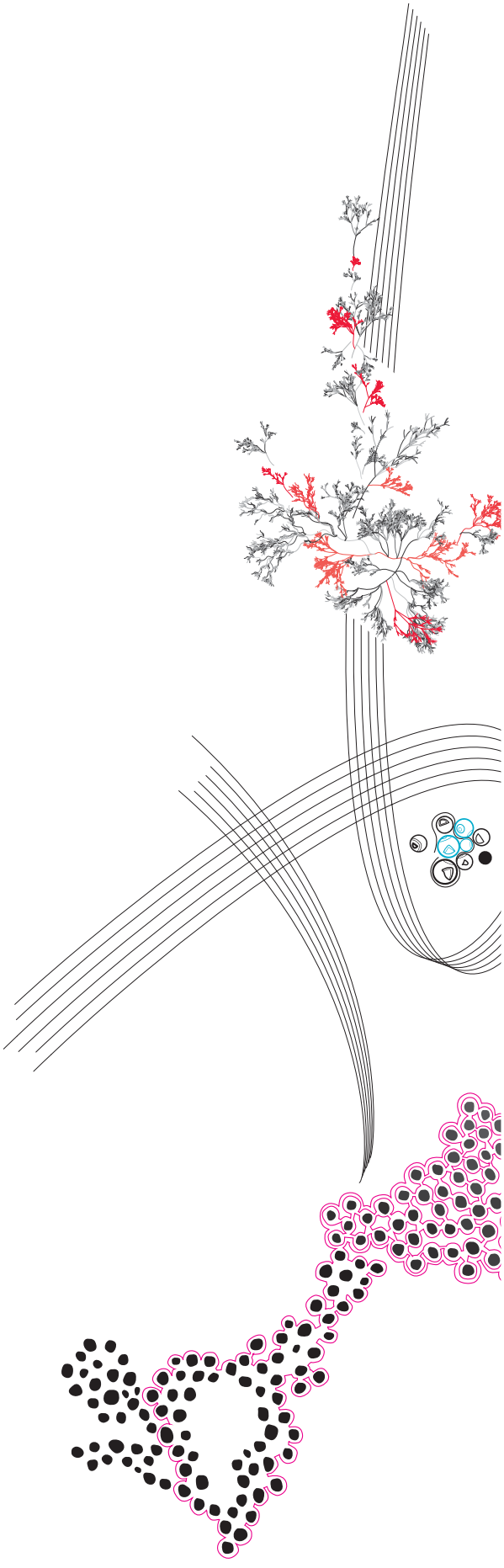

Thesis report

Investigating limitations of computer vision structural health monitoring: A case study on the UT campus footbridge

Mikolaj Turkow
S-2309688

Supervisor: dr. ing. R. Kromanis

Co-supervisor: ing. M. Xofi

Final Report

08 July 2022

BSc-thesis CE
2021-202000085-2B

UNIVERSITY OF TWENTE.

Preface

The research carried out in this report concludes my three-year long journey with University of Twente and BSc Civil Engineering study. The bachelor thesis is titled *Investigating limitations of computer vision-based structural health monitoring: A case study on UT Campus footbridge.* The project has been initiated in consultation with dr. ing. Rolands Kromanis in winter of 2022. After many meetings and discussions regarding potential Bachelor Assignments, we have decided that I will support research efforts in the field of computer vision and structural health monitoring at the University of Twente.

The project has been carried out both from the Netherlands and Czech Republic. The data has been collected on the UT campus at the Dynamic's Lab of Mechanical Engineering Department as well as on the UT Campus footbridge located between U Parkhotel and UTrack Sintelbaan. Majority of data processing and report writing has been completed remotely from Czech Republic. Unfortunately, my work on the project has been interrupted due to contracting COVID-19; however despite a small delay I was able to complete the research project on time.

I would like to thank my supervisor Rolands Kromanis, who over the many meetings has introduced me to the topic of computer vision-based SHM, provided constructive feedback and advice throughout the entire project. I would also like to say thank you to my co-supervisor Maria Xofi, who was always eager to help, both with data collection as well as with shaping the report. At the same time I want to acknowledge the help of Axel Lok with lab measurements and Dr Dario Di Maio from Computer Vision, Laser Vibrometry and DIC group. I also owe a special thank you to Rubi, Mihai and Mohsen, who helped with 'exciting' the UT campus footbridge on an exceptionally hot day.

I would like to dedicate this work to my parents who made this whole adventure happen in the first place.

Mikolaj Turkow
July 2022, Prague, CZ

Table of Contents

Abstract

Computer vision-based solutions are a promising tool for structural health monitoring (SHM) and asset management of civil engineering infrastructure. Computer vision (CV) systems offer many advantages over existing sensor technologies: non-contact, long-distance and precise measurements without complex installations, with lower cost and labor intensity. However, the robustness of CV-SHM remains significantly unexplored. Insights into limitations and challenges of various CV-SHM methods must be investigated, before automated input-output systems can be developed for quick analysis and asset management. The paper presents the results of literature review about applications of computer vision in short and long term bridge monitoring and damage detection. Moreover, paper presents findings from two sets of measurements collected with a use of consumer grade cameras (*GoPro Hero 5/8*) to extract structural displacement of 1) aluminum beam in a laboratory set-up and 2) steel girder footbridge subjected to a variety of different loading scenarios. The cameras are modified with a mounted varifocal zoom lenses recording at high resolution (720 x 1280 px) and high frame rate (240 FPS). Based on the data collection process and analysis of the results a set of conclusions is drawn, especially with regard to limitations of proposed computer vision-based structural health monitoring methodology and its potential in dynamic analysis, namely identification of modal shapes. The proposed methodology has proven successful in identification of three natural frequencies in laboratory setting within 0 - 120 Hz frequency spectrum. In the field conditions, the researcher was able to identify the natural frequencies at lower bandwidth (0 – 5 Hz). Both in laboratory and field environment, it has been proven successful to extract the dynamic signature of the structure in the form of 1st and 2nd mode shapes. The results obtained via image processing, namely template matching, are compared against accelerometry findings. No significant differences have been found between accelerometers and image processing. Furthermore, limitations in terms of hardware specification, environmental factors, operational conditions, and others are discussed.

Keywords: Computer vision • Structural health monitoring • Bridge deflection • Mode identification • Experimental modal analysis • Limitations

1 Introduction

Civil infrastructure such as buildings, bridges, wind turbines towers, dams or others are exposed to a variety of external loads throughout their lifetime. Although both live and dead loads are taken into account during their design, effects of loads over long periods of time might result in structural damages such as corrosion, cracking, overall degradation among others. Through the use of SHM overall structural health can be monitored and operational safety can be ensured by providing real-time data for structural assessment. Application of SHM solutions are especially desirable for condition assessment and management of bridges. Majority of bridges in the Netherlands and Europe was built decades ago, notably between 1950 and 1970 (Lourens et al. n.d.). Many of these bridges have already reached end of their designed life or will soon do. Furthermore, the design standards and requirements based on which these bridges have been designed have changed considerably since then, as today's traffic loads from EN 1991-2, no longer reflect the loading scenarios from the 1950-1970 period (NEN Connect 2019).

The SHM has been evolving over the decades and maturing with development and lowering of costs of relevant technologies. One of the strategies to capture bridge behavior characteristics is a CV based measurement approach. The approach for measuring displacement-based structural characteristics has gained a lot of attention and publicity within the academia and in practice in the recent years (Kromanis and Kripakaran 2021; Xu, Brownjohn, and Kong 2018; Lydon et al. 2019). Video cameras can be easily and inexpensively used to collect information remotely from structures, even through the use of modern smartphones that are easily accessible in the current era of technology. Computer and vision-based monitoring offers many advantages relative to conventional inspection methods, which are costly and often infeasible due to difficult access to certain structural elements of the bridge or requiring interference to the daily operation of the structure. Furthermore, deploying a dense network of sensors throughout the bridge can be quite costly and comes with tedious maintenance work besides the troublesome installation and reoccurring malfunctions. Huge amount of data collected on daily basis can be a challenge in itself, as the 'big data' has to be properly managed and analyzed in order to draw relevant and informative conclusions or maintenance recommendations. Vision-based approach provides the opportunity of long distance, non-contact, low cost and low labor intensity as complementary and supportive tool for short-term maintenance planning, design verification, assessment of bridge's performance or prediction of structure's potential lifespan (Dong and Catbas 2021).

The planned thesis project has been proposed by University of Twente (UT) with aim to support research efforts concerning computer vision-based (CV) structural health monitoring (SHM) systems. The project has been initiated in consultancy with dr.ing. Rolands Kromanis, assistant professor at the UT with research focus on resilient

engineering, computer vision, and SHM data collection and interpretation. The research process is overseen by ing. M. Xofi, PhD researcher at UT working on bridge monitoring and development of sensor technologies for structural condition and damage assessment of bridges in the Netherlands.

This paper studies dynamic bridge response in order to extract its dynamic properties (natural frequency, mode shapes) from data collected through CV approach. The method is first applied in laboratory setting on the aluminum beam. The data is analyzed and insights gained are later utilized on real-life footbridge (UT Campus bridge) case in order to investigate its dynamic properties and potential limitations of modal analysis of associated method in a field setting.

1.1 Problem statement

Recent advancements in CV SHM approaches including development of low-cost and high-quality cameras, image processing algorithms and computer vision software present a promising method for bridge characterization and asset management. However, limitations of aforementioned approach exist. In order to develop a standardized strategy for vision-based SHM and make informed decisions based on reliable data, influence factors and relations between camera set-up and resulting displacement-derived structural characteristics (modal shapes) must be explored. At the current stage, the opportunities and limitations of CV-SHM have not been fully investigated in the field conditions on operational bridges. The research aims to support this direction of CV-SHM.

1.2 Goals & research objectives

The planned research and accompanying experiments will be carried out on the UT campus, specifically on the footbridge. Furthermore the research not only combines in-situ methods, but also makes use of experimental methods conducted in controlled environment of the UT Department of Mechanical Engineering's Dynamics lab.

The goal of the research is to establish and investigate limitations of CV SHM for measuring bridge's dynamic response and its potential in modal shape identification. The UT campus bridge will serve as the test-bed for this study, and aluminum beam in the laboratory is used for method validation purposes.

The overall motivation for this study is to support vibration field tests for modes shapes identification and characterization. The research objectives can be summarized as presented below:

1) Identify main natural frequencies and associated vibration modes in:
    a. laboratory aluminum beam,
    b. UT Campus bridge.

2) Determine accuracy relative to accelerometry of the CV hardware system in controlled environment.

3) Investigate limitations of CV approach in the identification of vibration mode and frequencies.

1.3 Research Questions

The following research questions follow from the stated objectives, which the presented report tries to answer:

- Which vibration modes with their natural frequencies can be identified through proposed computer vision-based SHM method?
    - in a lab setting
    - in UT Campus bridge

- What are potential limitations of vision-based SHM method on operational footbridge measurements?

- To what extent can modes of vibration be identified in operational footbridges?

1.4 Thesis scope and organization

The scope of the thesis revolves around remote displacement measurements, which are used to derive dynamic characteristic of bridges – namely, the modal parameters. The research focuses on studies related to identification of natural frequencies and their associated mode shapes. During the analysis dynamic response is of the main interest.

The report is organized into six chapters. A literature review is presented in Chapter 2. The review is split into three main sections, first one introduces the topic of structural dynamics, next past and recent advancements in CV related to displacement measurement are presented and the most important findings are summarized in the last section – conclusions. Chapter 3 describes the proposed methodology and explanation of image processing technique. A detailed experimental setup together with results of the two studied cases – aluminum beam in a lab and UT Campus footbridge are presented in Chapters 4 and 5 respectively. The chapters introduce both of the cases as well as present the results. In Chapter 5, a discussion is presented. Finally, Chapter 6 summarizes the findings as set of conclusions as well as points out future research directions.

2 Literature Review

The study combines testing of structures both in laboratory and field conditions as well as a literature review with respect to applications of computer vision in structural health monitoring and dynamic analysis of civil infrastructure, especially bridges. This chapter first introduces the topic of structural dynamics, then presents advancements in CV approaches in SHM, especially modal analysis.

2.1 Structural dynamics

Structural dynamics describes and studies the effects of external and dynamic forces and loads that induce high acceleration vibrations on structural systems. Forces and loads that vary in time are referred to as dynamic forces. Dynamic loads acting on a bridge can be induced by people, traffic, winds, or earthquakes. The need for dynamic assessment and analysis due to moving loads was motivated by the rise of rail transport, and erection of many bridges that came with it. Moving loads (such as a pedestrian or a truck traversing through the structure) have a great effect on dynamic stresses within the structure and result in intensive vibrations (Fryba 1999). The effects of footsteps or traversing vehicles are dissipated within the bridge's structure in form of vibrations, through resonant vibrations. Resonance vibrations can cause significant damage to the structure and shorten their lifespan if not designed appropriately. In the past a number of fatal failures has occurred caused by resonance response, e.g. the infamous collapse of Tacoma highway bridge in 1940 or Chester rail bridge in 1947 among others.

Nowadays, structural dynamics and analysis are very important with applications in multiple fields from car manufacturing ensuring sufficient comfort of driving, through design of earthquake resistant buildings, to safe operation of aircrafts, or deep sea oil platforms withstanding wave action. In these and many other fields of engineering, structural dynamics and analysis play a key role in design, testing and functioning of various products and solutions, through process known as structural characterization. Structural characterization refers to the process in which physical quantities such as accelerations, displacements, or strains are measured in order to obtain a qualitative and quantitative assessment of the tested structure. One of the key dynamic properties that provide insight into condition and behavior of a structure are modal parameters – natural frequency with an associated mode shape and damping value.

Except for the direct dynamic loads, the bridge response is driven by environmental conditions such as humidity and temperature. The same modal parameters can be sensitive to changes in environmental factors, especially temperature. This adds to the complexity of measuring and defining the response of the bridge in form of natural frequencies and mode shapes, which can not only be affected by local damage or change in boundary conditions, but also temperature, which in most of the global regions changes quite significantly on average with seasons (Zolghadri et al. 2015).

Nowadays, measurement of bridge vibration is incorporated in regulations and standards of many countries – natural frequency, mode shape, the dynamic coefficient, and damping. All of the aforementioned properties describe the dynamic bridge characteristic. These properties, thanks to years of research, can be measured and verified via in-situ experimental methods. Traditionally the response of the structure due to natural or experimental excitements was measured using dense sensor networks (accelerometers, strain gauges, etc.); however in recent years with growing interest and development of image processing technologies, a new direction of bridge monitoring and testing is on the rise – computer vision-based SHM (CV-SHM).

2.2 Computer vision

When studying civil infrastructure, especially long spanning structures such as towers or bridges, deflections are of key interest when measuring the structural health and behavior. Through deflection measurements, one can investigate both the static and dynamic responses of the structure. Static response is caused by e.g. pedestrian or vehicle moving through the bridge.

As mentioned before bridge's are subjected to continuous loading and environmental conditions. With time the damage, aging, and environmentally-induced deterioration may result in change of the physical and structural properties of a bridge (e.g. damping, stiffness or deflection). The fundamental idea behind measuring bridge's vibrations is to track these changes or abnormal responses of bridge's vibration in aforementioned properties. When such (significant) change is detected, inspection and appropriate maintenance works can be scheduled to manage a given asset.

However, when it comes to modal identification, through CV based approaches, there remains a significant knowledge gap; possibilities as well as limitations of such methods have not been fully researched and identified. Modal shapes and their corresponding frequencies are known based on series on experiments conducted in the past for structures with well-defined boundary conditions. For instance, pin-pin beam modal vibrations have been identified and normalized to better visualize and understand their shapes and relations, as shown in Figure 1 below. When modal shapes are known, annual measurements can be conducted with purpose of damage identification by means of modal shape analysis. In this study, the research focuses on vibration measurements and visualization of modal shapes of footbridges through CV methods.

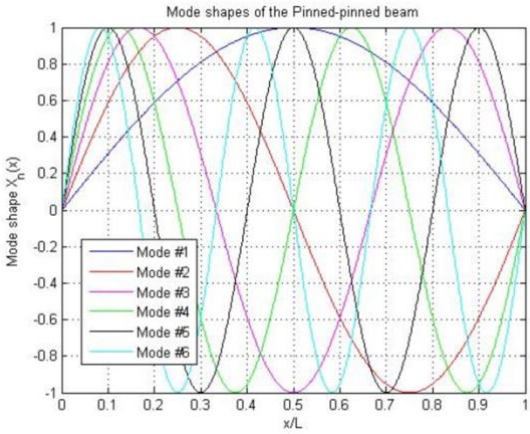

Figure 1 Mode shapes of the pinned-pinned beam, from (Eschkabilov, 2011)

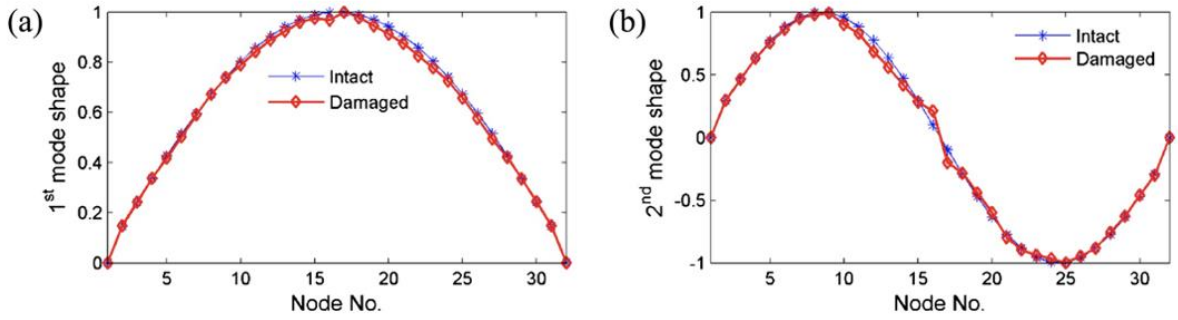

Figure 2 Identified two mode shapes: (a) 1st mode shape and (b) 2nd mode shape, from (Feng, 2016)

There exists a number of non-contact sensing methods, which include Laser Doppler Vibrometry (LDV), synthetic aperture radar (SAR), ultra-sound or camera based systems (Shang and Shen 2017). In this study camera based system is considered. Initially the most common CV-SHM methodology required manual installation of artificial markers, combined with various tracking algorithms. The principle is quite simple, a marker (of known dimensions and shape) is manually attached to relevant structural component and a camera mounted on a tripod is focused on the target. Since the size and geometry of the marker are known, the displacements of the bridge's structure can be directly measured from the marker. This approach requires less advanced tracking algorithms and offers confident results; however as it requires manual installation of the markers at measured locations, the benefit of ease and convenience that non-contact methods offer is lost, especially at difficult to reach parts of the structure (Jiang, Jáuregui, and White 2008). New direction of research towards bridge engineering and monitoring uses distinctive features of the structure to track and measure its displacement. As tracking algorithms, image processing and hardware technologies have matured over the years, possibilities to extract 2D displacement no longer requires use of artificial markers. Distinctive features of the bridge's structure are chosen as targets such as bolts, deck components or cables among others. This method offers truly non-contact procedure for extracting bridge's deflection under one significant condition – it requires evident and notable movements

of chosen targets. This can be a major drawback as some types of bridges depending on design have lower deflection limits, while others have higher degree of deflection (e.g. cable stayed bridges). Through the use of CV methods even the tiniest structural responses can be tracked; when it comes to footbridges it can be as small as a fraction of a millimeter, resulting in sub-pixel level changes (Kromanis 2021).

As mentioned in the previous paragraph, CV-SHM methods make use of various image-processing and signal interpretation technologies. In this section the overview of main image processing algorithms is presented. Through research and field tests, two most prevalent algorithmic methods have been identified: digital image correlation (DIC) and feature matching (Lydon et al. 2019; Xu and Brownjohn 2018). Both methods require fixed camera locations and notable distortion deformations or displacements. DIC compares two digital images of the selected target at different stages of displacement by tracking blocks of pixels and computing the difference between them (McCormick and Lord 2010). Feature matching technique uses existing characteristic features in the bridge's structure to track pixel movements, which are later converted to engineering units. In this study template matching is deployed, used in previous studies (Kromanis, 2021; Voordijk & Kromanis, 2022).

The final data collection step concerns camera calibration, which is the main source of the measurement error in the CV based approaches. The camera is sensitive to background vibrations, which result in camera movements. Actual structural response must be separated from noisy measurements. This can be achieved by applying various statistical techniques and decomposing camera movements from the measurements to determine actual structural deflection at the target location.

Recorded displacement data is represented as time series. In order to gain insights into dynamic properties of the structure, the data must be first transformed from the time domain to the frequency domain. This is achieved using fast Fourier Transformation (FFT) to obtain frequency spectrum, represented by power spectral density function (PSD). From the plot main frequencies and frequency range of bridge vibrations can be identified. Since the expected displacements are really tiny – as small as fraction of a millimeter, logarithmic scale is applied in order to identify dominant frequencies.

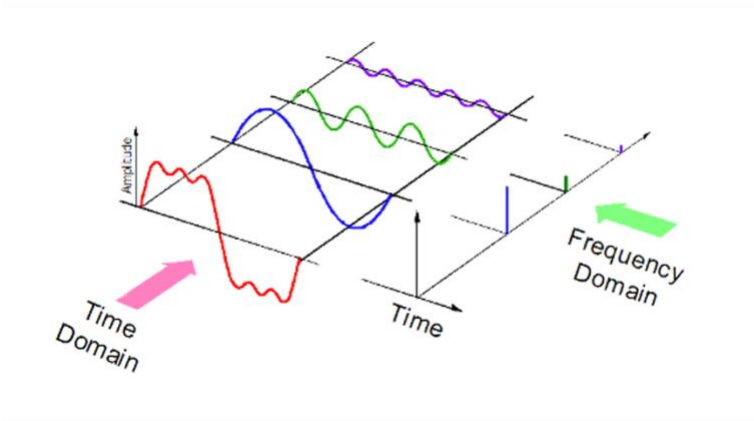

Figure 3 Fourier transform visualized, from (Chen, 2021)

The research makes use of mode-shape-based methods, which can be used for structural health monitoring purposes. Modal shape is an unique characteristic of all structures, which can determined based on dynamic vibrations measurements. Mode shapes are distinctive to their 'natural frequency' and represent a particular displacement pattern of the structure through its entire length, thus providing a global structural health characteristic. Mode shapes can be used to assess structure's health and identify damage at early stage. If a structure deforms under the loads and stresses or suffers fatigue or other type of damage, it is expected to see changes in the modal shape in vicinity of the damage location (Dackermann 2009). The method described in the next section aims to capture dynamic vibrations and resulting modal shapes.

The natural frequency is relatively simple to identify, it is represented by distinctive peak in PSD function plot. Such plot is visualized via *MATLAB* graphing tools, a peak is determined and corresponding frequency and its amplitude are recorded. The frequency range and resolution determine the accuracy of frequency measurement and thus potential for modal shape identification. Some modes of vibration occur at lower frequencies, while others at higher magnitudes; generally civil structures such as bridges have lower natural frequencies, which are easier to capture (Allada, Saravanan, and Mariani 2021). The lower the natural frequency range, the lower the required rate of recording of the hardware used to capture the vibrations. Similarly the higher the rate of recording (more FPS), the more detailed and accurate response measurement and greater range of frequency domain.

2.3 Conclusions

The review of the technical literature of the computer vision-based laboratory and field applications presented in this paper led to the formulation of the following overview and conclusions.

Published works in the field of CV-SHM can be split into two main categories: (1) field measurements of displacement time histories or (2) measurements and validation of displacement time histories of various small scale models or prototypes in controlled laboratory environments. The capabilities and accuracy of the vision-based systems has been thoroughly tested in various experiment setups proving the CV successful in recording of displacement measurements and modal analysis.

Although many different image processing algorithms and approaches coexist within the literature with various advantages over each other, the most prevalent and commonly deployed image processing algorithm are template matching, feature matching, and optical flow.

The most commonly used method of validation of CV-SHM systems is through comparison against data obtained through deployment of conventional contact sensors such as accelerometers or linear variable differential transformer (LVDT). Accelerometers are deployed at the same location as targets tracked through the cameras, and resulting measurements histories are transformed (usually via FFT) to a frequency domain. In majority of the reviewed studies the comparisons of results between image processing (various algorithms) and accelerometers have shown no significant differences; Zona has reviewed sixteen case studies, which have shown good correlation between image processing algorithm and other sensing technologies (mainly accelerometers and strain gauges), only in one case the differences were not negligible (2020). Furthermore, accelerometers were used to assess the error due to camera movements by attaching one directly onto a camera and later removing the camera noise, to obtain actual structural displacement.

With respect to deployed cameras, which together with image processing algorithm are one of the most important part of CV system. Cameras act as sensors, which transform light into electrical signal, which is displayed as image. Researchers have used different types of cameras ranging from regular smartphone cameras, through low-cost action cameras, to more expensive specialized super slow motion cameras.

Another way to split recent advancements in CV-SHM is to distinguish between local and global levels of monitoring. At local level CV-SHM can be used for early damage detection of various structural components; such detection may include: cracking, spalling, rusting, or loose bolts (Dong and Catbas 2021). SHM at global level refers to measurement of parameters that reveal whole structure behavior, especially related to structural response in form of displacement and displacement-derived data.

3 Methodology

The methodology borrows and builds upon research methods and strategies from existing literature on CV field testing approaches, especially (Lydon et al. 2019; Voordijk and Kromanis 2022; Kromanis 2021).

An overview of basic methodology process is visualized in figure 4. A detailed description of experimental setup together with equipment list and procedural steps for data collection can be found in Chapters 4 & 5. Once the displacement histories have been extracted and transformed into engineering units via template matching technique, the plots are further transformed into frequency domain via FFT in order to reveal patterns, which are difficult to be distinguished with simple visual inspection. In cases where in all four measurement locations a given natural frequency with associated signal peaks is identified, it is possible to reveal the modal shapes. In this study the shapes are determined by polynomial curve fitting methods. Once the curve is fitted to the four data points and its equation is known with resulting data set points, it is normalized between 0 and 1 based on the following formula:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where: z_i is the ith normalized value in the dataset
x_i is the ith value in the dataset
min(x) is the minimum value in the dataset
max(x) is the maximum value in the dataset

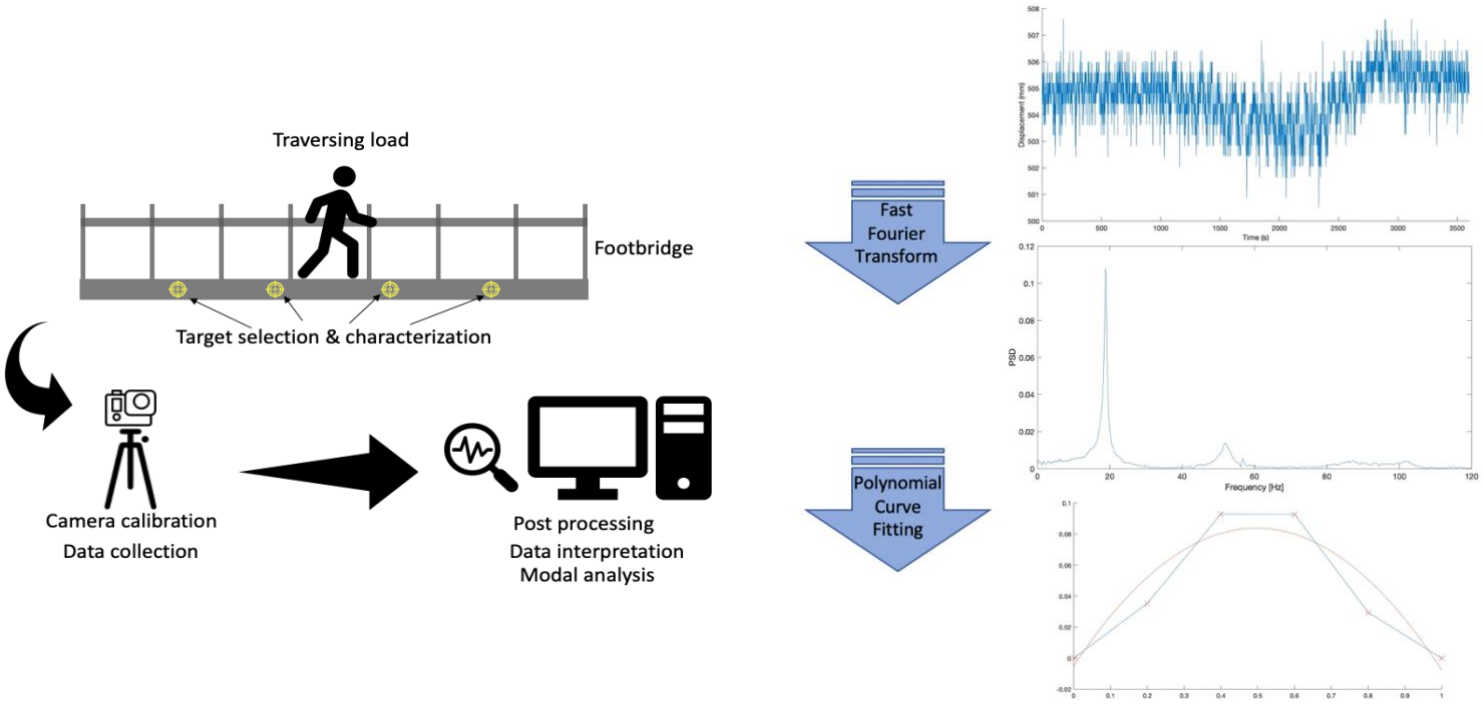

Figure 4 (left) Basic process for CV-SHM and (right) steps for modal shape identification

3.1 Image processing

As mentioned previously a number of different image processing methods exist with various advantages and shortcomings between them. Different computer vision studies have compared their accuracies and reliability based on analysis of different measurements. In this study in particular image processing algorithm by (Kromanis) is used. The algorithm has been created in *MATLAB* environment and requires installation of *Computer Vision Toolbox* for its operation (MATLAB 2021.). Furthermore, the algorithm makes use of template matching technique. First a starting reference frame is selected from the video. It is desired for the reference frame to represent the structure at its resting state. The position of the 'template box' is tracked frame by frame, and corresponding x and y coordinates are saved as time series history, and later used for further processing and analysis. Image processing steps for generating displacement are visualized in Figure 5. An example of target characterization for template matching technique is shown in Figure 6.

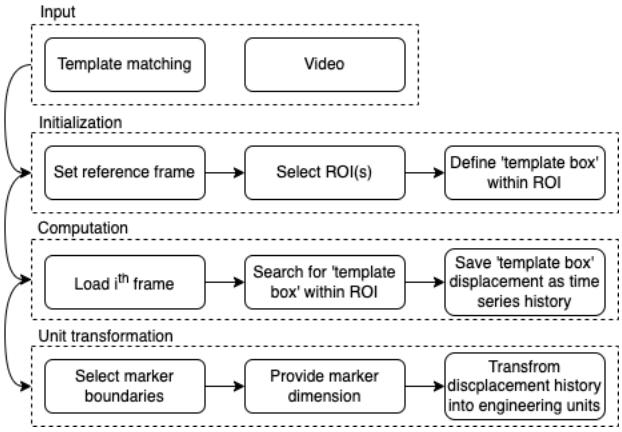

Figure 5 Flowchart of template matching

Template matching manifests limitation in terms of computational resource intensity. The high computational cost is a result of two factors. Often, due to complexity of recorded image and objects, multiple templates must be selected within the frame to arrive with accurate measurements. Secondly, the resolution of the recorded video, which refers to the number of pixels per x and y direction; the larger the number of pixels (higher resolution), the greater the computational cost of running the algorithm (Brunelli 2009). The computation time of the algorithm is dependent on few factors, related to both the video characteristics and configuration of template matching. Firstly, the length of the video, as the longer the video or the period of interest the more frames, one by one, must be processed. Next, the size of the selected ROI, as the larger the ROI, the bigger the size of the image that the algorithm must 'search' through for the selected 'template box', frame by frame. Finally, the algorithm's search resolution can put a significant strain in terms of computational cost.

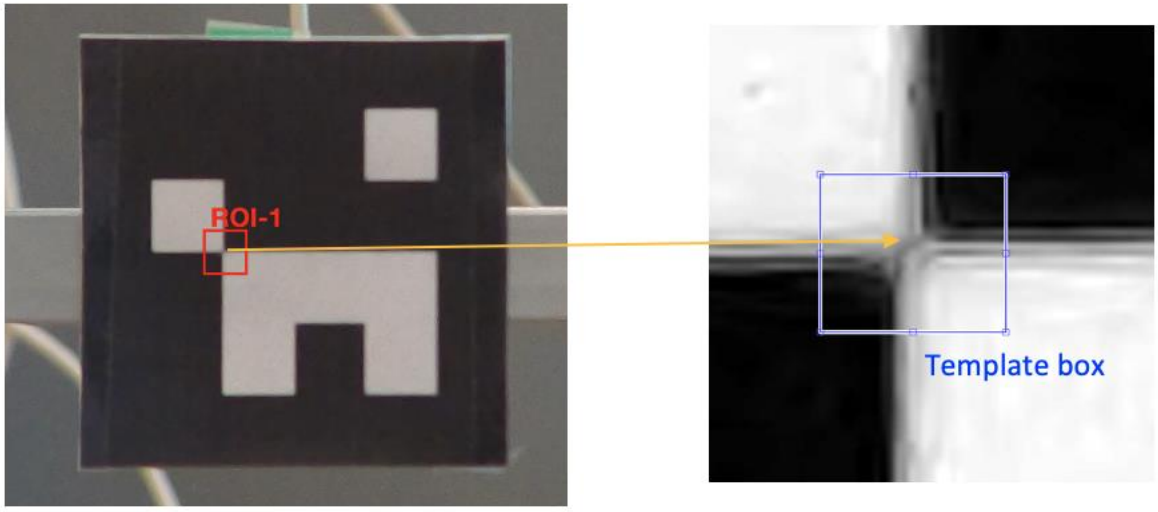

Figure 6 Image characterization: (left) reference frame with chosen ROI and (right) selected template box within ROI for template matching processing

4 Aluminum beam in a lab

This section provides overview of the experimental set-up and methodology for measurements on the aluminum channel section beam in a laboratory setting.

4.1 Experimental set-up

The first stage of this study involves experimental investigation on the aluminum channel section beam which is fixed via makeshift clamp supports. The support is meant to resemble pin-roller or pin-pin support. The laboratory set-up acts as testing grounds with purpose of laying a basis for field testing; it provides grounds for validation of proposed CV method. The experimental set-up can be seen in figure 7 below.

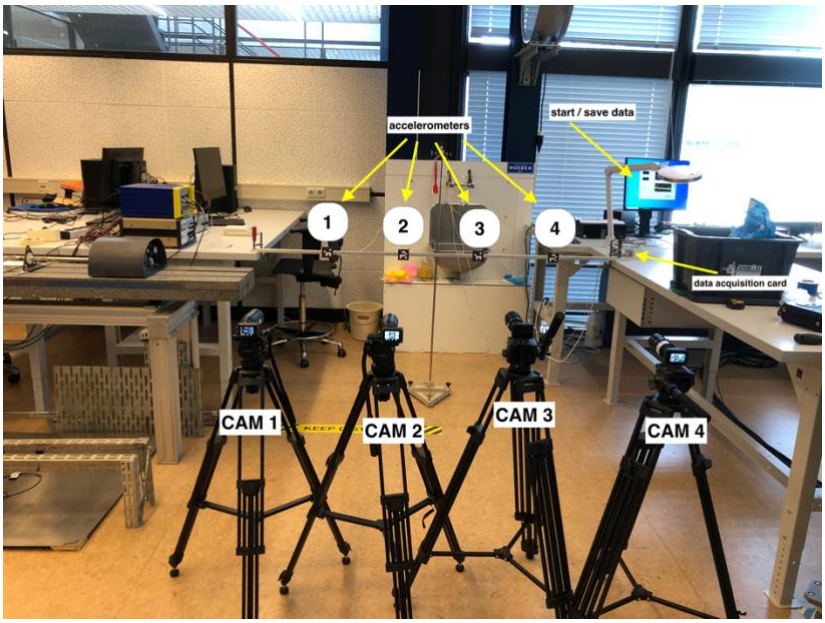

Figure 7 Experimental test set-up

The markers are first attached on the side of the beam section, with equal spacings every 400 mm. Next, four accelerometers are connected to *National Instruments* data acquisition unit and attached on the beam with honey wax at the same locations as markers. The aluminum section beam is firmly fastened at both ends to two tables with two clamps. The zoom cameras are set-up approx. 1.5 m away on tripods at the same height level as the beam section. The beam is struck with a rubber end of metal rod between the support and first marker. The cameras recording is synced via Remote Control, the accelerometers measurement is automatically initiated upon beam excitation.

4.1.1 Aluminum beam

The tested channel section beam is made out of aluminum with dimensions 20 mm by 30 mm by 2,000 mm. Aluminum is a light and ductile material: it can undergo large deformations before a fracture occurs. Aluminum's modulus of elasticity is 70GPa, whereas for steel it is three times larger – 210GPa. This means that "in elastic stage the deformations of an aluminum structure are about three times as large as deformations of the same [steel] structure" (Hartsuijker & Welleman, 2007, p. 9). This makes it a perfect experimental material, as responses due to the forces applied (and induced vibrations) are the largest, i.e. easier to detect via CV approach and investigate potential limitations in modal analysis and consequently in identification of vibration modes shapes. Furthermore, the channel section vaguely resembles the structure of bridge deck with railing on each side.

4.1.2 Accelerometers

Acceleration of the beam was measured with four low cost isotron accelerometers of model 256-100 with sensitivity of 9.191, 10.20, 9.41, 10.62 mV/m/s^2 respectively. The accelerometers are attached in same locations as artificial markers in order to record the dynamic response at the same location for data quality assessment and validation. Accelerometers measure the response of the beam due to induced vibrations, namely acceleration forces, both static (constant force of gravity) and dynamic (moving or vibrating the accelerometer) components. Processed acceleration data is transformed into frequency domain via Fast Fourier Transform (FFT) with automatic built in feature from *National Instruments* data collection and acquisition system.

4.1.3 Computer vision system

The computer vision system consists of two *GoPro HERO 8* with Computar H6Z0812 8-48mm 1:1.2 lens and two *GoPro 5* with Computar MEGAPIXEL f=25-155mm 1:1.8 1/1.8C lens. All four of the cameras are mounted on a tripod. The camera recording synchronization is achieved by pairing the cameras together via *Smart Control* remote. Moreover the system consists of four artificial markers obtained from existing *ArUco* library ("Online ArUco Markers Generator" 2022). The markers are printed in black & white with dimensions of 57 x 57 mm, attached to a piece of firm cardboard. As the measurements are conducted indoors the use firm cardboard is sufficient to ensure stability of the marker during the measurements. The markers are equally spaced out along the length of the beam. As four markers are used, they are distributed every 400 mm on the side of the beam, as shown in figure 4. In order to achieve optimal zoom and image sharpness, the camera-to-beam distance is 1,5 m. The cameras are positioned at the same height level as the markers via tripod regulation. The horizontal stability is ensured via built in level scale in the tripods. The final element of the CV system is the image processing algorithm, which is explained in section 3.1 Image processing.

Table 1 Inventory specification, lab setting

Equipment	Quantity	Specifications
Channel section beam	1	2m long Channel approx. 20x30 mm Aluminum
GoPro HERO 8	2	240 FPS 1080p
GoPro Hero 5	2	240 FPS 720 x 1280 px
GoPro Smart Remote	1	-
Varifocal zoom lens	4	2x Computar MEGAPIXEL f=25-155mm 1:1.8 1/1.8C 2x Computar H6Z0812 8-48mm 1:1.2
Tripod	4	-
Isotron Accelerometer	4	92.28 mV/g (or 9.410mV/m/s^2) Model 256-100
Data acquisition system	1	by *National Instruments*
Modal hammer	1	-
Artificial marker	4	From ArUco library 57 x 57 mm look *Appendix A*
Image processing system	1	look *Image Processing*

4.2 Results

A total of five experiments was carried out on the aluminum beam, with a use of four cameras to track the motion of four equally spaced out markers, resulting in analysis of twenty videos in total. The total size of the 20 lab videos was 959,2 MB with an average length of approximately 6 seconds per video.

The displacement histories generated via template matching and scale transformation to engineering units (mm) for experiment 4 are presented in figure 4 for cameras one to four (markers one to four). From the displacement histories, the moment of excitation is clearly visible, indicated with a downward drop (as the beam is struck from above). Then the beam dynamic response can be traced, until it dampens to pre-excitation noise levels.

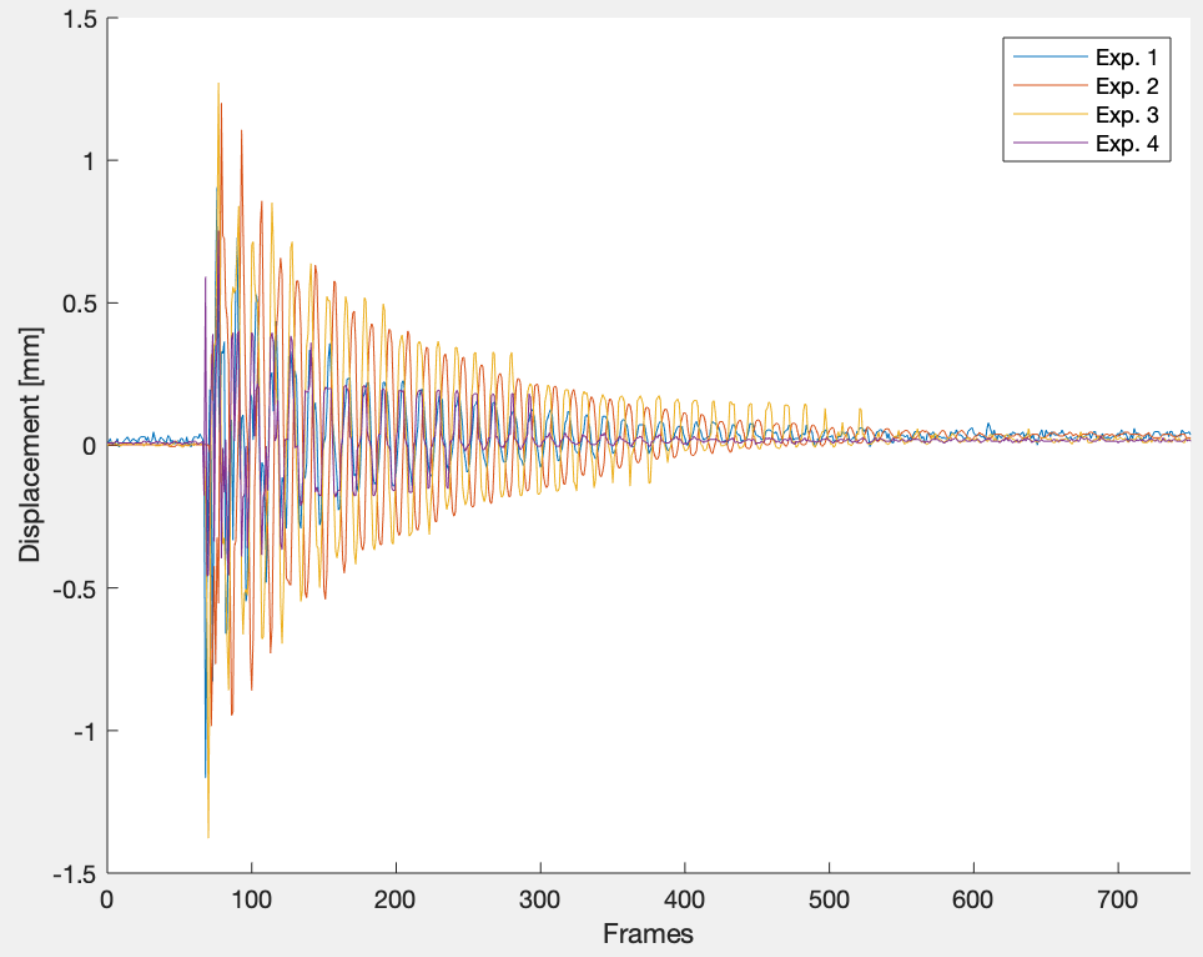

Figure 8 Displacement histories per camera for experiment 4

From the displacement history dampening of the vibration pattern can be observed. Through further analysis and transformation of displacement measurements into a frequency domain, particularly via FFT, more information can be accessed with regard to displacement patterns. The total duration of recorded signals is around 5 seconds.

The frequency bandwidth of the generated PSD plot is 120 Hz. On the specified range, three distinct signal peaks can be observed, corresponding to the first three modes of vibration. The frequency values associated with the peaks, represent natural frequencies of the tested aluminum beam. Figure 9 shows all of the individual signal peaks, which indicate natural frequencies - 18.75 Hz, 51.75 Hz and 102.8 Hz respectively. The data of identified natural frequencies for each experiment per camera is shown in Appendix B. As can be noted from these tables, the recorded frequency tables exhibit value variability, when theoretically all of the frequency values for a given peak should be the same. By analyzing the identified frequencies with basic descriptive statistics, the standard deviation was determined for modes 1 to 3 – 0.118 Hz, 0.192 Hz and 0.530 Hz as summarized in Table 2. It can be noted, that the frequency differences between experiments are larger when the mode is higher. This means that higher natural frequencies and higher modes are more difficult to be accurately measured.

Table 2 Identified Natural Frequencies from the laboratory, based on 5x4 measurements

Frequency [Hz]	Standard deviation [Hz]	# of successful identifications (max 20)
18,75	0,118	20
51,75	0,192	20
102,8	0,530	13

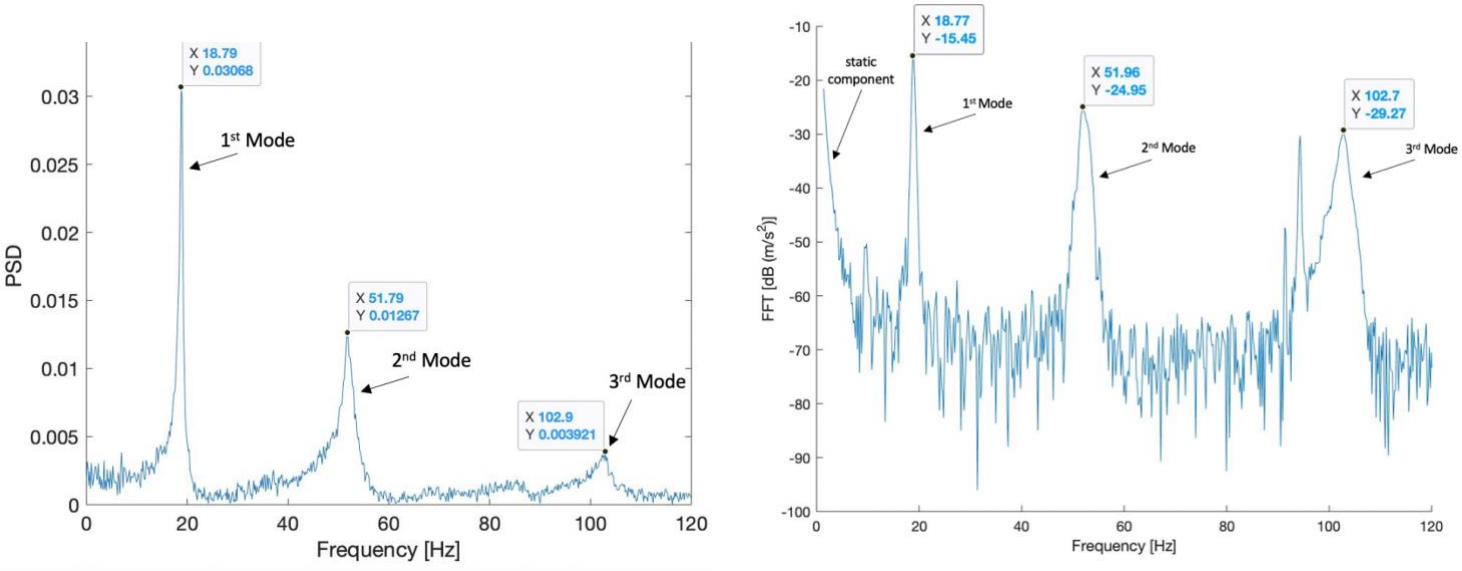

Figure 9 PSD plots generated from (left) displacement and (right) acceleration histories via FFT, experiment 3

From figure 9, it can be noted that both the image processing and accelerometers have successfully identified first three natural frequencies. A direct comparison is made of natural frequencies values identified from PSD plots between accelerometry and image processing data, and an absolute difference is computed. This is summarized in table 3, from the table it can be observed that the natural frequency value differ in all of the cases. However, the differences are quite low and comparable to the differences between various experiment trials obtained with image processing. The average absolute difference for identified natural frequencies from third experiment, between template matching and accelerometry is 0,16 Hz, which can be considered negligible difference. For the first frequency this is only a relative error of 0,85%, for the higher frequencies this would be even smaller. The maximum and minimum difference is 0,19 Hz and 0,02 Hz respectively.

Table 3 Natural frequencies identified, image processing vs accelerometry, laboratory

Frequency	Image processing [Hz]	Accelerometry [Hz]	Abs. difference [Hz]
f_1	18.79	18.97	0.18
	18.75	18.77	0.02
	18.75	18.77	0.02
	19	18.77	0.23
f_2	51.79	51.96	0.17
	51.75	51.96	0.21
	51.75	51.96	0.21
	51.75	51.96	0.21
f_3	102.9	102.7	0.2
	x	102.7	x
	x	102.7	x
	102.8	102.7	0.1

In the experimental cases, where in all four marker locations, the frequency peaks have been identified from FFT plots, the line scatter plots are generated, which reveal the outlines of the modal shapes. For f_3, the mode shape was not identified as the PSD plots generated from displacement history, has not picked up the natural frequency and its amplitude in all four marker location in all experiments. Similarly, for the 2nd mode shape, the first experiment has proven unsuccessful in picking up the f_2 and its amplitude, thus only experiments two to four are presented. In the figure 10, an example of polynomial curve fitting is presented. The four data points are plotted with a line scatter, which reveals the rough outline of the mode shape. In the next step a polynomial trendline is fitted to the data, for first mode this is a 2nd degree polynomial, for second this is 3rd degree, etc. Based on the equation of the curve, a new data set is generated, which is later normalized on 0 – 1 interval based on equation 1 from Chapter 3: Methodology.

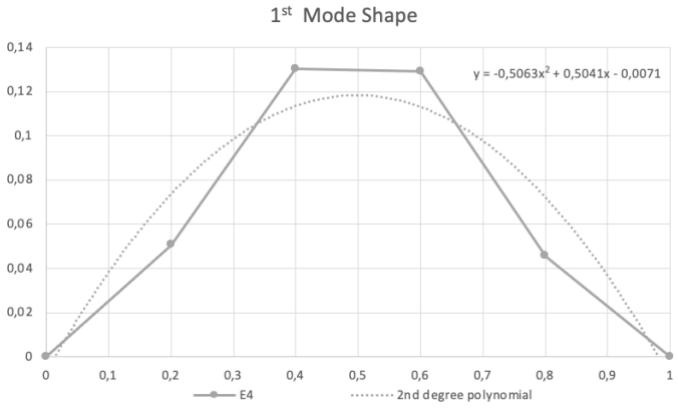

Figure 10 Polynomial curve fitting for 1st Mode, experiment 4

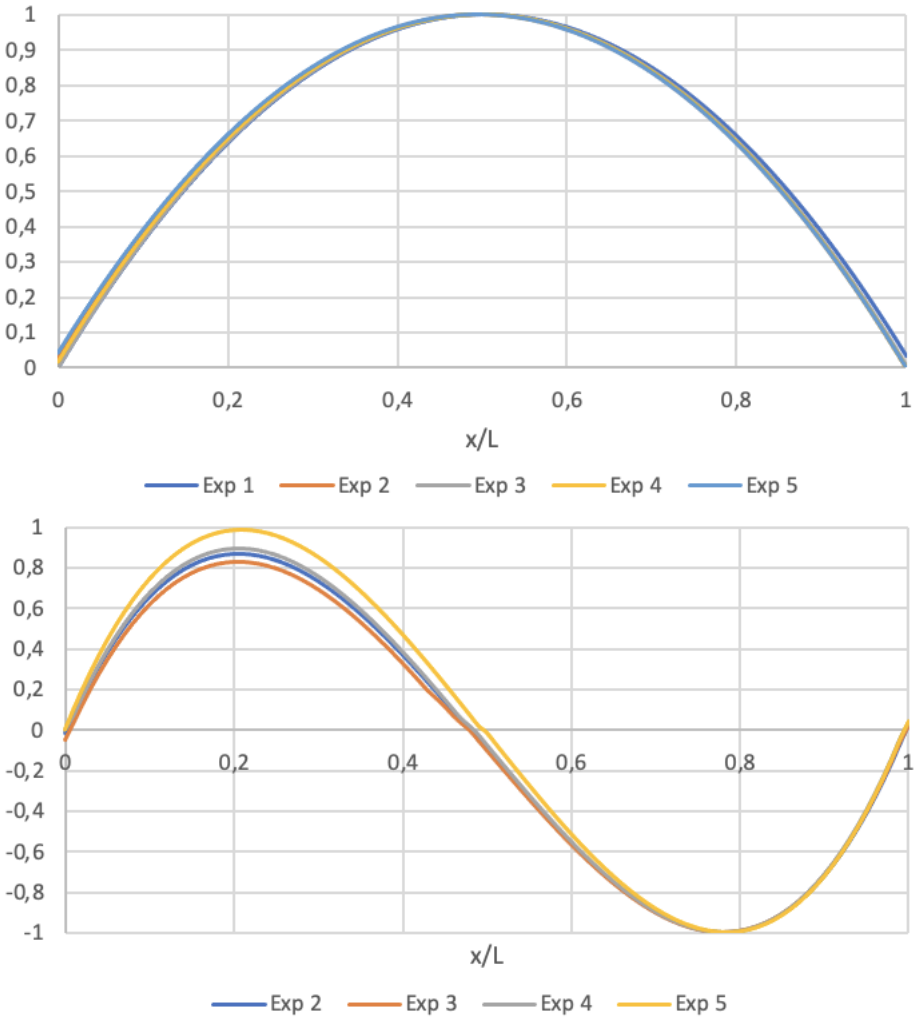

Figure 11 Normalized 1st & 2nd mode shapes of fixed-fixed beam

In ideal theoretical scenario, the resulting normalized mode shapes for all five experiments should look exactly the same; however from figure 11 it is apparent, that it is not the case. For the first vertical mode shape, the results are quite close to achieve it - all five normalized curves almost overlap each other, there is very small difference between all five experiments. For the second vertical mode shape, these differences are much larger, especially on the 0 – 0,5 interval. Moreover, only the mode shape resulting from experiment 5 (yellow curve) attains both the values of 1 and -1; the other three experiments have much lower degree of horizontal symmetry.

5 UT Campus footbridge

The second stage of the methodology involves field testing, where the UT Campus footbridge serves as the test-bed. The bridge has been a study subject in Kromanis (2021) and (Voordijk and Kromanis 2022; Kromanis 2021) as well as testing grounds for MSc and PhD students from the UT with focus on Smart infrastructure and CV-SHM.

5.1 Experimental setup

Figure 12 (left) UT Campus footbridge from above and (right) up-close

The UT campus bridge is a simple steel girder and timber deck and railing bridge. The structure consists of three steel girders, on which rests the timber deck. A total of twenty-two timber railing posts is bolted to both of outer girders. The footbridge is 2 m wide and spans 27 m over a man-made canal. During the measurements the temperature was approximately between 25°C and 27°C.

The very same computer vision system is used as in the laboratory, with exception for artificial markers, which in this case are bigger (10 cm x 10 cm). Artificial markers are attached to a small right angle iron plate via magnets. The four plates are then firmly attached with zip ties to the railing posts of the footbridge. Similarly to laboratory set-up, the goal was to distribute the markers as evenly as possible, the markers were attached at 5.5 m, 11 m, 16.6 m, and 22.1 m of footbridge's length, which corresponds to normalized lengths of 0.20, 0.41, 0.61, and 0.82. This is almost identical marker placement as the one implemented on the aluminum beam in laboratory. The cameras were located next to each other on South-Eastern side, pointed at markers one to four, where the first marker was the closest to the camera location. The deck of the

footbridge is split into three axis – left side, center, and right side as shown in Figure 12. The markers are attached to the railing posts on the right side.

Table 4 Inventory specification, UT campus footbridge

Equipment	Quantity	Specifications
GoPro HERO 8	2	240 FPS 1080p
GoPro HERO 5	2	240FPS 720p
Smart Control remote	1	-
Varifocal zoom lens	4	2x Computar MEGAPIXEL f=25-155mm 1:1.8 1/1.8C 2x Computar H6Z0812 8-48mm 1:1.2
Tripod	4	-
Large zip ties	8	-
Iron angle	4	-
Magnet	4	-
Scale	1	Simple bathroom scale was used
Marker	4	1000 x 1000 mm look *Appendix A*
Image processing algorithm		look *Image Processing*

A total of four experimental cases was carried out as summarized in Table 5. All of the cases were conducted on each of the axis of the bridge as shown in figure 12 – left side, center, and right side. The loading in kilograms was measured with a simple bathroom scale. The crossings are initiated in a sequence one after another, from both sides of bridge, from northeastern and southwestern sides. The cameras are synced via Smart Remote and recording is initiated few second before the subject enters the footbridge and is terminated few seconds after the subject has passed through the footbridge.

Table 5 Footbridge load cases

Case #	*Loading [kg]*	*Description*
1	88.75	Jogging left, centre, right
2	108.10	Cycling left, centre, right
3	243.75	Jogging as group left, center, right
4	243.75	Jumping as group at midspan left, centre, right

5.2 Results

A total of thirty-six videos from the UT Campus footbridge have been analyzed with the template matching image processing algorithm. As a result three natural frequencies, corresponding to vertical modes, were identified, within 5 Hz bandwidth. The displacement time histories for four selected cases are presented on the next page in figure 13. The displacement patterns differ significantly between the different cases, meaning the bridge's dynamic response varies depending on excitation method. In majority of the analyzed videos, the static component is difficult to extract, with the exception for cycling cases, in which the static response can be quite easily spotted, especially if a moving average was to be applied; however the static response was not in the scope of this study. When more load is applied (jogging group) the oscillating displacements are of much higher magnitude, relative to single person jogging through the structure. From the simple visual inspection the displacement pattern seems quite stochastic, in order to reveal the response patterns, the displacement time histories are again, analyzed with FFT to obtain PSD plot on frequency domain. Based on first few initial transformations and visual inspection of the resulting signals, it has been found that only at lower frequency, clear signal peaks are picked up. Therefore, the presented PSD plots are terminated at 5 Hz. The first two signal peaks correspond to the first two vertical modal shapes – 1st Mode at 2.60 Hz, and 2nd Mode at 3,23 Hz. The frequencies in the table 6 are chosen based on the median value. A PSD plot in which all three natural frequencies have been identified is presented in figure 14.

Table 6 Identified Natural Frequencies on UT Campus bridge, based on analysis of 9x4 measurements

Frequency [Hz]	Standard deviation [Hz]	# of successful identifications (max 36)
2,60	0,183	27
3,23	0,136	20
3,68	0,0537	12

In case of UT Campus footbridge, the trend is quite different, the first natural frequency represents the largest standard deviation of 0,183 Hz, the standard deviation then decreases with larger frequencies – opposite to what was observed in the laboratory, where the standard deviation increased with higher frequencies. In significant proportion of analyzed videos and resulting PSD plots, the first natural frequency was represented with two peaks almost at the same frequency (look fig. 14). This explains the largest standard deviation as well the largest range of 0,629 Hz. Jogging and cycling measurements were closer to median value of 2,60 Hz, for jumping, the peak was on average closer to 2,30 Hz, with lowest peak frequency at 2,101 Hz (Jumping group right). Detailed overview of all of the recorded natural frequencies can be found in *Appendix C*.

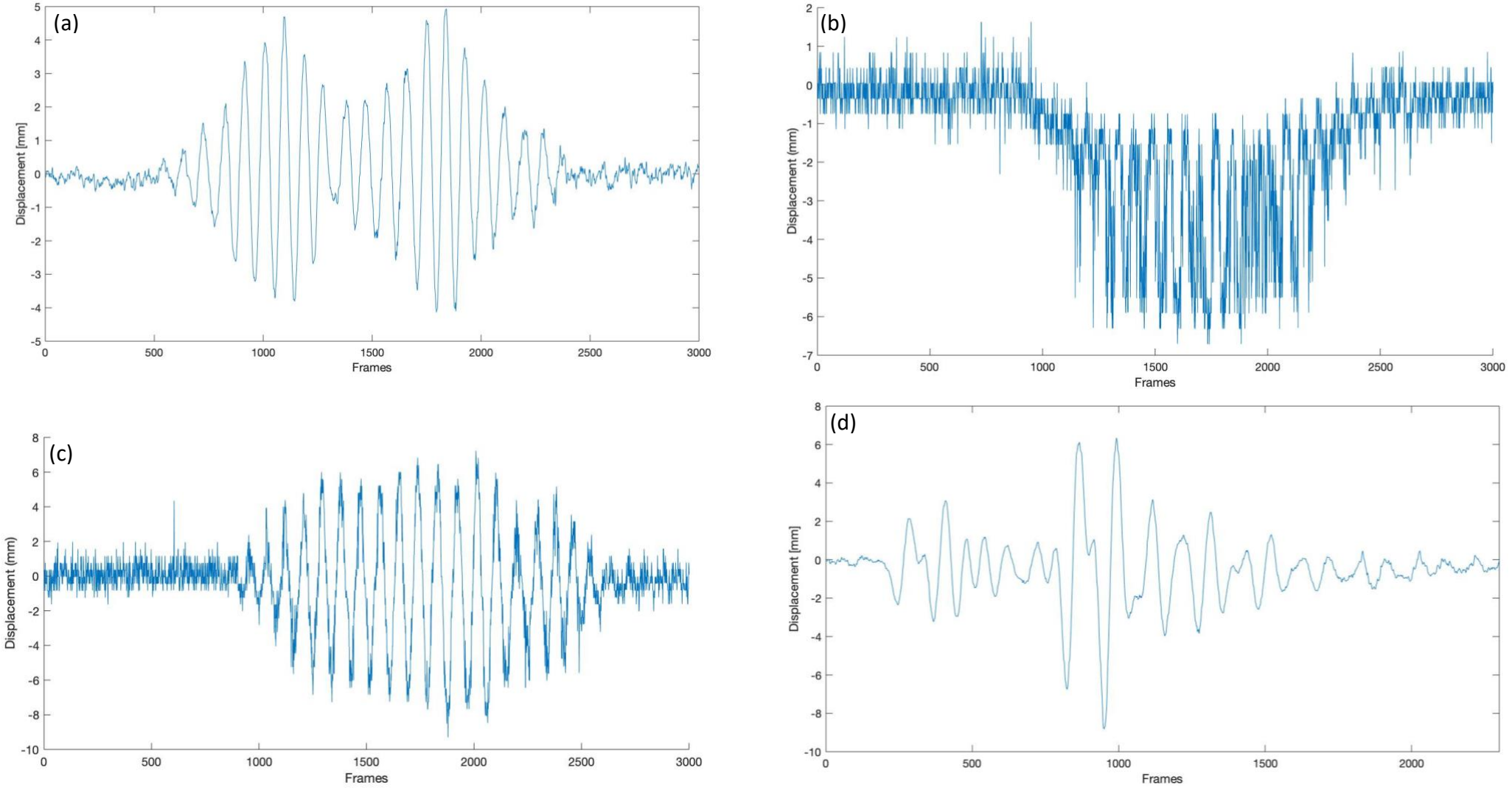

Figure 13 Displacement histories for selected cases, (a) Jogging Group Left (b) Cycling Right (c) Jogging Center (d) Jumping Group Right

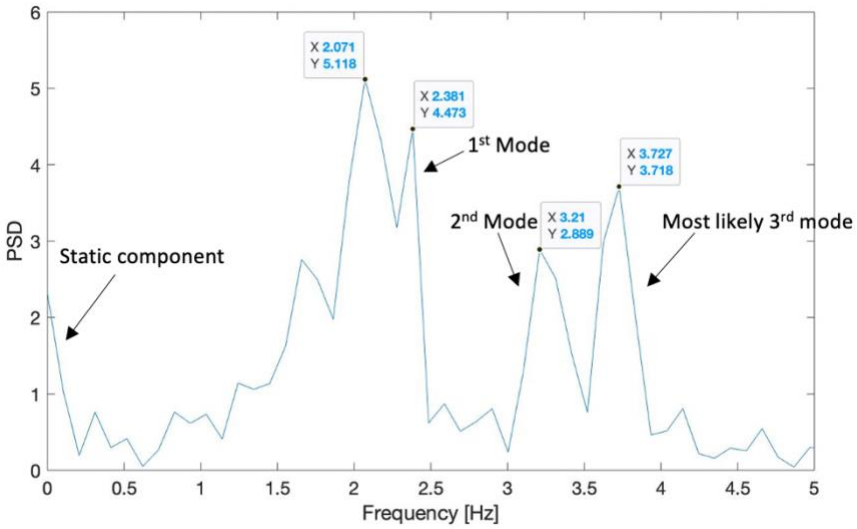

Figure 14 PSD plot generated via FFT from Jumping Group Right

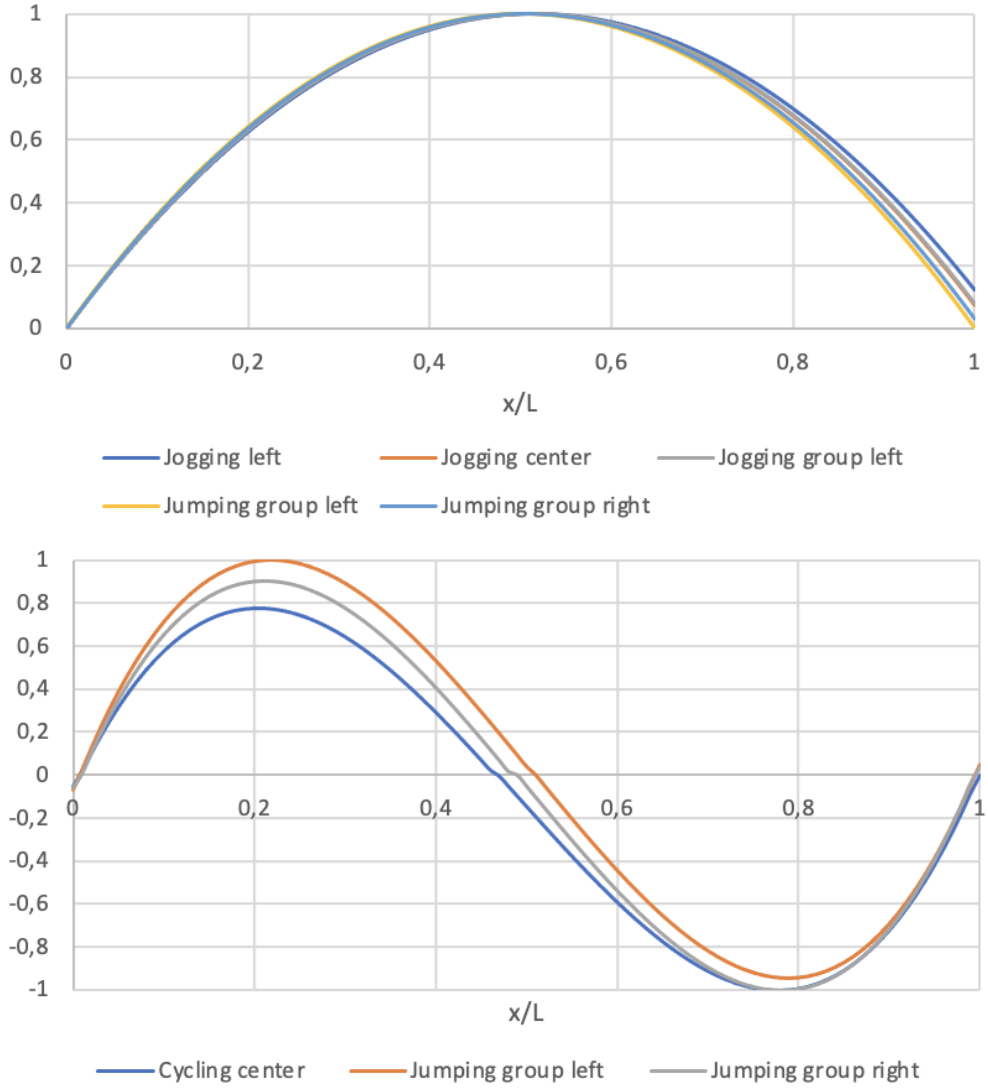

Figure 15 Normalized 1st & 2nd mode shapes of UT Campus footbridge

From all of the analyzed videos, the following cases have yielded identification of mode shapes. Similarly as in the lab, the normalized curves show discrepancies between each other, especially for the second vertical mode. However, the curves associated with the first modal shape can be considered as quite satisfying, especially on 0 – 0,5 interval. At support in all cases the value is 0 and the curves overlap almost perfectly over each other reaching the maximum value of 1. On the 0,5 – 1 interval, the curves no longer coincide and value at the support differs from 0. This interval is fitted based on measurements from third and fourth markers, which were located the furthest away from the camera position. Moreover, the image quality of the fourth marker was the worst out of all recorded cases. This could partially explain the differences on 0,5 – 1 interval for the first mode shape. The normalized curves for the second mode shape vary even more between each other – similarly as in the lab. Higher modes could be more difficult to identify, as they manifest themselves with lower displacement magnitudes relative to the first mode.

5.3 Comparison with parallel study

Parallel to this study fellow BSc Civil Engineering student has tested the potential of using a smartphone, namely its built in accelerometer feature, in identification of natural frequencies of the same UT Campus footbridge as part of his thesis work. The researcher would walk across the UT Campus footbridge with the smartphone. The obtained results, in form of identified natural frequencies are presented in Table 7. The table contains natural frequencies identified in all three planes – x, y, and z directions, frequencies in vertical direction are highlighted with green color. The results from smartphone accelerometer show greater potential in picking up the natural frequencies relative to CV method as the number of successful identifications is greater. Moreover, the researcher has distinguished between 2,41 Hz and 2,79 Hz as separate frequencies. The results from image processing for that frequency range were often overlapping between each other or in many cases only peaks closer to 2,79 Hz were identified, thus treated as single frequency, rather than two separate ones. The natural frequencies identified at 3,29 Hz and 3,73 Hz, are quite comparable to results obtained via image processing – 3,23 Hz and 3,68 Hz, with smaller standard deviations than the ones from table 7. This proves that modern computer vision-based modal testing can match the accuracy of accelerometers, as the differences between the two methods can be considered negligible.

Table 7 Identified Natural Frequencies from walking across the footbridge, smartphone accelerometer

Frequency [Hz]	Standard deviation [Hz]	# of successful identifications (max 30)
2.41	0.086	22
2.79	0.065	20
3.29	0.135	29
3.73	0.109	30
4.31	0.129	30
4.74	0.076	22
5.59	0.097	29

6 Discussion

During the data collection of experimental measurements on the UT Campus bridge, although it was a day with near perfect weather conditions, wind, with occasional stronger gusts was present. A phenomena that does not occur at the laboratory setting, can pose quite a challenge in the field conditions. A lot of bridges and other civil infrastructure are situated at locations known for windy conditions, such as bodies of water (rivers, canals), highways among open fields or hilly landscapes, or within cities, where due to high-rise buildings, the wind can be accelerated significantly. This can cause quite a nuisance, or even all together prevent data collection, as cameras are sensitive to unwanted movement or drift caused by weather conditions.

In order to achieve measurement synchroneity, the cameras must be synced and paired together to ensure recording time alignment. This was achieved via the use of GoPro's *Smart Remote*. Although usually the cameras are recording in close proximity to one another, in operational conditions, it could be desirable to place the cameras on both ends of the bridge, in order to be closer to targets/features distributed along the structure's length, to acquire a sharper image quality. Although the manufacturer claims operational range of approximately 60 m, in this study, after multiple efforts, it was not possible to pair the cameras between both sides of the canal banks (approx. 30m) (GoPro 2022). Thus, the cameras were located on one side of the footbridge.

Personal data protection and privacy laws may interfere with data collection process. Depending on the structure, environment around, and set-up, the cameras may capture not only the specific parts of the bridge, but also people and commuters in the fore/background. In this study, it was possible to wait for moments, where no one was crossing the bridge; however this is usually not the case, especially in urban zones and cities. Presented experimental cases and recorded subjects, provided a permission to store the videos for analysis purposes; thus there was no privacy conflict.

Collecting a series of different measurements can be a matter of day's work, however the management and analysis of collected data can take much longer and is an overall tedious and complex process. In order to extract and obtain an accurate representation of bridge's signature many experimental cases must be carried out and recorded, which can results in hundreds of videos that require a thorough analysis. This poses a limitation, as such data can be not only troublesome to process, but also store. As highlighted by Spencer et. al. with regard to SHM of infrastructure, "big data needs big data management" (2019, pg. 16).

6.1 Data quality

The baseline for data quality assessment of the proposed CV approach is built upon comparison against data obtained from the accelerometers attached to the aluminum beam as well as comparison to a parallel study conducted on the UT Campus footbridge with a use of smartphone's built in accelerometer. The results and direct comparison between image processing and accelerometry is presented separately for both cases in Chapters 4 and 5.

Another factor to consider is the image quality, which plays an important role in effectiveness of the image processing algorithms. As one would expect, the more sharp and clear image, the more effective the template matching. When the video is blurry or grainy it creates interference of the 'real' image and reducing the quality and resolution of the image frames processed by the algorithm. In the figure 16, the disparity in image quality between the two frames can be easily observed. This image quality difference translates into poorer performance of template matching technique. The potential of natural frequency identification from fourth marker (worse quality image), is smaller relative to other three markers – out of the twenty-seven analyzed cases, the natural frequencies for the fourth marker were picked up twelve times. For first, second, and third markers this was sixteen successful identifications.

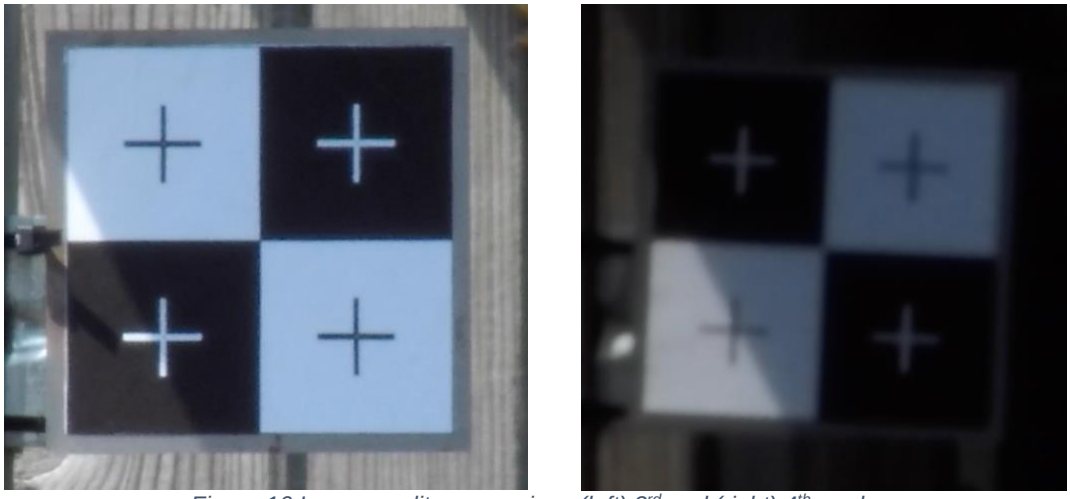

Figure 16 Image quality comparison (left) 3rd and (right) 4th marker

From the displacements histories the error due to camera movement can be easily obtained by analyzing the first few seconds of the recorded videos; the interval between the start of video and moment of excitation provides insight into measurement error due to background vibrations of the camera. This was performed for one video each from lab and field measurements. The error (in mm) due to camera movement is presented in the figure 17 below. By looking at the y-axis of both plots, it is apparent that the error associated with the camera movement is larger in the field conditions, than in the laboratory setting. In the laboratory, the maximum displacement due to camera movement is 0.015 mm, for bridge this is 1.5 mm. However, the camera

movement mainly oscillates with a magnitude of approximately 1 mm – between -0.5 mm up to 0.5 mm. Camera shaking in the lab can be explained by background vibrations of the floor due movement of people throughout the building. In the field conditions the camera movement are of larger magnitude due to greater number of factors that can potentially influence the stability of camera. There are many contributors that can occur in the field setting - wind and other background and environmental vibrations e.g. nearby traffic (car, bus, train) or wind (Zhuang et al. 2022). In the study location there was no significant traffic nearby; however slight wind with occasional stronger gusts was present, which could explain the camera movement captured in the displacement histories.

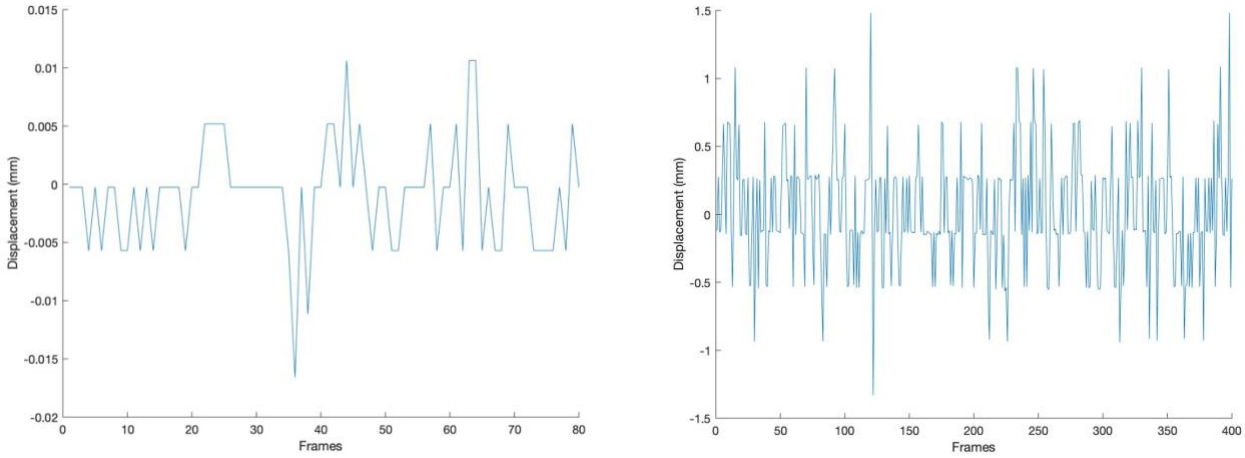

Figure 17 Camera movement extracted from (left) lab and (right) field conditions

7 Conclusions and further research

The proposed computer vision-based structural health monitoring methodology has been proven successful in identification of natural frequencies in both of the tested scenarios – aluminum beam in laboratory and UT Campus footbridge.

In the laboratory three natural frequencies associated with vertical modes were identified on entire recording bandwidth of 120 Hz. The first mode at 18.75 Hz, the second at 51.75 Hz and third at 102.7 Hz. The analysis of measurements from the UT Campus bridge has shown that only lower frequency modes can be identified, no clear identifiable signal peaks were detected above 5 Hz. The identified natural frequencies of UT Campus footbridge are 2.60 Hz (1st mode), 3.29 Hz (2nd mode) and 3,68 Hz. In both of the studied cases the researcher was able to determine the first two vertical modal shapes.

The proposed CV-SHM methodology for modal parameters identification has a number of challenges and limitations. Based on the findings from literature review and experiences gained through conducted measurements, the following limitations are highlighted:

- Computer vision-based monitoring of displacement histories can be easily affected by environmental and site conditions, while conducting in-situ tests. Various influencing factors will lead to unwanted change in camera's position or image quality degradation, these may include nearby traffic or weather conditions (wind, rain, fog).

- The system is susceptible to changes and variations in illumination levels. Illumination leads to degradation of image quality or even failure to produce image sharp enough for further processing. The shadows casted by nearby buildings, trees or the studied structure itself can reduce the image processing capabilities. This has been observed in this study, where the 4th marker from UT Campus footbridge, captured by cameras was significantly darker, as a result less natural frequencies were identified relative to other markers. The proposed system is not robust to various weather and environmental factors, which can be a significant limitation.

- Various limitations due to hardware specifications exist and can restrict the potential of modal analysis. Important camera specifications include recording resolution (number of pixels), recording rate (frames per second) and zoom capabilities. All of these three variables are desired to be as high as possible. Greater resolution results in better image quality, higher recording rate provides insight into larger frequency range (thus higher-frequency modes) and more

powerful zoom lens allows to zoom in closer on the target/feature from further distance.

- As shown in this study, not all recordings will yield a successful identification of natural frequencies (and associated mode shapes). The displacements associated with bridge structures are often very small and are simply difficult to capture. Moreover, the possibility to place the camera in an optimal spot – with good viewing position, stability, little to no perspective distortions, and outlook on vital structural elements, are some of the most important limiting factors.

7.1 Further research

Significant progress has been made by research community in the field of CV-SHM and modal analysis over the past 20-30 years, yet a number of technical and operational obstacles remains, before the development of automated CV-SHM can be fully realized and implemented as standard SHM procedures.

Further research should focus on quantifying the effects of influence factors and robustness of main image processing techniques, especially with regard to illumination effects. If an illumination correction model or coefficients are developed, the CV-SHM could be realized at any time of the day, no matter the weather or season. Additionally, a lot of structures cast a shadow on itself, including vital structural components, this was the case on UT Campus footbridge, in all four marker locations, the shadows of balusters were partially obstructing the measurement.

Another direction for CV-SHM is research into identification of higher-frequency modes of operational bridges with the use of super slow motion cameras, which can capture higher frequency domain, due to faster recording rate. This and other studies have proven that lower natural frequencies and lower-frequency mode shapes can be identified via proposed method; however what is the limit of CV-SHM, up until which mode, can the shapes be accurately determined? The more natural frequencies and associated mode shapes that can be identified via presented methodology, the more extensive and detailed signature of a bridge can be drawn.

References

Allada, Vishal, Thiyagarajan Jothi Saravanan, and Stefano Mariani. 2021. “Computer Vision Technique for Blind Identification of Modal Frequency of Structures from Video Measurements.” *Engineering Proceedings 2021, Vol. 10, Page 12* 10 (1): 12. https://doi.org/10.3390/ECSA-8-11298.

Brunelli, Roberto. 2009. “Computational Aspects of Template Matching.” *Template Matching Techniques in Computer Vision*, April, 201–19. https://doi.org/10.1002/9780470744055.CH10.

Dackermann, Ulrike. 2009. “Dynamic Based Damage Identification Methods for Civil Engineering Structures Using Artificial Neural Networks.” University of Technology Sydney. https://opus.lib.uts.edu.au/handle/10453/20303.

Dong, Chuan Zhi, and F. Necati Catbas. 2021. “A Review of Computer Vision–Based Structural Health Monitoring at Local and Global Levels.” *Structural Health Monitoring* 20 (2): 692–743. https://doi.org/10.1177/1475921720935585.

Fryba, Ladislav. 1999. “Vibration of Solids and Structures under Moving Loads.”

GoPro. n.d. “The Remote - Action Cameras Remote Control | GoPro.” Accessed June 24, 2022. https://gopro.com/en/us/shop/mounts-accessories/the-remote/ARMTE-003.html.

Hartsuijker, C., and J.W. Welleman. 2007. *Engineering Mehcnaics Volume 2: Stresses, Strains, Discplacements*. Springer Netherlands.

Jiang, Ruinian, David V. Jáuregui, and Kenneth R. White. 2008. “Close-Range Photogrammetry Applications in Bridge Measurement: Literature Review.” *Measurement: Journal of the International Measurement Confederation*. Elsevier. https://doi.org/10.1016/j.measurement.2007.12.005.

Kromanis, Rolands. 2021. “Characterizing Footbridge Response from Cyclist Crossings with Computer Vision-Based Monitoring.” In *Lecture Notes in Civil Engineering*, 156:83–95. Springer, Cham. https://doi.org/10.1007/978-3-030-74258-4_5.

Kromanis, Rolands, and Prakash Kripakaran. 2021. “A Multiple Camera Position Approach for Accurate Displacement Measurement Using Computer Vision.” *Journal of Civil Structural Health Monitoring* 11 (3): 661–78. https://doi.org/10.1007/S13349-021-00473-0/FIGURES/21.

Lourens, Eliz-Mari, Pim van der Male, Andreas Hartmann, and Irina Stipanovic. n.d. “Smart Monitoring of Bridge Performance.” 4TU.Built Environment. Accessed June 18, 2022. https://www.4tu.nl/bouw/Projects/Smart Monitoring of Bridge Performance/.

Lydon, Darragh, Myra Lydon, Su Taylor, Jesus Martinez Del Rincon, David Hester, and James Brownjohn. 2019. “Development and Field Testing of a Vision-Based Displacement System Using a Low Cost Wireless Action Camera.” *Mechanical Systems and Signal Processing* 121 (April): 343–58. https://doi.org/10.1016/J.YMSSP.2018.11.015.

MATLAB. n.d. “Computer Vision Toolbox - MATLAB & Simulink.” Accessed July 4, 2022. https://nl.mathworks.com/products/computer-vision.html.

McCormick, Nick, and Jerry Lord. 2010. “Digital Image Correlation.” *Materials Today* 13 (12): 52–54. https://doi.org/10.1016/S1369-7021(10)70235-2.

“NEN Connect - NEN-EN 1991-2+C1:2015/NB:2019 Nl.” n.d. Accessed July 5, 2022. https://connect.nen.nl/Standard/Detail/3622776?compId=16755&collectionId=0.

“Online ArUco Markers Generator.” n.d. Accessed May 10, 2022.

https://chev.me/arucogen/.
Shang, Zhexiong, and Zhigang Shen. 2017. “Multi-Point Vibration Measurement for Mode Identification of Bridge Structures Using Video-Based Motion Magnification.” *ArXiv*.
Spencer, Billie F., Vedhus Hoskere, and Yasutaka Narazaki. 2019. “Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring.” *Engineering* 5 (2): 199–222. https://doi.org/10.1016/J.ENG.2018.11.030.
Voordijk, Hans, and Rolands Kromanis. 2022. “Technological Mediation and Civil Structure Condition Assessment: The Case of Vision-Based Systems.” *Civil Engineering and Environmental Systems*. https://doi.org/10.1080/10286608.2022.2030318.
Xu, Yan, James Brownjohn, and Dali Kong. 2018. “A Non-Contact Vision-Based System for Multipoint Displacement Monitoring in a Cable-Stayed Footbridge.” *Structural Control and Health Monitoring* 25 (5): e2155. https://doi.org/10.1002/STC.2155.
Xu, Yan, and James M.W. Brownjohn. 2018. “Review of Machine-Vision Based Methodologies for Displacement Measurement in Civil Structures.” *Journal of Civil Structural Health Monitoring* 8 (1): 91–110. https://doi.org/10.1007/S13349-017-0261-4/TABLES/7.
Zhuang, Y ;, W ; Chen, T ; Jin, B ; Chen, H ; Zhang, W A Zhang, Yizhou Zhuang, et al. 2022. “Citation: A Review of Computer Vision-Based Structural Deformation Monitoring in Field Environments.” https://doi.org/10.3390/s22103789.
Zolghadri, Navid, Marvin W Halling, Paul J Barr, Nick Foust Masters, and Nick Foust. 2015. “Effects of Temperature on Bridge Dynamic Properties.”
Zona, Alessandro. 2020. “Vision-Based Vibration Monitoring of Structures and Infrastructures: An Overview of Recent Applications.” *Infrastructures 2021, Vol. 6, Page 4* 6 (1): 4. https://doi.org/10.3390/INFRASTRUCTURES6010004.

Appendix A: Markers

Figure 18 ArUco Markers (57x57 mm) used in laboratory set-up

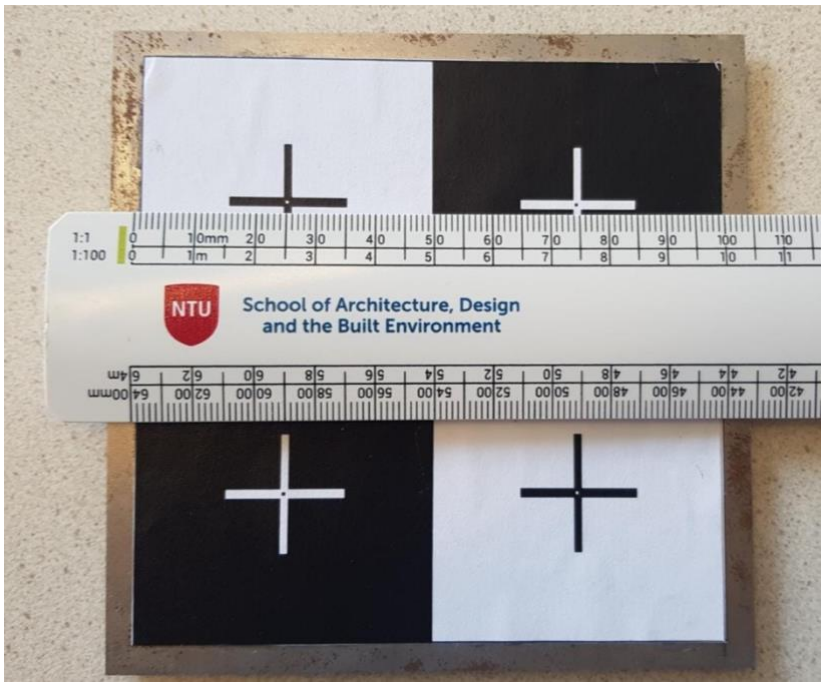

Figure 19 Marker (10 x 10 cm) used on UT Campus footbridge

Appendix B: data lab

The tables 8 to 10 below show the exact values of first three natural frequencies identified for all marker location, for all five of the carried out experiments.

Table 8 Peak frequencies identified for given camera per experiment, frequency 1 [Hz]

Exp. #	CAM-1	CAM-2	CAM-3	CAM-4
1	18,80	18,88	19	19
2	18,92	18,75	19	18,75
3	18,79	18,75	18,75	19
4	18,80	18,75	18,75	18,75
5	19	18,75	19	19

Table 9 Peak frequencies identified for given camera per experiment, frequency 2 [Hz]

Exp. #	CAM-1	CAM-2	CAM-3	CAM-4
1	51,49	x	51,75	51,75
2	51,75	52	51,75	52
3	51,79	51,75	51,75	51,75
4	52	52,25	52	52,25
5	51,75	51,75	52	51,75

Table 10 Peak frequencies identified for given camera per experiment, frequency 3 [Hz]

Exp. #	CAM-1	CAM-2	CAM-3	CAM-4
1	103,1	x	x	102,8
2	102,1	102,5	x	103
3	102,9	x	x	102,8
4	102,2	101,2	x	102
5	102,2	102	x	102,5

Appendix C: data UT Campus footbridge

Table 11 Peak frequencies identified for given scenario per marker, frequency 1 [Hz]

Marker → / Case ↷	1st	2nd	3rd	4th
Cycling left	2,549	2,536	2,523	x
Cycling center	x	2,675	x	x
Cycling right	x	x	x	x
Jogging left	2,726	2,713	2,692	2,675
Jogging center	2,683	2,662	2,694	2,690
Jogging right	x	2,571	2,534	2,633
Jogging group left	2,625	2,637	2,610	2,603
Jumping group left	2,326	2,343	2,315	2,300
Jumping group right	2,101	2,097	2,393	2,381

Table 12 Peak frequencies identified for given scenario per marker, frequency 2 [Hz]

Marker → / Case ↷	1st	2nd	3rd	4th
Cycling left	3,238	3,27	3,34	x
Cycling center	3,41	3,477	3,423	3,407
Cycling right	3,213	3,263	3,26	x
Jogging left	x	x	x	x
Jogging center	x	x	x	3,322
Jogging right	x	x	x	x
Jogging group left	x	x	3,228	x
Jumping group left	3,041	3,013	3,059	2,997
Jumping group right	3,201	3,195	3,225	3,21

Table 13 Peak frequencies identified for given scenario per marker, frequency 3 [Hz]

Marker → / Case ↷	1st	2nd	3rd	4th
Cycling left	3,651	3,604	3,636	x
Cycling center	x	x	x	x
Cycling right	3,785	x	x	x
Jogging left	x	x	x	x
Jogging center	x	x	x	x
Jogging right	x	x	x	x
Jogging group left	x	x	x	x
Jumping group left	3,668	3,682	3,638	3,624
Jumping group right	3,702	3,694	3,745	3,727