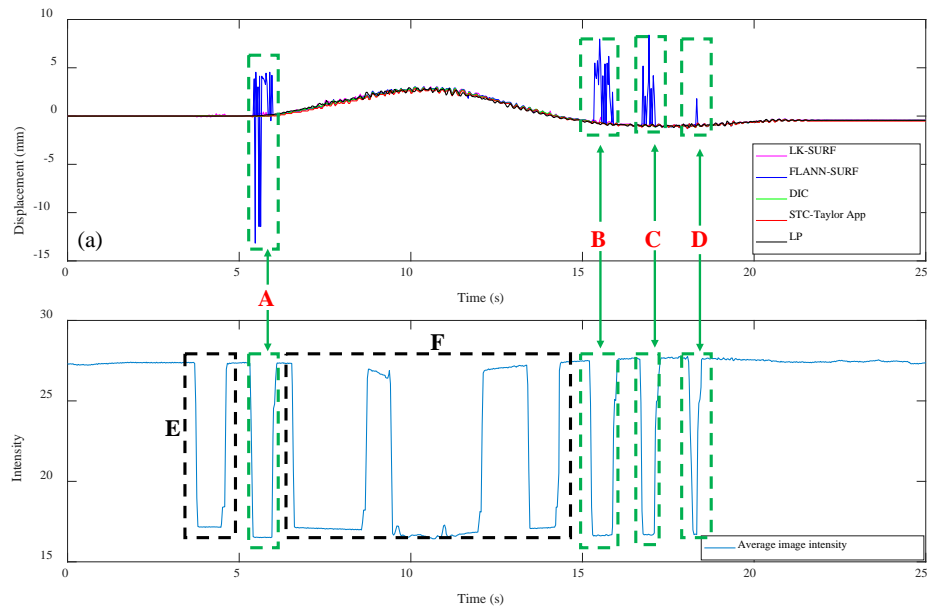


Figure 68 Case 2 (illumination change): (a) displacement time histories of P1 obtained from different methods and (b) average image intensity time history

The spikes in Figure 68a show that the vision-based method FLANN-SURF was apparently influenced by the illumination change, which means FLANN-SURF cannot handle this kind of situation compared to other vision-based methods. Figure 68b shows the average image light intensity time history and it is easy to see the illumination change (36 lux to 18 lux) from the tooth-shaped signal. The green, dashed boxes A, B, C, and D indicate the corresponding illumination changes when FLANN-SURF gives spikes in its displacement measurement result. It is suspected that for the illumination changes as shown in boxes E and F, FLANN-SURF shows good consistency with other displacement measurement methods, while the possible reasons are unknown.

As shown in Figure 69, the correlation coefficient of the time histories between that obtained from FLANN-SURF and the ground truth, LP, dropped to 0.84, while LK-SURF's and DIC's dropped from 0.99 to 0.98 and from 0.99 to 0.97 when compared with the correlation matrix obtained in Case 1, shown in Figure 66. However, the correlation coefficient of the time histories between that obtained from the proposed method, STC-Taylor App, and the ground truth, LP, was still 0.98 compared to that of Case 1.

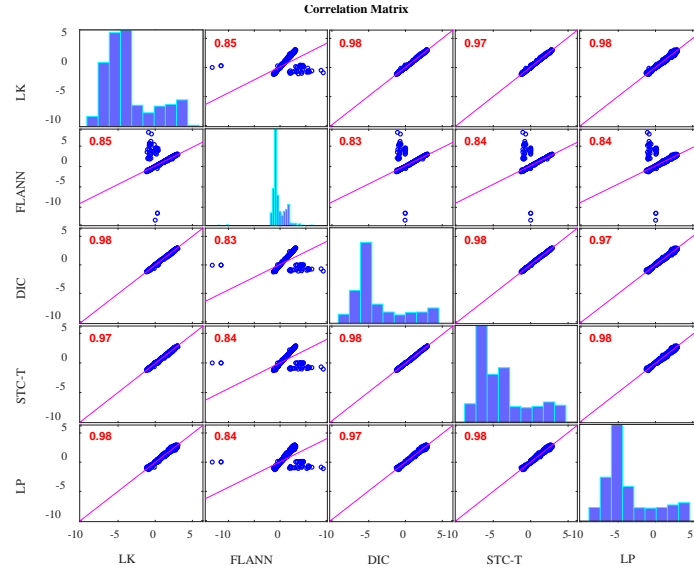


Figure 69 Correlation matrix of time displacement histories of Case 2 (illumination change)

From Figure 68 and Figure 69, it is indicated that the illumination change did have a significant negative effect on FLANN-SURF and might also influence LK-SURF and DIC slightly. On the other hand, the proposed method, STC-Taylor App, showed great robustness and was almost not influenced by the illumination change. This observation can also be found from the frequency domain, as shown in Figure 70. The method FLANN-SURF showed lots of pseudo vibration frequencies which were apparently not the structural vibration modes. The first frequency, 0.033 Hz, indicated the loading cycle, which was similar to Case 1. The structural vibration mode around 5 Hz can barely be seen here, which may be caused by the lower truck speed compared to Case 1. The FLANN-SURF method showed several big pseudo peaks around 5 Hz, while the other vision-based methods kept consistent with LP. The STC-Taylor App could be a good option for long-term vision-based displacement measurement since illumination change is a common problem in field applications.

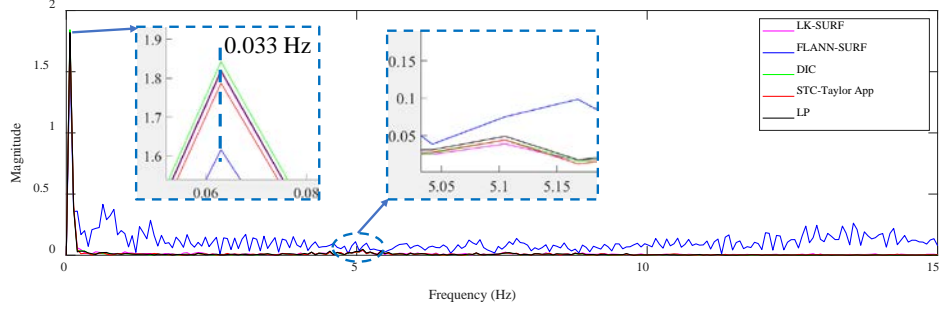


Figure 70 Displacement comparison in the frequency domain of Case 2 (illumination change)

6.3.4 Results Analysis and Comparative Study of Case 3 (Fog Interference)

This case was designated to verify the robustness of the proposed vision-based displacement method under the adverse environmental condition: fog. In this experiment, the fog was simulated by the mist produced by the humidifier, as shown in Figure 60. The fog not only affected the image quality but also contaminated the image features, which are the basic foundation of target recognition for visual tracking. In addition, the fog was not still but had a random motion. The DIC method may have performed undesirably because it highly relied on the image intensity to conduct pattern matching and the intensity would always change under this situation. Due to the random motion of the fog, a false optical flow would be added to the real target motion which causes the optical flow method (e.g., LK method) to fail. Even though feature points, e.g., Shi-Tomasi corners, SURF, SIFT, FREAK, etc., are very robust and distinctive, their use with feature-based methods (e.g., LK-SURF and FLANN-SURF) still can result in a bad performance due to the presence of bad matches. The mist could block features and induce more bad matches as shown in Figure 71. It causes displacement measurement to have errors, especially when there are not enough feature points to describe the tracking targets.

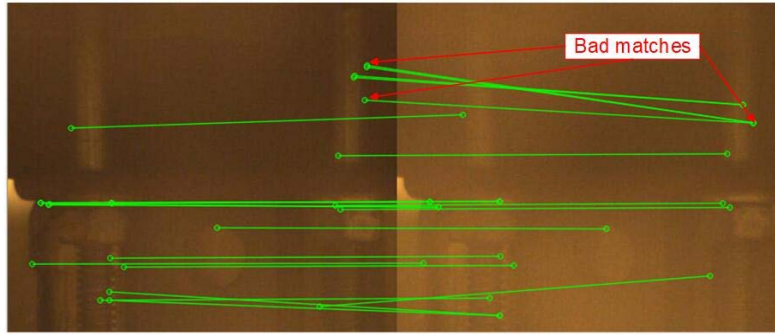


Figure 71 Poor matches when using feature-based methods (fog interference)

Figure 72 illustrates the time histories of P1 obtained from different vision-based methods and LP under fog interference. When the fog was imposed to the measurement environment, the displacement results obtained from LK-SURF and DIC provided very poor performance resulting in many spikes appearing in the displacement time histories. Only the results from the proposed method (STC-Taylor App) and FLANN-SURF showed satisfactory performance. Figure 73 zooms in on the purple, dashed line box area of Figure 72 and provides more details. In this figure, except the spikes, some data were also lost from the displacement time history, because the visual tracker based on LK-SURF lost the targets due to the fog interference. In general view, even though FLANN-SURF gives a good result, it still has outliers. Figure 73 shows an example of an outlier when using FLANN-SURF. The outlier caused more than a one-millimeter error compared with the ground truth and the result from STC-Taylor App. Statistically, it can be removed.

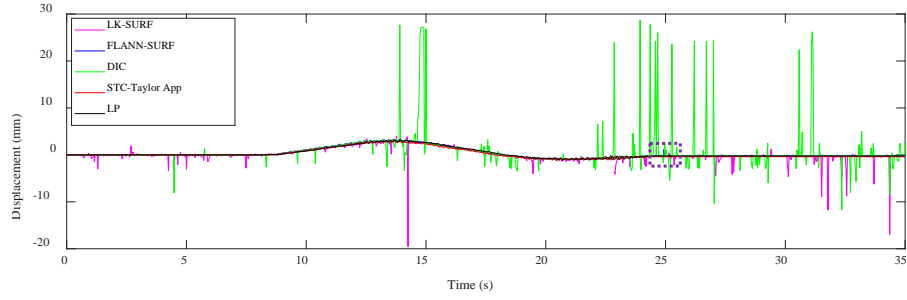


Figure 72 Case 3 (fog interference): displacement time histories of P1 from different methods

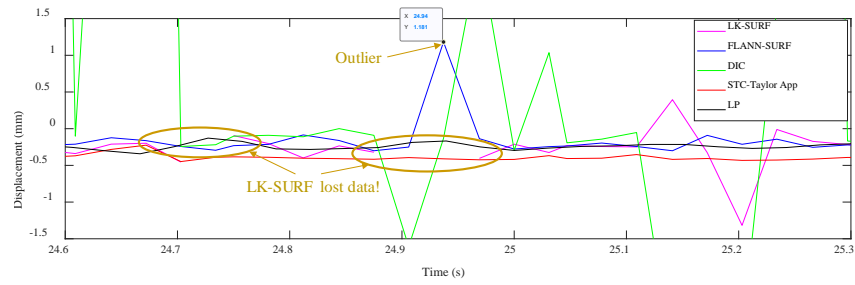


Figure 73 Zoomed in section of the horizontal displacement time histories of P1

Figure 74 shows the correlation matrix between the vision-based methods and the conventional displacement method, i.e., LP. The correlation coefficient between LK-SURF and LP dropped from 0.99 (in Case 1) to 0.84, which means the measurement error of LK-SURF increased. It was even worse for DIC, for which the correlation coefficient dropped from 0.99 (in Case 1) to 0.78. The linear fit plots between LK-SURF and LP and that between DIC and LP are hard to be interpreted as correlation. The correlation coefficient between the proposed method, STC-Taylor App, and LP also dropped from 0.98 (in Case 1) to 0.92. Considering the initial status, it is a little bit better than that of FLANN-SURF, since the correlation coefficient between the proposed method, STC-Taylor App, and LP also dropped from 0.99 (in Case 1) to 0.92. The outlier in the displacement time history obtained using FLANN-SURF also showed correlation in matrix plot. The fog indeed had undesirable effects on all of these vision-based methods at different levels.

These bad effects might not be easy to find or be quantified in the time histories, but they apparently reveal themselves in correlation matrix.

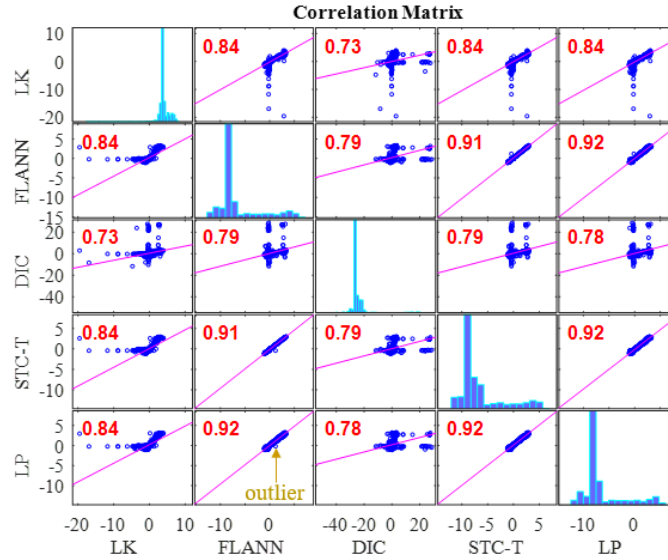


Figure 74 Correlation matrix of time displacement histories of Case 3 (fog interference)

Figure 75 gives the comparison in the frequency domain of Case 3. The poor performance of LK-SURF and DIC caused by fog can also easily be seen from the pseudo peaks (purple and green curves) in the frequency domain. Similar to Case 1 and Case 2, the first set of peaks indicate the loading procedure and that there should be a structural vibration mode frequency peak around 5 Hz; however, it can barely be seen. Except for LK-SURF and DIC, the proposed method (STC-Taylor APP) and FLANN-SURF seemed to match very well with LP, which is consistent with the observations from Figure 72 and Figure 74.

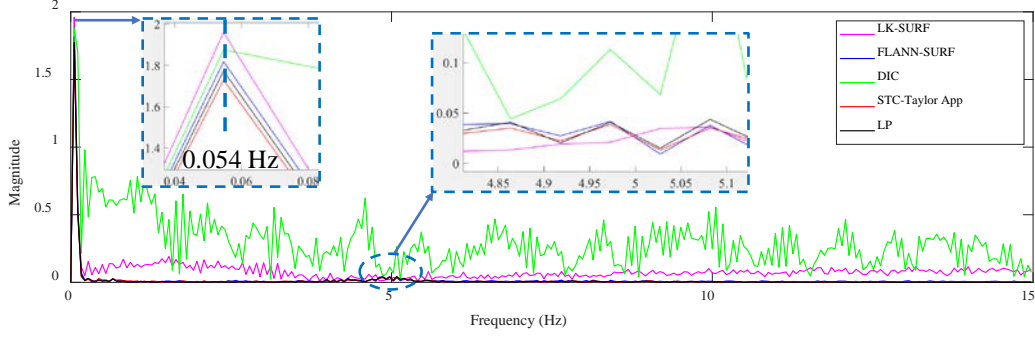


Figure 75 Displacement comparison in frequency domain of Case 3 (fog interference)

Compared to the other three vision-based methods, the proposed method provided the best performance. The proposed vision-based displacement measurement method, i.e., STC-Taylor App, showed great robustness under fog interference. The STC-Taylor App could be a good option for long-term vision-based bridge displacement measurement since fog is a common weather problem in field application, especially when the bridge crosses a river and during foggy seasons. Considering the results analysis of Case 2 and Case 3, the proposed method shows the best performance under the two adverse environmental factors.

6.4 Summary

In this study, a robust non-contact displacement measurement method using spatio-temporal context learning and Taylor approximation was proposed. This study aimed to resolve the adverse effects induced by the environmental factors such as illumination change and fog interference when using vision-based methods to conduct displacement measurements without adding manual markers or artificial light source for long-term bridge monitoring. The first method that was proposed, namely, spatio-temporal context learning, leveraged the advantage of images with high-

resolution spatial and temporal aspects, which can be used for long-term bridge monitoring. Then, as an extension, the Taylor approximation technique was implemented into the proposed method to improve the accuracy of the displacement at the subpixel level without sacrificing the processing speed. The performance of the proposed subpixel estimation method was compared with general image upsampling techniques and the results showed that the proposed subpixel estimation method was faster than the general image upsampling techniques by about 50 times. Also, the precision of the proposed method was much better than the general image upsampling technique. To validate the feasibility, stability, and robustness of the proposed method, a series of experiments on a two-span three-lane bridge in laboratory under the adverse environmental factors such as illumination change and fog interference were conducted. The illumination change was achieved by turning on and off the light switches in the room and the fog interference was simulated with a humidifier which could produce mist. The results from the proposed method showed that:

(1) In Case 1, there were no adverse environmental factors and the measurement condition was desirable for vision-based systems. The correlation coefficients of the LK-SURF, FLANN-SURF, and DIC with the ground truth, i.e., LP, were all 0.99, while the correlation coefficient between the proposed method, STC-Taylor App, and LP was 0.98, which was also quite good. It indicated that that at least in a desirable measurement environment, the proposed method is a strong competitor of the current methods.

(2) In Case 2, with the illumination change, the correlation coefficient of the time histories between that obtained from FLANN-SURF and the ground truth, LP, dropped to 0.84, while LK-SURF's and DIC's only dropped from 0.99 to 0.98 and from 0.99 to 0.97, respectively, compared with the

correlation matrix obtained from Case 1. However, the correlation coefficient of the time histories between that obtained from the proposed method, STC-Taylor App, and the ground truth, LP, was still 0.98 comparing to that of Case 1.

(3) In Case 3, with the fog interference, the correlation coefficient between LK-SURF and LP dropped from 0.99 to 0.84, while DIC's drops from 0.99 to 0.78, which was the worst performance. The FLANN-SURF's dropped from 0.99 to 0.92 and the proposed method, STC-Taylor App, dropped from 0.98 to 0.92.

Combining the results analysis of the experiments, the proposed method showed the best performance under the two adverse environmental factors, and it provided an accuracy at the subpixel level without sacrificing the processing speed. By considering the spatial and temporal context learning processes, the proposed method in this paper successfully mitigated the effects induced by illumination change and fog interference. The poor performances of FLANN-SURF in Case 2 and of LK-SURF and DIC in case 3 resulted in spikes in the displacement measurements, which can be removed by using low-pass filtering. However, this would limit the usability of these methods at higher frequencies. The proposed method seems to provide accurate displacements without the need of filtering the results.

Although, the benefits of the proposed method to address other real-world challenges is not explored in this paper, the proposed method may be applied to solve other adverse influencing factors, such as motion blur, rain, object occlusion, out of plane movement, orientation of the camera relative to the bridge and camera motion, etc., by taking advantage of the high spatio-

temporal resolution. The computer vision-based approach along with the proposed method can be a good alternative and a complementary approach to conventional structural health monitoring practices. In the future, more studies will be carried out on real bridges to validate the feasibility of the proposed method and also to investigate other relevant challenges for long-term bridge monitoring using computer vision. Besides, in this study, only one camera was used, and the proposed method was verified by tracking the motion of the bridge deck on a two-dimensional (2D) plane, which is a limitation. In future studies, the feasibility of 3D motion tracking using the proposed method will be investigated and will be tested on other applications such as long span bridge monitoring and cable vibration monitoring. Also, the effects of illumination inhomogeneity and non-linear illumination changes to the measurement performance of different vision-based methods will be explored.

CHAPTER SEVEN: DISPLACEMENT MONITORING OF LONG SPAN BRIDGES USING VISION-BASED METHODS

7.1 Introduction

In the previous chapters, different vision-based structural response methods with various types of algorithms are introduced and verified on different applications such as grandstand, stadium, small bridge model, railway bridge, footbridge. The application scenarios are in close range distance. When the vision-based approaches are applied to large structures and within long distances measurement, more problems such as image quality, blur, haze, heatwave etc. occur and would affect the structural measurement. In this chapter, applications of large structures (Three long span suspension bridges in Turkey, First Bosphorus Bridge, Second Bosphorus Bridge and Osman Gazi Bridge) in long distance are introduced and problems and issues are discussed.

7.2 Methodology

In long distance monitoring, feature points are not easy to extract so that optical flow using feature points and feature matching are not suitable for this case. Full field optical flow using deep learning and tracking using spatio-temporal learning are also difficult to achieve good performance. In chapter, the normalized correlation coefficient based template matching is implemented to track structural motions. The basics of normalized correlation coefficient based template matching is reviewed in Section 2.4.1.2., and they are not repeated here.

7.3 Experiment on the First Bosphorus Bridge

The First Bosphorus Bridge is the first bridge that was built in Istanbul to cross the Bosphorus Strait and connect the two continents: Europe and Asia. The bridge has the main span length of 1074 m and two approach-spans with the length of 231 m at the European side and 255 m at the Asian side, respectively (Bas et al. 2018). The main span is suspended by main cable and vertical hangers, whereas the side spans are supported at the base.

The experiment setup is shown in Figure 76. A camera with the resolution of 4K, speed of 30 FPS and a 75-300 mm zoom lens was setup on the European side, north of the bridge. The distance from the camera to the midspan of the bridge is around 755 m.

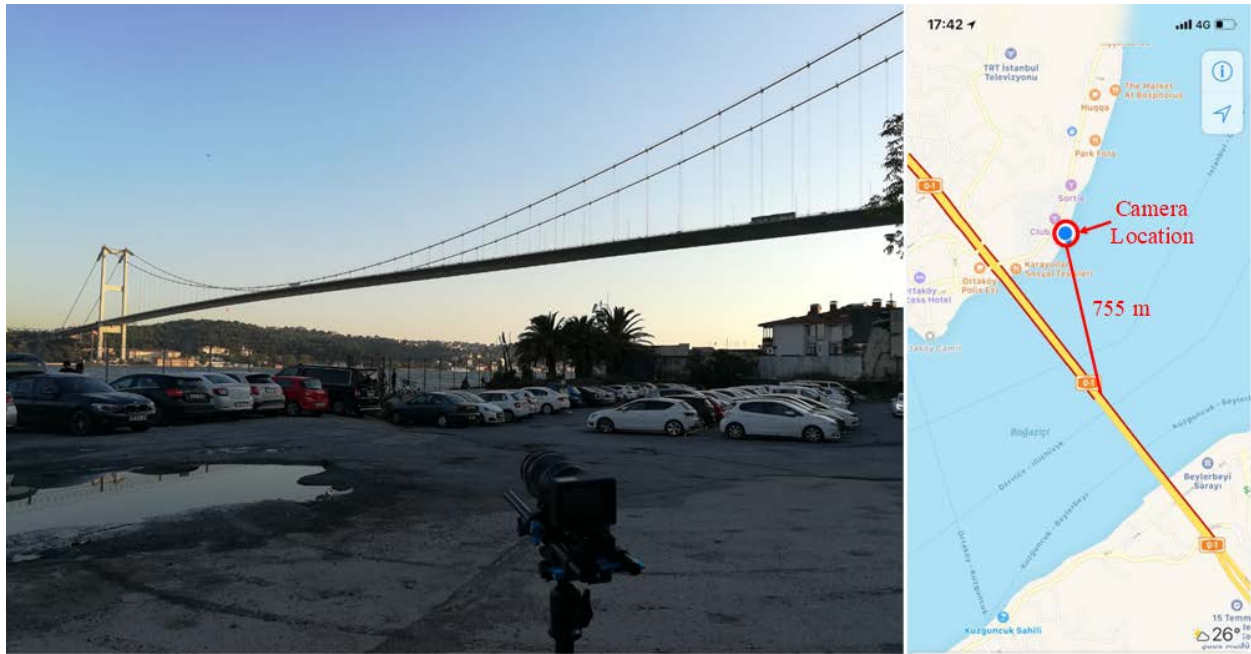


Figure 76 Experimental setup of the First Bosphorus Bridge



Figure 77 Measurement point and static point

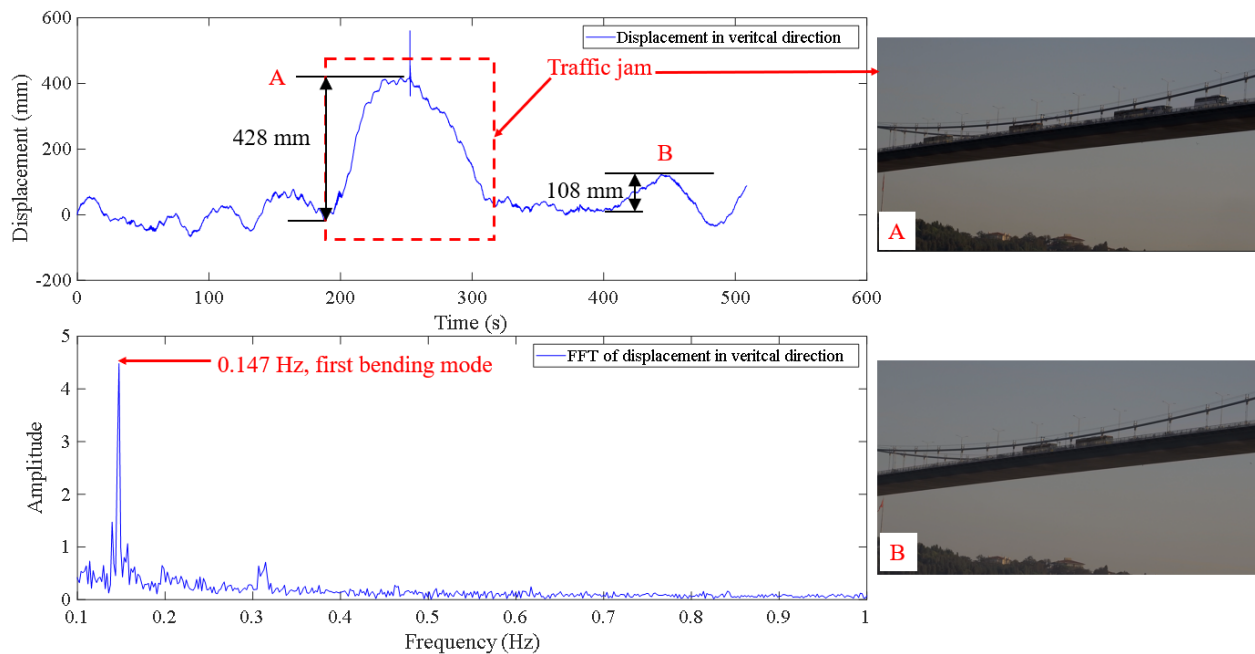


Figure 78 Displacement results of the mid-span of the First Bosphorus Bridge

The guardrail and edge of the deck on midspan of the northern side of the bridge was selected as the tracking target as shown in Figure 77. During the experiment, the wind is a big problem and it kept the camera shaking all the time, so that a building on the background was regarded as static

point and selected to mitigate the camera motion problem by using background subtraction. The experiment was conducted on a Saturday afternoon and at that time (around 5pm to 6pm), the traffic was becoming very busy and the light was getting dark. Tracking the rails in a long distance is very difficult. As shown in Figure 78, in first 180 seconds of the displacement time history, the maximum displacement is around 100 mm while from the 200 s, traffic jam occurred, and vehicles began to accumulate and stop on the bridge. In this traffic jam (marked as A) induced a large deflection which is around 428 mm while this deflection is within the limit defined by $L/800$. L is the length of the span. From the 260 s, the traffic flow was getting better and the deflection started to decrease to the normal level. Around the 450 s (marked B), there were two big buses at the midspan which also induced a large deflection (108 mm). After converting to frequency domain, a peak in the frequency spectrum is found and it is 0.147 Hz. Since the measurement point is at the midspan and this is the only peak, it might be the first bending mode. In addition, the frequency 0.14 Hz is very close to the first symmetric bending mode reported by (Soyoz et al. 2017).

7.4 Experiment on the Second Bosphorus Bridge

The Second Bosphorus Bridge is the second bridge in Istanbul, which was built to cross the Bosphorus Strait. The bridge is located on the north side of the First Bosphorus Bridge. The bridge has the main span with the length of 1090 m and two side spans with the length of 210 m. The main span is also supported by main cable and vertical hangers.

In this experiment, the same camera and lens were used to monitor the displacement of the midspan. As shown in Figure 79, the camera was setup on the European side and northern side of the bridge.

The distance from the midspan to the camera is around 600 m. The experiment was conducted on a Saturday afternoon around 2pm. Guardrails and edge of the midspan were selected as tracking target and a building on the background was selected as the static point to illuminate the camera motion.



Figure 79 Experimental setup of the Second Bosphorus Bridge

Figure 80 shows the displacement results of the mid-span of the Second Bosphorus Bridge. The largest displacement in the time history is 88 mm. There was no traffic jam during this monitoring. At the time points marked as A, B and C, there were large bus, trucks and other vehicles crossing the midspan respectively. Comparing the deflections with the First Bosphorus Bridge, it was much smaller. In the frequency domain analysis, two frequencies, 0.156 Hz and 0.286 Hz are identified as operational modes.

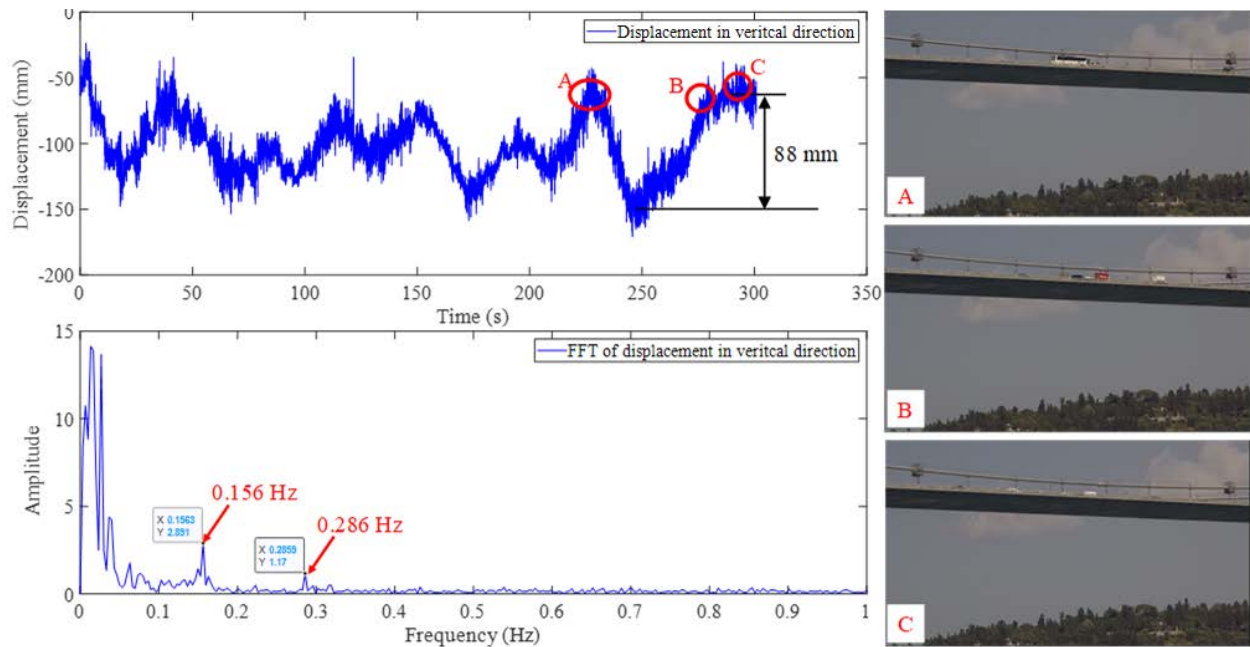


Figure 80 Displacement results of the mid-span of the Second Bosphorus Bridge

7.5 Experiments on the Osman Gazi Bridge

The Osman Gazi bridge is located in Izmit Bay and crosses the Marmara Sea. The bridge has three spans which are 566 m (approach span) + 1550 m (main span) + 566 m (approach span). In addition, a transition span with the length of 120 m is connected with each approach span. The Osman Gazi Bridge is the fourth longest bridge in the world. The approach spans and main span are supported by main cable and vertical hangers. Two experiments were conducted on the Osman Gazi Bridge.

7.5.1 The First Experiment

As shown in Figure 81, in the first experiment, the camera was located on the bank close the northern end of the bridge. The distance from the camera to the midspan is around 1.35 km, which

is the longest measurement distance by using cameras in literature. A 4K camera with a Navitar 24X zoom extender lens was employed to measure the displacement of the midspan. Because it was too far from the camera to the midspan of the bridge, it is very difficult to find the target on the bridge to track. A wood chessboard (967 mm \times 676 mm) was installed on the midspan to be the tracking target. Two accelerometers were installed on the midspan to measure the vertical and horizontal (transverse) vibrations. Figure 81(a) shows the images captured by the 4K camera. The image is not clear and the boundary is blur. From the video it can also be seen that the camera was shaking a lot during the experiment. Here the left red box in Figure 81(a) is to select the part of the tower as the static point to illuminate the camera motion problems. The right red box is the installed wood chessboard.



Figure 81 Experimental setup of the first experiment of the Osman Gazi Bridge: (a) image from camera, (b) camera location in map; (c) wood chessboard, (d) accelerometers in vertical and horizontal direction, (e) computer for data acquisition, (f) camera setup

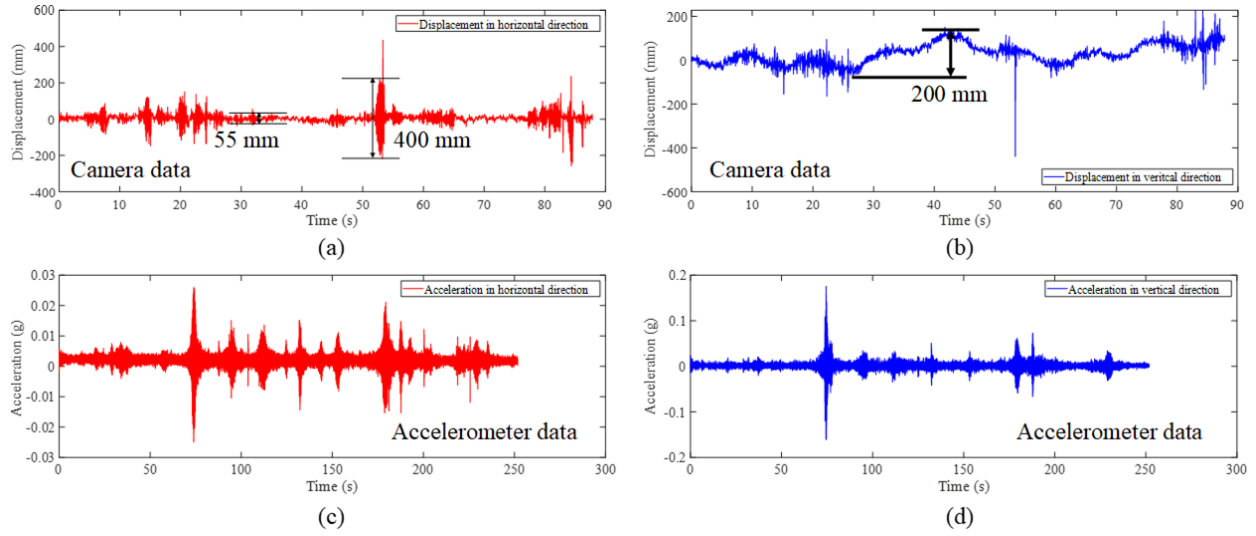


Figure 82 Results of the first experiment on the Osman Gazi Bridge in time domain: (a) horizontal displacement from vision-based method, (b) vertical displacement from vision-based method, (c) horizontal acceleration from accelerometer, (d) vertical acceleration from accelerometer

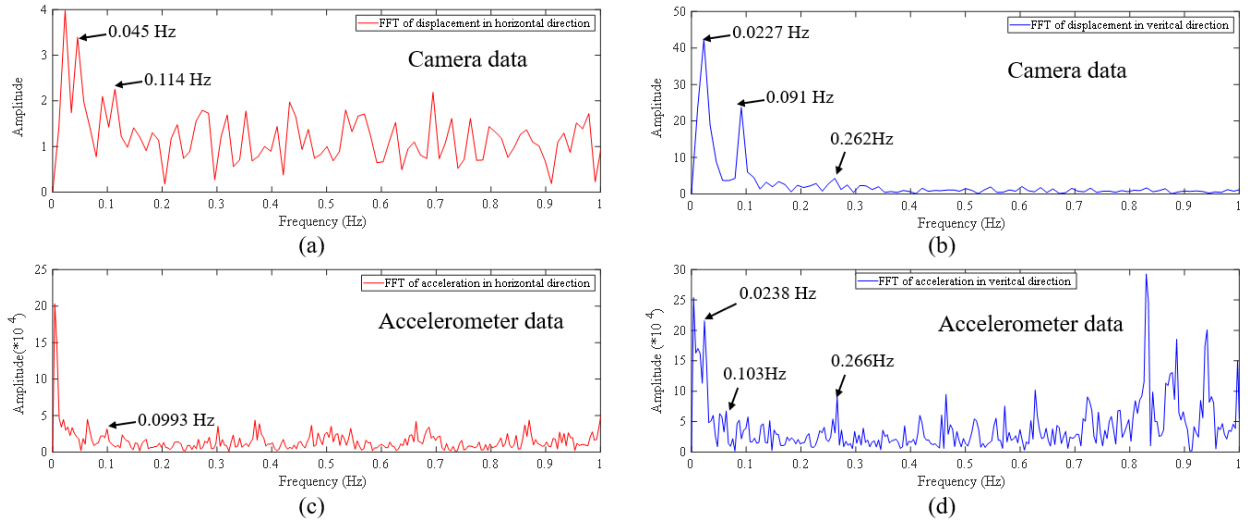


Figure 83 Results of the first experiment on the Osman Gazi Bridge in frequency domain: (a) FFT of horizontal displacement from vision-based method, (b) FFT of vertical displacement from vision-based method, (c) FFT of horizontal acceleration from accelerometer, (d) FFT of vertical acceleration from accelerometer

Figure 82 shows the displacement and acceleration results obtained from camera and accelerometers. From Figure 82(a) it can be seen that the maximum displacement in horizontal direction was 400 mm and in vertical direction was 200 mm. In Figure 83, the frequency domain data is shown. From Figure 83(a) and (c), it can be seen that the first vibration mode frequency in horizontal direction extracted from camera data is 0.114 Hz which is very close the value obtained from accelerometer, 0.0993 Hz. From Figure 83(b) and (d), the first two vibration mode frequencies (0.091 Hz and 0.262 Hz) in vertical direction obtained from camera data are very close to the one obtained from the accelerometers (0.103 Hz and 0.266 Hz).

7.5.2 The Second Experiment

Figure 84 shows the experimental setup of the second experiment. The camera was located on the tower close to Izmit side. A 4K camera with a zoom lens (75-300 mm) was employed to record the video of the motion of the wood chessboard installed at the midspan. During the experiment, the camera was vibrated by the wind load. The buildings in the background marked by the blue box was selected as the static point to illuminate the camera motion. The distance from the camera to the midspan is 750 m.

Figure 85 shows the results of the second experiment on the Osman Gazi Bridge in time domain. It can be seen that the average range of the horizontal displacement is 36 mm from Figure 85(a). In Figure 85(b), at the time points marked as A and B, the displacements are 247 mm and 410 mm respectively. The large displacement can be verified from the video frames as shown in Figure 86. In Figure 86(a), there was a truck crossing the midspan and it induced the displacement of 247 mm

and in Figure 86(b), there were two trucks crossing the midspan and it induced the displacement of 410 mm.

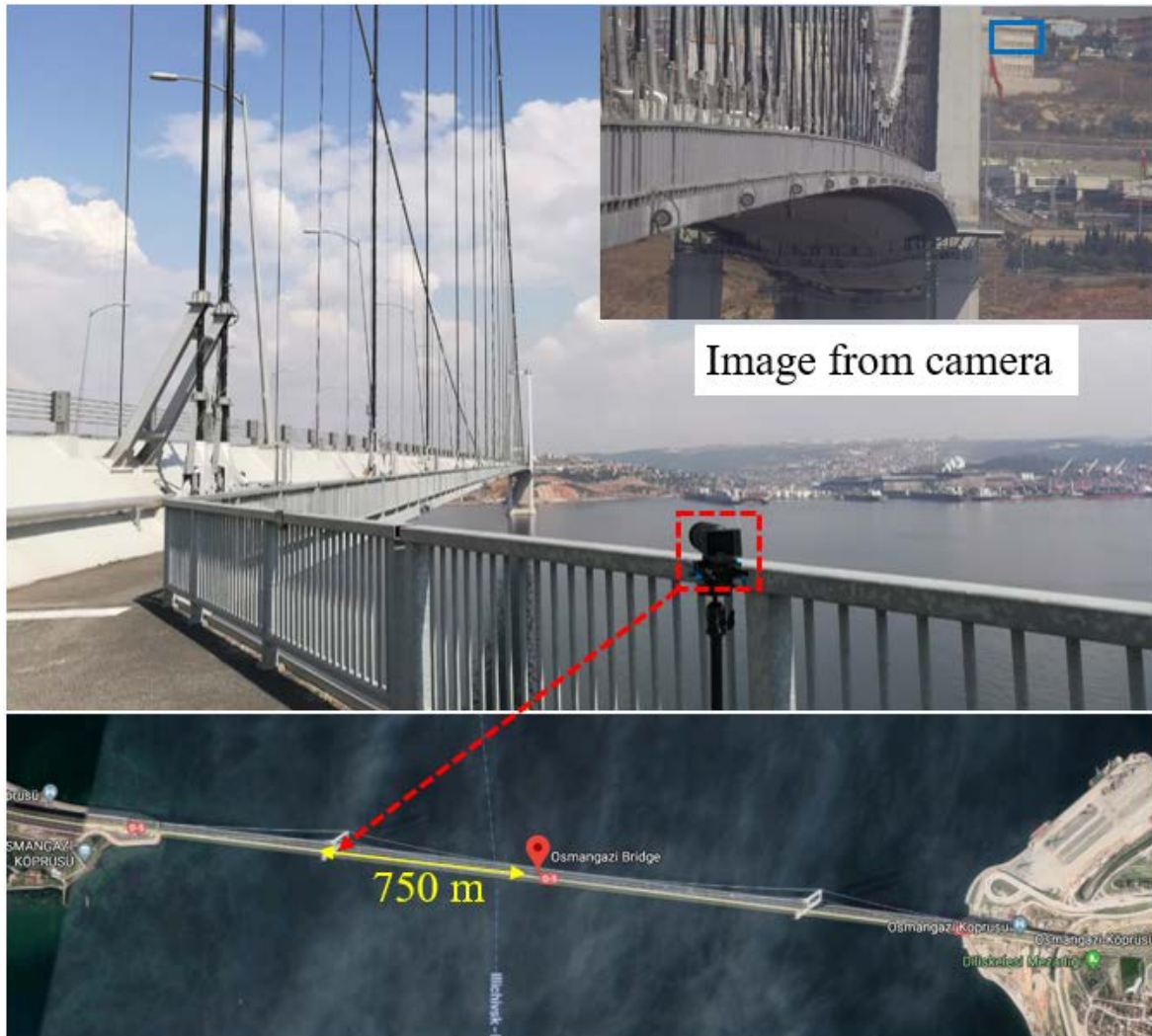


Figure 84 Experimental setup of the second experiment of the Osman Gazi Bridge

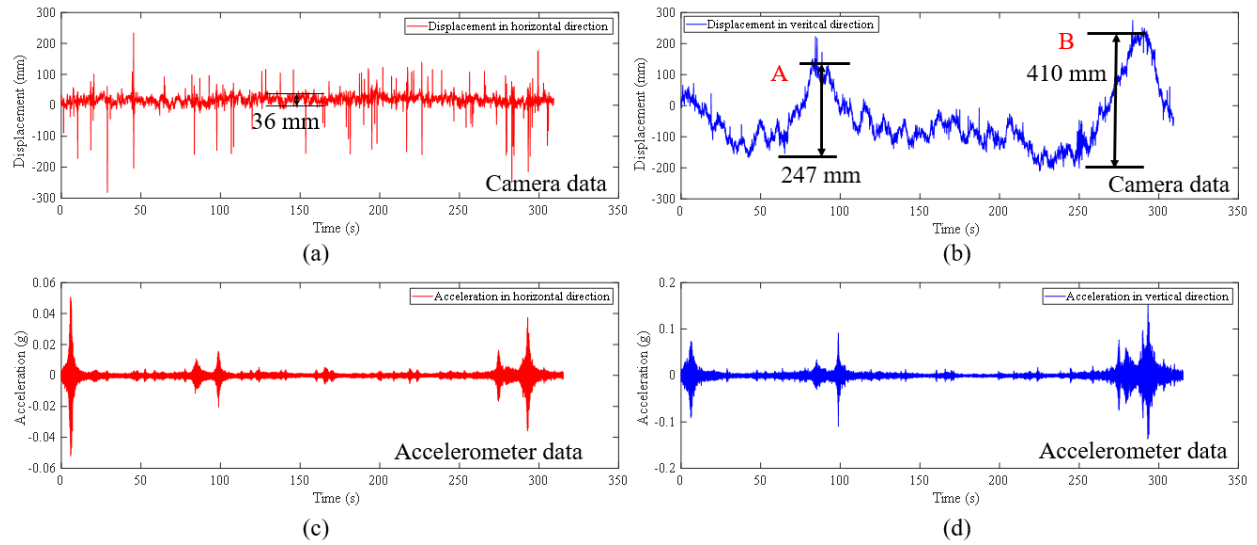


Figure 85 Results of the second experiment on the Osman Gazi Bridge in time domain: (a) horizontal displacement from vision-based method, (b) vertical displacement from vision-based method, (c) horizontal acceleration from accelerometer, (d) vertical acceleration from accelerometer



Figure 86 Vehicles on the bridge at time A and B

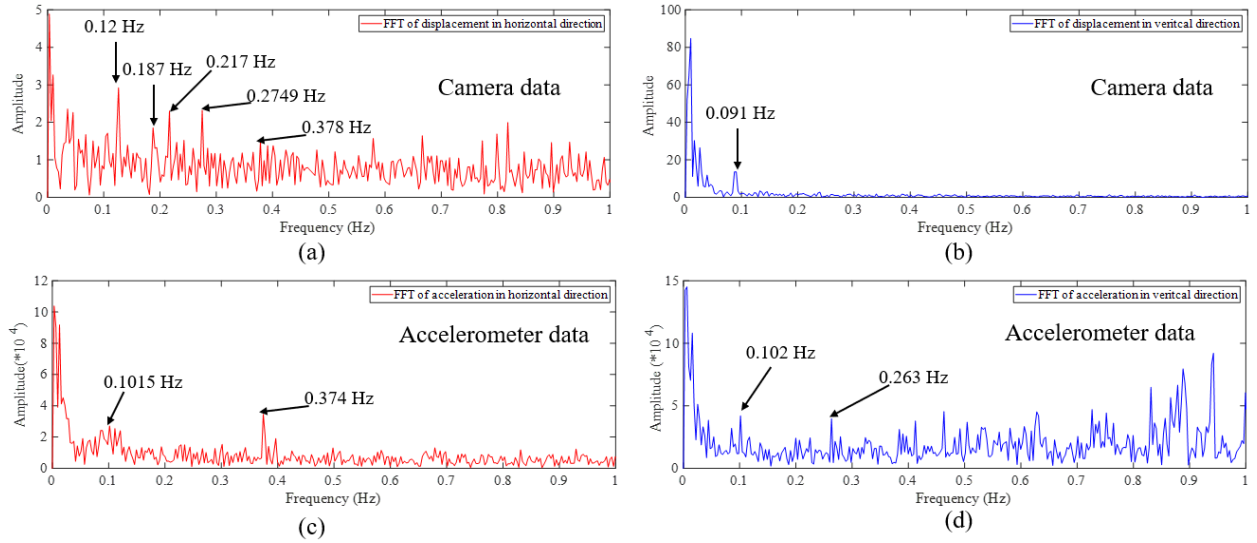


Figure 87 Results of the first experiment on the Osman Gazi Bridge in frequency domain:
(a) FFT of horizontal displacement from vision-based method, (b) FFT of vertical displacement from vision-based method, (c) FFT of horizontal acceleration from accelerometer, (d) FFT of vertical acceleration from accelerometer

From Figure 87(a) and (c), peaks can be identified from the spectra of the data from camera and accelerometer. The first vibration mode frequency in horizontal direction obtained from the camera data is 0.12 Hz which is close the one from accelerometer (0.1015 Hz) and also close to the values obtained in the first experiment (0.114 Hz-camera and 0.0993 Hz-accelerometer). From Figure 87(b) and (d), the first vibration mode frequency in vertical direction obtained from the camera data is 0.091 Hz which is close the one from accelerometer (0.102 Hz) and also close to the values obtained in the first experiment (0.091 Hz-camera and 0.103 Hz-accelerometer). While from Figure 87(b), the second mode frequency in vertical direction is not identified by the data from camera. It is identified from the acceleration data and consistent with the results from the first experiment.

7.6 Summary

In this chapter, field applications on large structures (three long span bridges in Turkey) using vision-based methods with long measurement distance are presented. Problems such as low image quality, haze problem, camera motion and blur etc. are encountered during the experiments. In this long-distance measurement, a manual target is necessary to improve the accuracy. To illuminate the camera motion problem, a static point on the background is required. While it is difficult to find a static point on the background in some cases, especially for these long distance measurement and large structures.

CHAPTER EIGHT: INVESTIGATION OF VIBRATION SERVICEABILITY OF A FOOTBRIDGE USING COMPUTER VISION- BASED METHODS

8.1 Introduction

With the development of high performance structural materials and aesthetic requirements for structures, longer and slender footbridges have attracted great public attention and a large amount of modern footbridges with lightweight and lively structures have been constructed in the last several decades (Živanović et al. 2005). This trend makes a large achievement of infrastructures in the progress of smart cities. However, it also causes critical issue: excessive vibrations of slender footbridges caused by pedestrian live load. These excessive vibrations may cause another problem, i.e., human comfort, since the main function of footbridges are to convey pedestrians. In this respect, it means that the estimated dynamic response of the footbridges has to be evaluated against human comfort level (Živanović and Pavia 2009). In general vibration produced by human-induced loads is a structural vibration serviceability problem rather than a structural safety (structural damage) problem. A famous example related to vibration serviceability problems of the footbridges is the Millennium Bridge over Thames River in England. In 2000, at the opening of the newly built footbridge, excessive vibrations which were described as “swaying violently” was reported when a group of pedestrians crossed the footbridge (BBC 2000). It took £5 million and about eight months to solve the problem, while the original cost of the “Wobbly” bridge was £18.2 million (BBC 2002). Examples about footbridge collapse incidents due to soldiers marching

movement in unison over the bridge such as Broughton Footbridge in England and Angers Bridge in France can date back to nineteenth century (Setareh 2016).

The vibration serviceability has become a hot topic of research and practice in the community of structural engineering and large amount of work has been done to assess the footbridge vibration level and to mitigate the excessive levels of vibrations in slender footbridges. Živanović et al. (2005) summarized the main research focus of footbridge vibration serviceability which generally includes: (1) vibration source related work such as human force estimation and modelling, (2) vibration path related work such as stiffness, mass and damping, (3) human perception related work such as estimation of human comfort level, (4) human-structure interaction, (5) design guidelines and (6) vibration reduction measures. Among these topics, estimation of human comfort level is one of the most direct and intuitive way to assess the vibration serviceability considering that the main function of footbridge is to convey pedestrian. Organizations and agencies established standards and codes such as ISO 10137 by International Organization for Standardization (ISO 2007), Euro code 5 by European Committee for Standardization (ECS) (1997), BS 5400 by British Standards Institution (BSI 1978) and Setra code by French Technical Department for Transport, Roads And Bridges Engineering and Road Safety (Setra 2006). In most of the standards and codes, the acceleration related indices combined with vibration frequencies are used to define the serviceability limit. For example, ISO 10137 uses the maximum 1-second running RMS value of the frequency-weighted acceleration time histories. Euro code 5, BS5400 and Setra selected the peak value of acceleration at the fundamental frequencies of structures. In addition to the standards and codes established by the organizations and agencies above, researchers also proposed different serviceability assessment criteria. Mackenzie et al. (2005)

proposed the serviceability assessment by defining the acceleration limits as a function of footbridge height, parapet height and route redundancy etc. Kasperski (2006) conducted a series of experiments on a footbridge, and based on the ratio of pedestrians alarmed by vibration, he presented that the RMS limit recommended by ISO 10137 is excessive and should be reduced to 60% for footbridge serviceability. Barker (2007) recommended using the root-mean-squared (RMS) of acceleration to assess the vibration serviceability instead. Živanović and Pavia (2009) proposed a probabilistic approach for the assessment of vibration serviceability based on the acceleration measurements. Setareh (2016) investigated the relationships between various evaluation parameters such peak value of acceleration, peak value of weighted acceleration, RMS and vibration dose value (VDV) and found that based on the relationships, VDV can also be a good index for the definition of serviceability limit. Dey et al. (2017, 2018) conducted a series of experiment on a large scale aluminum pedestrian bridge in laboratory and evaluated and calibrated various guidelines for the serviceability based on the experimental data. Feng et al. (2019) analyzed the correlation between the peak acceleration of footbridge vibration and the pedestrian comfort level collected by doing pedestrian questionnaire and based on the correlation they proposed a procedure of using acceleration data to assess footbridge vibration serviceability.

The current research and practice for vibration serviceability is based on the estimation of vibration level collected from conventional sensors such as an accelerometer. The drawbacks of using conventional sensors are traffic closure, setup time, cost and labor force to deal with the cable wiring work. It is not convenient to conduct such experiments, especially for field application. With the development of imaging devices and computer vision technology, vision-based approaches for vibration monitoring is gathering increasing attention in the field of structural

monitoring due to the advantages such as non-contact, long distance, low cost, time saving, and ease of use (Dong et al. 2018, 2019a; b, e; c; Dong and Catbas 2019; Feng et al. 2015a; Xu et al. 2018). By using cameras to record the structures and tracking the motion of targets on them, the displacement of structures can be easily obtained. To fit the standards and codes of vibration serviceability assessment, acceleration data can be calculated from displacement data with numerical differentiation.

This study proposes a non-contact approach for footbridge serviceability assessment by using computer vision-based methods and combining with current standards and codes. With a non-contact, cost-effective and time-saving way, the vision-based approach can overcome the drawbacks of using conventional sensors for the experimental assessment of vibration serviceability of footbridges. Živanović et al. (2005) in their review paper summarized that in the early stage of research work on vibration serviceability, using displacement and velocity to assess human perception is a more direct and intuitive way and was recommended by lots of researchers. By using vision-based methods, the displacement and velocity are easier to obtain compared with conventional displacement sensor and velocimeter. With the displacement data from vision-based methods, it may supply a possible alternative to define the vibration serviceability criteria based on displacement and velocity.

8.2 Vibration Limit for Serviceability Assessment in Current Standards and Codes

In this section, the current standards and codes employed by different countries and areas for vibration serviceability assessment of footbridges are discussed. Figure 88 illustrates the

acceleration limits for footbridge vibration serviceability in different standards and codes. The footbridge vibration serviceability assessment is based on the human comfort level of structures. Except for ISO 10137, the others employed the peak value of accelerations as the limits for serviceability. ISO 10137 use RMS value of the frequency weighted accelerations. It provides a base curve using RMS value for serviceability limit and recommends that for footbridges the serviceability limit is calculated by multiplying the base curve with the factor 60 for active pedestrians. To give a comparative view with other standards and codes, in this study the RMS value provided by ISO 10137 is converted to the equivalent acceleration peak value by multiplying by the factor $\sqrt{2}$. The peak acceleration limit of the standards and codes mentioned above are shown in Figure 88. Detailed formulas to determine the curves in Figure 88 are listed in Table 13. For ISO 10137, BS 5400 and Euro code 5, a unique curve of serviceability limit of peak acceleration against fundamental frequencies of the structures is provided. The region under the curve is acceptable for serviceability level and the region above the curve is unacceptable. For Setra standard as shown in Figure 88, it gives three curves to indicate the vibration serviceability level such as min comfort level, mid (medium) comfort level and max (extreme) comfort level.

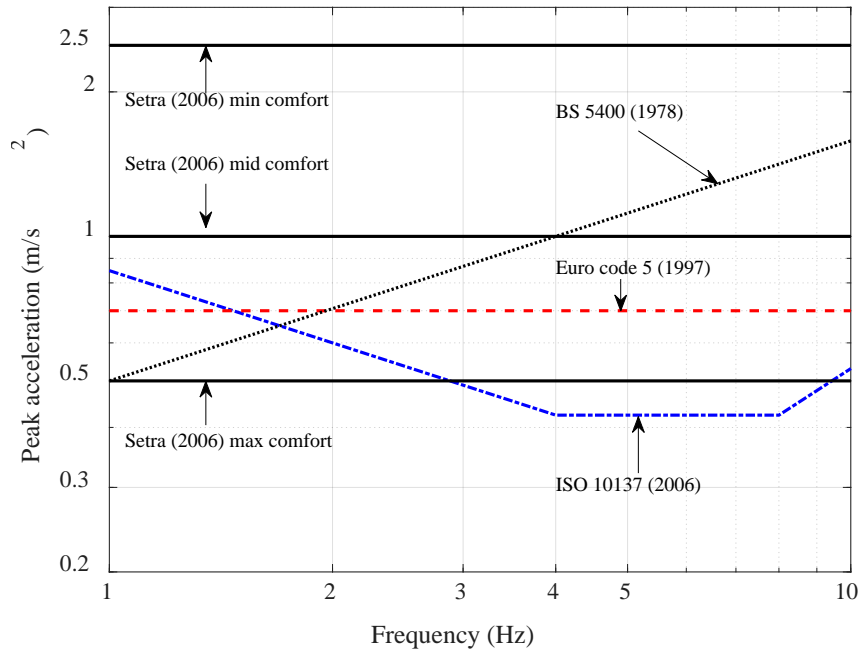


Figure 88 Acceleration limits for footbridge vibration serviceability in different standards and codes

Table 13 Detailed acceleration limits for footbridge vibration serviceability in different standards and codes

Code	Vertical peak acceleration (m/s)
ISO 10137 (2007)	$\begin{cases} a_{\text{limit}} = -0.1414f + 0.9899, 1 < f \leq 4 \\ a_{\text{limit}} = 0.4243, 4 < f \leq 8 \\ a_{\text{limit}} = 0.053f, 8 < f \leq 80 \end{cases}$
Euro code 5 (European committee for Standardization, 1997)	$a_{\text{limit}} = 0.7$
BS5400 (British Standards Institution, 1978)	$a_{\text{limit}} = 0.5f^{0.5}$
BS5400 (British Standards Institution, 1978)	$\begin{cases} a_{\text{limit}} < 0.5, \text{Extreme} \\ 0.5 < a_{\text{limit}} < 1.0, \text{Medium} \\ 1.0 < a_{\text{limit}} < 2.5, \text{Low} \end{cases}$

8.3 Vision-based Displacement/Velocity Measurement Using Feature Matching

In general, there are four steps to extract structural displacement/velocity from video or image sequence. In this study, a vision-based displacement/velocity measurement method using feature matching is employed. Figure 89 shows the procedure of the proposed method.

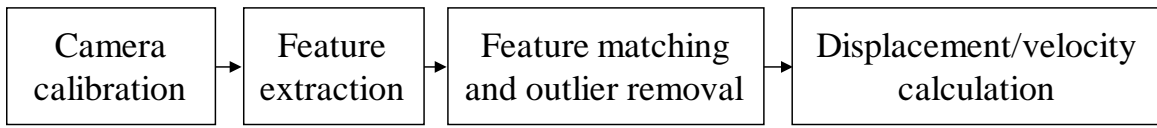


Figure 89 Procedure for proposed vision-based displacement/velocity measurement method using feature matching

Firstly, the camera is calibrated to estimate the relationship between the image coordinates and the real-world coordinates. Here the scale ratio is used which calculate the ratio between the actual dimension in physical unit (e.g. millimetre) and the image dimension in pixel (Dong et al. 2019b). For example, if the actual height of an object in real world is D mm, and the height of the object in image is d pixel, the scale ratio, SR , will be

$$SR = \frac{D}{d} \quad (66)$$

The scale ratio expressed in Eq. 66 is only suitable for the case when the axis of the camera and lens is perpendicular to the motion plane of the measurement target. For the cases that there is an inclination between them, (Dong et al. 2019b) gives a detailed discussion. Here it is not repeated.

Secondly, the camera records the video or image sequence of the structural motion. Feature points (also called key points, or kps) are extracted from the region of interest (ROI) of each image. A ROI is generally a sub region of an image that represent the measurement target of a structure. As shown in Figure 90, ROI 1 and ROI 2 are parts of a beam of a bridge. The feature point means a small image patch of the ROI with distinction such as corner, texture and gradient. On the right-top of Figure 90, extracted feature points are marked with circles in different color. In general, two components are required to define a feature point: feature detector and descriptor. Detector is to locate the region of the feature in an image and descriptor is a vector to describe the feature in mathematical language. In this study, SIFT (Scale-invariant feature transform) detector is applied to locate the feature point and VGG (Visual Geometry Group) descriptor is employed to describe the SIFT feature (Dong and Catbas 2019). It is noted in (Dong and Catbas 2019), that using SIFT detector and VGG descriptor performs better than using the original SIFT feature method (SIFT detector and SIFT descriptor). The ROI selected in Figure 90 is with the size of 209×210 pixels and 120 feature points (kps) are extracted.

Thirdly, the feature points extracted in two images are matched based on the similarity of the feature points. The similarity of two feature points can be calculated by the distance of the descriptors of them. The feature match pairs are selected by the one with the best similarity, i.e., with smallest distance. After the initial feature matching, there might be some wrong matches as shown in Figure 90. The RANSAC (RANdom SAmple Consensus) method is implemented to remove the outliers (Dong and Catbas 2019). On the right-bottom of Figure 90, it shows that three wrong matches are removed from the initial match. Depending on whether displacement or velocity is required in the end, the feature matching is performed between different image pairs.

Figure 91 shows two different feature matching strategies. The top of the Figure 91 illustrates the feature matching between the ROI of the first frame in the image sequence and the ROI of Frame i ($i = 2, 3, 4, \dots$). In other words, the first frame is not updated during the feature matching. The location change between the two ROIs is the relative displacement at the time of Frame i to the initial frame (Frame 1) and this is exactly the concept of displacement. While on the bottom of Figure 91, the feature matching is performed between the consecutive frames such as Frame 1 and Frame 2, Frame 2 and Frame 3, and Frame $i-1$ and Frame i , etc. The frame is always updated during the feature matching. The location change between the two ROIs of consecutive frames is the incremental of the displacement at the time of the Frame $i-1$ to Frame i and this refers to the concept of velocity. Using update frame strategy might be better to get more good matches when the structure moves over time and the light condition or surface feature changes along with the motion. The number of matched feature pairs shows the performance of the strategy. A higher number means the high matching quality. Figure 92 gives an example to show the comparison of the performance of the feature matching using two different strategies. In this case, 120 feature points are extracted in the ROI of each image, and the number of matched pairs using frame update strategy within an image sequence is more than the one without using frame update. It means that feature matching using frame update give better performance. (Dong et al. 2019c) summarized the pros and cons of using the two different strategies. In this study, whether the former or the latter is used depends on the requirement of whether displacement or velocity is necessary. In the experimental section, more details will be discussed according to the results.

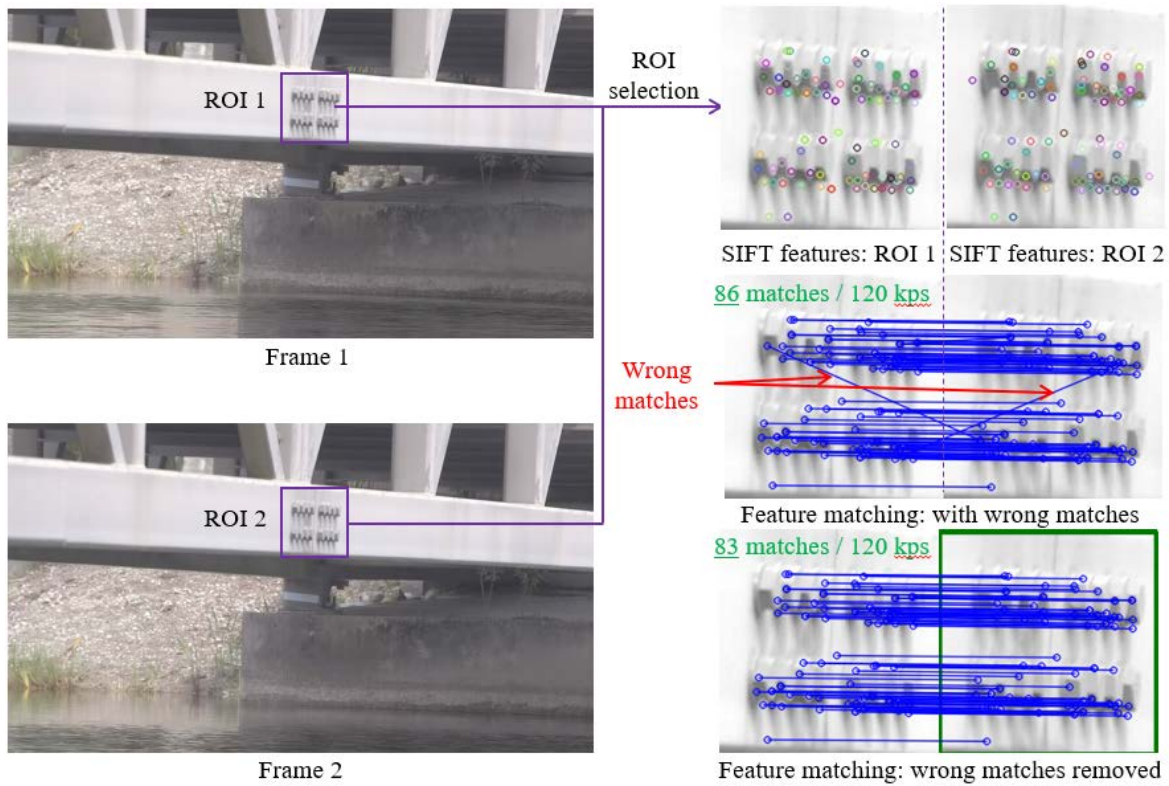


Figure 90 Feature extraction and outlier removal

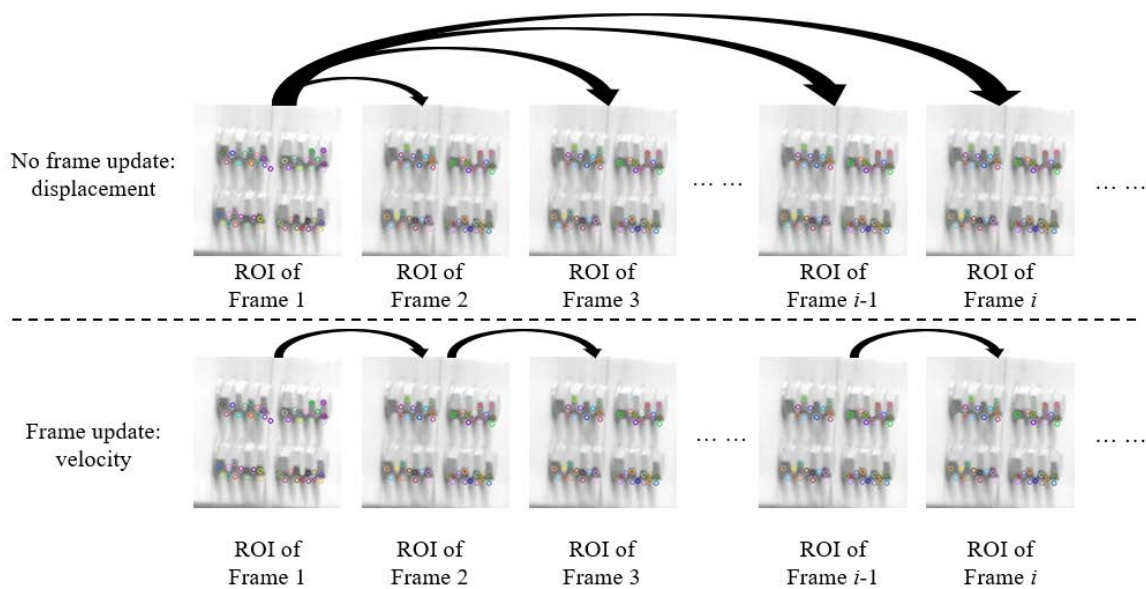


Figure 91 Feature matching for displacement and velocity

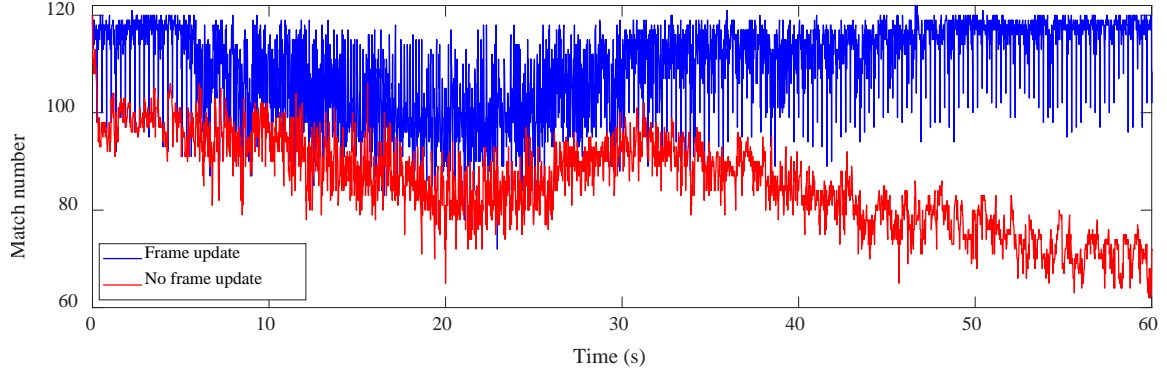


Figure 92 Matched pairs of an image sequences using different feature matching strategy

At last, the displacement/velocity can be calculated by taking the average of the location change of the matched feature points in two images. The displacements in x and y direction, X and Y , in physical unit can be calculated by

$$\begin{cases} X_i = SR_x \frac{\sum_{j=1}^n (x_i^j - x_1^j)}{n} \\ Y_i = SR_y \frac{\sum_{j=1}^n (y_i^j - y_1^j)}{n} \end{cases} \quad (67)$$

where (x_i^j, y_i^j) and (x_1^j, y_1^j) are the image coordinates of the j th matched feature point of between the ROIs of Frame i and Frame 1, n is the total number of the matched feature point between the ROIs of Frame i and Frame 1, and SR_x and SR_y are the scale ratio in x and y direction. The velocity in x and y direction, V_x and V_y , in physical unit can be calculated by

$$\begin{cases} V_x = SR_x \frac{\sum_{j=1}^m (x_i^j - x_{i-1}^j)}{n \cdot \Delta t} \\ V_y = SR_y \frac{\sum_{j=1}^m (y_i^j - y_{i-1}^j)}{n \cdot \Delta t} \end{cases} \quad (68)$$

where (x_i^j, y_i^j) and (x_{i-1}^j, y_{i-1}^j) are the image coordinates of the j th matched feature point of between the ROIs of Frame i and Frame $i-1$, and Δt is time interval of the image sampling which is the reciprocal of the sampling rate.

8.4 Experimental Verification and Field Application

8.4.1 Experimental setup

A series of experiments were conducted on a footbridge on a campus in the southeast of the United States. As shown in Figure 93, the footbridge comprises of 19.5m long vertical truss frames which are connected via splice connection in the middle and spans an entire length of 39m over a pond. The width of the bridge is 4.17 m. The vertical truss members on the left and the right side have HSS 10×10×3/8 top and bottom chords and are stabilized with HSS 6×4×3/8 type vertical and HSS 4×4×1/4 type diagonal steel members. The lateral stability is provided by another truss frame that is 3.65m wide which is constructed with HSS 3×3×1/4 type diagonal cross braces, W12×22 type lateral members. Two separate spans are spliced in the middle and the entire frame holds a thin layered aluminum-concrete composite deck (Dong et al. 2019a). In general, the bridge is under light pedestrian traffic loads and small vehicles such as golf carts. The fundamental frequency of the footbridge is 2.54 Hz as presented in the authors previous publication (Celik et al. 2019b). In

this experiment, a camera with the resolution of 1920×1080 pixels and the speed of 60 frame per second was employed to monitor the vibration of the mid-span, marked as P1. An accelerometer was also installed at the mid-span to record the vibration of the footbridge. The sampling rate of the accelerometer was 200 Hz.



Figure 93 Experimental setup

Table 14 Experimental cases

Case	Loading form	Speed (bpm, beat per minute)	Frequency (Hz)
1	Seven people, walking with a metronome	101	1.68
2	Seven people, walking with a metronome	120	2.0
3	Seven people, walking with a metronome	201	3.35
4	Seven people, random walking	Random speed and beat	--
5	Seven people, running with a metronome	180	3.0
6	Seven people, random running	Random speed and beat	--
7	Seven people, random jumping	150	2.5
8	Seven people, random jumping	Random speed and beat	--

Eight people were employed to conduct this experiment. One person stood on the mid span as the passive subject (footbridge bystander) and the other seven people moved as active subjects (pedestrians) by walking, running and jumping in different cases respectively. The reason why eight people was employed is that ISO 10137 states that a group size of 8 to 15 people are the average pedestrian flow based on the daily occurrence rate. For this footbridge, the daily occurrence rate as observed is smaller than the range stated in ISO 10137. While here the low bound of 8 to 15 was selected. Table 14 lists the experimental cases conducted in this experiment. In Case 1 to 3, the seven people walked on the footbridge in a group by following the beats played by a metronome. The speeds are 101, 120 and 201 beats per minute (bps), respectively which are equivalent to 1.68, 2.0 and 3.35 Hz. 1.68 Hz and 3.35 Hz are close to the value calculated by subtracting and adding the fundamental frequency of the footbridge with one third of it. In

references (Matsumoto et al. 1978; Živanović et al. 2005), it is presented that the frequency people walking on footbridges follows a normal distribution with a mean pacing rate of 2.0 Hz and standard deviation of 0.173 Hz. That is the reason that 2.0 Hz was chosen in this experiment. In Case 4, the seven people randomly walked on the footbridge with their normal speeds. In Case 5, the seven people ran with the speed of 180 bpm (3 Hz) and the frequency 3 Hz for running is in the range 2.0-3.5 Hz defined by (Bachmann et al. 1995). In Case 6, the seven people randomly ran on the footbridge. In Case 7, the seven people jumped on the footbridge following the speed of 150 bpm (2.5 Hz) which is close to the fundamental frequency of the structure (2.54 Hz). In Case 8, the seven people jumped on the footbridge randomly. During the experiment, the camera and accelerometer both recorded the vibrations of all cases.

8.4.2 Result analysis

8.4.2.1 Comparison of Displacement Results from Two Different Feature Matching Strategies

In this section the results from Case 5 is selected to compare the performance of displacement measurement using the vision-based methods with frame update and no frame update strategies. With no frame update, the displacement can be calculated directly from Eq. 67. With frame update, the displacement is calculated by accumulating the displacement change between the two consecutive images since the location change in two consecutive images actually is the incremental value of displacement at the time of current frame.

Figure 94 shows the displacements from two different matching strategies. The correlation coefficient between the two time histories is 91.68%. As stated in (Dong et al. 2019c), the displacement by using the strategy of frame update has the accumulating errors when converting the velocity/acceleration to displacement and this phenomena also occurs in Figure 94a. While in frequency domain, both of them give the consistent frequencies: (1) 2.534 Hz (close to fundamental frequency, 2.54 Hz) and (2) 3.001 Hz (frequency of people running beat, 3.0 Hz). From this comparison, it is recommended that if displacement is required for serviceability assessment, the vision-based method with no frame update should be used.

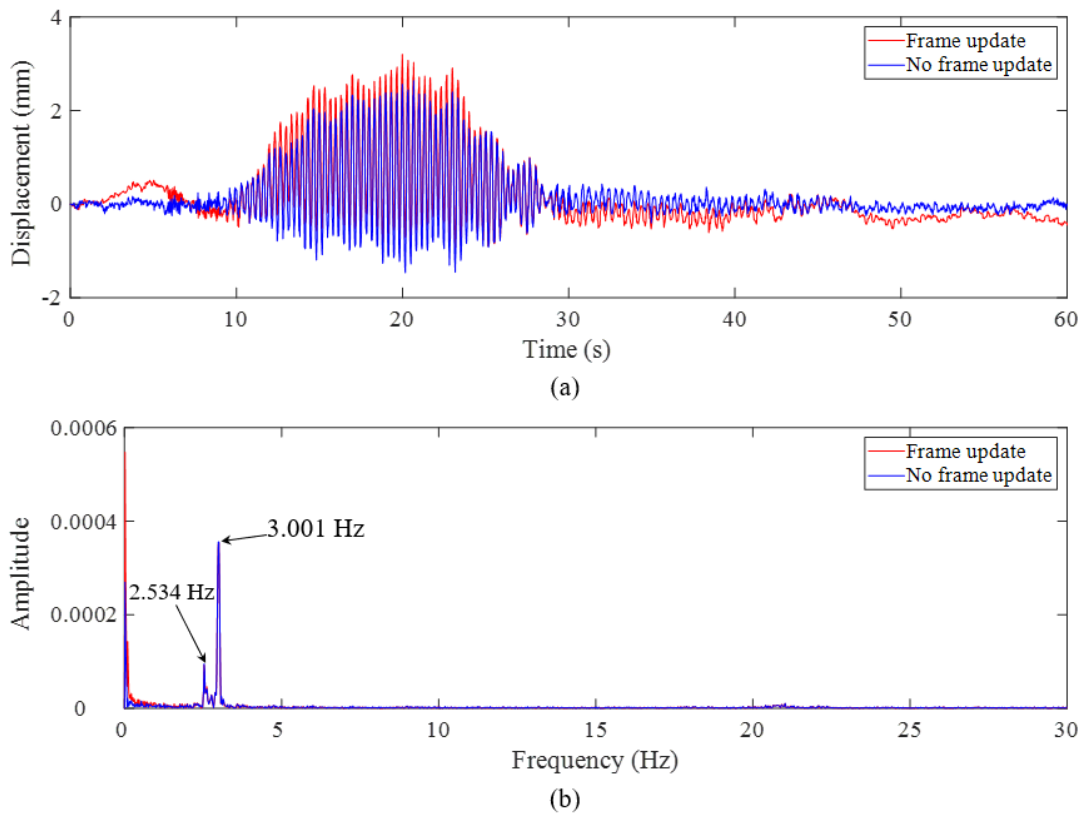


Figure 94 Comparison of displacement results from two different feature matching strategies of Case 5: (a) comparison in time domain, and (b) comparison in frequency domain

8.4.2.2 Comparison of Velocity Results from Two Different Feature Matching Strategies

Here Case 5 is also selected to do the comparison: to compare the performance of velocity measurement using the vision-based methods with frame update and no frame update strategies. With frame update, the displacement can be calculated directly from Eq. (3). With frame update, the displacement is obtained by calculating the numerical differentiation of the displacement which is calculated by Eq. 67. Figure 95 shows the velocities from two different matching strategies. The correlation coefficient between the two time histories is 97.03%. Comparing to the results of displacement (correlation coefficient is 91.68%), the velocity results show higher consistency, which means calculating velocity from displacement using numerical differentiation gives more reliable results. In frequency domain, the two frequencies are also 2.534 Hz and 3.001 Hz which are consistent with those in Section 8.4.2.1. From this comparison, it is suggested that if velocity is required for serviceability assessment, the vision-based method with either frame update or no frame update is good.

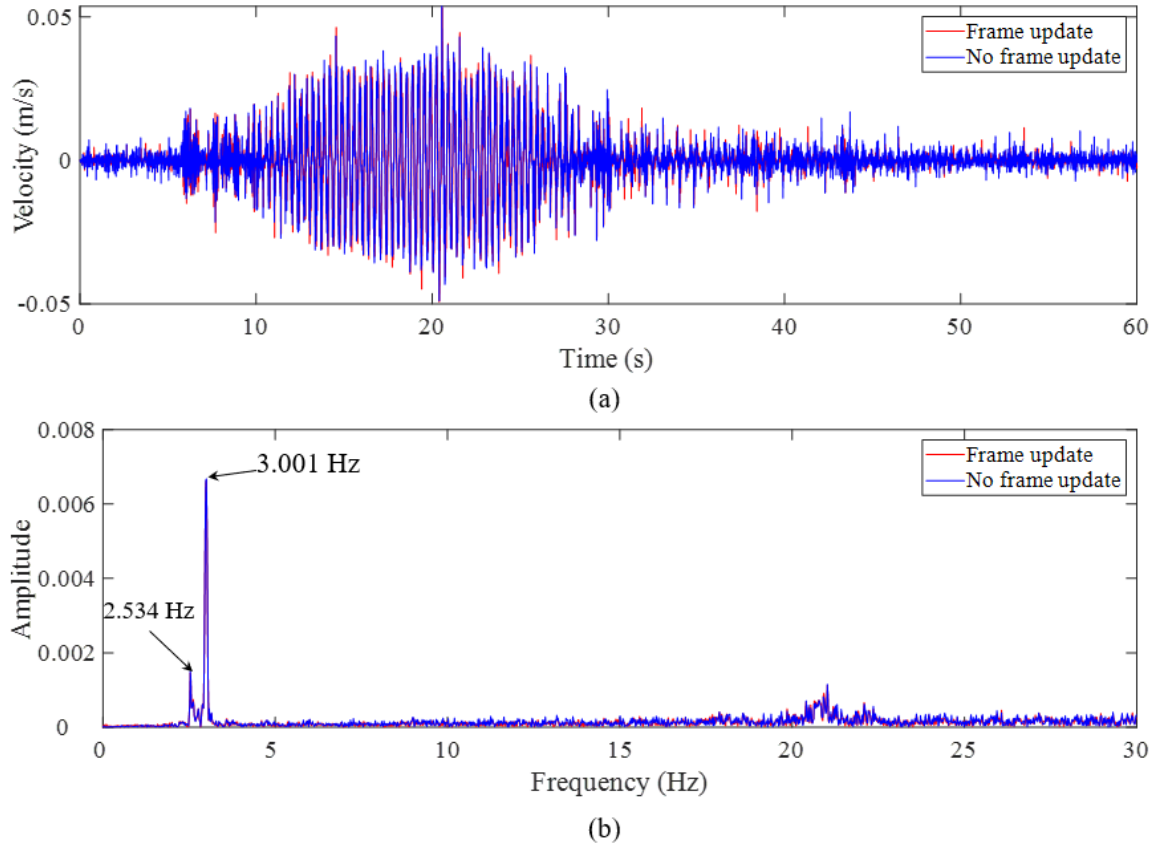


Figure 95 Comparison of velocity results from two different feature matching strategies of Case 5: (a) comparison in time domain, and (b) comparison in frequency domain

8.4.2.3 Comparison of Acceleration Results from Two Different Feature Matching Strategies and Accelerometer

To get the acceleration from displacement/velocity data, numerical differentiation operations are necessary. The acceleration can be calculated by taking the second derivative of the displacement using numerical methods and can also be calculated by taking the first derivative of the velocity using numerical methods. Figure 96 and Figure 97 shows the acceleration data (raw data, without filtering) directly calculated from displacement (no frame update) and velocity (frame update). Figure 98 shows the acceleration data collected by accelerometer. Comparing with Figure 96a,

Figure 97a and Figure 98a, it can be seen that the ranges of the raw acceleration data obtained from vision-based methods are both from -2 m/s^2 to 2 m/s^2 , while the range of the acceleration collected by accelerometer is from -0.5 m/s^2 to 0.5 m/s^2 . Comparing Figure 94b, Figure 95b, Figure 96b, Figure 97b, and Figure 98b, it is indicated that high order modes are mixed into the acceleration time histories when converting displacement to acceleration and velocity to acceleration. The reason is thought due to the numerical differentiation. It might also induce the large spikes in acceleration data and a large range, e.g. from -2 m/s^2 to 2 m/s^2 .

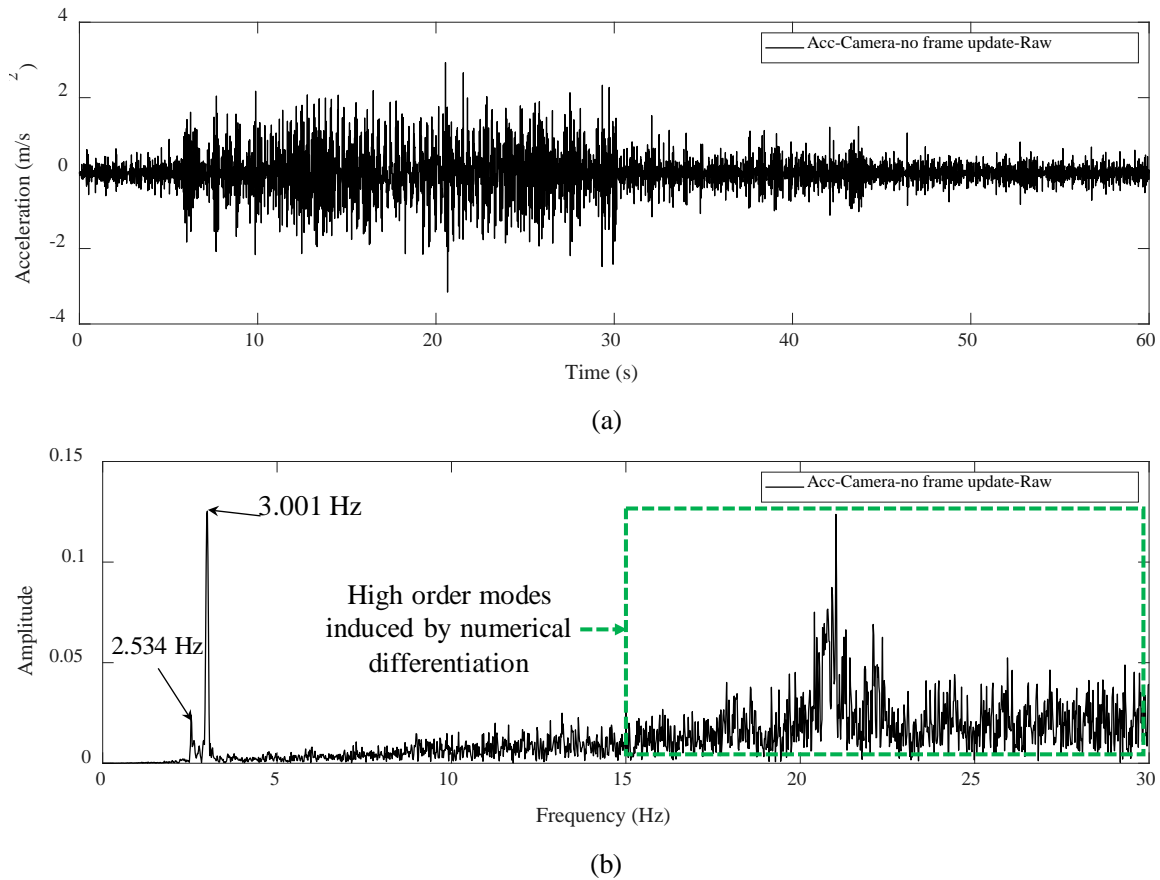


Figure 96 Acceleration (raw data) directly calculated from the displacement data by using vision-based method with no frame update of Case 5: (a) in time domain, and (b) in frequency domain

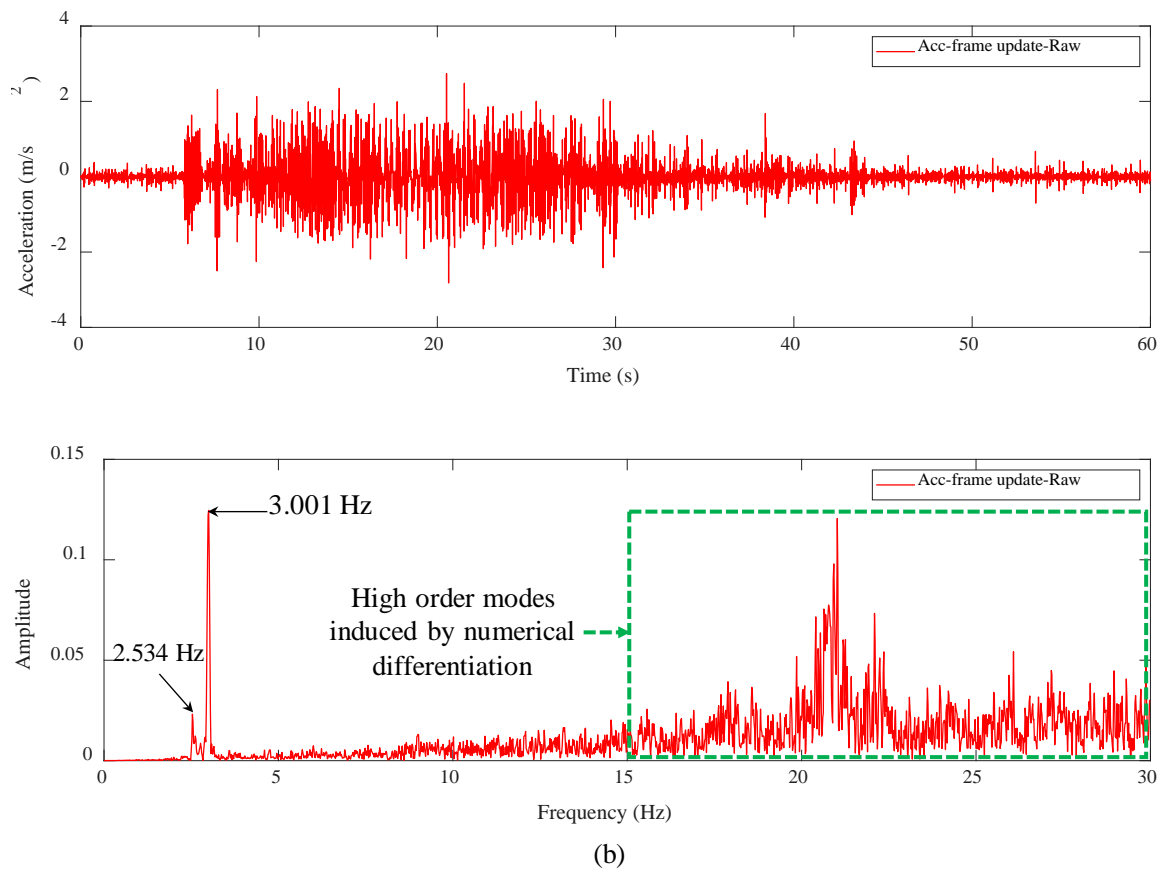


Figure 97 Acceleration (raw data) directly calculated from the velocity data by using vision-based method with frame update of Case 5: (a) in time domain, and (b) in frequency domain

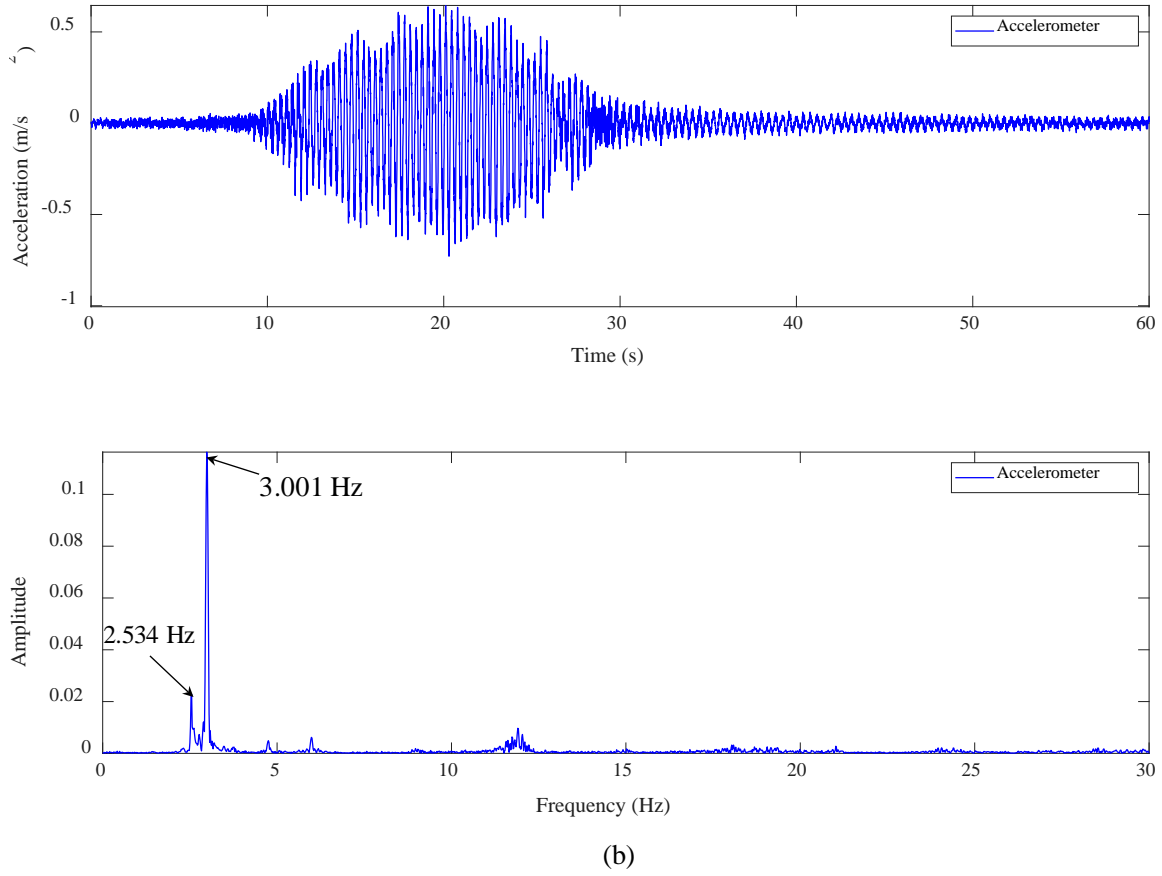


Figure 98 Acceleration data collected by accelerometer of Case 5: (a) in time domain, and (b) in frequency domain

To eliminate the high order modes mixed in the acceleration data obtained from vision-based methods in Figure 96 and Figure 97, a low-pass filtering is required. Figure 99 shows the comparison of filtered acceleration obtained from vision-based methods with the data from accelerometer. From Figure 99a, it is suggested that the three acceleration time histories are consistent with each other very well. Figure 99b shows the zoomed window of Figure 99a within the range of 18s to 22s. Large spikes in the raw acceleration data obtained from vision-based methods shown in Figure 96a and Figure 97a are removed and the amplitude range is within -0.5 m/s^2 to 0.5 m/s^2 . The high order modes in Figure 96b and Figure 97b are eliminated as shown in

Figure 99c. This is important because in current standards and codes for vibration serviceability assessment, the peak acceleration value or RMS is used to define the serviceability limits. Spikes in the raw acceleration time histories due to the numerical differentiation would cause higher peak value and RMS, which result in underestimated serviceability assessment results. Table 15 lists the correlation matrix of time acceleration histories obtained from vision-based methods (camera) with low-pass filtering and accelerometer. It can be seen that the correlation coefficients between the data extracted from camera and accelerometer are 96.07% (frame update) and 95.72% (no frame update), which gives a quite consistent verification. Also, the correlation coefficient between the data obtained from two feature matching strategies is very high, 99.38%. If the acceleration collected by the conventional sensor, i.e., accelerometer is regarded as the ground truth, considering the high consistencies between the three acceleration time histories, the acceleration obtained from both the two feature matching strategies are suitable for serviceability assessment.

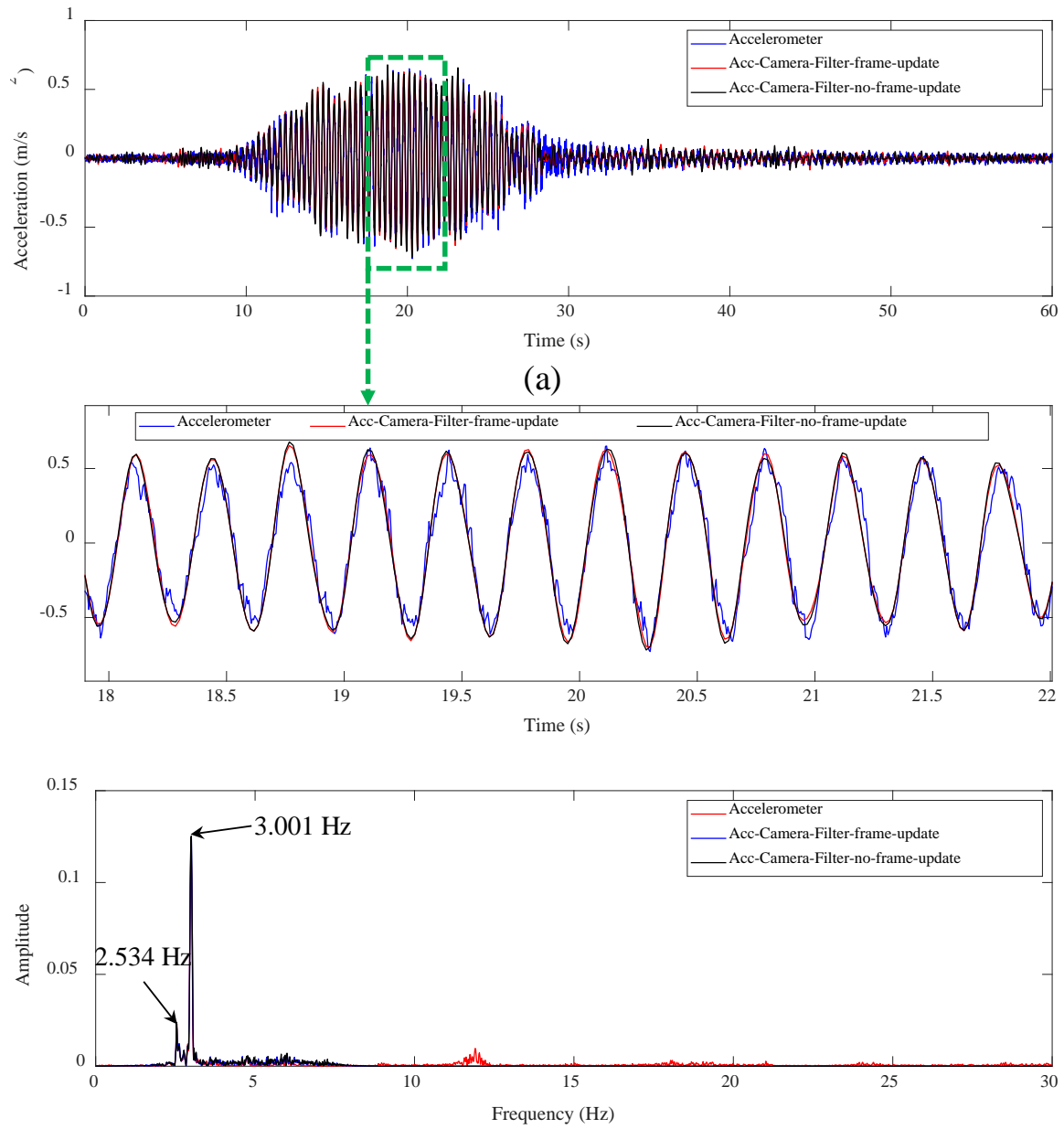


Figure 99 Comparison of filtered acceleration obtained from vision-based methods with the data from accelerometer of Case 5: (a) in time domain, and (b) in frequency domain

Table 15 Correlation matrix of time acceleration histories

	Accelerometer	Camera-frame update	Camera-no frame update
Accelerometer	1	96.07%	95.72%
Camera-frame update	96.07%	1	99.38%
Camera-no frame update	95.72%	99.38%	1

8.4.2.4 Vibration Serviceability Assessment Based on Human Comfort Level

In this study, considering the comparisons of displacement, velocity and acceleration of using two different feature matching strategies, using no frame update to obtain displacement data and then convert to velocity and acceleration would be applied to assess the vibration serviceability. Table 4 listed the RMS and peak value of the acceleration data (a_{peak}). Here to the RMS is also converted to the equivalent peak acceleration value ($a_{\text{peak-RMS}}$) as indicated in Figure 88. $a_{\text{peak-RMS}}$ and a_{peak} are applied to assess the vibration serviceability comparing with the serviceability limits based on human comfort shown in Figure 88. In Table 16, the column titled by “Cam” is the value calculated by the displacement data obtained from the vision-based method with no frame update and the column titled by “Accl” is the value calculated by the acceleration obtained from accelerometer.

Table 16 RMS and peak value of the acceleration data

Case	RMS (m/s ²)		$a_{\text{peak-RMS}}$ (m/s ²)		a_{peak} (m/s ²)	
	Cam	Accl	Cam	Accl	Cam	Accl
1	0.0324	0.0341	0.0458	0.0482	0.099	0.124
2	0.0564	0.0389	0.0798	0.055	0.176	0.118
3	0.147	0.110	0.207	0.156	0.341	0.467
4	0.0373	0.0351	0.0527	0.0497	0.0857	0.109
5	0.369	0.330	0.522	0.468	0.659	0.650
6	0.474	0.495	0.671	0.701	1.051	1.178
7	0.904	0.892	1.279	1.262	1.919	1.836
8	0.6151	0.584	0.870	0.826	1.500	1.426

Table 17 listed the results of the vibration serviceability assessment of different cases using vision-based methods and accelerometer. The results of two different approaches are very consistent. It can be seen that the level of vibration serviceability of Cases 1 to 4 is within the limits of ISO 10137, Euro code 5 and BS 5400. For the Setra code, it is within the maximum comfort level. Cases 1 to 4 are the cases of human walking with different paces (1.68 Hz, 2.0 Hz, 3.35Hz and random beat). Among the four cases, walking with the frequency of 3.35 Hz (Case 4) causes the largest RMS and a_{peak} .

Table 17 Serviceability assessment of different cases

Case	Serviceability Assessment							
	ISO 10137		Euro code 5		BS5400		SETRA	
	Cam	Accl	Cam	Accl	Cam	Accl	Cam	Accl
1	Under limit	Under limit	Under limit	Under limit	Under limit	Under limit	Max comfort	Max comfort
2	Under limit	Under limit	Under limit	Under limit	Under limit	Under limit	Max comfort	Max comfort
3	Under limit	Under limit	Under limit	Under limit	Under limit	Under limit	Max comfort	Max comfort
4	Under limit	Under limit	Under limit	Under limit	Under limit	Under limit	Max comfort	Max comfort
5	Under limit	Under limit	Under limit	Under limit	Under limit	Under limit	Mid-Max comfort	Mid-Max comfort
6	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Min-Mid comfort	Min-Mid comfort
7	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Min-Mid comfort	Min-Mid comfort
8	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Exceed limit	Min-Mid comfort	Min-Mid comfort

In Case 5, seven people running with the frequency of 3 Hz caused larger RMS and a_{peak} than walking cases, but the vibration serviceability level is still within the limits of ISO 10137, Euro code 5 and BS 5400. For the Setra code, it is in the range between mid and maximum comfort level. However, when the seven people ran with random paces (Case 6), the vibration serviceability level exceeds the limits defined by ISO 10137, Euro code 5 and BS 5400. Also, for Setra code, it moves to the level between minimum and mid comfort level. The reason for the increase of RMS and a_{peak} might be that during the random running, as shown in Figure 100f the seven people's random running induced a dominant frequency, 2.548 Hz which is the resonant frequency of the

footbridge and the component takes a large part in the time history. The resonant effect induced larger amplitude of acceleration responses and larger RMS and a_{peak} . While in Case 5 shown in Figure 100e, the dominant frequency of running with the frequency of 3.0 Hz is 3.001 Hz which is not the resonant frequency.

Similar to Case 6, the vibration serviceability levels of Case 7 and Case 8 also exceed the limit defined by ISO 10137, Euro code 5 and BS 5400. For Setra code, it also moves to the level between minimum and mid comfort level. The loading of Case 7 and Case 8 are jumping.

In Case 7, the seven people jumped with the 2.5 Hz which is close to the fundamental (resonant) frequency of the footbridge and it induced the largest amplitudes of the acceleration responses. In Case 8, the seven people jumped with random beat but still it induced a frequency of 2.52 Hz which is close to the fundamental (resonant) frequency of the footbridge, which is very similar to Case 6: random running. Also, the similar phenomena occur in Case 4: random walking. It is indicated that when people move (walk, run and jump) with random beat on the footbridge, it has high chances to excite the fundamental frequency. While it is also observed from Figure 100 that the fundamental frequency of the footbridge changes slightly over the types of people's motion (walking, running and jumping) and the beat speed. The range is within 0.1 Hz. The slight change might affect the serviceability level since the serviceability limit defined in standards and codes is related the fundamental frequency.

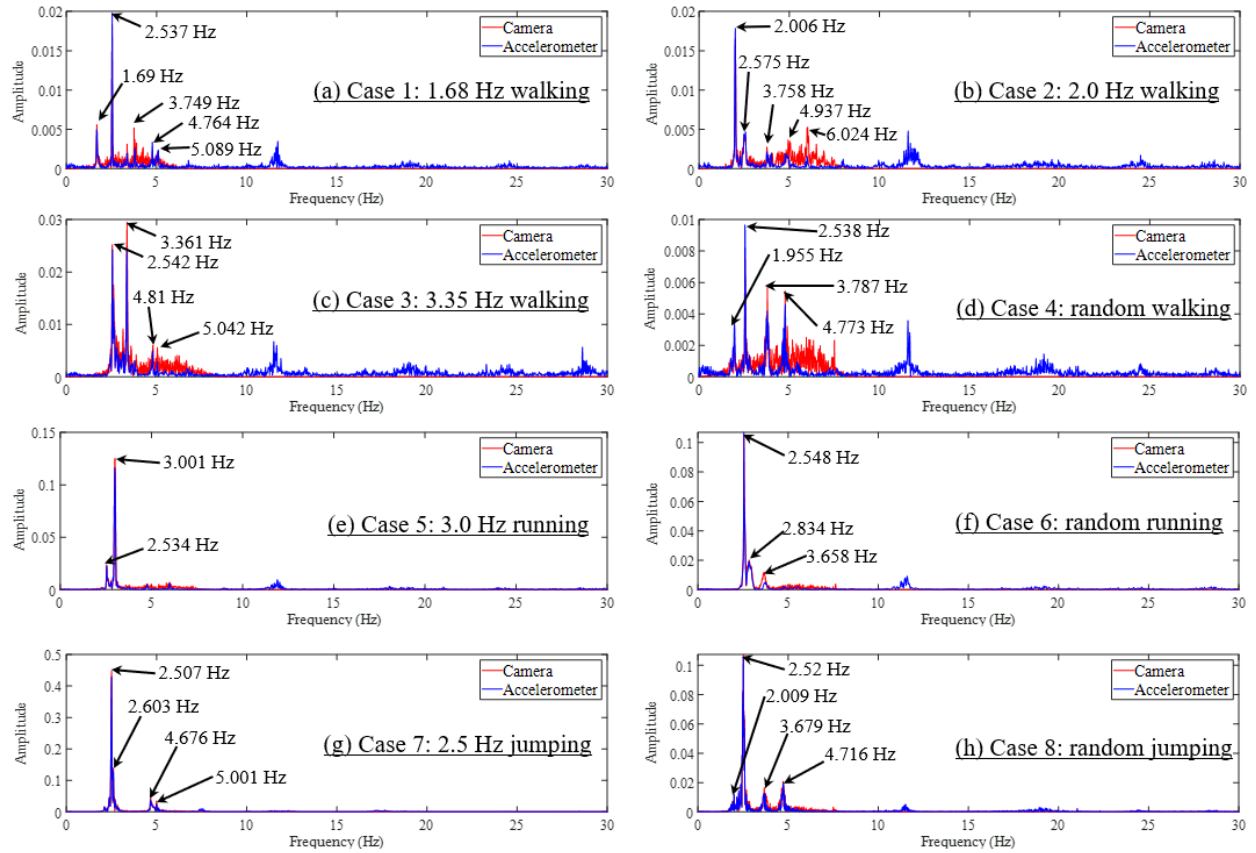


Figure 100 Frequency spectra of the acceleration time histories obtained from vision-based method and accelerometer of all cases

8.5 Summary

In this study, the vibration serviceability assessment based on human comfort of a footbridge using computer vision techniques is investigated. A series of experiments of footbridge under different types of human loading including walking, running and jumping with different speeds (beats and frequencies) are conducted to verify the proposed approach. The main approaches, findings, and conclusions are as follows:

(1) The vibration serviceability assessment criteria in current standards and codes are discussed and ISO 10137, Euro code 5, BS 5400 and Setra are taken as the reference for serviceability assessment;

(2) A vision-based displacement/velocity monitoring approach is proposed by using feature matching. Two different feature matching strategies such as matching between first frame and current frame (no frame update) and matching between consecutive frames (frame update) are compared and it is suggested that the way of feature matching with no frame update give good displacement, velocity and acceleration, while the way with no frame update perform not good in terms of displacement measurement.

(3) The feature matching with no frame update is applied to estimate the displacement from image sequence and the displacement is first converted to acceleration and then indicators such RMS and a_{peak} are calculated for the serviceability assessment compared with the conventional accelerometer.

(4) The random running, random jumping and jumping with the frequency that is close to the fundamental frequency of the footbridge can induce the serviceability level to exceed the limit defined by the current standards and codes.

(5) The human loads of different types and speeds can induce a slight change (within 0.1 Hz in this experiment) of the fundamental frequencies of the footbridge and this would affect the serviceability assessment since the serviceability limit defined in standards and codes is related the fundamental frequency.

Although in current standards and codes, acceleration is still the most common data type for serviceability assessment, displacement and velocity can also be other alternatives. The serviceability assessment in current standards and codes is based on the human comfort and human comfort is a concept of perception. It is stated in literature that both the displacement and velocity can trigger the perception of human (Živanović et al. 2005). For example, the minimum displacement that human can perceive is 0.001mm (Pretlove and Rainer 1995). Reference (Smith 1969) applied displacement as the indicator for human comfort and serviceability. It is also noted that in Japan, velocity is used as the index to assess footbridge serviceability (Yoneda 2002). In reference (Živanović et al. 2005), it was stated that “*Usually, acceleration response was measured because it was established as the best parameter for describing people’s reaction to vibrations and, also, it was easy to measure it using widely available accelerometers.*” This might be the reason why acceleration is used in most of the standards and codes for serviceability assessment. Considering the advantages of vision-based methods such as non-contact, long distance, low cost, time saving, and ease of use compared to conventional accelerometers, using vision-based methods to collect displacement, velocity and acceleration would be a better choice for vibration serviceability assessment and it gives high possibilities for the serviceability assessment using different data types.

CHAPTER NINE: A PRACTICAL APPROACH FOR THE ESTIMATION OF BRIDGE DISTRIBUTION FACTOR AND LOAD RATING FACTOR

9.1 Introduction

Bridge load rating is one of the main tools for the condition assessment. Resulting in load rating factor (RF), load-testing aims at determining operational and extreme traffic load bearing capacity/level of bridges. During load rating, the estimation of the distribution of live loads on bridges is a critical step. (AASHTO 2014) provides a standard procedure to calculate the distribution factor (DF) and the rating factor. By conducting rapid experimental test on a reinforced concrete bridge population, (Catbas et al. 2012a) proposed a method to determine the moment DFs for single-span-T-beam bridges. They depicted that the new approach relatively predicted live load increases when compared to standard girder analysis given in the (AASHTO 2014) code. Based on the load and resistant factor rating (LRFR) approach, load rating factors according to three methods (standard, experimental strain data and FEM) were obtained for a fully instrumented bridge. Standard approach was resulted to approximate lower rating factor than the other ones (Sanayei et al. 2016). The standard approach for load rating gives conservative results. Conducting experiment for load rating test costs large amount of time and labor force. FEM based methods need experimental data to update the modal.

In this study, a practical approach for the estimation of bridge live load distribution factor and load rating is proposed. The DF is calculated with the experimental bridge deflection results by using computer vision approach during load testing or normal traffic and then the RF is calculated with

the obtained DF and AASHTO codes. The proposed approach is verified with a case study of a prestressed concrete highway bridge.

9.2 Methodology

Figure 101 shows the schematic diagram of the proposed method for distribution factor estimation. In this example, a prestressed concrete highway bridges with five I shape girders are considered. The bridge has two lanes. Here, a camera (Camera 1) is employed to record the traffic flow of the bridge and the other cameras (Camera 2, Camera 3, Camera 4, ...) are employed to monitor the response of the girders in the same cross section. Depending whether a truck (T1) is crossing the bridge on Lane 1 (L1) or Lane 2 (L2) or two trucks on the two lanes at same time with same location, Figure 101b, c or d is selected to calculate the distribution factor (DF) as:

$$DF_i = \frac{d_i}{\sum_{j=1}^n d_j} \quad (69)$$

where DF_i is the distribution factor of the i th girder, n is the total number of bridge girders and d_i is the displacement of the i th girder. Here, the displacement of each girder is obtained from cameras. In Figure 101d, it is very rare that two exact same truck are on the same cross section at the same time so that to calculate the distribution factor considering the cases of two trucks presence, linear superposition of the two cases shown in Figure 101b and Figure 101c is used.

After the calculation of the distribution factor, the load rating factor is calculated by using the equation below:

$$RF = \frac{\phi_c \phi_s \phi R - \gamma_{DC} DC - \gamma_{DW} DW \mp \gamma_p P}{\gamma_L (LL + IM)} \quad (70)$$

Where ϕ is the load and resistance factor design (LRFD) resistant factor, ϕ_s is the system factor for redundancy, R is the structure resistance, DC is the dead load, DW is the wearing surface load, P is the prestress load, LL is the live load, IM is the impact effect, and γ_{DC} , γ_{DW} , γ_p , and γ_L are the factors for different loads. Detailed explanation can be found in (AASHTO 2014).

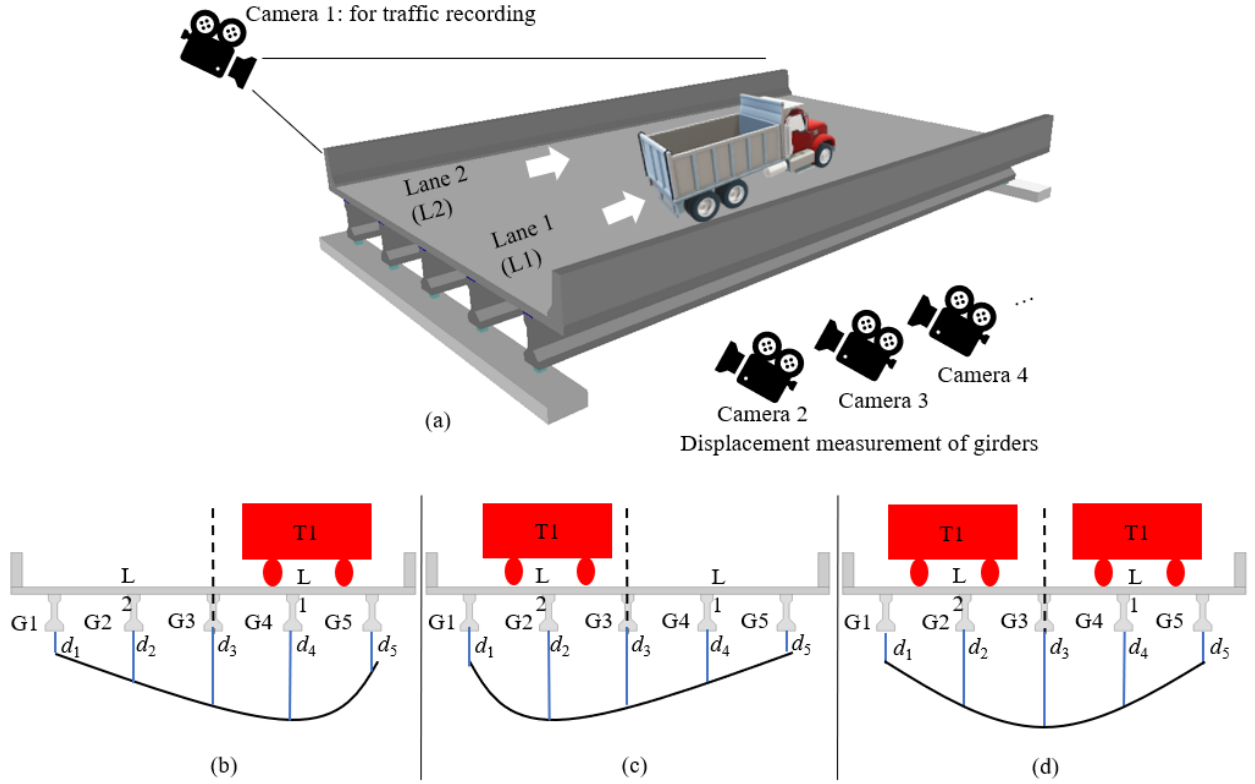


Figure 101 Schematic diagram of the proposed method for distribution factor estimation

9.3 Application on a Real Bridge

9.3.1 General Features of the Bridge

As shown in Figure 102, the bridge in the study is a multi-span prestressed concrete bridge. The bridge was constructed in 1964 and has a total length of 2993 ft (912 m). Each span consists of five pre-stressed I-beam (AASHTO Type II girders). The total length of each span is 52 ft (15.85 m) and the width is 33.08 ft (10.08 m). The girders are spaced at 6.5 ft (1.98 m). The thickness of the slab is 7 inches (17.8 cm). In this experiment, only the first span is considered, and all the experiments were conducted for the first span.



Figure 102 The prestressed concrete highway bridge in this study

9.3.2 Experimental Setup

As shown in Figure 103, two types of sensors were installed on the bridge. Five displacement sensors (i.e., potentiometers) were installed at the mid span of each girder to measure the displacement. Three cameras (Z CAM E1, 4K, 30 FPS, 75-300 mm zoom lens) were employed to measure the displacements at the same location. The first camera recorded the motion of Girder 1 (G1) and Girder 2 (G2), the second camera recorded the motion of Girder 3 (G3) and Girder 4 (G4), and the third camera recorded the motion of Girder 5 (G5). Five strain gauges were installed at the 1/4 span of each girder. One camera (Canon VIXIA HF R42) was employed to record the traffic footage. The sensors and cameras are shown in Figure 104, and all the sensors were installed at the bottom of the girders as shown in Figure 105. Manual markers were attached on the side surface of the girder, and they were regarded as targets for visual tracking when using vision-based displacement measurement methods. Figure 106 shows the overview of the instrumentation.

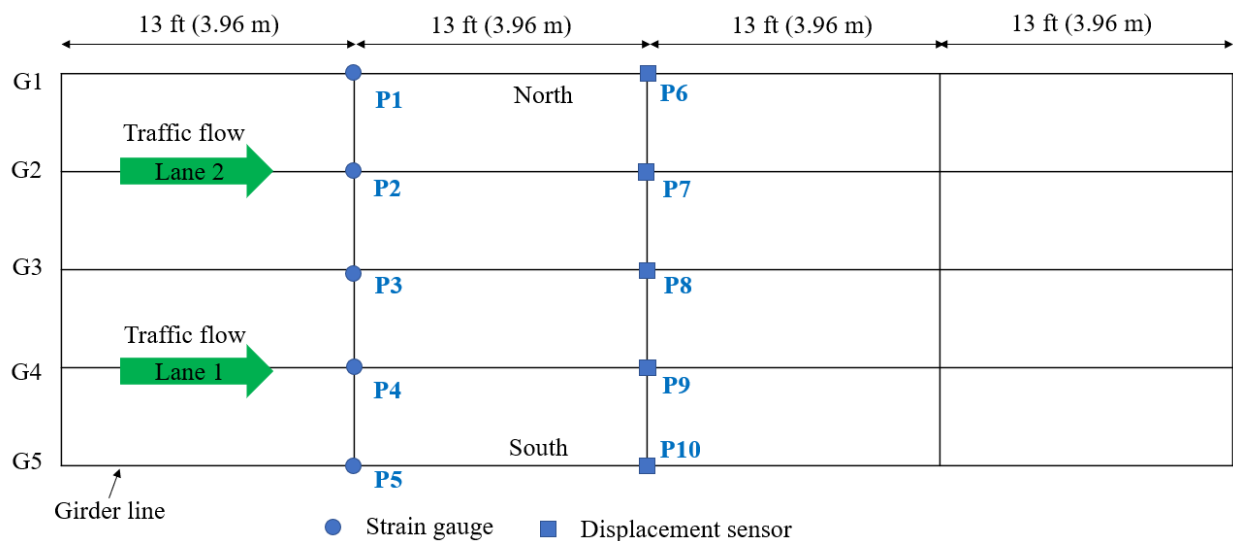


Figure 103 Sensor instrumentation plan



Figure 104 Sensors and cameras used in the test: (a) displacement sensor, (b) strain gauge, (c) camera for traffic monitoring, and (d) camera for displacement measurement



Figure 105 Sensors installed on the bridge

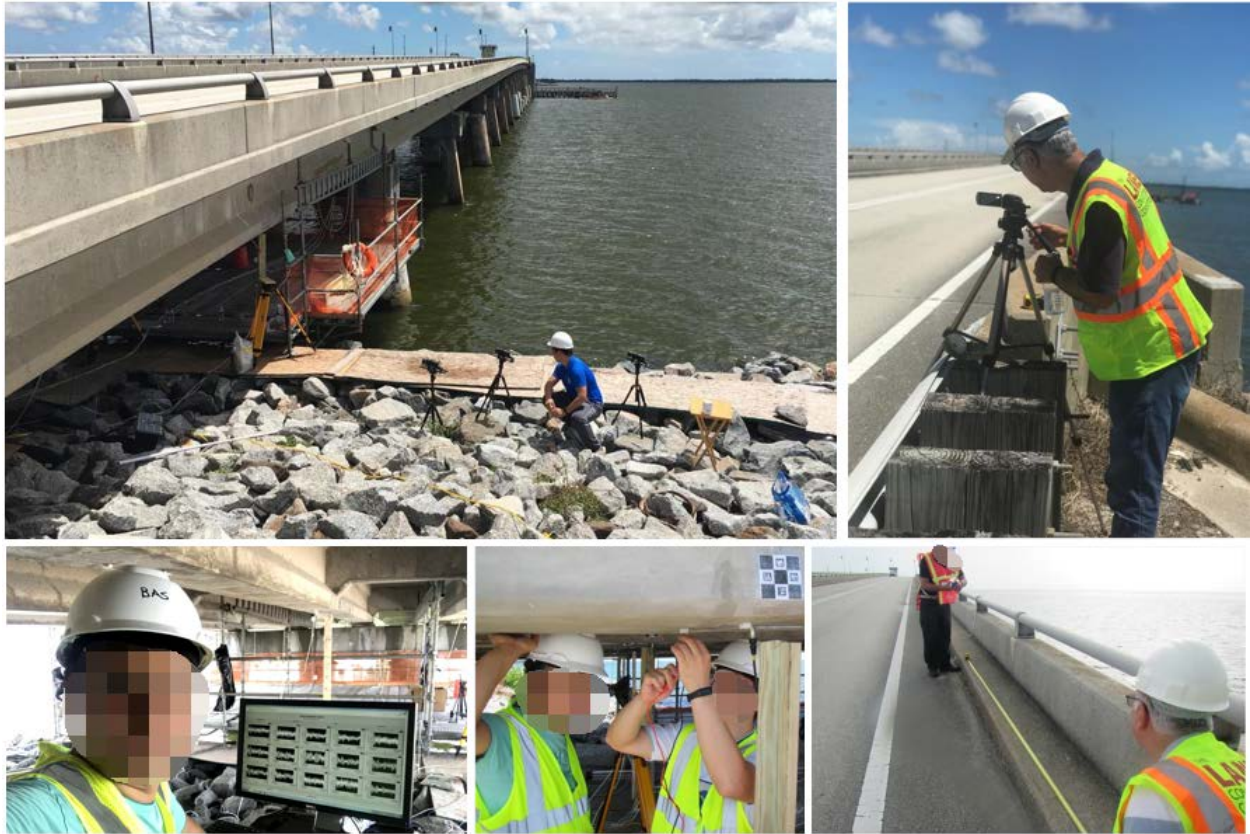


Figure 106 Overview of the instrumentation

Table 18 Experimental cases

Case	Case Name	Lane	Loading Type
1	T1L1	Lane 1 (L1)	Static, four-step
2	T1L2	Lane 2 (L2)	Static, four-step
3	T1L1-35	Lane 1 (L1)	Dynamic, 35 mph
4	T1L2-35	Lane 2 (L2)	Dynamic, 35 mph



Figure 107 Truck loading on the bridge

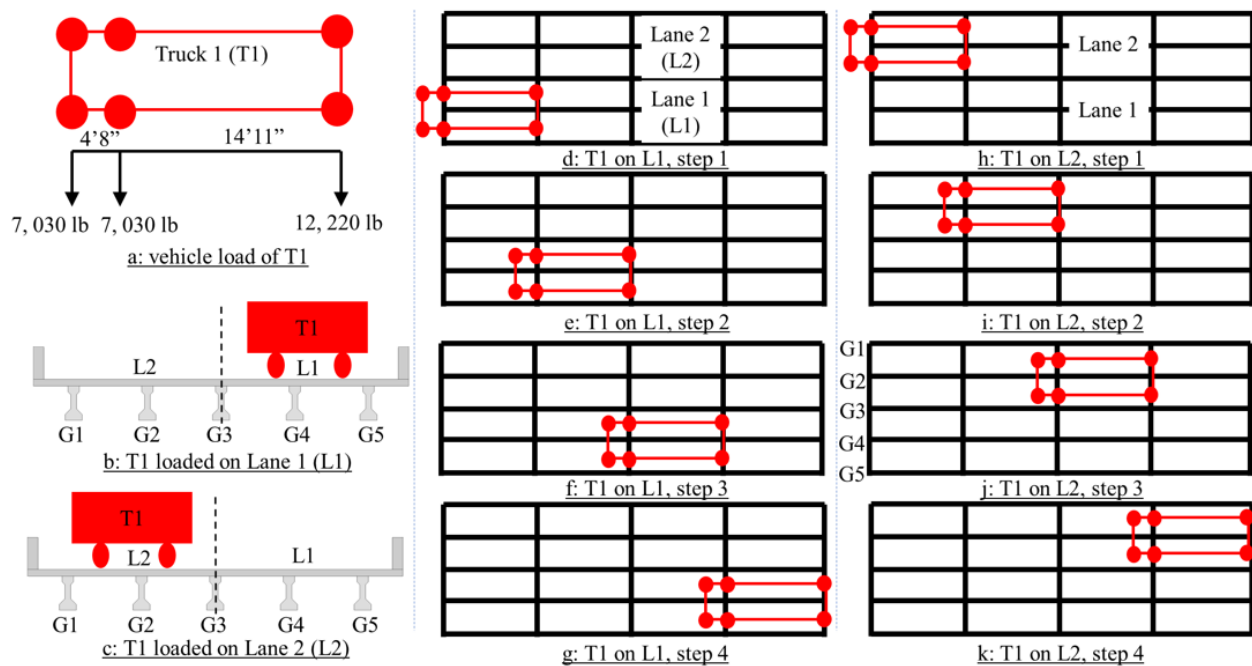


Figure 108 Loading plan of the static test

The experiment consisted of a series of static loading and a series of dynamic loading. The truck employed in this experiment is shown in Figure 107. Table 18 listed the experimental cases. The first two cases are static loading cases and the truck (T1) was loaded at four different locations on

the first span of the bridge with four steps as shown in Figure 108. In each step, the truck stopped for certain time and then slowly moved to the next location. In Case 1 (T1L1) the truck was loaded on Lane 1 (L1) and in Case 2 (T1L2) the truck was loaded on Lane 2. Case 3 and Case 4 are dynamic test. The truck crossed the first span of the bridge with a speed of 35 miles per hour (mph) on Lane 1 and Lane 2, respectively. In the dynamic test, the truck moving load is regarded as the normal traffic load. During the experiment, the sensors and cameras recorded the responses of the first span of the bridge, and the videos from Canon camera were regarded as reference to check the truck loading.

9.3.3 Result analysis

Figure 109 and Figure 110 show the displacement results of the mid-span of Case 1 (T1L1, static) and the strain results of the 1/4 span of Case 1 (T1L1, static), respectively. The displacement data is obtained from camera and displacement sensor (potentiometer). The procedure of vision-based displacement measurement the same with the one presented in Section 2.4.1.1 and the tracking algorithm is normalized correlation coefficient-based template matching (NCCTM). During the experiment, the potentiometers installed at Girder 1 and Girder 5 didn't work. The comparison between the displacement data from cameras and potentiometer are only within Girder 2, Girder 3 and Girder 4. From Figure 109, it can be seen that the differences between cameras and potentiometer are very small. In this study, the distribution factor calculation by displacement method is conducted by using the displacement data from cameras. Since this is a four-step static loading test, the deflections at each step can be extracted to calculate the distribution factor. In this study, the deflections at the second step (marked as green bounding box in Figure 109 and Figure

110) are choose to calculate the distribution factor. The distribution factor using displacement data is calculated by Eq. 69. The distribution factor using displacement data is expressed as

$$DF_i = \frac{\varepsilon_i}{\sum_{j=1}^n \varepsilon_j} \quad (71)$$

where ε_i is the strain of the i th girder. The other parameters are the same with those in Eq. 69. To compare with the standard distribution factor, the equations from AASHTO codes are also used, as listed in Table 19.

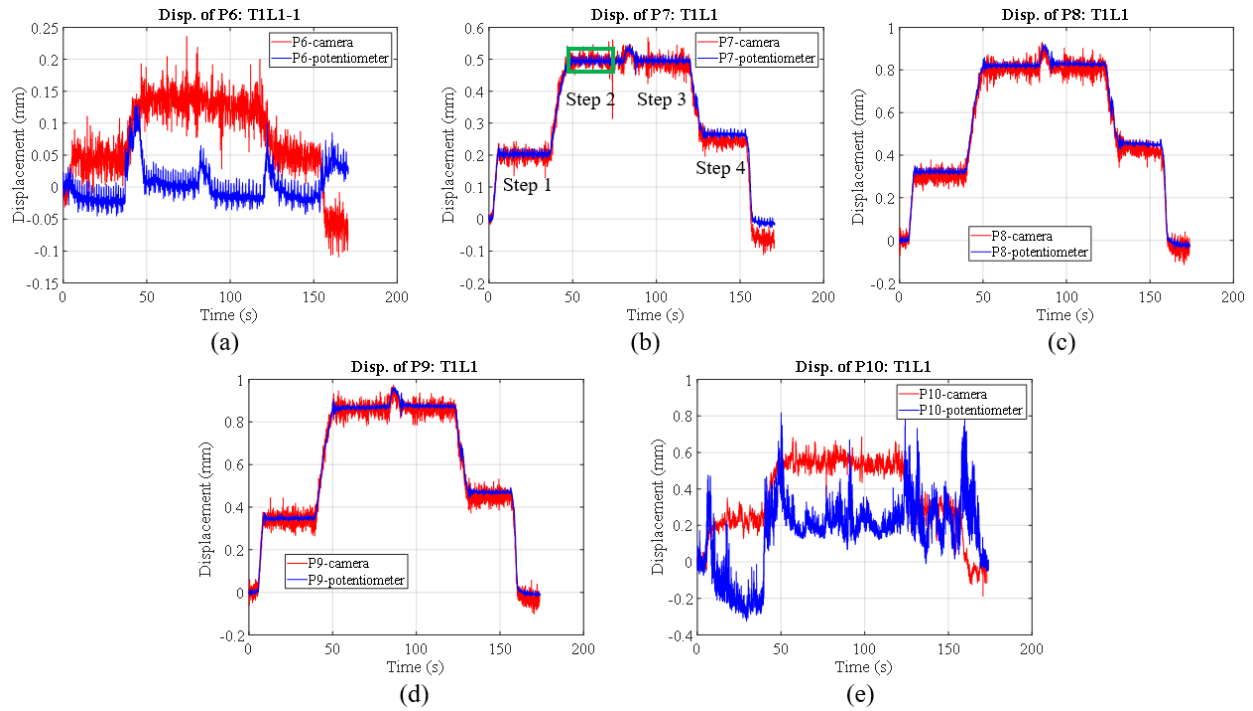


Figure 109 Displacement results of the midspan of Case 1 (T1L1, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5

Table 19 Distribution factor calculation in AASHTO

DF for moment, interior girder	DF for moment, exterior girder
<p>One design lane loaded:</p> $mg_{moment}^{SI} = 0.06 + \left(\frac{S}{14}\right)^{0.4} \left(\frac{S}{L}\right)^{0.3} \left(\frac{K_g}{12Lt_s^3}\right)^{0.1}$ <p>Two or more (multiple) design lanes loaded:</p> $mg_{moment}^{MI} = 0.075 + \left(\frac{S}{9.5}\right)^{0.6} \left(\frac{S}{L}\right)^{0.2} \left(\frac{K_g}{12Lt_s^3}\right)^{0.1}$ <p>AASHTO 4.6.2.2.2b-1</p>	<p>One design lane loaded:</p> $mg_{moment}^{SE} = \frac{5.5}{S}$ <p>Two or more (multiple) design lanes loaded:</p> $mg_{moment}^{ME} = e(mg_{moment}^{MI})$ $e = 0.77 + \frac{d_e}{9.1} \geq 1.0$ <p>d_e is positive if girder is inside of barrier, otherwise negative</p> <p>AASHTO 4.6.2.2.2.1d-1</p>
<p>S = girder spacing (ft); L = span length (ft); t_s = slab thickness (in.); $K_g = n(I_g + e_g^2 A)$; n = modular ratio of girder and deck; I_g = moment of inertia of girder (in.⁴), e_g = girder eccentricity which is the distance from girder centroid to middle centroid of slab, (in.); A = girder area (in.²)</p>	

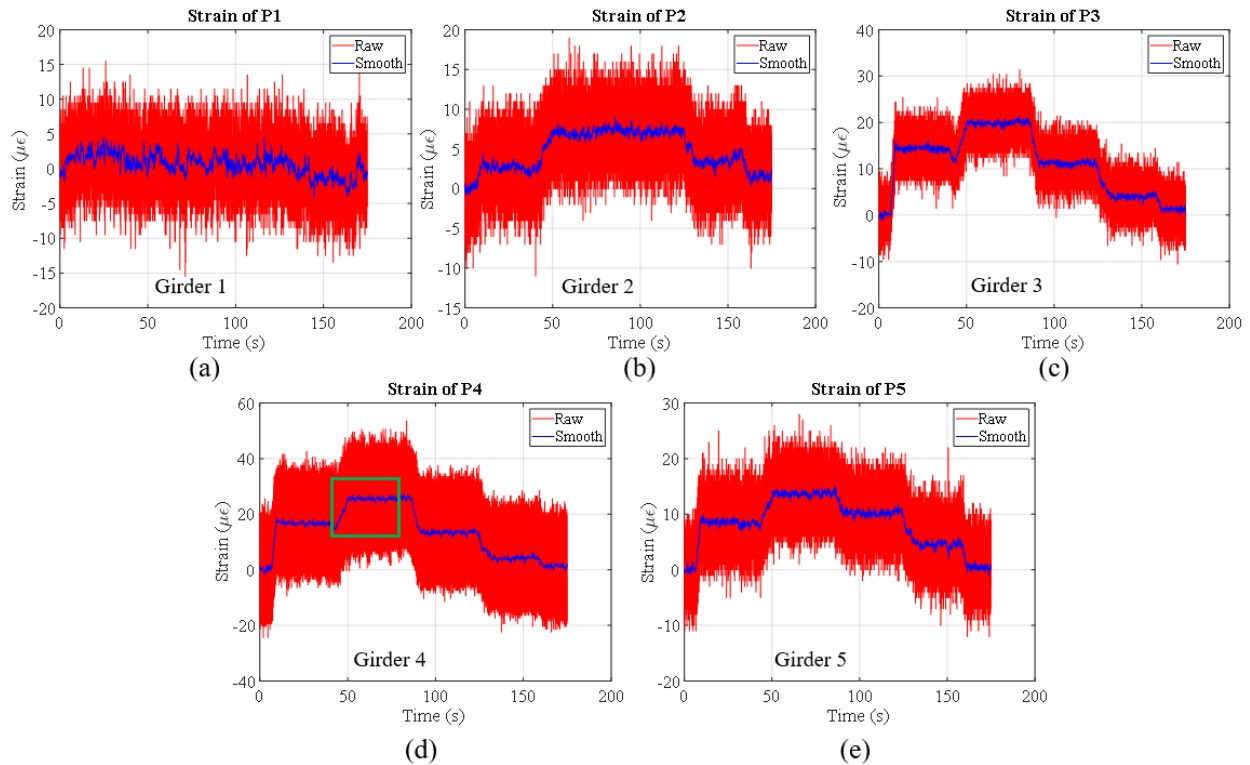


Figure 110 Strain results of the 1/4 span of Case 1 (T1L1, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5

Figure 111 and Figure 112 show the displacement results of the mid-span of Case 2 (T1L2, static) and the strain results of the 1/4 span of Case 2 (T1L2, static). The procedure of distribution factor calculation is the same with Case 1.

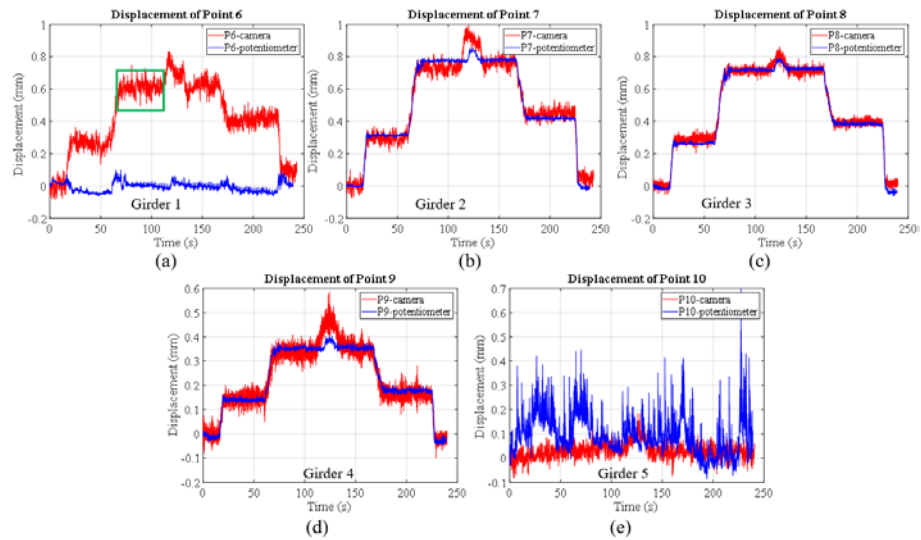


Figure 111 Displacement results of the midspan of Case 2 (T1L2, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5

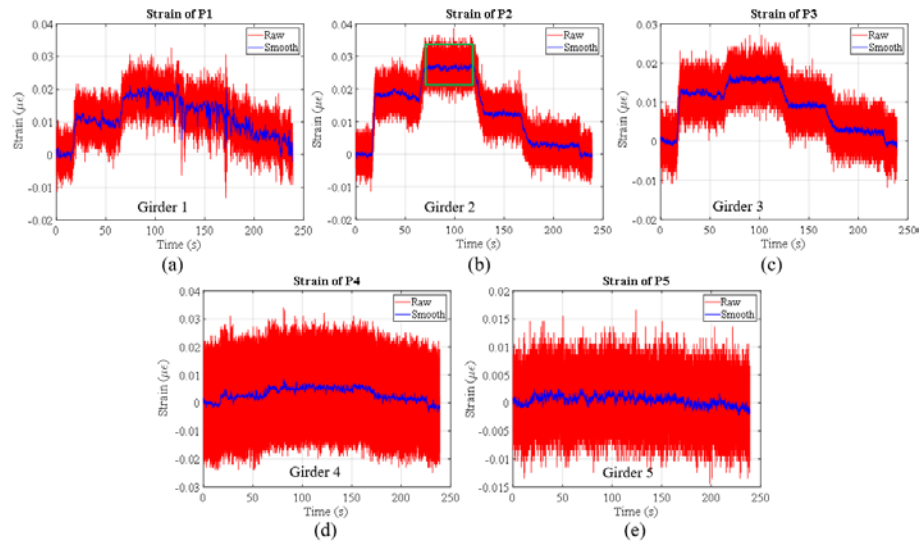


Figure 112 Strain results of the 1/4 span of Case 2 (T1L2, static): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5

Figure 113 shows the distribution factors calculated from the static load cases. Figure 113a and b are calculated from the cases when only one lane is loaded by a truck. By using superposition method, the distribution factors when two trucks are loaded on two lanes in the same cross section at the same time are obtained, as shown in Figure 113c. Figure 113c also refers to the case in AASHTO code “Two or more (multiple) design lanes” which is mentioned in Table 19. Table 20 lists the distribution factors of the static cases. Here “disp” represents the value calculated by displacement data, “str” represents the value calculated by strain data, and the “m” in the item “DF-disp-m” means “Two or more (multiple) design lanes”. From Figure 113 and Table 19 it can be seen that when using the AASHTO codes to calculate the distribution factor, it gives more conservative results and distributed more live load on each girder, especially the girders away from the girder in the center line (Girder 3). The distribution factor calculated from strain data and displacement data (camera) shows very consistent trend and gives close results.

With the extracted distribution factor in Figure 113 and Table 19, the bridge load rating factors can be calculated by following the AASHTO standard (Eq. 70). In the calculation of rating factor, the truck is HL93. Figure 114 and Table 21 shows the load rating factors extracted from the static load cases. The rating factors using the distribution factors calculated from AASHTO codes underestimate the structural load carrying capacity compared to those using strain data and displacement data (camera). And the differences between the results from strain gauges and cameras are very small. It proves that it is feasible to use cameras and vision-based displacement measurement method to estimate the distribution factors and perform load rating.

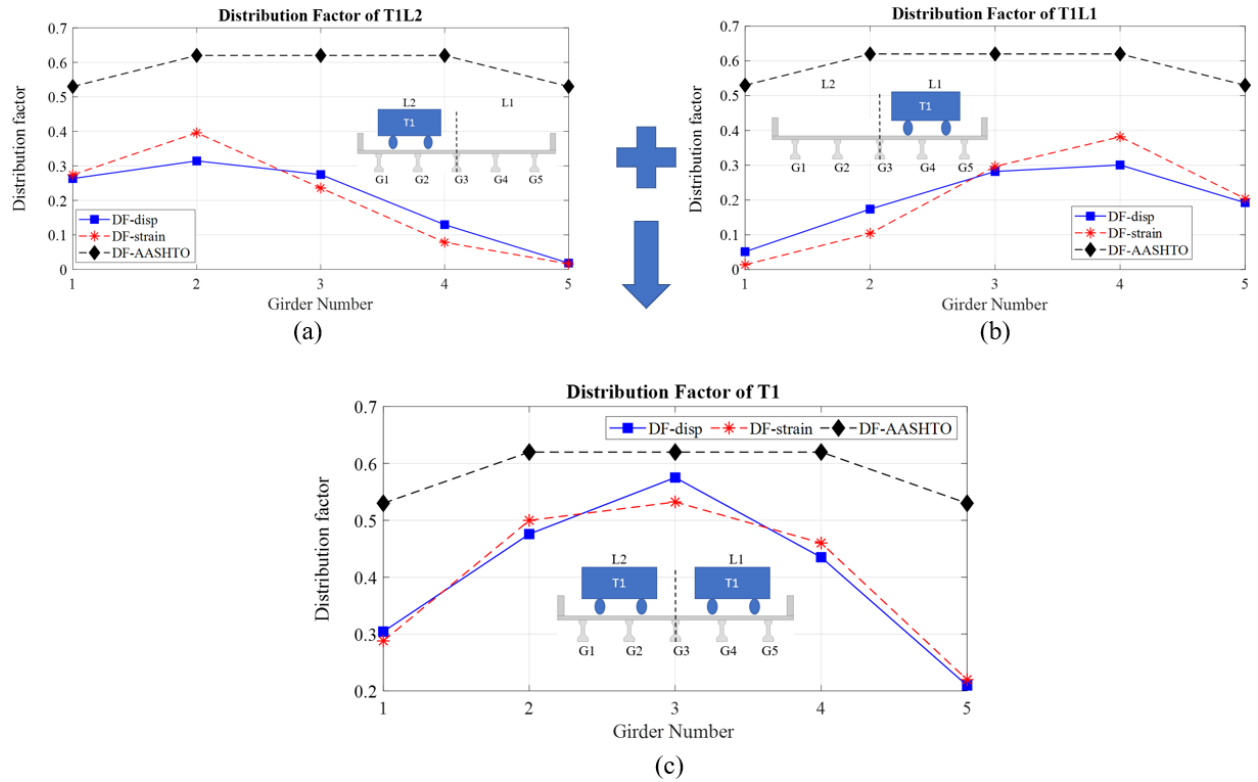


Figure 113 Distribution factor calculated from the static test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (e) two same trucks in both lanes

Table 20 Distribution factor calculated from the static test cases

Girder	DF-disp-T1L1	DF-disp-T1L2	DF-str-T1L1	DF-str-T1L2	DF-disp-m	DF-str-m	DF-AASHTO
1	0.051	0.253	0.014	0.274	0.304	0.288	0.530
2	0.174	0.302	0.104	0.396	0.476	0.500	0.620
3	0.282	0.293	0.297	0.236	0.575	0.532	0.620
4	0.301	0.134	0.382	0.078	0.435	0.460	0.620
5	0.192	0.018	0.204	0.016	0.210	0.220	0.530

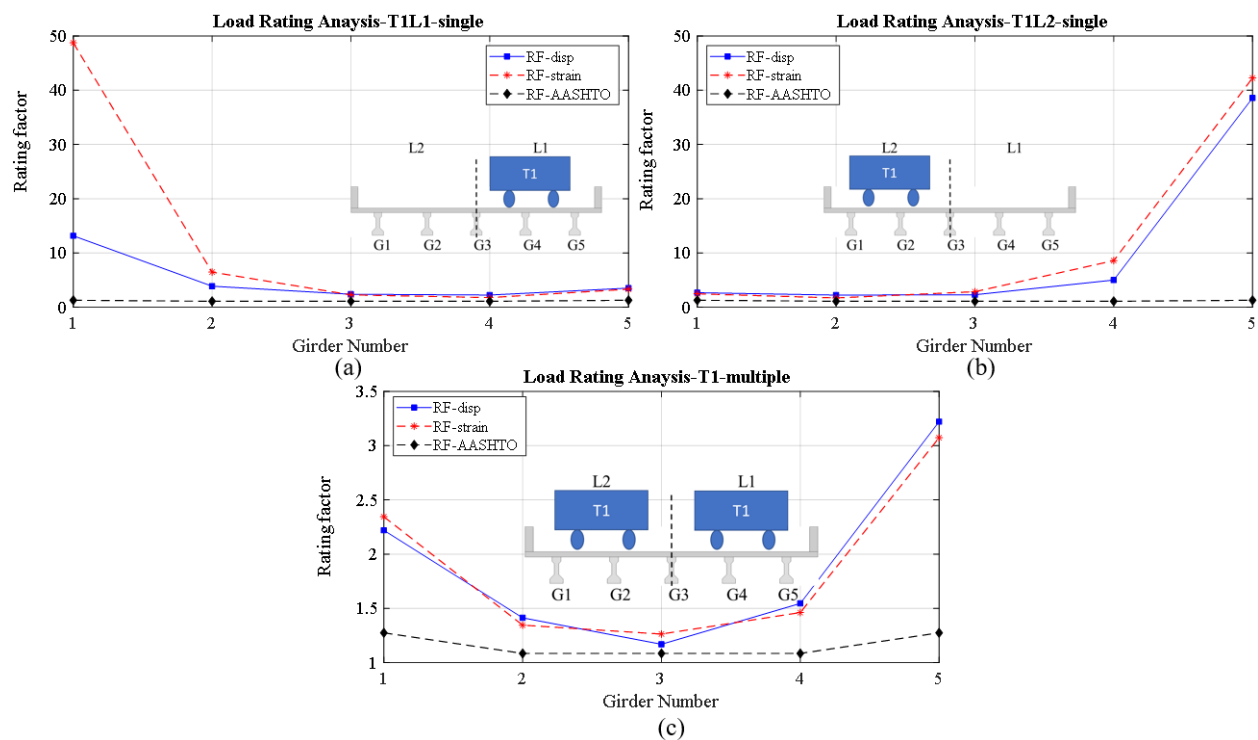


Figure 114 Load rating factor calculated from the static test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (c) two same trucks in both lanes

Table 21 Load rating factor calculated from the static test cases

Girder	RF-disp-T1L1	RF-disp-T1L2	RF-str-T1L1	RF-str-T1L2	RF-disp-m	RF-str-m	RF-AASHTO
1	13.16	2.67	48.74	2.46	2.22	2.35	1.27
2	3.87	2.23	6.46	1.70	1.41	1.35	1.08
3	2.38	2.29	2.27	2.85	1.17	1.26	1.08
4	2.24	5.00	1.76	8.58	1.55	1.46	1.08
5	3.52	38.58	3.31	42.26	3.22	3.07	1.27

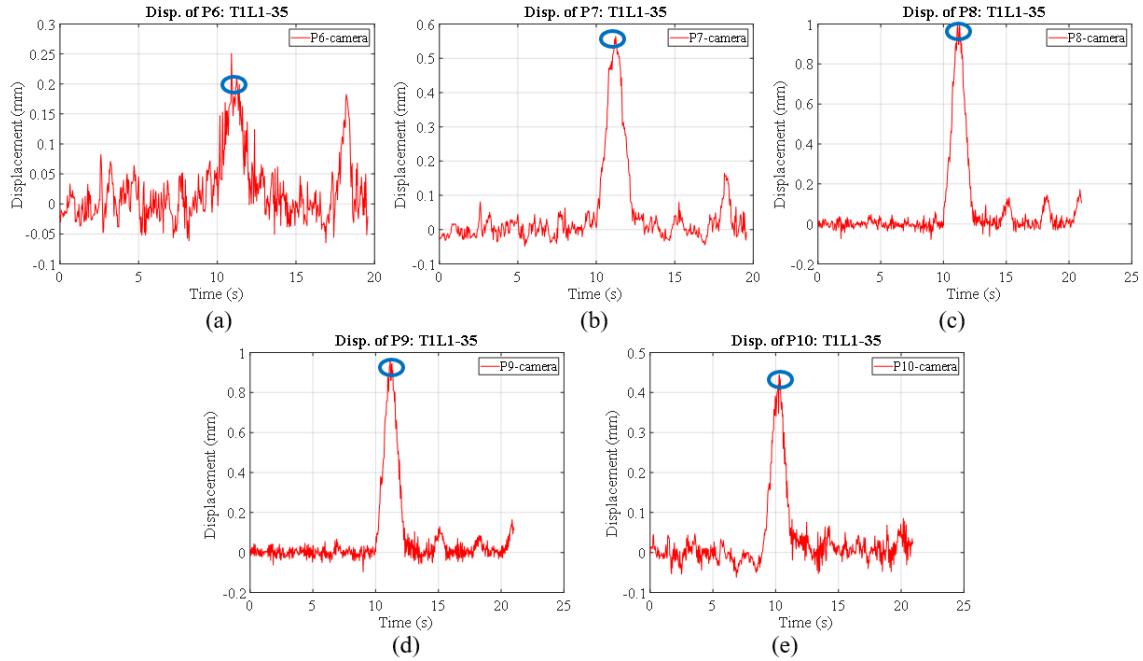


Figure 115 Displacement results of the midspan of Case 3 (T1L1-35, dynamic): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5

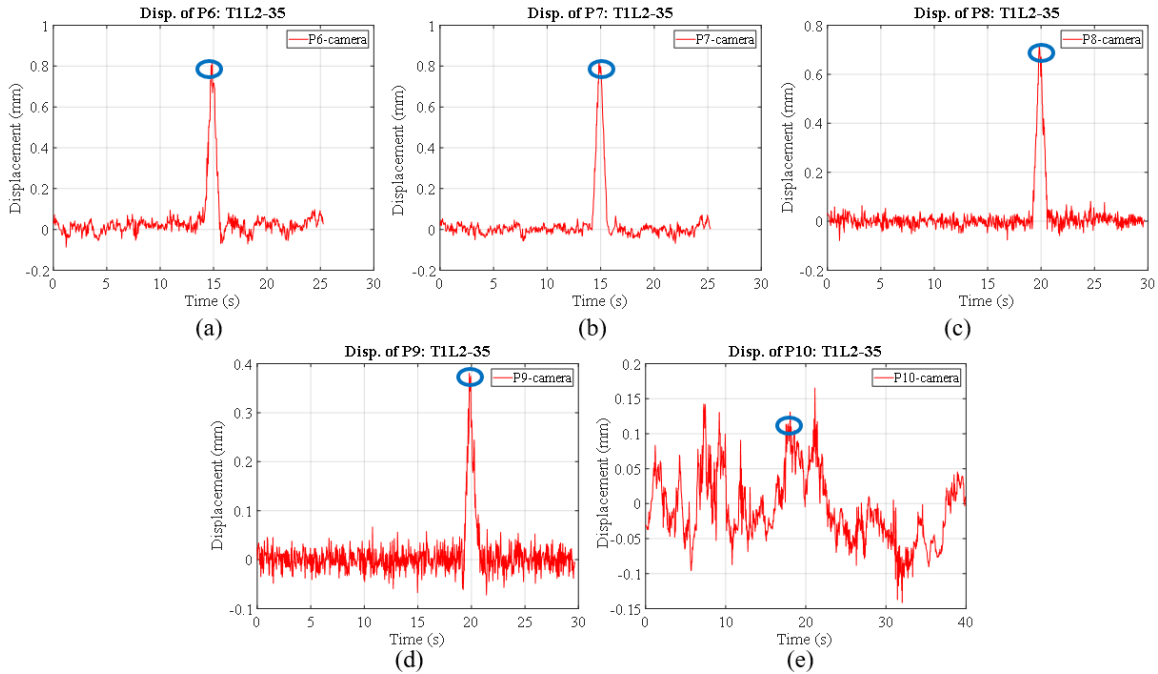


Figure 116 Displacement results of the midspan of Case 4 (T1L2-35, dynamic): (a) Girder 1, (b) Girder 2, (c) Girder 3, (d) Girder 4 and (e) Girder 5

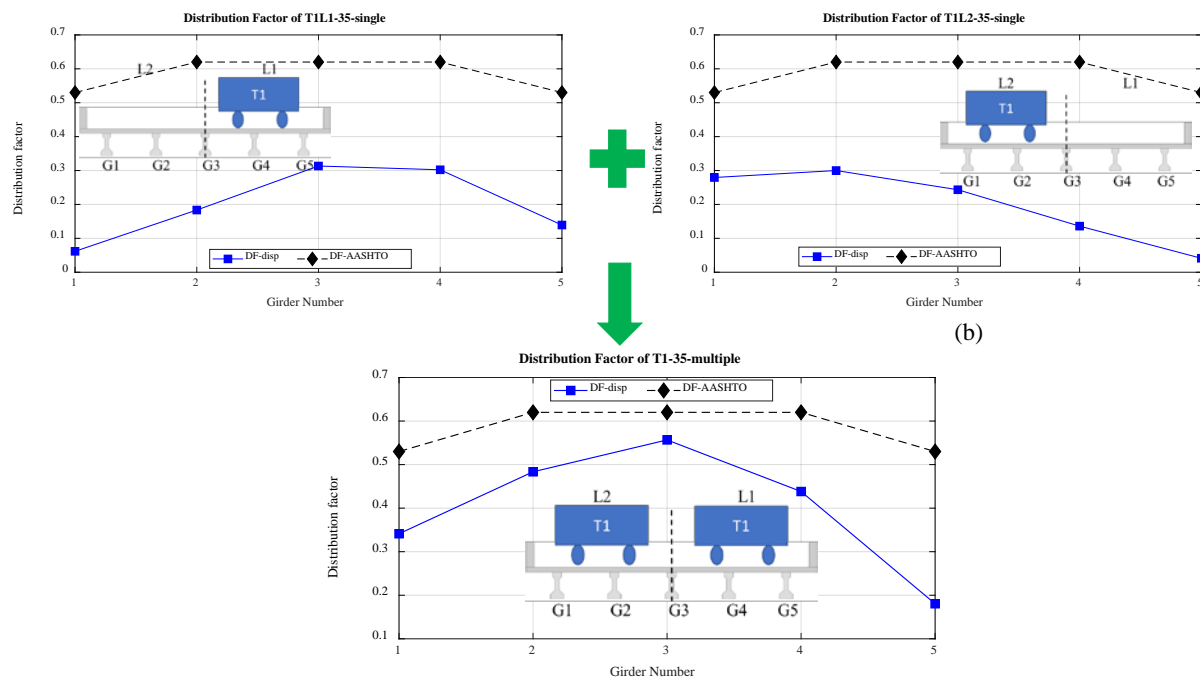


Figure 117 Distribution factor calculated from the dynamic test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (c) two same trucks in both lanes

Table 22 Distribution factor calculated from the dynamic test cases

Girder	DF-disp-T1L1	DF-disp-T1L2	DF-disp-m	DF-AASHTO
1	0.062	0.279	0.341	0.530
2	0.183	0.300	0.483	0.620
3	0.313	0.243	0.557	0.620
4	0.302	0.136	0.438	0.620
5	0.139	0.041	0.180	0.530

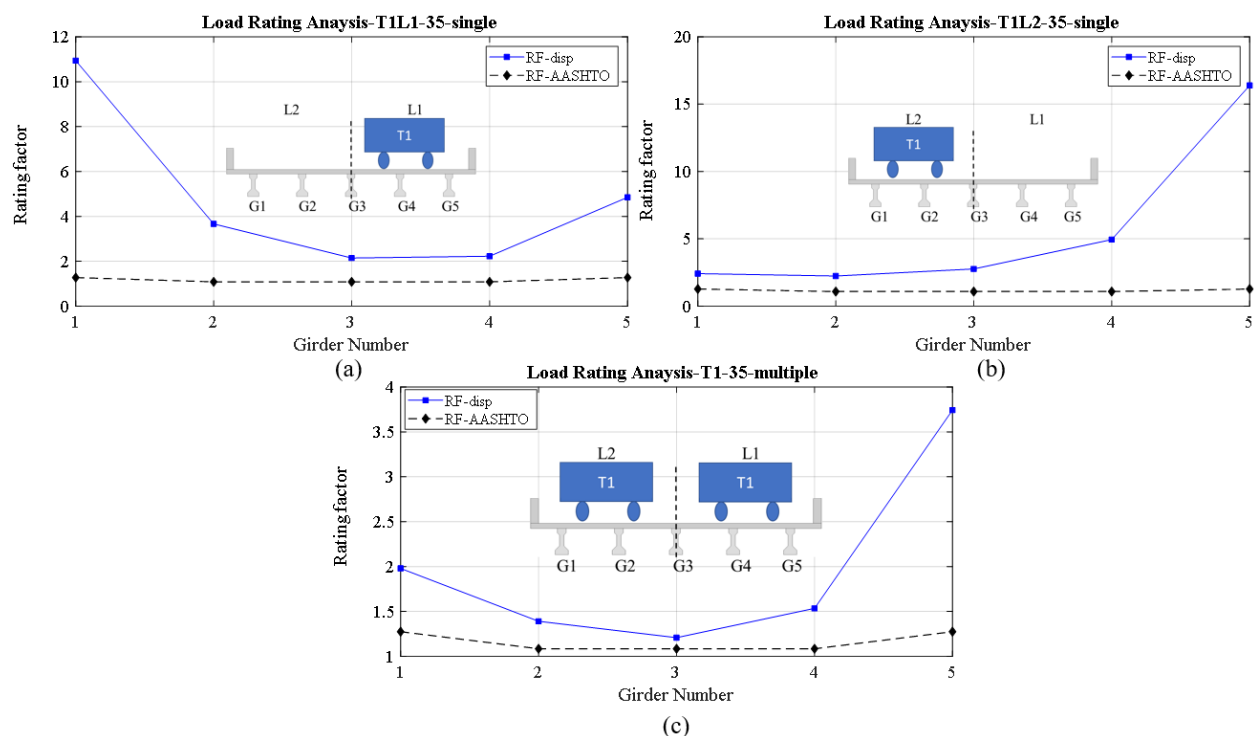


Figure 118 Load rating factor calculated from the dynamic test cases: (a) truck in Lane 2, (b) truck in Lane 1, and (c) two same trucks in both lanes

Table 23 Load rating factor calculated from the dynamic test cases

Girder	RF-disp-T1L1	RF-disp-T1L2	RF-disp-m	RF-AASHTO
1	10.94	2.42	1.98	1.27
2	3.66	2.24	1.39	1.08
3	2.15	2.76	1.21	1.08
4	2.22	4.94	1.53	1.08
5	4.85	16.39	3.74	1.27

Figure 115 and Figure 116 shows the displacement data of the midspan of the dynamic cases. By picking the maximum values of the displacement time histories, the distribution factors can also

be calculated as shown in Figure 117 and Table 22. The distribution factors obtained from the dynamic test are very similar to the static test. And the distribution factors calculated from AASHTO codes are more conservative in comparison with the one calculated from the displacement data obtained by cameras. In the dynamic test, the strain gauges were not used. Figure 118 and Table 23 show the load rating factors calculated with the extracted distribution factors. It is indicated that the load rating factor calculated by using the distribution factors from AASHTO codes underestimates the bridge carrying capacity.

9.4 Summary

In this chapter, a practical approach for the estimation of the bridge distribution factor and load rating factor is proposed. The distribution factor is calculated by using the displacement data extracted from cameras, and it is very consistent with the value calculated from the conventional sensors such as strain gauges in the static and dynamic tests on a prestressed concrete bridge. The results of the static and dynamic tests also show that the distribution factors calculated from AASHTO codes are more conservative than that calculated by experimental data (displacement data from cameras and strain data from strain gauges). The load rating factors are calculated by following the AASHTO codes and using the distribution factors from experimental data and AASHTO codes. The load rating results by using the proposed method and strain data are very consistent and the load rating factors calculated by using AASHTO codes underestimate the bridge load carrying capacity. Comparing to the conventional testing method using strain gauges, the proposed approach using cameras and computer vision to estimate distribution factor and load rating factor is more convenient and does not need too much cost, time and labor force to do sensor

instrumentation and cable wiring work. It is very promising in the future to apply the proposed approach for bridge load rating.

CHAPTER TEN: A COMPLETELY NON-CONTACT RECOGNITION SYSTEM FOR BRIDGE UNIT INFLUENCE LINE USING PORTABLE CAMERAS AND COMPUTER VISION

Previously published as Dong, C. Z., Bas, S., and Catbas, F. N. (2019a). “A completely non-contact recognition system for bridge unit influence line using portable cameras and computer vision.” *Smart Structures and Systems*, 24(5), 617-630.

10.1 Introduction

Bridge structures are important components of the transportation systems, and it is important to keep them in safe working condition to ensure the normal operation of the transportation network. With daily traffic and other external effects, bridges are undergoing with structural changes, deterioration and damages over time. Currently, human visual inspection is still a common approach to detect defects and most of the decisions are made by inspectors’ experiences (Catbas et al. 2017). For safe operation, timely maintenance and convenient management in aspect of structural problems, effective sensing technologies and analytical approaches are necessary to detect the structural changes and damages and give reliable condition assessment and performance evaluation timely and sufficiently (Zaurin et al. 2015). To achieve this goal, in last two decades structural health monitoring (SHM) has been widely explored and implemented on bridges all over the world. SHM systems can collect massive valuable information including structural input (loads and other external effects) and structural output (responses such as displacement, strain and acceleration) and make diagnosis and prognosis to support the structural safety and decision

making.

With the benefits of interdisciplinary integrations, various advanced sensing technologies such as elasto-magneto-electric (EME) sensor for in-service steel cable forces measurement (Duan et al. 2016), wireless sensors for dynamic monitoring (Celik et al. 2018c), fiber Bragg grating (FBG) sensor for strain monitoring (Ye et al. 2016e), LiDAR scanning for structural condition assessment (Chen et al. 2012b), skin-type sensor for strain measurement (Kong et al. 2018), infrared thermography for automated concrete deck inspection (Catbas et al. 2017) and vision-based bridge monitoring at global level (Catbas et al. 2018), etc. have been employed in current research and practice. Among these technologies, vision-based approaches are gathering increasing attention in the field of SHM (Dong and Catbas 2019; Ye et al. 2016c) due to the advantages such as non-contact, long distance, low cost, time saving and ease of use. Generally, the studies and practices of vision-based monitoring are divided into two aspects: 1) inspection and condition assessment at local level such as crack, spalling (Karaaslan et al. 2018) and delamination detection (Hiasa et al. 2017d) and 2) structural monitoring at global level such as vibration and deflection monitoring (Dong et al. 2015, 2018, 2019d; Xu and Brownjohn 2018; Ye et al. 2015, 2016a; f), cable force monitoring (Feng et al. 2017; Ye et al. 2016b), modal analysis (Chen et al. 2018a; Hoskere et al. 2019; Yang et al. 2017b), load estimation (Celik et al. 2018a), load rating (Catbas et al. 2012b) and load capacity estimation (Lee et al. 2006a) etc. With vision-based inspection at local level, the condition assessment is carried out when damages already appear and are visible and large enough. It is very hard to estimate the tiny deteriorations of structures and give further prediction. Vision-based monitoring at global level mostly collects the structural responses and make evaluation of structural performance and safety based on the time histories such as displacement, acceleration

and strain. However, the research about the identification of operating traffic loads as structural input doesn't receive sufficient attention. Even though weigh-in-motion (WIM) systems are installed as parts of the SHM system on bridges, only the weight of vehicles can be estimated and the position information is hard to obtain. Both vehicle loads and position information on the bridge are quite important to structural identification at global level. If only response data are used for structural identification without knowing input force, the structural change and damage has to be large enough to induce significant change to the output responses. Feature extraction from output responses for damage detection is very difficult.

To comprehensively and sufficiently evaluate the structural performance, assess the condition, predict safety and remaining life, the monitoring of structural input and output are necessary. (Zaurin and Catbas 2010a; b) combined cameras and conventional sensors such as strain gages to extract the strain unit influence line (UIL) and recognized damages using statistical outlier detection from UIL vector sets and also conduct load rating (Catbas et al. 2012b). Their work was validated with laboratory experiments on large-scale bridge model and field application on real life bridges. (Khuc and Catbas 2018) integrated camera and displacement sensors to obtain displacement unit influence surface and proposed a statistical approach to detect bridge damages. Damage cases were simulated by changing the boundary condition and connection of bridge components of a bridge model in laboratory. Both the two examples extracted the static structural properties as damage features and used cameras for input monitoring and conventional sensors for output. There are other studies focusing on input-output data and evaluating structural dynamic properties. (Tian et al. 2018) conducted impact test on a small-scale beam in laboratory using camera to capture the human input and accelerometers to collect the output responses. The impact

test was also validated on a footbridge and modal parameters such as frequency, mode shape and scaling factor were extracted. In another research, (Tian et al. 2019) conducted impact test on a small scale beam with moveable camera to collect the beam outputs and impact hammer to give excitation and record the inputs. The major difference between two studies done by Tian et al. is just to switch the data collection approaches for input-output data sources. The studies above including static and dynamic structural properties estimation were carried out by combining cameras and conventional sensors. The drawbacks of using conventional sensors are traffic closure, setup time and labor force to deal with the cable wiring work. It is not convenient to conduct such experiments, especially for field application. The synchronization between cameras and sensors are also a big challenge.

In this chapter, the study of structural identification using input-output data will further advanced from combining cameras and conventional sensors to a completely non-contact recognition system just using cameras. The input and output data are both obtained from portable cameras and computer vision techniques are employed to process the images and track the structural behaviors. UIL is an effective and sensitive index for monitoring bridge behavior under identified loading conditions and explicit structural feature for efficient structural evaluation and assessment. It is also very intuitive for engineers. The proposed recognition system will take UIL as the target parameter for structural identification and the proposed UIL extraction method can be extended to a fully non-contact damage detection approach.

10.2 Methodology

10.2.1 UIL

UIL as shown in Figure 119 indicates the variation of a response such as moment, force, displacement, strain and acceleration at a given position on a structure due to the imposition of a unit load at any point on the structure (Zaurin and Catbas 2010a). To generate a UIL, a unit load is imposed on the structure and moved on it. The response induced by the load at the selected position on the structure is calculated by structural analysis methods or measured by experimental approaches. The response values are then plotted against with the position of load on the structure to generate the UIL. Mathematically UIL of a selected position is the function of the position of the moving unit load on the structure. The detailed concept and calculations of UIL are discussed in elementary structural analysis courses as basics and here only the procedure of UIL extraction using experimental data is introduced.

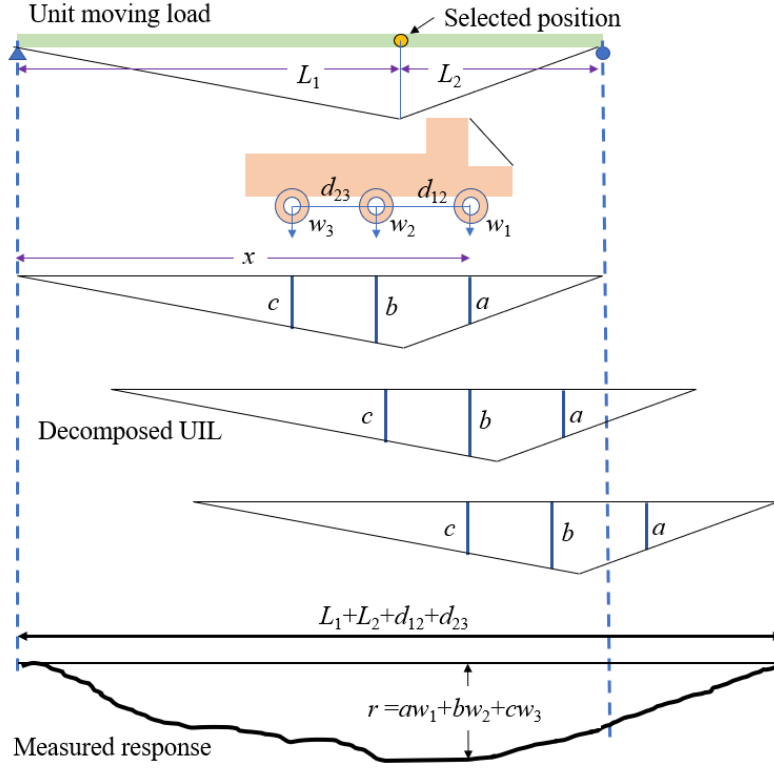


Figure 119 UIL decomposition

When UIL is extracted from experimental data by processing an inverse problem, UIL provides a signature with a normalized structural response for the selected critical locations instrumented by any type of sensors (selected positions in Figure 119). To extract UIL with experimental approaches, the weight and location of each axle of a vehicle crossing the bridge has to be known in advance and responses of the selected position are also measured with sensors. Then the UIL of the structure can be extracted using the following equation (Zaurin and Catbas 2010a):

$$\{r\} = [w]\{u\} \quad (72)$$

where $\{r\}$ is the vector containing the responses of the selected position induced by the moving load, $[w]$ is the matrix containing the axle weights with respect to the corresponding distances, and $\{u\}$ is the UIL vector. Figure 119 gives an example of the extraction of moment UIL. In this

example, a specific position determined by L_1 and L_2 is selected and a moving vehicle is imposed on the bridge. The axle weights of the vehicle are w_1 , w_2 and w_3 . The distance between axles are d_{12} and d_{23} . In this case, one element of Eq. 72 can be expressed as

$$r = aw_1 + bw_2 + cw_3 \quad (73)$$

When knowing any location of the vehicle on the bridge, Eq. (72) can be written as

$$\begin{Bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{Bmatrix}_{m \times 1} = \begin{bmatrix} w_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_2 & 0 & 0 & 0 & w_1 & 0 & 0 & 0 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_3 & 0 & 0 & w_2 & 0 & 0 & 0 & w_1 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & w_3 & 0 & 0 & w_2 & 0 & 0 & 0 & w_1 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_3 \end{bmatrix}_{m \times n} \times \begin{Bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{Bmatrix}_{n \times 1} \quad (74)$$

where n is the moving steps of the vehicle and also the number of discretized coefficients for unit influence along the actual length of the bridge, and m is the number of the samples of the measured responses. The UIL is calculated as an inverse problem using the equation below:

$$\{u\} = [w]^{-1} \{r\} \quad (75)$$

In the study, displacement UIL is targeted and the displacement is regarded as the response that can be measured by vision-based methods. The weight and distance between axles are predesignated and the location of the vehicle is estimated by vehicle tracking from images. In the followings, the vision-based structural input estimation (vehicle location) and vision-based structural output estimation method (displacement responses) are introduced respectively.

10.2.2 Vision-based Structural Input Estimation: Vehicle Location

10.2.2.1 General Procedure of Vehicle Tracking

To identify the vehicle location on bridge surface, in general there are four steps as shown in Figure 120. At first, the camera is calibrated to rectify the distortions such as projective distortion caused by camera pose and radial distortion caused by lenses. Then the object detection algorithms are implemented to detect the category of the vehicles and give the initial bounding boxes of detected vehicles and they will be regarded as the tracking targets. The tracking targets can also be selected manually. In the third step, the visual tracking algorithms are implemented to tracking the detected or selected vehicles and the vehicle location in each frame of the image sequence or video can be estimated. At last, the vehicle location in the image coordinates is transformed to the real-world coordinates to estimate the vehicle location on bridges. The planar homography matrix method is applied for the camera calibration which is introduced in Section 2.2. The other three steps are presented as following.

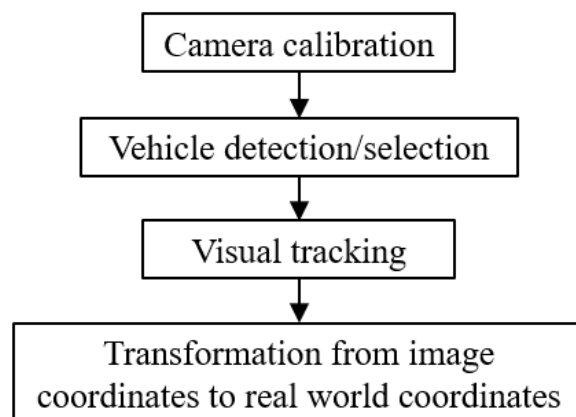


Figure 120 General procedure of vehicle tracking

10.2.2.2 Vehicle Detection/Selection

There are various algorithms for automated vehicle detection. (Zaurin and Catbas 2010a) applied background subtraction to detect the vehicles and made classifications. (Khuc and Catbas 2018) implemented AdaBoost technique and Cascade classifier using histograms of oriented gradients (HOG) features to train and detect vehicle types. With the application of deep learning in computer vision, deep learning-based visual tracking has made great progress. The classical studies in this area are R-CNN (Regions with Convolutional Neural Networks) (Girshick et al. 2012) and its successors such as Fast R-CNN (Girshick 2015), Faster R-CNN (Ren et al. 2017), Mask R-CNN (He et al. 2017), YOLO (You only look once) (Redmon et al. 2015) and SSD (Single shot multibox detector) (Liu et al. 2016a). As stated in Section 10.2.2.1, the vehicle targets can also be selected manually. It all depends on the experimental requirement and application scenarios. If during the time of the experiment there is only one vehicle crossing the bridge, manual selection is good enough to deal with this work. While, if multiple vehicles crossing, pretraining and using deep learning-based vehicle detection algorithms are the more convenient options. In this study, the demonstration is designated for the experiments of UIL extraction and predefined vehicles are selected for the experiments, so that the tracking targets are manually selected from images. Also, using one vehicle to extract UIL faces fewer influencing problems that would happen in multiple vehicle cases. In the real bridge application, automated vehicle detection should be applied to adopt the cases of multiple vehicle crossing bridges.

10.2.2.3 Visual Tracking

Once the vehicle is detected or selected in the first frame of the video or image sequence, visual tracking is necessary to track the location of the vehicle in the successive images. Up to now in the field of computer vision, there are many algorithms for visual tracking and more studies are developed every year (Kristan et al. 2018). However not all the algorithms are suitable for the vehicle localization on bridge for load distribution information extraction. As illustrated in Figure 121a and b, due to the camera angle and view depth, the scale of the vehicle and the view changes from the beginning to the end even the camera is stationary. In addition, since this is a truss bridge, the vehicle is occluded during crossing the bridge. The visual tracking algorithm has to satisfy the requirements of scale invariant and view robustness and can predict target location when occlusion happens.

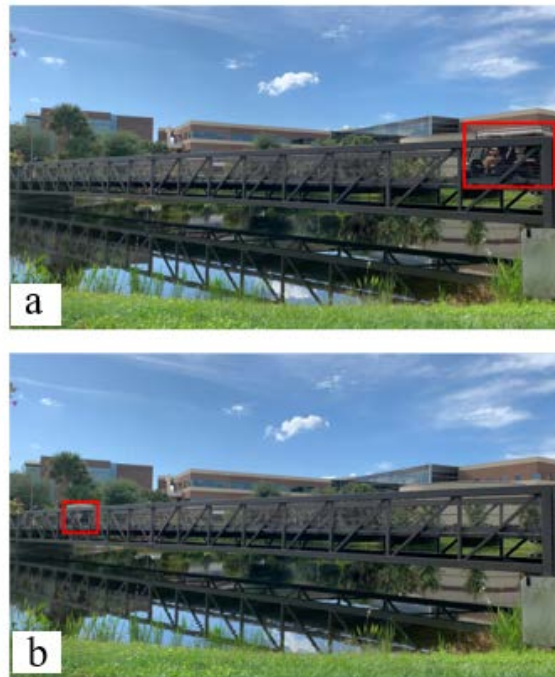


Figure 121 Vehicle tracking example: (a) Status of vehicle at the beginning of a truss bridge; (b) Status of vehicle at the end of a truss bridge

In this research, the Discriminative Correlation Filter tracker with Channel and Spatial Reliability (CSR-DCF, also called CSRT tracker) (Lukezic et al. 2017) is employed to do vehicle visual tracking. CSRT is one of the algorithms using discriminative correlation filter (DCF) which shows great performance. In CSRT, channel and spatial reliability maps are implemented, and a learning progress is applied to update the filter during tracking. This enlarges the search region and improves tracking accuracy of non-rectangular objects. The channel reliability map considers multiple features such as Histogram of Oriented Gradient (HOG), color names and grayscale template to learn and update better filter and spatial reliability map reflect weighting effects in target localization. With the integration of channel and spatial reliability, CSRT achieves state-of-art performance in various popular datasets for visual tracking (Lukezic et al. 2017). CSRT satisfies the requirements aforementioned and it is implemented for vehicle tracking in this study.

10.2.2.4 Coordinate Transformation

After the vehicle location (coordinates) in the image is estimated, it needs to be transformed to the real-world coordinates. In this study, the bridge deck is assumed as a plane so that the question is to transform the vehicle from image plane to the deck plane. As shown in Figure 122, the real-world objects (the bridge and the vehicle) are projected to the image plane. As a result, the shape that is determined by the four points (A, B, C, D) on the real-world plane is distorted due to the projection. The transformation is completed by the planar homography matrix method introduced in Section 2.2.

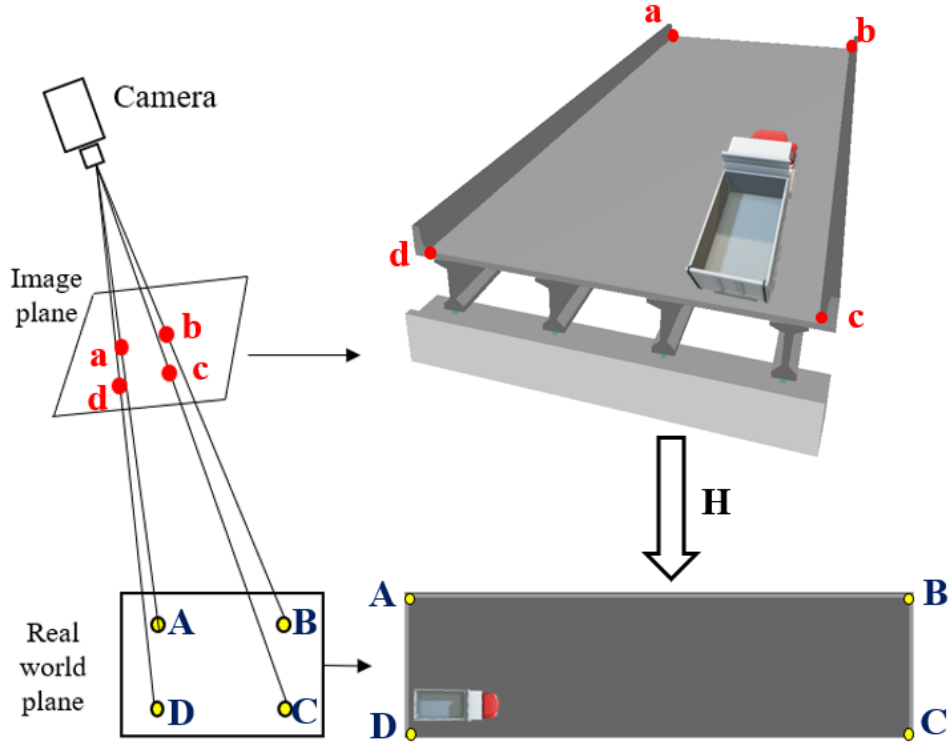


Figure 122 In-plane transformation using Homography matrix

10.2.3 Vision-based Structural Output Estimation: Displacement Responses

The structural output estimation carried out in this study is vision-based displacement measurement. The procedure for the vision-based displacement is similar to that presented in Section 2.4.1.1. While the difference is that here normalized cross-correlation coefficient using edge map (NCCEM) for template matching is implemented.

Unlike vehicle tracking, the monitoring target of vision-based displacement measurement is simpler and limited in a specific region. The view and scale do not change too much. Although the tracking for displacement measurement is much easier than that cases in vehicle tracking, it needs much higher accuracy. In general, tracking result should be in sub-pixel level. In previous studies,

the authors proposed keypoint matching-based methods (Khuc and Catbas 2016, 2017) and optical flow with keypoints methods (Dong et al. 2019b) to achieve the subpixel level results. All of them showed good measurement results comparing with the conventional displacement sensor. While the processing speed is too slow when using keypoint-based methods. NCCEM is an improved version of digital image correlation (DIC) based template matching methods. The most popular DIC based template method is the normalized cross-correlation coefficient (NCC) method (Chen et al. 2018b; Zhong et al. 2019b, 2018a, 2019c; Zhong and Quan 2018b). The NCC coefficient is expressed as:

$$R(x, y) = \frac{\sum_{x', y'} (T'(x', y') \cdot I'(x + x', y + y'))}{\sqrt{\sum_{x', y'} T'^2(x', y') \sum_{x', y'} I'^2(x + x', y + y')}} \quad (76)$$

where

$$T'(x', y') = T(x', y') - \frac{\sum_{x'', y''} T(x'', y'')}{wh} \quad (77)$$

$$I'(x + x', y + y') = I(x + x', y + y') - \frac{\sum_{x'', y''} I(x + x'', y + y'')}{wh} \quad (78)$$

In these formulas above, T is the grayscale image intensity of selected template, I is the grayscale image intensity of the image searching region, (x, y) , (x', y') and (x'', y'') represent the location coordinates in image searching regions and template. The NCC coefficient is normalized with image mean value and standard deviation so that it assures the matching result is not affected by the light changing.

As shown in Figure 123a, when the NCC coefficient achieves the maximum, the matching target is located, i.e., at the peak of the map of NCC coefficient. However, the regular NCC coefficient method is not accurate sometimes. For example, in Figure 123a, there are some pseudo peaks in the map of NCC coefficient, which make distractions of accurate matching. In this chapter, the grayscale image is replaced with edge map before template matching using NCC coefficient.

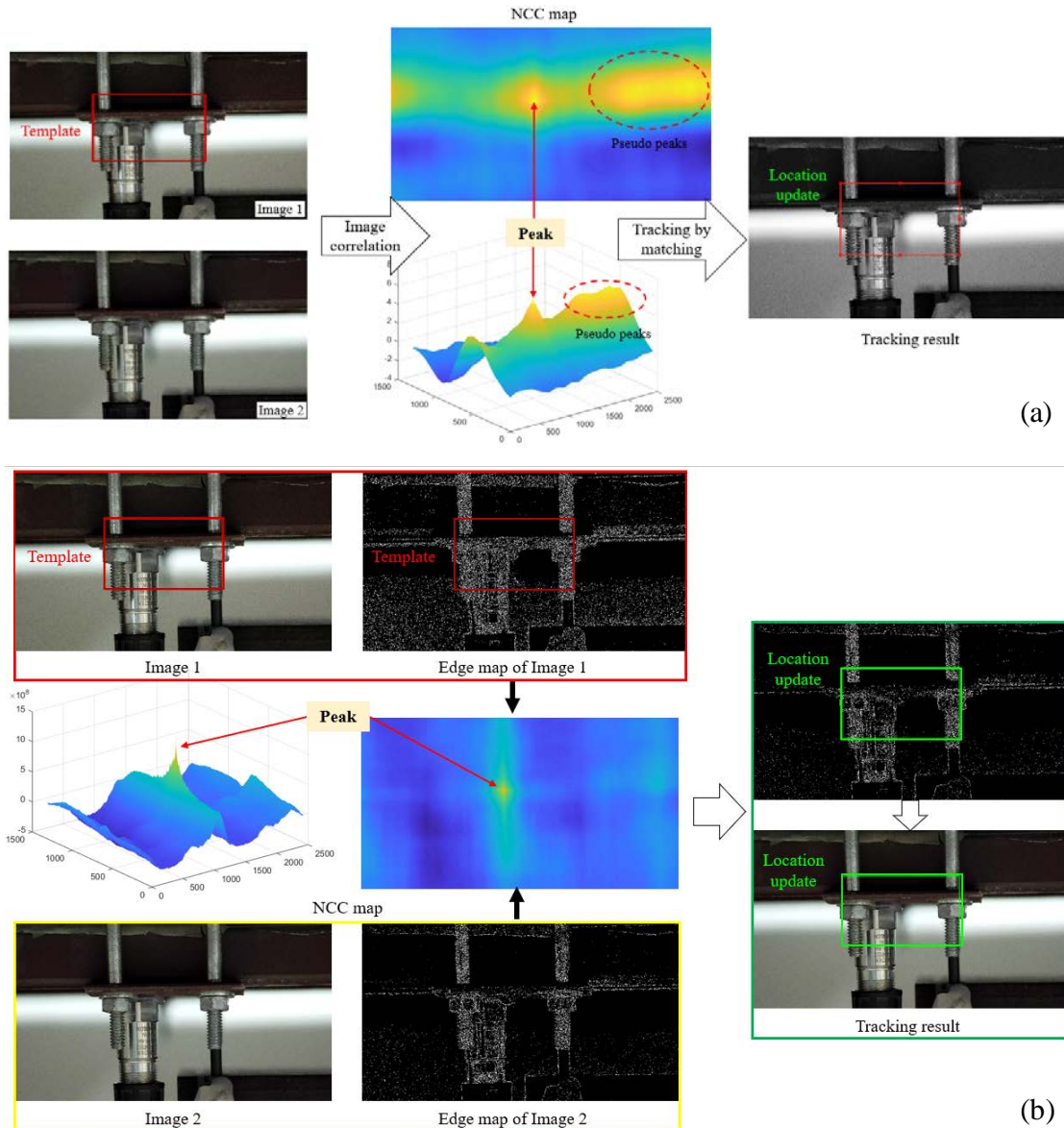


Figure 123 Digital correlation-based template matching: (a) Normalized cross-correlation coefficient of grayscale image; Normalized cross-correlation coefficient of Canny edge map

The edge map is extracted from grayscale image using Canny operator (Canny 1986). Figure 123b shows the NCC coefficient of Canny edge map, which can be seen that the peak in it is very clear and there is no pseudo peak. With the peak in NCC map of Canny edge map, the matching target is first located in the Canny edge map of the image searching region. Then it is updated in the image searching region. The comparison of NCC maps in Figure 123a and Figure 123b indicates that using edge map gives more accurate, reliable and robust results. The NCC methods including regular one and the one using edge map, only give measurement results in pixel level. To achieve the sub-pixel level, a refined searching progress is necessary. In this study, the local pixel upsampling and interpolation operations are applied to do searching refinement. Depends on the required accuracy, a specific iteration number needs to be preset.

After getting the matching of the target template in consecutive images within sub pixel level, the centre of the template is regarded as the tracking location, (x, y) . Assuming the initial location is (x_0, y_0) and with the scale ratio, SR , the displacement of the selected target is $SR \times (x - x_0)$ in horizontal direction and $SR \times (y - y_0)$ in vertical direction.

10.3 System Configuration

The proposed system consists of a set of portable cameras, synchronization modules, a computer and a suite of processing software. Figure 124 shows the system configuration. The portable cameras are divided into two groups, one is for bridge displacement measurement, and the other is for the vehicle tracking. The synchronization modules are applied to synchronize the image

sequences captured from different cameras. All the image sequences are transferred to the computer and processed by the predesignated software. At the end, the UIL is obtained as output.

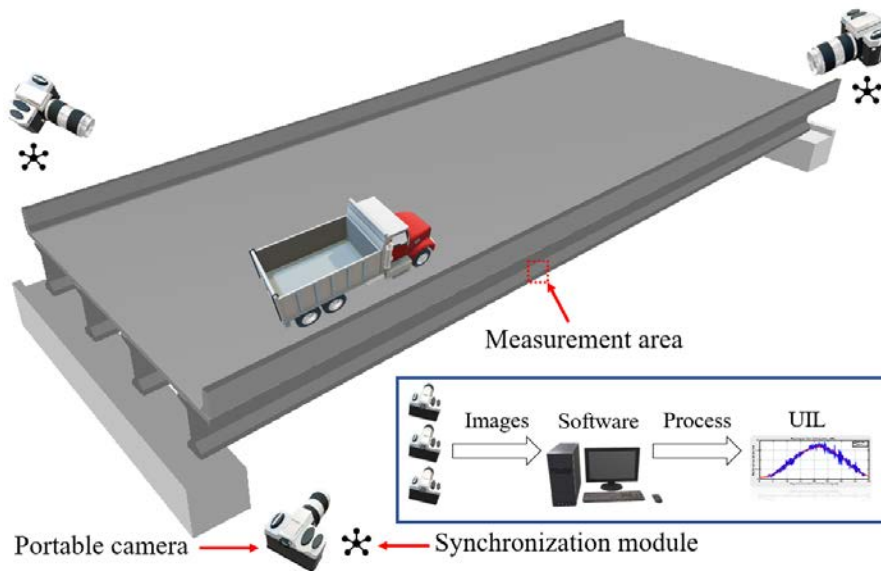


Figure 124 System configuration

10.4 Laboratory Demonstration

10.4.1 Experimental Setup

The proposed system is verified on the two-span bridge model (UCF two-span bridge) constructed in the University of Central Florida's Civil Infrastructure Technologies for Resilience and Safety (CITRS) Experimental Design and Monitoring (EDM) laboratory. As shown in Figure 125, The bridge is a scaled down model of a mid-sized real-life structure and toy trucks with variable weights are used to model moving loads. The bridge consists of two 300-cm main continuous spans. The bridge deck includes a 3.18-mm steel sheet at 120 cm wide, which makes the deck 600

cm long by 120 cm wide.

To view the whole bridge deck and track the vehicle during the whole loading process, a fisheye camera is mounted on the tripod which is 2 m the middle of the bridge. The fisheye camera used here is a Raynic 4K Sports Action Camera with 170-degree wide angle lens. The camera can be connected with a smart phone through the Ez iCam App for remote controlling. This fisheye camera can capture full 1080p (1920×1080 pixels) high-density (HD) video clips at a speed of 30 frames per second (30 FPS). The reason why the fisheye camera is used is that fisheye camera provides a wide angle and can broaden the field of view to assure whole bridge is in the image.

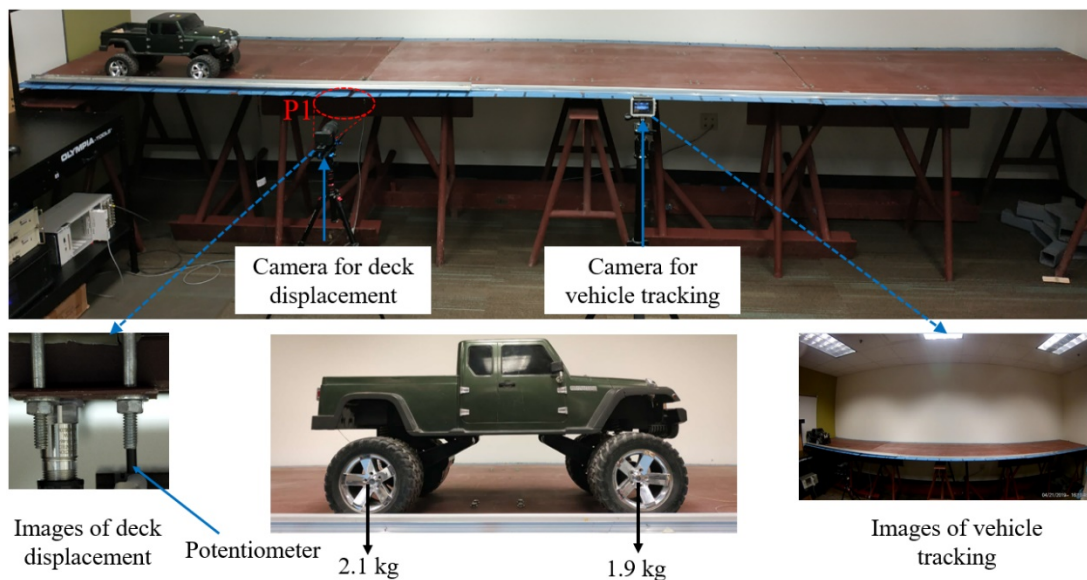


Figure 125 Experimental setup in laboratory

Another portable camera is mounted on the tripod which is close to the midspan of the left span of the bridge to measure the bridge displacement. The distance from the camera to the measurement region, P1, is around 0.8 m. The camera used here is a Z-CAM E1 action camera with a 75-300

mm zoom lens. The camera can also be connected with a smart phone through the Z-CAM official application. The video format set here is 4K (3840×2160 pixels) resolution at a speed of 30 FPS. A potentiometer is mounted under the deck to measure the displacement of P1 and is assumed as the ground truth. The model No. of the potentiometer is BEI Duncan 9615. The sampling rate of the data acquisition system for the potentiometer is 200 Hz, which is then downsampled to 30 Hz during post processing. During the experiments, the toy truck moves from one side of the bridge to the other while the potentiometer and the camera record the motion of P1 (midspan of the left span) synchronously.

As shown in Figure 126, since images captured by the fisheye camera have a severe radical distortion and the straight bridge in the image becomes a curved bridge.



Figure 126 Image from fisheye camera

The fisheye camera has to be calibrated. The calibration procedure follows the steps presented in Section 2.2 and a white black chess board as shown in Figure 127 is employed to complete the calibration. The intrinsic parameters of the fisheye camera are (1) mapping coefficients are [1.03

$\times 10^3$, -2.49×10^{-4} , -4.89×10^{-7} , 3.55×10^{-10}]; (2) distortion center is [951.95, 577.98]; and (3) stretch matrix is [1, 0; 0, 1]. The extrinsic parameters, i.e., \mathbf{H} is [-0.100, -1.982, 1107.9; 0.0136, -1.240, 684.14; 1.86×10^{-5} , -0.002, 1]. Figure 128 shows the rectified image after camera calibration.

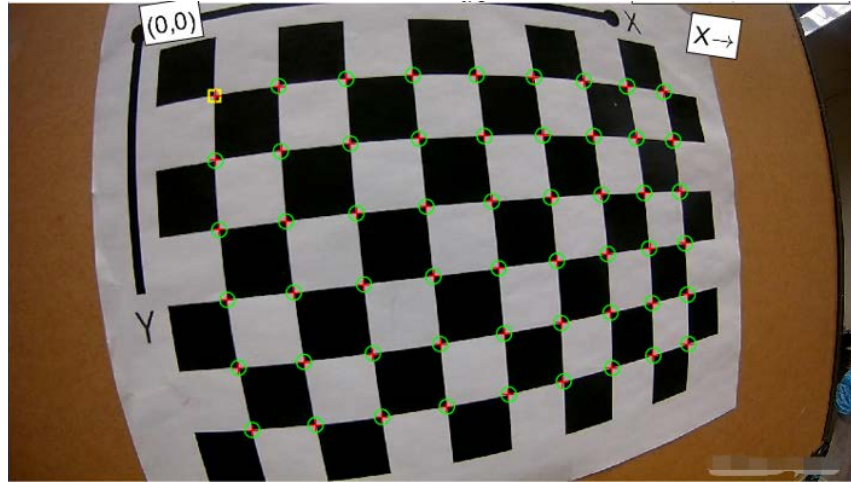


Figure 127 Camera calibration using a black white chessboard



Figure 128 Rectified image after camera calibration

In this experiment, the video recorded by the two different portable cameras are synchronized by using cross-correlation based pattern matching of audio signals. During the video recording, the

portable cameras also record the audio signals and within the same camera, the images and audio signals are synchronized by the internal clock. As shown in Figure 129, the audio signals of the two cameras start at different time as the two cameras started recording with different smart phones. During the recording, a voice, “Start”, is called at the beginning of the experiment and a voice, “Stop”, is called at the end of the experiment. The two audio signals are first realigned and synchronized with one dimensional normalized cross-correlation based pattern matching. The pattern is the signal “Start”. Then the signal synchronization is validated by another signal pattern “Stop”. Finally, the two videos are synchronized with the synchronized audio signals.

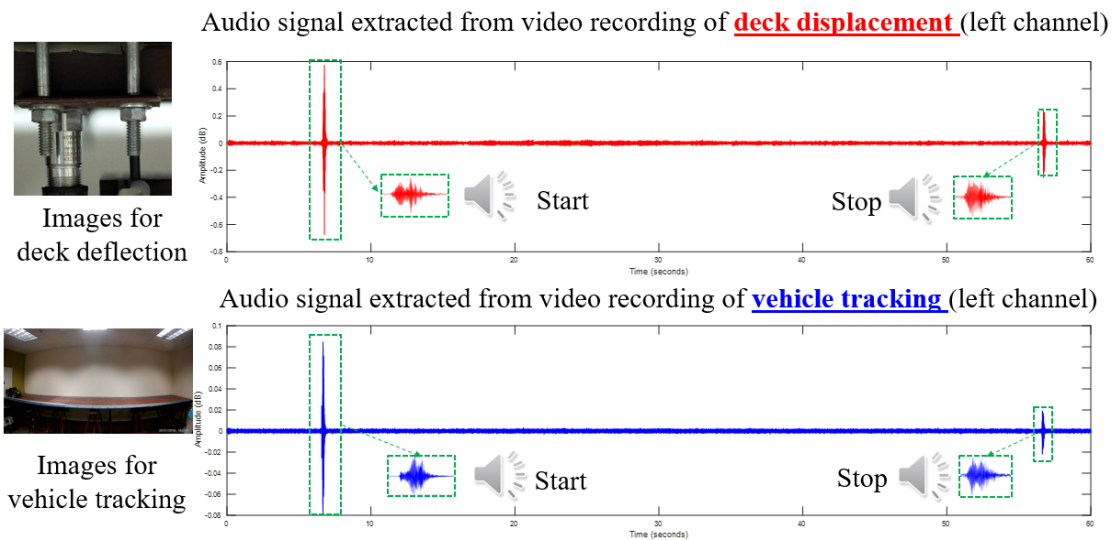


Figure 129 Video time synchronization using normalized cross-correlation based pattern matching of audio signals

10.4.2 Result Analysis

Figure 130 shows the tracking results of the toy truck in the rectified images obtained from the fisheye camera after calibration. During the loading process, although the view and scale of the

truck changes, the CSRT tracker can still successfully estimate the location of the toy truck in each image. And eventually the locations in the rectified images are converted to the location on the bridge deck using homography matrix.

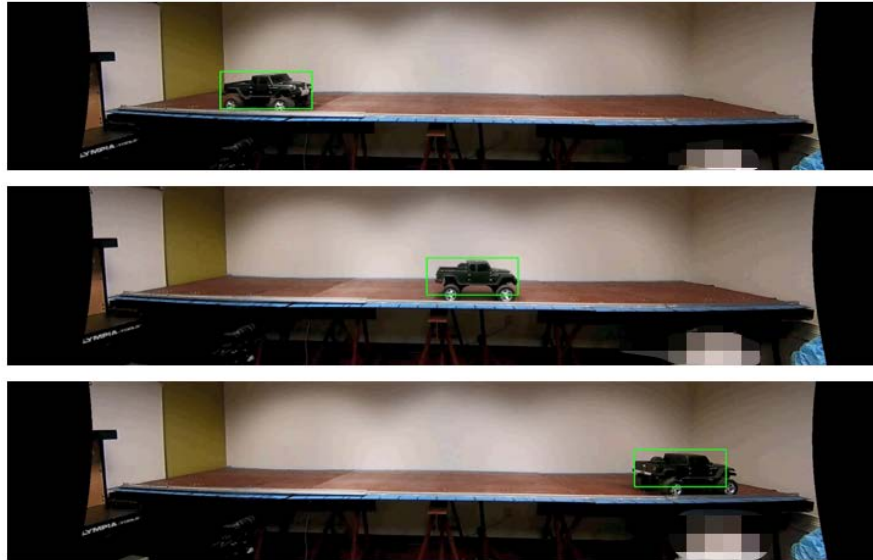


Figure 130 Vehicle tracking in the rectified images from fish camera

Figure 131 shows the displacement comparison between the proposed vision-based method and the potentiometer. The calibration method for the camera used for displacement measurement is scale ratio and in this experiment, it is 0.0316 mm/pixel. From Figure 88, it is easy to see that the result obtained from the proposed method is quite consistent with those obtained from the potentiometer. The normalized cross-correlation (NCC) (Dong et al. 2019c) is calculated to evaluate the similarities between them. The NCC between the two methods is 99.91%, which shows a very high fit of goodness between the test method (the proposed vision-based method) and the ground truth (potentiometer).

Combining the displacement obtained from the vision-based method and the vehicle location information estimated using vehicle tracking, the UIL is built. In Figure 132, the blue curve (UIL-raw) is the extracted UIL without any post-processing and filtering. As this bridge displacement is the response under the moving load, it also includes the high vibration modes in the response signal. By applying the Fourier filter, the high vibration modes are removed and the final UIL is shown the red curve (UIL-Fourier). The maximum value of the UIL is 0.16 mm/kg and minimum value is -0.047 mm/kg. Here the downward direction of deck motion is the positive direction and it means the displacement has a positive sign. The negative portion of UIL is obtained when the truck is located on the other span next to the one has the measurement point.

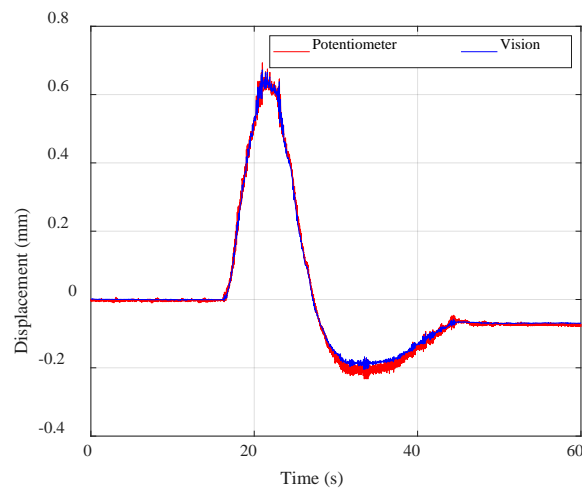


Figure 131 Displacement comparison between the proposed vision-based method and the potentiometer

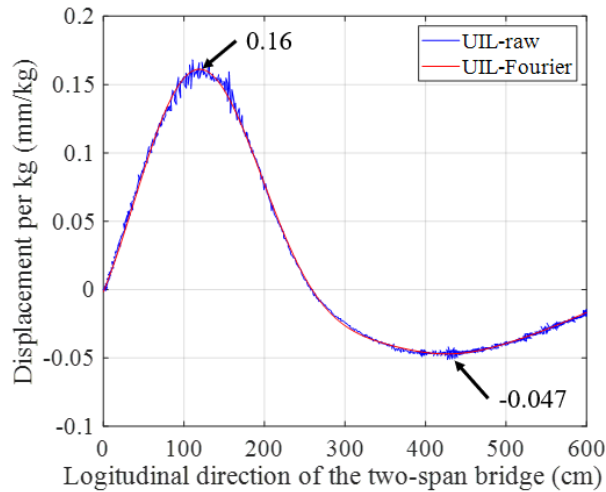


Figure 132 Extracted UIL using the proposed system

10.5 Field Application

10.5.1 Experimental Setup

A field application is demonstrated on a footbridge under small scale vehicle (golf cart) load as shown in Figure 133. The bridge comprises of 19.5m long vertical truss frames that are connected via splice connection in the middle and spans an entire length of 39m over a pond. The width of the bridge is 4.17 m. The vertical truss members on the left and the right side have HSS 10×10×3/8 top and bottom chords and are stabilized with HSS 6×4×3/8 type vertical and HSS 4×4×1/4 type diagonal steel members. The lateral stability is provided by another truss frame that is 3.65m wide which is constructed with HSS 3×3×1/4 type diagonal cross braces, W12×22 type lateral members. Two separate spans are spliced in the middle and the entire frame holds a thin layered aluminum-concrete composite deck. The bridge is located at a university campus and is generally under a light human traffic and small-scale vehicles.

The experimental setup is shown in Figure 133. To track the vehicle during the whole loading process in this experiment, the iPhone XS MAX is employed. The homography transform matrix of the iPhone camera, \mathbf{H} , is $[3.27, 0.15, 289.09; 0.59, 1.23, -398.85; 5.94 \times 10^{-4}, 4.75 \times 10^{-5}, 1]$. The camera used for displacement measurement of the midspan is also Z-CAM E1 camera with a 75-300 mm lens, the same with the one in the laboratory experiment. The scale ratio of this camera is 0.302 mm/pixel. The video formats of both cameras are 4K resolution at a speed of 30 FPS. The videos from the two different cameras are also synchronized with the normalized cross-correlation based pattern matching of audio signals introduced in Figure 86.

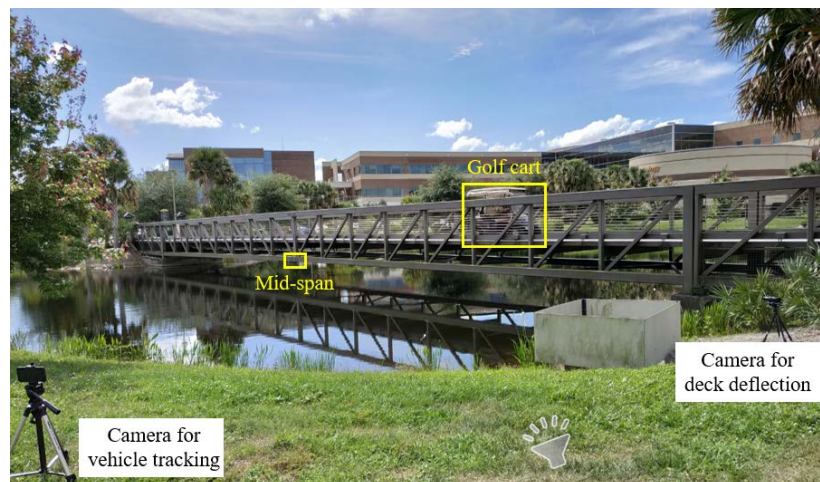


Figure 133 Experimental setup of a footbridge

A golf cart with three people including the driver drove through the bridge back and forth from one end to the other. The weight of the golf cart is 496.69kg and the weights of the three people are 94.34 kg, 78.47 kg, 75 kg respectively. Before starting the golf cart, there was a group of people coming and crossing the bridge. The cameras also recorded this event.

10.5.2 Result Analysis

Figure 134 shows the tracking results of the golf cart in the rectified images obtained from the iPhone camera after planar transformation. The original images captured by the iPhone camera are shown in Figure 121. The scale and view angle of the golf cart changes during the loading process because the iPhone camera is not perpendicular to the longitudinal direction of the footbridge. The images in Figure 121 are first transformed to the fashion in Figure 134 using homography matrix. Then the vehicle tracking is performed in the transformed images. During the loading process, even though the view and scale of the truck changes and the truss part of the footbridge occludes the golf cart, the CSRT tracker can still successfully estimate the location of the golf cart in each image.

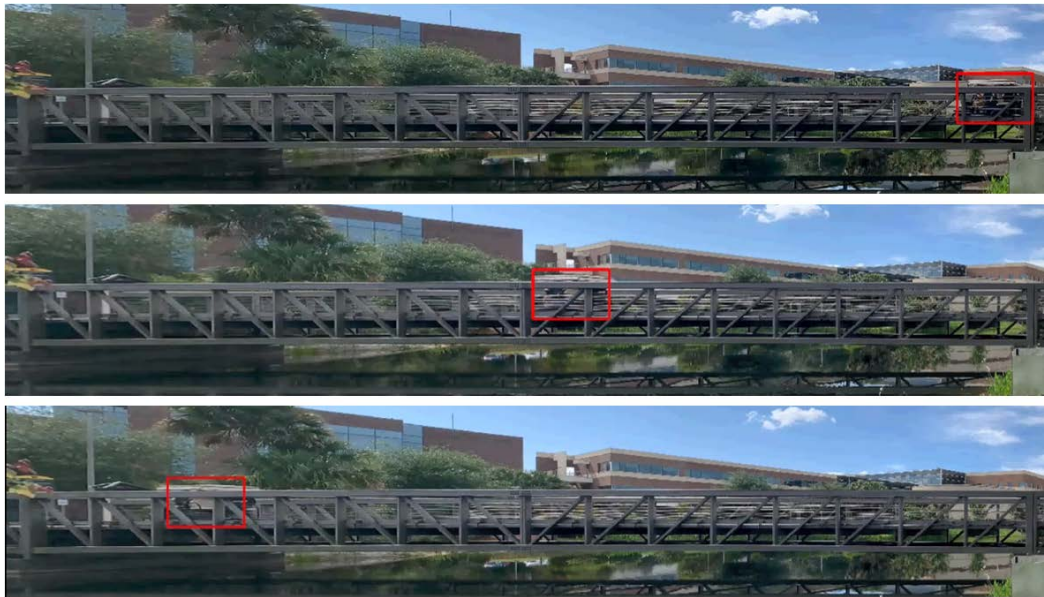


Figure 134 Vehicle tracking in the planar transformed images: (top) tracking when the vehicle starts from the left end of the footbridge; (middle) tracking when the vehicle is at the midspan; (bottom) tracking when the vehicle arrives the right end

The displacement of the midspan during loading process is shown in Figure 135. The proposed vision-based method successfully recognized the pedestrian loads when a group of people crossed the bridge before the starting of the golf cart. The maximum static displacement response (removing the high vibration modes) of the midspan under the pedestrian load is about 0.6 mm. From the displacement time histories, it also shows the displacement responses when the golf cart crossed the bridge back and forth and both of them are very similar with almost the same maximum static response (removing the high vibration modes), around 1.35 mm. It is reasonable because the weight of the golf cart is constant during the experiment and enough time is spent to let the vibration of the bridge attenuate after the golf cart drove from one end to the other.

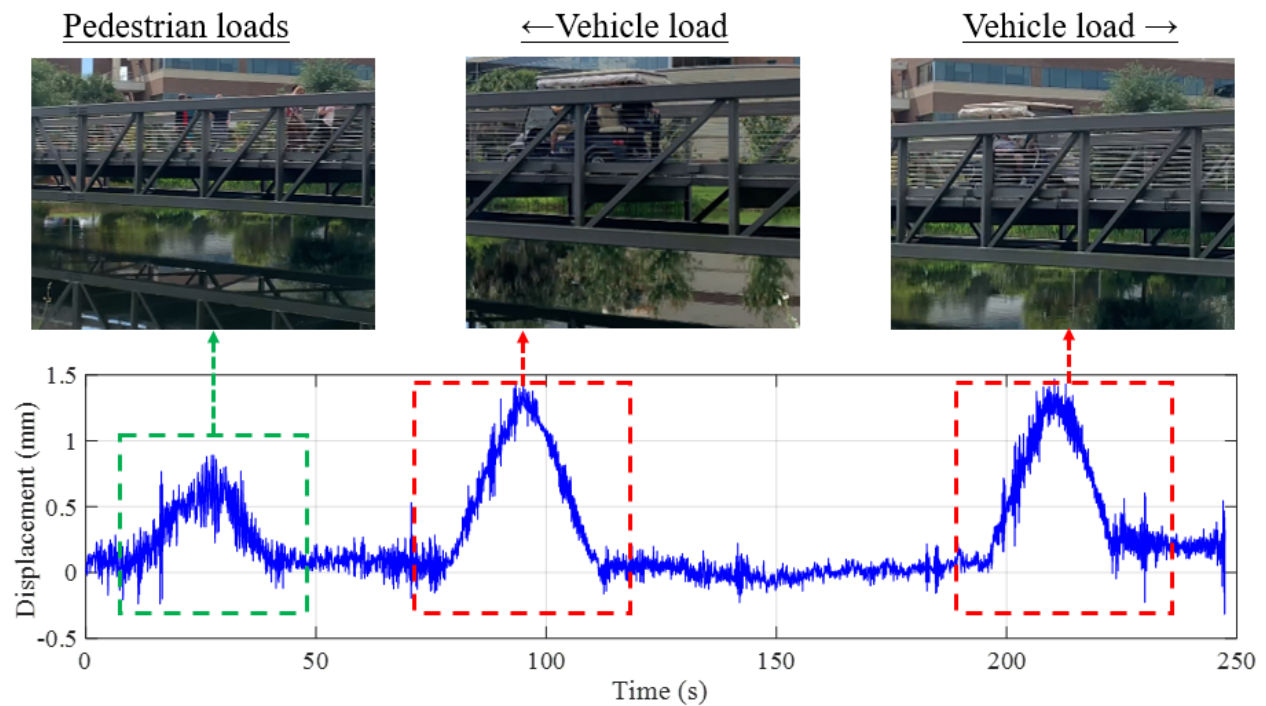


Figure 135 Displacement of the midspan under different loads

In this experiment, the video clips when the golf cart from left end to the right are used to extract the UIL. As shown in Figure 136, the original UIL is represented with the blue curve and removing the high vibration modes with Fourier filter, the UIL is obtained as represented as the red curve. The maximum value of the UIL is 1.70 mm/ton. With the UIL and the maximum static displacement of the bridge under pedestrian loads, the maximum static pedestrian loading is calculated, and it is 353 kg. Recalling the pedestrian loading event from the iPhone video, there are six middle-aged female people crossing the bridge. With the predicted total load, 353 kg, the average weight of each pedestrian is around 58.8 kg (129.6 lbs), which is acceptable.

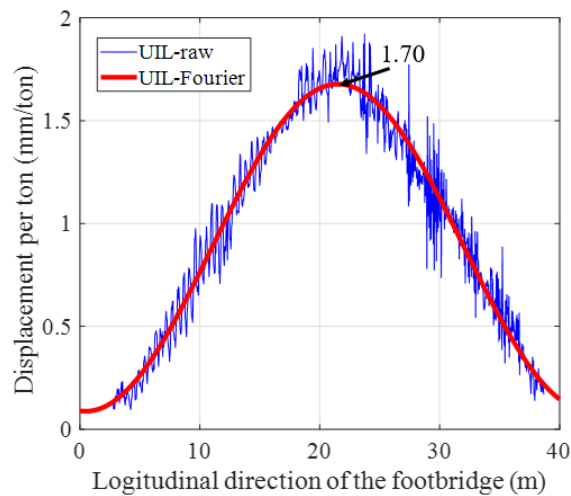


Figure 136 Extracted UIL of the midspan of the footbridge using the proposed system

10.6 Summary

To overcome the inconveniences and disadvantages of the conventional structural health monitoring practices such as high cost, excessive setup time, labor forces with cable wiring work, it would be important and useful to build a structural identification framework with a normalized

structural response indicator irrespective of the type and/or the loads for better decision making with a completely non-contact recognition system. In this study, bridge unit influence line (UIL) using only portable cameras and computer vision is proposed. The feasibility of the proposed method is verified through a comparative study of a series of laboratory experiments and a field application. The main approaches, findings, and conclusions are as follows:

(1) A five-step general procedure for vision-based structural input (vehicle location) recognition is presented. CSRT tracker is implemented to track the vehicle successfully even the scale and view changes and occlusion happens during the visual tracking process.

(2) To broaden the field of view of camera and to track the vehicle during the whole process, a fisheye camera with wide angle is employed and the full camera calibration is carried out to rectify the radial distortion for accurate vehicle localization.

(3) A normalized cross-correlation coefficient using edge map (Canny) for template matching is proposed to achieve reliable displacement measurement. The proposed method avoids the pseudo peaks in NCC map when doing template matching using the traditional NCC based template matching using grayscale images. The displacement results obtained from the proposed method have high consistency with that obtained from conventional displacement sensor with an NCC coefficient of 99.91%.

(4) The two video recordings from two different portable cameras are successfully synchronized by using the normalized cross-correlation based pattern matching of audio signals.

(5) The displacement UIL is successfully identified by combining the vehicle location estimated using visual tracking and homography transformation and the displacement record obtained with vision-based method. It makes the whole identification process in a completely non-contact fashion and UIL is extracted in daily traffic flow.

(6) The extracted displacement UIL is employed for pedestrian load estimation and the predicted weights of pedestrians are observed to be in acceptable ranges. It makes the proposed system work as non-contact weigh-in-motion (WIM) system as presented in this example.

The proposed UIL recognition system also shows great probability to detect damage by using statistical analysis of UILs, bridge load capacity evaluation by regarding UIL as a normalized structural performance indicator and load rating by extracting UILs with the daily traffic flow. The future work will focus on the investigation of them and extend to more possible aspects of structural condition assessment at global level.

CHAPTER ELEVEN: SUMMARY AND CONCLUSIONS

The main objective of this dissertation is to investigate computer vision concepts and methods for structural health monitoring and structural identification research and applications on civil infrastructure systems. The findings of this dissertation can be summarized as discussed in the following:

(1) Current state of art in research and engineering practice of image/computer vision-based methods, approaches and procedures for local and global structural health monitoring and identification for different kind of infrastructures are reviewed in detail;

(2) Structure-oriented (i.e. structure-specific) computer vision-based response monitoring frameworks with different algorithms are developed and demonstrated on different structures in laboratory and with real-life field structures. The detailed summaries including different scenarios for different structure types along with different computer vision-based output-only dynamic monitoring methods are presented in Table 24 and Table 25.

(3) A user-friendly implementation for engineering practice by using a deep learning-based full field optical flow method, i.e., FlowNet2 is demonstrated. This implementation is a viable solution for the civil engineers and researchers who would like to implement yet lack knowledge and experience in computer vision and image processing. Such an approach will leverage the convenience of vision-based monitoring for a broader impact for civil infrastructure community.

(4) The performances of computer vision-based methods such as Lucas-Kanade feature point tracking, feature matching, DIC-based template matching and systems under adverse environmental factors such as illumination change and fog interference are analyzed in a comparative fashion. A robust system by using Spatio-Temporal Context (STC) learning and Taylor approximation is developed to mitigate the adverse environmental factors abovementioned when computer-vision data are to be utilized. The proposed STC-Taylor App method performs better than the current vision-based methods under illumination change and fog interference.

(5) A computer vision-based output-only structural identification for decision making are demonstrated with special emphasis on bridges. In this approach, the conventional output-only structural identification tasks and practical engineering information such as structural response monitoring and deflection limit check, load distribution evaluation, load rating, structural carrying capacity assessment, modal identification and vibration serviceability can be extracted by using the proposed computer vision-based approaches.

(6) A computer vision-based structural identification framework by using input-output both from image data is developed and this framework can successfully extract a conceptual index, the unit influence lines of structures. In addition, the external loads imposed on the structures can also be evaluated with this framework. Based on this approach, the proposed computer vision-based input-output framework can be regarded as a non-contact type of Weigh-in-Motion system.

(7) Recommendations and considerations for field applications on different real-life structures are provided and summarized. The research and engineering community utilizing computer vision-based structural health monitoring and identification are expected to benefit from such the research study presented in this dissertation.

Although computer vision-based structural health monitoring and structural identification present a promising prospect, there are still a few challenges and issues that need further research such as the influences of other types of adverse factors (rain, haze, image blur), long term monitoring, subtle and fast detection from images, selection of markers, connections between structural health monitoring and identification at global and local levels, developments of industry certificate, specifications standards and codes for computer vision-based applications. These challenges and issues are to be considered in future research and engineering studies

Table 24 Applicable scenarios and restraints for the proposed computer vision-based structural output-only monitoring methods

Methods	Applicable Scenarios and Restraints
L-K feature tracking	Close range monitoring/measurement, distinct features on surface, no big lightness change, small motion. Realtime monitoring and post processing.
Feature matching	Close range monitoring/measurement, distinct features on surface
Full field optical flow	Close range monitoring/measurement, distinct features on surface, with the need of full field measurement, difficult to achieve real time monitoring and make decision instantly
STC+Taylor App tracking	Close range monitoring, spatial and temporal context in image sequences are necessary, real time monitoring/post processing
Template matching	Long distance and close-range monitoring

Table 25 Recommendations for different structure types and different computer vision-based structural output-only monitoring methods

Structures Type	L-K feature tracking	Feature matching	Full field optical flow	STC tracking	Template matching
Lab structure	√	√	√ Full field boundary	Light change Fog occurs	√ Need target
Footbridge	√	√	√ Full field boundary	Light change Fog occurs	√ Need target
Highway bridge	√ Need target	√ Need target	√ Full field boundary	Light change Fog occurs	√ Need target
Long-span bridge	×	×	×	Light change Fog occurs	√ Need target
Tower	×	×	×	Light change Fog occurs	√ Need target
Stadium	√	√	√ Full field boundary	Light change Fog occurs	√ Need target

REFERENCES

- AASHTO. (2014). *AASHTO LRFD Bridge Design Specifications*. American Association of State Highway and Transportation Officials, Washington, D.C.
- AASHTO. (2015). *2015 Interim Manual for Bridge Element Inspection*. Washington, DC.
- AASHTO. (2018). *The Manual for Bridge Evaluation*. American Association of State Highway and Transportation Officials, Washington, DC.
- Abdel-Qader, I., Abudayyeh, O., and Kelly, M. E. (2003). “Analysis of Edge-Detection Techniques for Crack Identification in Bridges.” *Journal of Computing in Civil Engineering*, 17(4), 255–263.
- Abdel-Qader, I., Pashaie-Rad, S., Abudayyeh, O., and Yehia, S. (2006). “PCA-Based algorithm for unsupervised bridge crack detection.” *Advances in Engineering Software*, 37(12), 771–778.
- Adhikari, R. S., Moselhi, O., and Bagchi, A. (2014). “Image-based retrieval of concrete crack properties for bridge inspection.” *Automation in Construction*, Elsevier B.V., 39, 180–194.
- Aktan, A. E., Grimmelsman, K. A., Barrish, R. A., Catbas, F. N., and Tsikos, C. J. (2000). “Structural Identification of a Long-Span Truss Bridge.” *Transportation Research Record*, 1696(1), 210–218.
- Alahi, A., Ortiz, R., and Vandergheynst, P. (2012). “FREAK: Fast retina keypoint.” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 510–517.
- Amhaz, R., Chambon, S., Idier, J., and Baltazart, V. (2016). “Automatic Crack Detection on Two-Dimensional Pavement Images: An Algorithm Based on Minimal Path Selection.” *IEEE*

- Transactions on Intelligent Transportation Systems*, 17(10), 2718–2729.
- ASCE. (2017). “2017 Infrastructure Report Card.” <https://www.infrastructurereportcard.org/making-the-grade/report-card-history/>.
- Bachmann, H., Pretlove, A. J., and Rainer, J. H. (1995). “Dynamic forces from rhythmical human body motions.” *Vibration problems in structures: practical guidelines*, Birkhauser, Basel.
- Bai, S., He, Z., Xu, T.-B., Zhu, Z., Dong, Y., and Bai, H. (2018). “Multi-hierarchical Independent Correlation Filters for Visual Tracking.”
- Barker, C. (2007). “Vibration Limits Background to Response Calculation.” *International Journal of Space Structures*, 22(1), 35–43.
- Bas, S., Apaydin, N. M., Ilki, A., and Catbas, F. N. (2018). “Structural health monitoring system of the long-span bridges in Turkey.” *Structure and Infrastructure Engineering*, Taylor & Francis, 14(4), 425–444.
- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). “Speeded-Up Robust Features (SURF).” *Computer Vision and Image Understanding*, 110(3), 346–359.
- BBC. (2000). “2000: Swaying Millennium Bridge closed.” *BBC News*, http://news.bbc.co.uk/onthisday/hi/dates/stories/june/10/newsid_2510000/2510839.stm.
- BBC. (2002). “‘Wobbly’ Millennium Bridge fixed.” *BBC News*, http://news.bbc.co.uk/2/hi/uk_news/england/1829053.stm.
- BEI. (2019). “BEI 9600 series sensors.” https://www.mouser.com/datasheet/2/657/linear-position-sensor_9600-1062977.pdf.
- Berenstein, C. A., Kanal, L. N., Lavin, D., and Olson, E. C. (1987). “A geometric approach to subpixel registration accuracy.” *Computer Vision, Graphics, and Image Processing*, 40(3), 334–360.

- Bharadwaj, K., Sheidaei, A., Afshar, A., and Baqersad, J. (2019). “Full-field strain prediction using mode shapes measured with digital image correlation.” *Measurement: Journal of the International Measurement Confederation*, 139(March), 326–333.
- Bhat, G., Johnander, J., Danelljan, M., Khan, F. S., and Felsberg, M. (2018). “Unveiling the Power of Deep Tracking.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11206 LNCS, 493–509.
- Bhattacharjee, S., and Deb, D. (2016). “Automatic detection and classification of damage zone(s) for incorporating in digital image correlation technique.” *Optics and Lasers in Engineering*, Elsevier, 82, 14–21.
- Black, M. J., and Anandan, P. (1996). “The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields.” *Computer Vision and Image Understanding*, 63(1), 75–104.
- Bouguet, J. (1999). “Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm.” *Intel Corporation, OpenCV Documents*.
- Bradski, G., and Kaehler, A. (2008). *Learning OpenCV:Oreilly*. O’Reilly Media, Sebastopol.
- Brownjohn, J. M. W., Xu, Y., and Hester, D. (2017). “Vision-Based Bridge Deformation Monitoring.” *Frontiers in Built Environment*, 3(April), 1–16.
- BSI. (1978). *Steel, Concrete and Composite Bridges—Part 2: Specification for Loads*. British Standards Association, London.
- Busca, G., Cigada, A., Mazzoleni, P., and Zappa, E. (2014). “Vibration Monitoring of Multiple Bridge Points by Means of a Unique Vision-Based Measuring System.” *Experimental Mechanics*, 54(2), 255–271.
- Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). “BRIEF: Binary Robust Independent Elementary Features.” *Computer Vision—ECCV 2010*, 6314, 778–792.

- Canny, J. F. (1986). "A Computational Approach to Edge Detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679–698.
- Catbas, F. N., and Aktan, A. E. (2002). "Condition and Damage Assessment: Issues and Some Promising Indices." *Journal of Structural Engineering*, 128(8), 1026–1036.
- Catbas, F. N., Brown, D. L., and Aktan, A. E. (2004). "Parameter estimation for multiple-input multiple-output modal analysis of large structures." *Journal of engineering mechanics*, 130(8), 921–930.
- Catbas, F. N., Dong, C. Z., Celik, O., and Khuc, T. (2018). "A Vision for Vision-based Technologies for Bridge Health Monitoring." *Maintenance, Safety, Risk, Management and Life-Cycle Performance of Bridges: Proceedings of the Ninth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2018)*, Melbourne, Australia, 54.
- Catbas, F. N., Gokce, H. B., and Gul, M. (2012a). "Practical Approach for Estimating Distribution Factor for Load Rating: Demonstration on Reinforced Concrete T-Beam Bridges." *Journal of Bridge Engineering*, 17(4), 652–661.
- Catbas, F. N., Gul, M., and Burkett, J. L. (2008a). "Conceptual damage-sensitive features for structural health monitoring: Laboratory and field demonstrations." *Mechanical Systems and Signal Processing*, 22(7), 1650–1669.
- Catbas, F. N., Gul, M., and Burkett, J. L. (2008b). "Damage assessment using flexibility and flexibility-based curvature for structural health monitoring." *Smart Materials and Structures*, 17(1), 015024.
- Catbas, F. N., Hiasa, S., Dong, C. Z., Pan, Y., Celik, O., and Karaaslan, E. (2017). "Comprehensive structural health monitoring at local and global level with vision-based technologies." *26th ASTN symposium*, Jacksonville.

- Catbas, N., Zaurin, R., Gul, M., and Gokce, H. B. (2012b). "Sensor Networks, Computer Imaging, and Unit Influence Lines for Structural Health Monitoring: Case Study for Bridge Load Rating." *Journal of Bridge Engineering*, 17(4), 662–670.
- Celik, O., Dong, C. Z., and Catbas, F. N. (2018a). "A computer vision approach for the load time history estimation of lively individuals and crowds." *Computers & Structures*, Elsevier Ltd, 200, 32–52.
- Celik, O., Dong, C. Z., and Catbas, F. N. (2018b). "Measurement of human loads using computer vision." *IMAC36: It's Not Just Modal Anymore: Engineering extremes: unifying concepts in shock, vibration, and nonlinear mechanics.*, Orlando.
- Celik, O., Dong, C. Z., and Catbas, F. N. (2019a). "Computer Vision Based Human Comfort Assessment of Stadia." *Journal of Performance of Constructed Facilities*, In press.
- Celik, O., Dong, C. Z., and Catbas, F. N. (2019b). "Investigation of human induced excitations using hybrid multivariate empirical mode decomposition." *Journal of Structural Engineering*, ASCE, *in Press*.
- Celik, O., Terrell, T., Gul, M., and Catbas, F. N. (2018c). "Sensor clustering technique for practical structural monitoring and maintenance." *Structural Monitoring and Maintenance*, 5(2), 273–295.
- Cha, Y.-J., Choi, W., and Büyüköztürk, O. (2017a). "Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks." *Computer-Aided Civil and Infrastructure Engineering*, 32(5), 361–378.
- Cha, Y.-J. J., Chen, J. G. G., and Büyüköztürk, O. (2017b). "Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters." *Engineering Structures*, Elsevier Ltd, 132, 300–313.

- Chambon, S., and Moliard, J. M. (2011). "Automatic road pavement assessment with image processing: Review and comparison." *International Journal of Geophysics*, 2011.
- Chan, S. H., Vo, D. T., and Nguyen, T. Q. (2010). "Subpixel motion estimation without interpolation." *2010 IEEE International Conference on Acoustics, Speech and Signal Processing.*, 722–725.
- Chang, C. C., and Ji, Y. F. (2007). "Flexible Videogrammetric Technique for Three-Dimensional Structural Vibration Measurement." *Journal of Engineering Mechanics*, 133(6), 656–664.
- Chen, A., Pan, Y., Pan, Z., and Ma, R. (2014a). "Vehicle load identification of long-span bridges based on dynamic image recognition: Theoretical research and application in Sutong Bridge." *4th International Symposium on Life-Cycle Civil Engineering*, Tokyo, Japan.
- Chen, C. C., Wu, W. H., Tseng, H. Z., Chen, C. H., and Lai, G. (2015a). "Application of digital photogrammetry techniques in identifying the mode shape ratios of stay cables with multiple camcorders." *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, 75, 134–146.
- Chen, F. C., and Jahanshahi, M. R. (2018). "NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion." *IEEE Transactions on Industrial Electronics*, IEEE, 65(5), 4392–4400.
- Chen, F. C., Jahanshahi, M. R., Wu, R. T., and Joffe, C. (2017). "A texture-Based Video Processing Methodology Using Bayesian Data Fusion for Autonomous Crack Detection on Metallic Surfaces." *Computer-Aided Civil and Infrastructure Engineering*, 32(4), 271–287.
- Chen, J. G., Adams, T. M., Sun, H., Bell, E. S., and Büyüköztürk, O. (2018a). "Camera-Based Vibration Measurement of the World War I Memorial Bridge in Portsmouth, New Hampshire." *Journal of Structural Engineering*, 144(11), 04018207.

- Chen, J. G., Wadhwa, N., Cha, Y. J., Durand, F., Freeman, W. T., and Buyukozturk, O. (2015b). “Modal identification of simple structures with high-speed video using motion magnification.” *Journal of Sound and Vibration*, Elsevier, 345, 58–71.
- Chen, P. H., and Chang, L. M. (2002). “Intelligent steel bridge coating assessment using neuro-fuzzy recognition approach.” *Computer-Aided Civil and Infrastructure Engineering*, 17(5), 307–319.
- Chen, P. H., Shen, H. K., Lei, C. Y., and Chang, L. M. (2011). “Fourier-Transform-based method for automated steel bridge coating defect recognition.” *Procedia Engineering*, Elsevier B.V., 14, 470–476.
- Chen, P. H., Shen, H. K., Lei, C. Y., and Chang, L. M. (2012a). “Support-vector-machine-based method for automated steel bridge rust assessment.” *Automation in Construction*, Elsevier B.V., 23, 9–19.
- Chen, P. H., Yang, Y. C., and Chang, L. M. (2009). “Automated bridge coating defect recognition using adaptive ellipse approach.” *Automation in Construction*, Elsevier B.V., 18(5), 632–643.
- Chen, P. H., Yang, Y. C., and Chang, L. M. (2010). “Illumination adjustment for bridge coating images using BEMD-Morphology Approach (BMA).” *Automation in Construction*, Elsevier B.V., 19(4), 475–484.
- Chen, S.-E., Liu, W., Bian, H., and Smith, B. (2012b). “3D LiDAR Scans for Bridge Damage Evaluations.” *Sixth Congress on Forensic Engineering*, San Francisco, California, United States, 487–495.
- Chen, Y., Joffre, D., and Avitabile, P. (2018b). “Underwater Dynamic Response at Limited Points Expanded to Full-Field Strain Response.” *Journal of Vibration and Acoustics*, 140(5), 051016.
- Chen, Y., Logan, P., Avitabile, P., and Dodson, J. (2019). “Non-Model Based Expansion from

- Limited Points to an Augmented Set of Points using Chebyshev Polynomials.” *Experimental Techniques*, 1–23.
- Chen, Y., Zhang, B., and Chen, S. (2014b). “Model Reduction Technique Tailored to the Dynamic Analysis of a Beam Structure under a Moving Load.” *Shock and Vibration*, 2014, 1–13.
- Chen, Y., Zhang, B., Zhang, N., and Zheng, M. (2015c). “A Condensation Method for the Dynamic Analysis of Vertical Vehicle–Track Interaction Considering Vehicle Flexibility.” *Journal of Vibration and Acoustics*, 137(4), 041010.
- Chen, Z., Li, H., Bao, Y., Li, N., and Jin, Y. (2016). “Identification of spatio-temporal distribution of vehicle loads on long-span bridges using computer vision technology.” *Structural Control and Health Monitoring*, 23(3), 517–534.
- Cheng, H. D., Shi, X. J., and Glazier, C. (2003). “Real-Time Image Thresholding Based on Sample Space Reduction and Interpolation Approach.” *Journal of Computing in Civil Engineering*, 17(4), 264–272.
- Daize, F., Micu, E. A., and Brien, E. J. O. (2018). “Using Images to Estimate Traffic Loading on Long-Span Bridges.” *The 9th International Conference on Bridge Maintenance, Safety and Management*, Melbourne, Australia.
- Dan, D., Ge, L., and Yan, X. (2019). “Identification of moving loads based on the information fusion of weigh-in-motion system and multiple camera machine vision.” *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, 144, 155–166.
- Dawood, T., Zhu, Z., and Zayed, T. (2017). “Automation in Construction Machine vision-based model for spalling detection and quantification in subway networks.” *Automation in Construction*, 81(August 2016), 149–160.
- Dey, P., Narasimhan, S., and Walbridge, S. (2017). “Evaluation of design guidelines for the

- serviceability assessment of aluminum pedestrian bridges.” *Journal of Bridge Engineering*, 22(1), 1–15.
- Dey, P., Narasimhan, S., and Walbridge, S. (2018). “Calibrating Pedestrian-Bridge Standards for Vibration Serviceability.” *Journal of Bridge Engineering*, 23(10), 1–17.
- Dong, C. Z., Bas, S., and Catbas, F. N. (2019a). “A completely non-contact recognition system for bridge unit influence line using portable cameras and computer vision.” *Smart Structures and Systems*, 24(5), 617–630.
- Dong, C. Z., and Catbas, F. N. (2019). “A non-target structural displacement measurement method using advanced feature matching strategy.” *Advances in Structural Engineering*, 22(16), 3461–3472.
- Dong, C. Z., Celik, O., and Catbas, F. N. (2017). “Marker Free Development of Computer Vision based Structural Dynamic Monitoring and Modal Analysis.” *The 2017 Engineering Mechanics Institute Conference (EMI 2017)*, San Diego.
- Dong, C. Z., Celik, O., and Catbas, F. N. (2019b). “Marker free monitoring of the grandstand structures and modal identification using computer vision methods.” *Structural Health Monitoring*, 18(5–6), 1491–1509.
- Dong, C. Z., Celik, O., Catbas, F. N., O’Brien, E. J., and Taylor, S. (2019c). “Structural displacement monitoring using deep learning-based full field optical flow methods.” *Structure and Infrastructure Engineering*, 16(1), 51–71.
- Dong, C. Z., Celik, O., Catbas, F. N., O’Brien, E. J., and Taylor, S. E. (2019d). “A Robust Non-Contact Displacement Measurement Method for Long-Term Bridge Monitoring using Spatio-Temporal Context Learning and Taylor Approximation.” *Sensors*, 19(14), 3197.
- Dong, C. Z., Celik, O., Catbas, F. N., O’Brien, E., and Taylor, S. (2019e). “A Robust Vision-Based

- Method for Displacement Measurement under Adverse Environmental Factors Using Spatio-Temporal Context Learning and Taylor Approximation.” *Sensors*, 19(14), 3197.
- Dong, C. Z., Ye, X. W., and Jin, T. (2018). “Identification of structural dynamic characteristics based on machine vision technology.” *Measurement: Journal of the International Measurement Confederation*, 126, 405–416.
- Dong, C. Z., Ye, X. W., and Liu, T. (2015). “Non-contact structural vibration monitoring under varying environmental conditions.” *Vibroengineering Procedia*, 5, 217–222.
- Dorafshan, S., Thomas, R. J., and Maguire, M. (2018). “Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete.” *Construction and Building Materials*, Elsevier Ltd, 186, 1031–1045.
- Dosovitskiy, A., Fischer, P., Ilg, E., Philip, H., Hazırbas, C., Golkov, V., Smagt, P. Van Der, Cremers, D., and Brox, T. (2015). “FlowNet: Learning Optical Flow with Convolutional Networks.” *IEEE International Conference on Computer Vision*, 2758–2766.
- Duan, Y. F., Zhang, R., Dong, C. Z., Luo, Y. Z., Or, S. W., Zhao, Y., and Fan, K. Q. (2016). “Development of Elasto-Magneto-Electric (EME) sensor for in-service cable force monitoring.” *International Journal of Structural Stability and Dynamics*, 16(04), 1640016.
- Dung, C. V., and Anh, L. D. (2019). “Autonomous concrete crack detection using deep fully convolutional neural network.” *Automation in Construction*, Elsevier, 99(December 2018), 52–58.
- ECS. (1997). *Eurocode 5: design of timber structures—part 2: bridges (ENV 1995- 2)*. European Committee for Standardization.
- Ehrhart, M., and Lienhart, W. (2016). “Accurate Measurements with Image-Assisted Total Stations and Their Prerequisites.” *Journal of Surveying Engineering*, 143(2), 04016024.

- Ellenberg, A., Kontsos, A., Moon, F., and Bartoli, I. (2016). “Bridge deck delamination identification from unmanned aerial vehicle infrared imagery.” *Automation in Construction*, Elsevier B.V., 72, 155–165.
- Farneback, G. (2003). “Two-Frame Motion Estimation Based on Polynomial Expansion.” *Lecture Notes in Computer Science*, 2749(1), 363–370.
- Fei, Y., Wang, K. C. P., Zhang, A., Chen, C., Li, J. Q., Liu, Y., Yang, G., and Li, B. (2019). “Pixel-Level Cracking Detection on 3D Asphalt Pavement Images Through Deep-Learning-Based CrackNet-V.” *IEEE Transactions on Intelligent Transportation Systems*, IEEE, PP, 1–12.
- Felipe-Sesé, L., and Díaz, F. A. (2018). “Damage methodology approach on a composite panel based on a combination of Fringe Projection and 2D Digital Image Correlation.” *Mechanical Systems and Signal Processing*, 101, 467–479.
- Feng, D., and Feng, M. Q. (2015). “Model Updating of Railway Bridge Using In Situ Dynamic Displacement Measurement under Trainloads.” *Journal of Bridge Engineering*, 20(APRIL), 1–12.
- Feng, D., and Feng, M. Q. (2016). “Vision-based multipoint displacement measurement for structural health monitoring.” *Structural Control and Health Monitoring*, 23(5), 876–890.
- Feng, D., and Feng, M. Q. (2017). “Experimental validation of cost-effective vision-based structural health monitoring.” *Mechanical Systems and Signal Processing*, Elsevier, 88(June 2016), 199–211.
- Feng, D., and Feng, M. Q. (2018). “Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection – A review.” *Engineering Structures*, Elsevier, 156(January), 105–117.
- Feng, D., Feng, M. Q., Ozer, E., and Fukuda, Y. (2015a). “A vision-based sensor for noncontact

- structural displacement measurement.” *Sensors (Switzerland)*, 15(7), 16557–16575.
- Feng, D., Scarangelo, T., Feng, M. Q., and Ye, Q. (2017). “Cable tension force estimate using novel noncontact vision-based sensor-.” *Measurement: Journal of the International Measurement Confederation*, 99(September), 44–52.
- Feng, M. Q. M. Q., Fukuda, Y., Feng, D., and Mizuta, M. (2015b). “Nontarget Vision Sensor for Remote Measurement of Bridge Dynamic Response.” *Journal of Bridge Engineering*, 20(12), 4015023.
- Feng, P., Wang, Z., Jin, F., and Zhu, S. (2019). “Vibration Serviceability Assessment of Pedestrian Bridges Based on Comfort Level.” *Journal of Performance of Constructed Facilities*, 33(5), 1–10.
- Fioriti, V., Roselli, I., Tatì, A., Romano, R., and De Canio, G. (2018). “Motion Magnification Analysis for structural monitoring of ancient constructions.” *Measurement: Journal of the International Measurement Confederation*, Elsevier, 129(July), 375–380.
- Fleet, D. J., and Jepson, A. D. (1990). “Computation of component image velocity from local phase information.” *International Journal of Computer Vision*, 5(1), 77–104.
- Foroosh, H., Zerubia, J. B., and Berthod, M. (2002). “Extension of phase correlation to subpixel registration.” *IEEE Transactions on Image Processing*, 11(3), 188–199.
- Fuchs, P. A., Washer, G. A., Chase, S. B., and Moore, M. (2004). “Laser-based instrumentation for bridge load testing.” *Journal of performance of constructed facilities*, 18(November), 213–219.
- Fujita, Y., and Hamamoto, Y. (2011). “A robust automatic crack detection method from noisy concrete surfaces.” *Machine Vision and Applications*, 22(2), 245–254.
- Fukuda, Y., Feng, M. Q., Narita, Y., and Tanaka, T. (2013). “Vision-Based Displacement Sensor

- for Monitoring Dynamic Response Using Robust Object Search Algorithm.” *IEEE Sensors Journal*, 13(12), 4725–4732.
- Fukuda, Y., Feng, M. Q., and Shinozuka, M. (2010). “Cost-effective vision-based system for monitoring dynamic response of civil engineering structures.” *Structural Control and Health Monitoring*, 17, 918–936.
- Gao, Y., and Mosalam, K. M. (2018). “Deep Transfer Learning for Image-Based Structural Damage Recognition.” *Computer-Aided Civil and Infrastructure Engineering*, 33(9), 748–768.
- Gautama, T., and Van Hulle, M. a. (2002). “A phase-based approach to the estimation of the optical flow field using spatial filtering.” *IEEE Transactions on Neural Networks*, 13(5), 1127–1136.
- Gavilán, M., Balcones, D., Marcos, O., Llorca, D. F., Sotelo, M. A., Parra, I., Ocaña, M., Aliseda, P., Yarza, P., and Amírola, A. (2011). “Adaptive road crack detection system by pavement classification.” *Sensors*, 11(10), 9628–9657.
- German, S., Brilakis, I., and Desroches, R. (2012). “Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments.” *Advanced Engineering Informatics*, Elsevier Ltd, 26(4), 846–858.
- Ghosn, M., Frangopol, D. M., Mcallister, T. P., Shah, M., Diniz, S., Ellingwood, B. R., Manuel, L., Biondini, F., Catbas, N., Strauss, A., and Zhao, Z. L. (2016). “Reliability-based structural performance indicators for structural members.” *Journal of Structural Engineering*, 142(9), 1–13.
- Girshick, R. (2015). “Fast R-CNN.” *International Conference on Computer Vision*, 1440–1448.
- Girshick, R., Donahue, J., Darrell, T., Berkeley, U. C., and Malik, J. (2012). “(r-cnn) Rich feature hierarchies for accurate object detection and semantic segmentation.” *2014 IEEE Conference*

on Computer Vision and Pattern Recognition, 2–9.

Goodfellow, I., Bengio, Y., Courville, A., Goodfellow, I., and Courville, A. (2017). *Deep learning*. MIT Press.

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., and Agrawal, A. (2017). “Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection.” *Construction and Building Materials*, Elsevier Ltd, 157, 322–330.

Grant, I. (1997). “Particle image velocimetry: A review.” *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 211, 55–76.

Gul, M., and Catbas, F. N. (2008). “Ambient Vibration Data Analysis for Structural Identification and Global Condition Assessment.” *Journal of Engineering Mechanics*, 134(8), 650–662.

Gul, M., and Catbas, F. N. (2009). “Statistical pattern recognition for Structural Health Monitoring using time series modeling: Theory and experimental verifications.” *Mechanical Systems and Signal Processing*, Elsevier, 23(7), 2192–2204.

Gul, M., and Catbas, F. N. (2011a). “Damage Assessment with Ambient Vibration Data Using a Novel Time Series Analysis Methodology.” *Journal of Structural Engineering*, 137(12), 1518–1526.

Gul, M., and Catbas, F. N. (2011b). “Structural health monitoring and damage assessment using a novel time series analysis methodology with sensor clustering.” *Journal of Sound and Vibration*, Elsevier, 330(6), 1196–1210.

Gul, M., Dumlupinar, T., Hattori, H., and Catbas, N. (2014). “Structural monitoring of movable bridge mechanical components for maintenance decision-making.” 1(3), 249–271.

Guo, J., and Zhu, C. (2016). “Dynamic displacement measurement of large-scale structures based on the Lucas-Kanade template tracking algorithm.” *Mechanical Systems and Signal*

- Processing*, Elsevier, 66–67, 425–436.
- Harris, C., and Stephens, M. (1988). “A Combined Corner and Edge Detector.” *Proceedings of the Alvey Vision Conference 1988*, 23.1-23.6.
- Hartley, R., and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). “Mask R-CNN.” *The IEEE International Conference on Computer Vision (ICCV)*, 2961–2969.
- Henke, K., Pawlowski, R., Schregle, P., and Winter, S. (2015). “Use of digital image processing in the monitoring of deformations in building structures.” *Journal of Civil Structural Health Monitoring*, 5(2), 141–152.
- Hiasa, S., Birgul, R., and Catbas, F. N. (2016a). “Infrared thermography for civil structural assessment: demonstrations with laboratory and field studies.” *Journal of Civil Structural Health Monitoring*, Springer Berlin Heidelberg, 6(3), 619–636.
- Hiasa, S., Birgul, R., and Catbas, F. N. (2017a). “Investigation of Effective Utilization of Infrared Thermography (IRT) through Advanced Finite Element Modeling.” *Construction and Building Materials*, in press.
- Hiasa, S., Birgul, R., and Catbas, F. N. (2017b). “Effect of Defect Size on Subsurface Defect Detectability and Defect Depth Estimation for Concrete Structures by Infrared Thermography.” *Journal of Nondestructive Evaluation*, 36(3).
- Hiasa, S., Birgul, R., Matsumoto, M., and Necati Catbas, F. (2018). “Experimental and Numerical Studies for Suitable Infrared Thermography Implementation on Concrete Bridge Decks.” *Measurement*.
- Hiasa, S., Birgul, R., and Necati Catbas, F. (2017c). “A data processing methodology for infrared

- thermography images of concrete bridges.” *Computers & Structures*, 190, 205–218.
- Hiasa, S., Catbas, F. N., Matsumoto, M., Mitani, K., and Necati Catbas, F. (2017d). “Considerations and Issues in the Utilization of Infrared Thermography for Concrete Bridge Inspection at Normal Driving Speeds.” *Journal of Bridge Engineering*, 22(11), 04017101.
- Hiasa, S., Catbas, F. N. N., Matsumoto, M., and Mitani, K. (2016b). “Monitoring Concrete Bridge Decks using Infrared Thermography with High Speed Vehicles.” *Structural Monitoring and Maintenance, An International Journal*, Techno-Press, Daejeon, Korea, 3(3), 277–296.
- Ho, H.-N., Lee, J.-J. J.-H., Park, Y.-S., and Lee, J.-J. J.-H. (2012). “A Synchronized Multipoint Vision-Based System for Displacement Measurement of Civil Infrastructures.” *The Scientific World Journal*, 2012, 1–9.
- Horn, B. K. P., and Schunck, B. G. (1981). “Determining optical flow.” *Artificial Intelligence*, 17(1–3), 185–203.
- Hoskere, V., Park, J., Yoon, H., and Spencer Jr., B. F. (2019). “Vision-based Modal Survey of Civil Infrastructure using Unmanned Aerial Vehicles.” *Journal of Structural Engineering*, 145(7), 04019062.
- Hosseini, A., Mostofinejad, D., and Hajialilue-Bonab, M. (2014). “Displacement and Strain Field Measurement in Steel and RC Beams Using Particle Image Velocimetry.” *Journal of Engineering Mechanics*, 140(11), 04014086.
- Hu, Q., He, S., Wang, S., Liu, Y., Zhang, Z., He, L., Wang, F., Cai, Q., Shi, R., and Yang, Y. (2017). “A High-Speed Target-Free Vision-Based Sensor for Bus Rapid Transit Viaduct Vibration Measurements Using CMT and ORB Algorithms.” *Sensors*, 17(6), 1305.
- Hu, Y., Zhao, C., and Wang, H. (2010). “Automatic Pavement Crack Detection Using Texture and Shape Descriptors.” *IETE Technical Review*, 27(5), 398.

- Huang, Y., and Xu, B. (2006). “Automatic inspection of pavement cracking distress.” *Journal of Electronic Imaging*, 15(1), 013017.
- Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., and Brox, T. (2017). “FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2462–2470.
- ISO. (2007). *Bases for design of structures -serviceability of buildings and walkways against Vibrations. ISO 10137*, International Organization for Standardization, Geneva, Switzerland.
- Iyer, S., and Sinha, S. K. (2006). “Segmentation of pipe images for crack detection in buried sewers.” *Computer-Aided Civil and Infrastructure Engineering*, 21(6), 395–410.
- Jahanshahi, M. R., and Masri, S. F. (2012). “Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures.” *Automation in Construction*, Elsevier B.V., 22, 567–576.
- Jang, K., Kim, N., and An, Y. K. (2019). “Deep learning–based autonomous concrete crack evaluation through hybrid image scanning.” *Structural Health Monitoring*.
- Jáuregui, D. V., White, K. R., Woodward, C. B., and Leitch, K. R. (2003). “Noncontact Photogrammetric Measurement of Vertical Bridge Deflection.” *Journal of Bridge Engineering*, 8(4), 212–222.
- Ji, Y. F., and Chang, C. C. (2008a). “Nontarget Stereo Vision Technique for Spatiotemporal Response Measurement of Line-Like Structures.” *Journal of Engineering Mechanics*, 134(6), 466–474.
- Ji, Y. F., and Chang, C. C. (2008b). “Nontarget image-based technique for small cable vibration measurement.” *Journal of Bridge Engineering*, 13(1), 34–42.
- Karaaslan, E., Bagci, U., and Catbas, F. N. (2018). “Artificial Intelligence Assisted Infrastructure

- Assessment Using Mixed Reality Systems.” *arXiv preprint*, arXiv:1812.05659.
- Karaaslan, E., Bagci, U., and Catbas, F. N. (2019). “Artificial Intelligence Assisted Infrastructure Assessment using Mixed Reality Systems.” *Transportation Research Record: Journal of the Transportation Research Board*, 036119811983998.
- Kasperski, M. (2006). “Vibration serviceability for pedestrain bridges.” *Proceedings of the Institution of Civil Engineers: Structures and Buildings*, 159(5), 273–282.
- Khaloo, A., and Lattanzi, D. (2017). “Pixel-wise structural motion tracking from rectified repurposed videos.” *Structural Control and Health Monitoring*, 24(11), 1–15.
- Khuc, T., and Catbas, F. N. (2016). “Computer vision-based displacement and vibration monitoring without using physical target on structures.” *Structure and Infrastructure Engineering*, Taylor & Francis, 13(4), 1–12.
- Khuc, T., and Catbas, F. N. (2017). “Completely contactless structural health monitoring of real-life structures using cameras and computer vision.” *Structural Control and Health Monitoring*, 24(1), e1852.
- Khuc, T., and Catbas, F. N. (2018). “Structural Identification Using Computer Vision-Based Bridge Health Monitoring.” *Journal of Structural Engineering*, 144(2), 04017202.
- Kim, B., and Cho, S. (2018). “Automated vision-based detection of cracks on concrete surfaces using a deep learning technique.” *Sensors (Switzerland)*, 18(10).
- Kim, C., Son, H., Hwang, N., and Kim, C. (2014). “Rapid and automated determination of rusted surface areas of a steel bridge for robotic maintenance systems.” *Automation in Construction*, Elsevier B.V., 42, 13–24.
- Kim, S.-W. S.-W., Jeon, B.-G. B.-G., Kim, N.-S. N.-S., and Park, J.-C. J.-C. (2013). “Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge.” *Structural*

- Health Monitoring: An International Journal*, 12(5–6), 440–456.
- Kim, S. C. B., Kim, H. K., Lee, C. G., and Kim, S. C. B. (2006). “A vision system for identifying structural vibration in civil engineering constructions.” *2006 SICE-ICASE International Joint Conference*, 5813–5818.
- Kitagawa, A., Hishida, K., and Kodama, Y. (2005). “Flow structure of microbubble-laden turbulent channel flow measured by PIV combined with the shadow image technique.” *Experiments in Fluids*, 38(4), 466–475.
- Ko, J. M., and Ni, Y. Q. (2005). “Technology developments in structural health monitoring of large-scale bridges.” *Engineering Structures*, 27(12 SPEC. ISS.), 1715–1725.
- Ko, J. M., Sun, Z. G., and Ni, Y. Q. (2002). “Multi-stage identification scheme for detecting damage in cable-stayed Kap Shui Mun Bridge.” *Engineering Structures*, 24(7), 857–868.
- Kohut, P., Holak, K., Uhl, T., Ortyl, L., Owerko, T., Kuras, P., and Kocierz, R. (2013). “Monitoring of a civil structure’s state based on noncontact measurements.” *Structural Health Monitoring*, 12(5–6), 411–429.
- Kong, X., and Li, J. (2018a). “Vision-Based Fatigue Crack Detection of Steel Structures Using Video Feature Tracking.” *Computer-Aided Civil and Infrastructure Engineering*, 33(9), 783–799.
- Kong, X., and Li, J. (2018b). “Image registration-based bolt loosening detection of steel joints.” *Sensors (Switzerland)*, 18(4).
- Kong, X., Li, J., Collins, W., Bennett, C., Laflamme, S., and Jo, H. (2018). “Sensing distortion-induced fatigue cracks in steel bridges with capacitive skin sensor arrays.” *Smart Materials and Structures*, 27(11), 115008.
- Kristan, M., Leonardis, A., Matas, J., Felsberg, M., Pfugfelder, R., Zajc, L. C., Vojir, T., Bhat, G.,

- Lukezic, A., Eldesokey, A., Fernandez, G., and et al. (2018). "The sixth Visual Object Tracking VOT2018 challenge results." *VOT2018 workshop, ECCV2018*, Munich, Germany.
- Lages Martins, L. L., Rebordão, J. M., and Silva Ribeiro, A. S. (2015). "Structural observation of long-span suspension bridges for safety assessment: Implementation of an optical displacement measurement system." *Journal of Physics: Conference Series*, 588(1).
- Lava, P., Cooreman, S., Coppieters, S., De Strycker, M., and Debruyne, D. (2009). "Assessment of measuring errors in DIC using deformation fields generated by plastic FEA." *Optics and Lasers in Engineering*, 47(7–8), 747–753.
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436–444.
- Lee, J. J., Cho, S., and Shinozuka, M. (2006a). "Evaluation of Bridge Load Carrying Capacity Based on Dynamic Displacement Measurement Using Real-time Image Processing Techniques." *Steel Structures*, 6, 377–385.
- Lee, J. J., and Shinozuka, M. (2006). "Real-Time Displacement Measurement of a Flexible Bridge Using Digital Image Processing Techniques." *Experimental Mechanics*, 46, 105–114.
- Lee, J., Jung, C., Woo, T., and Cheung, J. (2019). "Post-yielding tension stiffening of reinforced concrete members using an image analysis method with a consideration of steel ratios." 2, 117–126.
- Lee, J., Lee, K.-C. C., Cho, S., and Sim, S.-H. H. (2017). "Computer vision-based structural displacement measurement robust to light-induced image degradation for in-service bridges." *Sensors (Switzerland)*, 17(10), 2317.
- Lee, S. (2005). "Digital Image Processing Based Semi-automatic Data Acquisition System for Ongoing Construction Progress Assesment." Purdue University.
- Lee, S., Chang, L. M., and Chen, P. H. (2005). "Performance comparison of bridge coating defect

- recognition methods.” *Corrosion*, 61(1), 12–20.
- Lee, S., Chang, L. M., and Skibniewski, M. (2006b). “Automated recognition of surface defects using digital color image processing.” *Automation in Construction*, 15(4), 540–549.
- Leutenegger, S., Chli, M., and Siegwart, R. (2011). “BRISK: Binary Robust Invariant Scalable Keypoints.” *2011 IEEE international conference on computer vision (ICCV)*.
- Li, B., Wu, W., Wang, Q., Zhang, F., Xing, J., and Yan, J. (2018a). “SiamRPN++: Evolution of Siamese Visual Tracking with Very Deep Networks.”
- Li, G., He, S., and Ju, Y. (2013). “Image-based Method for Concrete Bridge Crack Detection.” *Journal of Information and Computational Science*, 10(8), 2229–2236.
- Li, G., He, S., Ju, Y., and Du, K. (2014). “Long-distance precision inspection method for bridge cracks with image processing.” *Automation in Construction*, Elsevier B.V., 41, 83–95.
- Li, G., Ji, T., and Chen, J. (2018b). “Determination of the dynamic load factors for crowd jumping using motion capture technique.” *Engineering Structures*, Elsevier, 174(July), 1–9.
- Li, G., Zhao, X., Du, K., Ru, F., and Zhang, Y. (2017). “Recognition and evaluation of bridge cracks with modified active contour model and greedy search-based support vector machine.” *Automation in Construction*, 78, 51–61.
- Liang, Z., Yin, B., Liu, H., Mo, J., and Wang, S. (2015). “Displacement measurement of specimen surfaces with damaged areas by digital image correlation.” *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, 76, 183–188.
- Liao, K. W., and Lee, Y. T. (2016). “Detection of rust defects on steel bridge coatings via digital image recognition.” *Automation in Construction*, Elsevier B.V., 71(Part 2), 294–306.
- Liu, B., and Zaccarin, A. (1993). “New fast algorithms for the estimation of block motion vectors.” *Circuits and Systems for Video Technology*, ..., 3(2), 148–157.

- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C. (2016a). "SSD: Single shot multibox detector." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9905 LNCS, 21–37.
- Liu, Y., Cho, S., and Fan, J. (2016b). "Concrete Crack Assessment Using Digital Image Processing and 3D Scene Reconstruction." *Journal of Computing in Civil Engineering*, 30(1), 1–19.
- Liu, Z., Suandi, S. A., Ohashi, T., and Ejima, T. (2003). "<title>Tunnel crack detection and classification system based on image processing</title>." *Machine Vision Applications in Industrial Inspection X*, 4664(March 2002), 145–152.
- Lomb, N. R. (1976). "Least-squares frequency analysis of unequally spaced data." *Astrophysics and Space Science*, 39(2), 447–462.
- Lowe, D. G. (2004). "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision*, 60(2), 91–110.
- Lu, W., Cui, Y., and Teng, J. (2017). "Structural displacement and strain monitoring based on the edge detection operator." *Advances in Structural Engineering*, 20(2), 191–201.
- Lucas, B. D., and Kanade, T. (1981). "An iterative image registration technique with an application to stereo vision." *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*.
- Lukezic, A., Vojir, T., Zajc, L. C., Matas, J., and Kristan, M. (2017). "Discriminative Correlation Filter Tracker with Channel and Spatial Reliability." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6309–6318.
- Luo, L., and Feng, M. Q. (2018). "Edge-Enhanced Matching for Gradient-Based Computer Vision Displacement Measurement." *Computer-Aided Civil and Infrastructure Engineering*, 33(12),

1019–1040.

Lydon, D., Lydon, M., del Rincon, J. M., Taylor, S. E., Robinson, D., O'Brien, E., Catbas, N., and Catbas, F. N. (2018). "Development and field testing of a time-synchronized system for multi-point displacement calculation using low cost wireless vision-based sensors." *IEEE Sensors Journal*, 1748(c), 9744–9754.

Lydon, D., Lydon, M., Taylor, S., Martinez, J., Rincon, D., Hester, D., and Brownjohn, J. (2019). "Development and field testing of a vision-based displacement system using a low cost wireless action camera." *Mechanical Systems and Signal Processing*, Elsevier Ltd, 121, 343–358.

Lydon, D., Taylor, S., Martinez del Rincon, J., Hester, D., Lydon, M., and Robinson, D. (2016). "Field investigation of contactless displacement measurement using computer vision systems for civil engineering applications." *Irish Machine Vision and Image Processing Conference - NUIG*, Galway, United Kingdom.

Ma, S., Pang, J., and Ma, Q. (2012). "The systematic error in digital image correlation induced by self-heating of a digital camera." *Measurement Science and Technology*, 23(2), 025403.

Mackenzie, D., Barker, C., Mcfadyen, N., and Allison, B. (2005). "Footbridge pedestrian vibration limits Part2: Human sensitivity." *Footbridge 2005 - Second International Conference*, Venice, Italy.

Mathavan, S., C.Eng, M. R., and Kamal, K. (2015). "Use of a self-organizing map for crack detection in highly textured pavement images." *Journal of Infrastructure Systems*, 21(3), 1–11.

Matsumoto, M., Mitani, K., and Catbas, F. N. (2013). "Bridge assessment methods using image processing and infrared thermography technology: on-site pilot application in Florida."

Transportation Research Board; 92nd Annual Meeting.

- Matsumoto, Y., Nishioka, T., Shiojiri, H., and Matsuzaki, K. (1978). “Dynamic Design of Footbridges.” *Proceedings of the International Association for Bridge and Structural Engineering*, 2(17), 1–15.
- Mazzoleni, P., and Zappa, E. (2012). “Vision-based estimation of vertical dynamic loading induced by jumping and bobbing crowds on civil structures.” *Mechanical Systems and Signal Processing*, Elsevier, 33, 1–12.
- Mei, Q., and Gül, M. (2019). “A Conditional Wasserstein Generative Adversarial Network for Pixel-level Crack Detection using Video Extracted Images.” *arXiv:1907.06014*.
- Moreu, F., Li, J., Jo, H., Kim, R. E., Scola, S., Spencer, B. F., and LaFave, J. M. (2016). “Reference-Free Displacements for Condition Assessment of Timber Railroad Bridges.” *Journal of Bridge Engineering*, 21(2), 04015052.
- National Instruments. (2016). “Color Pattern Matching.” <https://zone.ni.com/reference/en-XX/help/370281AC-01/nivisionconcepts/color_pattern_matching/>.
- National Instruments. (2017). “Multifunction I/O Device: USB-6343.” <<http://www.ni.com/en-us/support/model.usb-6343.html>>.
- Nayyeri, F., Hou, L., Zhou, J., and Guan, H. (2019). “Foreground–background separation technique for crack detection.” *Computer-Aided Civil and Infrastructure Engineering*, 34(6), 457–470.
- Nhat-Duc, H., Nguyen, Q. L., and Tran, V. D. (2018). “Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network.” *Automation in Construction*, 94(January), 203–213.
- Ni, F. T., Zhang, J., and Chen, Z. Q. (2019). “Pixel-level crack delineation in images with

- convolutional feature fusion.” *Structural Control and Health Monitoring*, 26(1), 1–18.
- Ni, Y. Q. Q., Ye, X. W., and Ko, J. M. (2010). “Monitoring-Based Fatigue Reliability Assessment of Steel Bridges : Analytical Model and Application.” *Journal of Structural Engineering*, 136(12), 1563–1573.
- Ni, Y. Q., Xia, H. W., and Ko, J. M. (2008). “Structural performance evaluation of tsing ma bridge deck using long-term monitoring data *.” *Modern Physics Letters B*, 22(11), 875–880.
- Ni, Y. Q., Xia, Y., Liao, W. Y., and Ko, J. M. (2009). “Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower.” *STRUCTURAL CONTROL AND HEALTH MONITORING*, 16, 73–98.
- Ni, Y. Q., Ye, X. W., and Ko, J. M. (2011). “Modeling of stress spectrum using long-term monitoring data and finite mixture distributions.” *Journal of Engineering Mechanics, ASCE*, 138(2), 175–184.
- Ni, Y. Q., Zhou, X. T., and Ko, J. M. (2003). “Modal and seismic damage identification of a tall building model.” *Structural Health Monitoring and Intelligent Infrastructure*, Z. S. Wu and M. Abe, eds., Balkema: Lisse, Netherlands, 575–583.
- O’Byrne, M., Ghosh, B., Schoefs, F., O’Donnell, D., Wright, R., and Akrashi, V. (2015). “Acquisition and Analysis of Dynamic Responses of a Historic Pedestrian Bridge using Video Image Processing.” *Journal of Physics: Conference Series*, 628, 012053.
- Ojio, T., Carey, C. H., OBrien, E. J., Doherty, C., and Taylor, S. E. (2016). “Contactless Bridge Weigh-in-Motion.” *Journal of Bridge Engineering*, 21(7), 04016032.
- Olaszek, P. (1999). “Investigation of the dynamic characteristic of bridge structures using a computer vision method.” *Measurement*, 25, 227–236.
- Omar, T., and Nehdi, M. L. (2017). “Remote sensing of concrete bridge decks using unmanned

- aerial vehicle infrared thermography.” *Automation in Construction*, 83(April), 360–371.
- Omar, T., Nehdi, M. L., and Zayed, T. (2018). “Infrared thermography model for automated detection of delamination in RC bridge decks.” *Construction and Building Materials*, 168, 313–327.
- Omidalizarandi, M., Kargoll, B., Paffenholz, J. A., and Neumann, I. (2018). “Accurate vision-based displacement and vibration analysis of bridge structures by means of an image-assisted total station.” *Advances in Mechanical Engineering*, 10(6), 1–19.
- OpenCV. (2019a). “OpenCV: Feature Matching.” *Open Source Computer Vision*, <https://www.docs.opencv.org/master/dc/dc3/tutorial_py_matcher.html>.
- OpenCV. (2019b). “OpenCV documents: feature detection and description.” <http://www.docs.opencv.org/master/df/d54/tutorial_py_features_meaning.html>.
- OpenCV. (2019c). “Feature Matching + Homography to find Objects.” <https://docs.opencv.org/master/d1/de0/tutorial_py_feature_homography.html>.
- OpenCV. (2019d). “OpenCV 4.1.0 document: Video Analysis.” <https://www.docs.opencv.org/master/d3/db0/tutorial_py_table_of_contents_video.html>.
- Oppenheim, A. V., Willsky, A. S., and Nawab, S. H. (1996). *Signals and Systems*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.
- Pan, B., Qian, K., Xie, H., and Asundi, A. (2009). “Two-dimensional digital image correlation for in-plane displacement and strain measurement: A review.” *Measurement Science and Technology*, 20(6).
- Pan, B., Tian, L., and Song, X. (2016). “Real-time, non-contact and targetless measurement of vertical deflection of bridges using off-axis digital image correlation.” *NDT and E International*, Elsevier, 79, 73–80.

- Pan, B., Xie, H. M., Xu, B. Q., and Dai, F. L. (2006). "Performance of sub-pixel registration algorithms in digital image correlation." *Measurement Science and Technology*, 17(6), 1615–1621.
- Pan, Y., Wang, D., Shen, X., Xu, Y., and Pan, Z. (2018). "A novel computer vision-based monitoring methodology for vehicle-induced aerodynamic load on noise barrier." *Structural Control and Health Monitoring*, 25(12), 1–19.
- Park, J.-W., Lee, J.-J., Jung, H.-J., and Myung, H. (2010). "Vision-based displacement measurement method for high-rise building structures using partitioning approach." *NDT & E International*, Elsevier, 43(7), 642–647.
- Park, J. H., Huynh, T. C., Choi, S. H., and Kim, J. T. (2015a). "Vision-based technique for bolt-loosening detection in wind turbine tower." *Wind and Structures, An International Journal*, 21(6), 709–726.
- Park, S. W., Park, H. S. H. S., Kim, J. H., and Adeli, H. (2015b). "3D displacement measurement model for health monitoring of structures using a motion capture system." *Measurement*, Elsevier Ltd, 59, 352–362.
- Poozesh, P., Baqersad, J., Niezrecki, C., Avitabile, P., Harvey, E., and Yarala, R. (2017). "Large-area photogrammetry based testing of wind turbine blades." *Mechanical Systems and Signal Processing*, Elsevier, 86, 98–115.
- Prasanna, P., Dana, K. J., Gucunski, N., Basily, B. B., La, H. M., Lim, R. S., and Parvardeh, H. (2016). "Automated Crack Detection on Concrete Bridges." *IEEE Transactions on Automation Science and Engineering*, 13(2), 591–599.
- Pretlove, A. J., and Rainer, J. H. (1995). "Human response to vibrations." *Vibration Problems in Structures: Practical Guidelines*, Birkhauser, Basel.

- Ramana, L., Choi, W., and Cha, Y. J. (2019). “Fully automated vision-based loosened bolt detection using the Viola–Jones algorithm.” *Structural Health Monitoring*, 18(2), 422–434.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2015). “You Only Look Once: Unified, Real-Time Object Detection.” *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 779–788.
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.
- Rosten, E., and Drummond, T. (2006). “Machine Learning for High Speed Corner Detection.” *Computer Vision -- ECCV 2006*, 1, 430–443.
- Rublee, E., and Bradski, G. (2011). “ORB - an efficient alternative to SIFT or SURF.” 2564–2571.
- Sajedi, S. O., and Liang, X. (2019). “A Convolutional Cost-Sensitive Crack Localization Algorithm for Automated and Reliable RC Bridge Inspection.” *arXiv preprint*.
- Sanayei, M., Reiff, A. J., Brenner, B. R., and Imbaro, G. R. (2016). “Load Rating of a Fully Instrumented Bridge: Comparison of LRFR Approaches.” *Journal of Performance of Constructed Facilities*, 30(2), 1–7.
- Schreier, H., Orteu, J.-J., and Sutton, M. A. (2009). *Image Correlation for Shape, Motion and Deformation Measurements. Image Correlation for Shape, Motion and Deformation Measurements*.
- Schumacher, T., and Shariati, A. (2013). “Monitoring of structures and mechanical systems using virtual visual sensors for video analysis: Fundamental concept and proof of feasibility.” *Sensors (Switzerland)*, 13(12), 16551–16564.
- Setareh, M. (2016). “Vibration Serviceability Issues of Slender Footbridges.” *Journal of Bridge*

- Engineering*, 21(11), 1–12.
- Setra. (2006). *Footbridges: Assessment of Vibrational Behavior of Footbridges Under Pedestrian Loading (Technical guide)*. Paris.
- Shan, B., Shen, Y., Su, X., and Xian, G. (2019). “Stereovision monitoring of deflection of concrete beam strengthened with ultraviolet-cured glass-fiber reinforced polymer in a destructive test.” *Journal of Low Frequency Noise Vibration and Active Control*, 0(0), 1–15.
- Shan, B., Yuan, W., Wang, H., and Zuo, Z. (2018a). “A parallel stereovision method used for monitoring the collapse of a three-story frame model subjected to seismic loading.” *International Journal of Distributed Sensor Networks*, 14(9).
- Shan, B., Yuan, W., Wang, H., Zuo, Z., and Li, S. (2018b). “Stereovision monitoring for entire collapse of a three-story frame model under earthquake excitation.” *Structural Control and Health Monitoring*, 25(6), 1–19.
- Shen, H. K., Chen, P. H., and Chang, L. M. (2013). “Automated steel bridge coating rust defect recognition method based on color and texture feature.” *Automation in Construction*.
- Shen, H. K., Chen, P. H., and Chang, L. M. (2018). “Human-visual-perception-like intensity recognition for color rust images based on artificial neural network.” *Automation in Construction*, Elsevier, 90(February), 178–187.
- Shi, J., and Tomasi, C. (1994). “Good Features to Track.” *Proceedings CVPR’94., 1994 IEEE Computer Society Conference*, 593–600.
- Shi, Y., Cui, L., Qi, Z., Meng, F., and Chen, Z. (2016). “Automatic road crack detection using random structured forests.” *IEEE Transactions on Intelligent Transportation Systems*, IEEE, 17(12), 3434–3445.
- da Silva, I., Ibañez, W., and Poleszuk, G. (2018). “Experience of Using Total Station and GNSS

- Technologies for Tall Building Construction Monitoring.” *Facing the Challenges in Structural Engineering, Sustainable Civil Infrastructures*, M. Remki, K. Fouad, and A. Kibboua, eds., Springer.
- Silva, W. R. L. da, and Lucena, D. S. de. (2018). “Concrete Cracks Detection Based on Deep Learning Image Classification.” *Proceedings*, 2(8), 489.
- Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). “Learning Local Feature Descriptors Using Convex Optimisation.” *Tpami*, 36(8), 1–14.
- Sinha, S. K., and Fieguth, P. W. (2006). “Automated detection of cracks in buried concrete pipe images.” *Automation in Construction*, 15(1), 58–72.
- Smith, J. W. (1969). “The vibration of highway bridges and the effects on human comfort.” University of Bristol, UK.
- Soyoz, S., Dikmen, U., Apaydin, N., Kaynardag, K., Aytulun, E., Senkardasler, O., Catbas, N., Lus, H., Safak, E., and Erdik, M. (2017). “System identification of Bogazici suspension bridge during hanger replacement.” *Procedia Engineering*, Elsevier B.V., 199, 1026–1031.
- Spencer, B. F., Hoskere, V., and Narazaki, Y. (2019). “Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring.” *Engineering*, Chinese Academy of Engineering, 5(2), 199–222.
- Srivastava, V., and Baqersad, J. (2019). “An optical-based technique to obtain operating deflection shapes of structures with complex geometries.” *Mechanical Systems and Signal Processing*, Elsevier Ltd, 128, 69–81.
- Stiros, S. C. (2008). “Errors in velocities and displacements deduced from accelerographs: An approach based on the theory of error propagation.” *Soil Dynamics and Earthquake Engineering*, 28(5), 415–420.

- Sun, D., Roth, S., and Black, M. J. (2010). "Secrets of optical flow estimation and their principles." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2432–2439.
- Sun, D., Roth, S., and Black, M. J. (2014). "A quantitative analysis of current practices in optical flow estimation and the principles behind them." *International Journal of Computer Vision*, 106(2), 115–137.
- Sun, J., Xie, Y., and Cheng, X. (2019). "A Fast Bolt-Loosening Detection Method of Running Train's Key Components Based on Binocular Vision." *IEEE Access*, IEEE, 7, 32227–32239.
- Sutton, M. A., Yan, J. H., Tiwari, V., Schreier, H. W., and Orteu, J. J. (2008). "The effect of out-of-plane motion on 2D and 3D digital image correlation measurements." *Optics and Lasers in Engineering*, 46(10), 746–757.
- Szeliski, R. (2011). "Computer vision: Algorithms and applications." *Springer London*.
- Tian, L., and Pan, B. (2016). "Remote bridge deflection measurement using an advanced video deflectometer and actively illuminated LED targets." *Sensors (Switzerland)*, 16(9), 1–13.
- Tian, Y., Zhang, J., and Yu, S. (2018). "Rapid Impact Testing and System Identification of Footbridges Using Particle Image Velocimetry." *Computer-Aided Civil and Infrastructure Engineering*, 0, 1–16.
- Tian, Y., Zhang, J., and Yu, S. (2019). "Vision-based structural scaling factor and flexibility identification through mobile impact testing." *Mechanical Systems and Signal Processing*, Elsevier Ltd, 122, 387–402.
- Tong, Z., Gao, J., Han, Z., and Wang, Z. (2018). "Recognition of asphalt pavement crack length using deep convolutional neural networks." *Road Materials and Pavement Design*, 19(6), 1334–1349.

- USDOT, and FHWA. (2000). *Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges*.
- VLFeat. (2018). "VLFeat tutorials: SIFT detector and descriptor." <http://www.vlfeat.org/overview/sift.html>.
- Wang, C., Wang, N., Ho, M., Chen, X., and Song, G. (2019). "Design of a New Vision-based Method for the Bolts Looseness Detection in Flange Connections." *IEEE Transactions on Industrial Electronics*, 0046(c), 1–1.
- Wang, Q., Zhang, L., Bertinetto, L., Hu, W., and Torr, P. H. S. (2018a). "Fast Online Object Tracking and Segmentation: A Unifying Approach."
- Wang, Q., Zhang, M., Xing, J., Gao, J., Hu, W., and Maybank, S. (2018b). "Do not lose the details: Reinforced representation learning for high performance visual tracking." *IJCAI International Joint Conference on Artificial Intelligence*, 2018-July, 985–991.
- Wang, T., Celik, O., and Catbas, F. (2016a). "Damage detection of a bridge model based on operational dynamic strain measurements." *Advances in Structural Engineering*, 19(9), 1379–1389.
- Wang, T., Celik, O., Catbas, F. N., and Zhang, L. M. (2016b). "A frequency and spatial domain decomposition method for operational strain modal analysis and its application." *Engineering Structures*, Elsevier Ltd, 114, 104–112.
- Watase, A., Birgul, R., Hiasa, S., Matsumoto, M., Mitani, K., and Catbas, F. N. N. (2015). "Practical identification of favorable time windows for infrared thermography for concrete bridge evaluation." *Construction and Building Materials*, Elsevier Ltd, 101, 1016–1030.
- Wilburn, B., Joshi, N., Vaish, V., Levoy, M., and Horowitz, M. (2005). "High-speed videography using a dense camera array." *Proceedings of ACM SIGGRAPH 2005*, Los Angeles, California,

765–776.

- Wu, L.-J. J., Casciati, F., and Casciati, S. (2014). “Dynamic testing of a laboratory model via vision-based sensing.” *Engineering Structures*, Elsevier Ltd, 60, 113–125.
- Xia, Y., Zhang, P., Ni, Y. Q., and Zhu, H. P. (2014). “Deformation monitoring of a super-tall structure using real-time strain data.” *Engineering Structures*, 67, 29–38.
- Xu, Y., Brownjohn, J., Hester, D., and Koo, K. Y. (2016). “Dynamic displacement measurement of a long span bridge using vision-based system.” *8th European Workshop On Structural Health Monitoring (EWSHM 2016)*, 5-8 July 2016, Spain, Bilbao, (July), 5–8.
- Xu, Y., Brownjohn, J., and Kong, D. (2018). “A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge.” *Structural Control and Health Monitoring*, 25(5), 1–23.
- Xu, Y., and Brownjohn, J. M. W. (2018). “Review of machine-vision based methodologies for displacement measurement in civil structures.” *Journal of Civil Structural Health Monitoring*, Springer Berlin Heidelberg, 8(1), 91–110.
- Xu, Y. L., Ko, J. M., and Zhang, W. S. (1997). “Vibration Studies Of Tsing Ma Suspension Bridge.” *ASCE Journal of Bridge Engineering*, 2(4), 149–156.
- Yamaguchi, T., and Hashimoto, S. (2010). “Fast crack detection method for large-size concrete surface images using percolation-based image processing.” *Machine Vision and Applications*, 21(5), 797–809.
- Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., and Ling, H. (2019). “Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection.” *IEEE Transactions on Intelligent Transportation Systems*, 1–11.
- Yang, X., Li, H., Yu, Y., Luo, X., Huang, T., and Yang, X. (2018a). “Automatic Pixel-Level Crack

- Detection and Measurement Using Fully Convolutional Network.” *Computer-Aided Civil and Infrastructure Engineering*, 33(12), 1090–1109.
- Yang, Y., Dorn, C., Mancini, T., Talken, Z., Kenyon, G., Farrar, C., and Mascarenas, D. (2017a). “Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification.” *Mechanical Systems and Signal Processing*, Elsevier, 85, 567–590.
- Yang, Y., Dorn, C., Mancini, T., Talken, Z., Nagarajaiah, S., Kenyon, G., Farrar, C., and Mascarenas, D. (2017b). “Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-Nyquist), video measurements.” *Journal of Sound and Vibration*, Elsevier, 390, 232–256.
- Yang, Y., Dorn, C., Mancini, T., Talken, Z., Theiler, J., Kenyon, G., Farrar, C., and Mascareñas, D. (2018b). “Reference-free detection of minute, non-visible, damage using full-field, high-resolution mode shapes output-only identified from digital videos of structures.” *Structural Health Monitoring*, 17(3), 514–531.
- Yang, Y., and Nagarajaiah, S. (2015). “Dynamic Imaging: Real-Time Detection of Local Structural Damage with Blind Separation of Low-Rank Background and Sparse Innovation.” *Journal of Structural Engineering*, 142(2), 04015144.
- Yang, Y. Sen, Wu, C. lin, Hsu, T. T. C., Yang, H. C., Lu, H. J., and Chang, C. C. (2018c). “Image analysis method for crack distribution and width estimation for reinforced concrete structures.” *Automation in Construction*, Elsevier, 91(May 2017), 120–132.
- Yang, Y. Sen, Yang, C. M., and Huang, C. W. (2015). “Thin crack observation in a reinforced concrete bridge pier test using image processing and analysis.” *Advances in Engineering Software*, Elsevier Ltd, 83, 99–108.

- Ye, X. W., Dong, C. Z., and Liu, T. (2016a). "Image-based structural dynamic displacement measurement using different multi-object tracking algorithms." *Smart Structures and Systems*, 17(6), 935–956.
- Ye, X. W., Dong, C. Z., and Liu, T. (2016b). "Force monitoring of steel cables using vision-based sensing technology : methodology and experimental verification." *Smart Structures and Systems*, 18(3), 585–599.
- Ye, X. W., Dong, C. Z., and Liu, T. (2016c). "A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications." *Journal of Sensors*, 2016.
- Ye, X. W., Dong, C. Z., and Liu, T. (2016d). "Environmental effect on vision-based structural dynamic displacement monitoring." *Proceedings of the Second International Conference on Performance-based and Life-cycle Structural Engineering (PLSE 2015)*, J. L. T. Dilum Fernando, Jin-Guang Teng, ed., Brisbane, Australia, 261–265.
- Ye, X. W., Jin, T., and Chen, P. Y. (2019). "Structural crack detection using deep learning-based fully convolutional networks." *Advances in Structural Engineering*.
- Ye, X. W., Liu, T., Dong, C. Z., and Chen, B. (2016e). "FBG-Based Structural Performance Monitoring and Safety Evaluation of a Composite Arch Bridge." *Structural Health Monitoring of Composite Structures Using Fiber Optic Methods*, CRC Press, 269–297.
- Ye, X. W., Liu, T., and Ni, Y. Q. (2017). "Probabilistic corrosion fatigue life assessment of a suspension bridge instrumented with long-term structural health monitoring system." *Advances in Structural Engineering*, 20(5), 674–681.
- Ye, X. W., Ni, Y. Q., Wai, T. T., Wong, K. Y., Zhang, X. M., and Xu, F. (2013). "A vision-based system for dynamic displacement measurement of long-span bridges: algorithm and verification." *Smart Structures and Systems*, 12(3–4), 363–379.

- Ye, X. W. W., Yi, T. H. T. H., Dong, C. Z. Z., and Liu, T. (2016f). "Vision-based structural displacement measurement: system performance evaluation and influence factor analysis." *Measurement: Journal of the International Measurement Confederation*, Elsevier B.V., 88, 372–384.
- Ye, X. W., Yi, T.-H., Dong, C. Z., Liu, T., and Bai, H. (2015). "Multi-point displacement monitoring of bridges using a vision-based approach." *Wind and Structures, An International Journal*, 20(2), 315–326.
- Yeum, C. M., and Dyke, S. J. (2015). "Vision-Based Automated Crack Detection for Bridge Inspection." *Computer-Aided Civil and Infrastructure Engineering*, 30(10), 759–770.
- Yi, T. H., Li, H. N., and Gu, M. (2013). "Experimental assessment of high-rate GPS receivers for deformation monitoring of bridge." *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, 46(1), 420–432.
- Yin, Z., Wu, C., and Chen, G. (2014). "Concrete crack detection through full-field displacement and curvature measurements by visual mark tracking: A proof-of-concept study." *Structural Health Monitoring*, 13(2), 205–218.
- Yokoyama, S., and Matsumoto, T. (2017). "Development of an Automatic Detector of Cracks in Concrete Using Machine Learning." *Procedia Engineering*, 171, 1250–1255.
- Yoneda, M. (2002). "A simplified method to evaluate pedestrian-induced maximum response of cable-supported pedestrian bridges." *Proceedings of the International Conference on the Design and Dynamic Behaviour of Footbridges*, Paris, France.
- Yoneyama, S., and Kitagawa, a. (2007). "Bridge deflection measurement using digital image correlation." *Experimental Techniques*, (February), 34–40.
- Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M., and Spencer, B. F. (2016). "Target-free

- approach for vision-based structural system identification using consumer-grade cameras.” *Structural Control and Health Monitoring*, 23(1), 1405–1416.
- Yoon, H., Hoskere, V., Park, J.-W., and Spencer, B. F. (2017). “Cross-Correlation-based structural system identification using unmanned aerial vehicles.” *Sensors (Switzerland)*, 17(9).
- Yoon, H., Shin, J., and Spencer, B. F. (2018). “Structural Displacement Measurement Using an Unmanned Aerial System.” *Computer-Aided Civil and Infrastructure Engineering*, 33(3), 183–192.
- Yu, S. N., Jang, J. H., and Han, C. S. (2007). “Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel.” *Automation in Construction*, 16(3), 255–261.
- Yu, Y., Zeng, W., Liu, W., Zhang, H., and Wang, X. (2019). “Crack Propagation and Fracture Process Zone (FPZ) of Wood in the Longitudinal Direction Determined Using Digital Image Correlation (DIC) Technique.” (Dic).
- Zaurin, R., and Catbas, F. (2010a). “Structural health monitoring using video stream, influence lines, and statistical analysis.” *Structural Health Monitoring*, 10(3), 309–332.
- Zaurin, R., and Catbas, F. (2010b). “Integration of computer imaging and sensor data for structural health monitoring of bridges.” *Smart Materials and Structures*, 19(1), 1–15.
- Zaurin, R., Khuc, T., and Catbas, F. N. (2015). “Hybrid Sensor-Camera Monitoring for Damage Detection: Case Study of a Real Bridge.” *Journal of Bridge Engineering*, 21(6), 1–27.
- Zhang, B., Zhou, L., and Zhang, J. (2019a). “A methodology for obtaining spatiotemporal information of the vehicles on bridges based on computer vision.” *Computer-Aided Civil and Infrastructure Engineering*, 34(6), 471–487.
- Zhang, D., Guo, J., Lei, X., and Zhu, C. (2016a). “A High-Speed Vision-Based Sensor for Dynamic Vibration Analysis Using Fast Motion Extraction Algorithms.” *Sensors*, 16(4), 572.

- Zhang, K., Zhang, L., Yang, M.-H., and Zhang, D. (2013). “Fast Tracking via Spatio-Temporal Context Learning.” *arXiv preprint arXiv:1311.1939*, 1–16.
- Zhang, L., Yang, F., Zhang, Y. D., and Zhu, Y. J. (2016b). “Road crack detection using deep convolutional neural network.” *IEEE International Conference on Image Processing*.
- Zhang, Y., Sun, X., Loh, K. J., Su, W., Xue, Z., and Zhao, X. (2019b). “Autonomous bolt loosening detection using deep learning.” *Structural Health Monitoring*.
- Zhang, Z. (2002). “A Flexible New Technique for Camera Calibration (Technical Report).” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.
- Zhao, X., Ri, K., and Wang, N. (2017). “Experimental Verification for Cable Force Estimation Using Handheld Shooting of Smartphones.” *Journal of Sensors*, 2017.
- Zheng, F., Shao, L., Racic, V., and Brownjohn, J. (2016). “Measuring human-induced vibrations of civil engineering structures via vision-based motion tracking.” *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, 83, 44–56.
- Zhong, F., Kumar, R., and Quan, C. (2019a). “RGB laser speckles based 3D profilometry.” *Applied Physics Letters*, 114(20), 1–5.
- Zhong, F., Kumar, R., and Quan, C. (2019b). “A cost-effective single-shot structured light system for 3D shape measurement.” *IEEE Sensors Journal*, IEEE, 19(17), 7335–7346.
- Zhong, F. Q., He, S. P., and Li, B. (2017). “Blob analyzation-based template matching algorithm for LED chip localization.” *International Journal of Advanced Manufacturing Technology*, The International Journal of Advanced Manufacturing Technology, 93(1–4), 55–63.
- Zhong, F. Q., Indurkar, P. P., and Quan, C. G. (2018a). “Three-dimensional digital image correlation with improved efficiency and accuracy.” *Measurement: Journal of the International Measurement Confederation*, Elsevier, 128, 23–33.

- Zhong, F. Q., and Quan, C. G. (2017). "Digital image correlation in polar coordinate robust to a large rotation." *Optics and Lasers in Engineering*, 98, 153–158.
- Zhong, F. Q., and Quan, C. G. (2018a). "A single color camera stereo vision system." *IEEE Sensors Journal*, IEEE, 18(4), 1474–1482.
- Zhong, F. Q., Shao, X. X., and Quan, C. (2018b). "3D digital image correlation using a single 3CCD colour camera and dichroic filter." *Measurement Science and Technology*, IOP Publishing, 29(4).
- Zhong, F. Q., Shao, X. X., and Quan, C. (2019c). "A comparative study of 3D reconstruction methods in stereo digital image correlation." *Optics and Lasers in Engineering*, 122(May), 142–150.
- Zhong, F., and Quan, C. (2018b). "Efficient digital image correlation using gradient orientation." *Optics and Laser Technology*, 106, 417–426.
- Zhong, X., Peng, X., Yan, S., Shen, M., and Zhai, Y. (2018c). "Assessment of the feasibility of detecting concrete cracks in images acquired by unmanned aerial vehicles." *Automation in Construction*, Elsevier, 89(February), 49–57.
- Zhu, Z., German, S., and Brilakis, I. (2011). "Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation." *Automation in Construction*, Elsevier B.V., 20(7), 874–883.
- Živanović, S., and Pavia, A. (2009). "Probabilistic assessment of human response to footbridge vibration." *Journal of Low Frequency Noise Vibration and Active Control*, 28(4), 255–268.
- Živanović, S., Pavic, A., and Reynolds, P. (2005). "Vibration serviceability of footbridges under human-induced excitation: A literature review." *Journal of Sound and Vibration*, 279(1–2), 1–74.

Zou, Q., Cao, Y., Li, Q., Mao, Q., and Wang, S. (2012). “CrackTree: Automatic crack detection from pavement images.” *Pattern Recognition Letters*, 33(3), 227–238.