Video-based vibration analysis for structural health monitoring in civil engineering

Bian Xiong

UNIVERSITÉ DE RENNES 1

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par
Bian XIONG

Video-based Vibration Analysis for Structural Health Monitoring in Civil Engineering

Thèse présentée et soutenue à Nantes , le 22 Décembre 2021
Unité de recherche : Inria Rennes – Bretagne Atlantique

Rapporteurs :

Pierre CHARBONNIER	Directeur de recherche Cerema
Joseph MORLIER	Professeur Isae-Supaero

Composition du Jury :

Président :	Marc FRANCOIS	Professeur Nantes Université
Examinateurs :	Vincent BALTAZART	Chargé de recherche UGE-IFSTTAR, Co-encadrant
	Dominique BEREZIAT	Maître de conférences Sorbonne-Université
	Pierre CHARBONNIER	Directeur de recherche Cerema
	Su RUAN	Professeur Université Rouen Normandie
Dir. de thèse :	Qinghua ZHANG	Directeur de recherche Inria Rennes Atlantique

Acknowledgements

First of all, I am indebted to my Thesis director, Dr. Qinghua ZHANG and my supervisor Dr. Vincent BALTAZART for their support and advice throughout the course of my research. I extend my gratitude to Qinghua and Vincent for guidance throughout the course of my research and for invaluable comments that helped me improve the manuscript. A special thank to Jean-Philippe TAREL for his background and advices in computer vision.

I am grateful to Région Bretagne, UGE-Cosys and Inria for funding this thesis. The Inria center and UGE-Cosys/SII laboratory provided an ideal research environment and sufficient equipments to conduct the experimental test.

A very special gratitude goes out to all the members of 'Comite de Suivi de Thèse', Prof. Marc FRANCOIS and Dr. Sébastien BOURGUIGNON, for providing me their valuable feedback and suggestions to improve my work.

A personal note of thanks to my dissertation rapporteurs: Prof. Joseph MORLIER and Dr. Pierre CHARBONNIER for taking their valuable time in providing their review and helpful remarks.

I extend my gratitude towards the thesis committee members, Prof. Su RUAN, Prof. Marc FRANCOIS and Prof. Dominique BEREZIAT for attending to my PhD thesis defense. A special thanks to Prof. Dominique BEREZIAT for my master training at Sorbonne University. He guided me during the master training in image processing, as I majored in electronics in bachelor and then switched to image processing in master.

And at last but by no means least, I really appreciate my colleagues at UGE-Ifsttar and I4S for directly or indirectly sharing their experience with me throughout the course of the thesis, in particular Jean-Philippe GOURDON and Fabrice BLAINEAU for technical support, Riccardo ARTONI for the free loan of the high speed camera, Jean DUMOULIN for the support of Desdemona project and Michaël DOHLER for introducing me to stochastic subspace identification techniques. A special thank to Boualem MERAINANI for his help in the experimental part. His sharing and experience helped me to complete the laboratory experiments and SHM tests.

Table of Contents

GLOSSARY

Throughout this dissertation, vectors and matrices are denoted by bold lowercase and bold uppercase symbols, respectively. Vectors are, by default, column vectors.

Mathematical Notations

— a: Scalar

— $\mathbf{a}$: Vector

— $\mathbf{A}$: Matrix

— $(.)^*$: Complex-conjugate operator

— $(.)^T$: Transpose operator

— $(.)^{-1}$: Inverse operator

— $(.)^H$: Conjugate-transpose operator

— $(.)^+$: Moore-Penrose inverse operator

— $*$: Convolution operator

— $\mathbf{I}$: Identity matrix

— $E[.]$: Expectation operator

— $Var(.)$: Variance

— diag($\mathbf{a}$): Diagonal matrix whose diagonal elements are those of the vector $\mathbf{a}$

— $||\mathbf{a}||^2$: Euclidean norm of the vector $\mathbf{a}$

— $|\mathbf{A}|$: Determinant of the matrix $\mathbf{A}$

— $\approx$: Approximately equal

— $\mathcal{N}(\mu, \sigma)$ Normal distribution with mean μ and variance σ

— $\mathbb{N}, \mathbb{R}, \mathbb{C}$: Set of natural, real, complex numbers

Acronyms and Abbreviations

— SHM : Structural Health Monitoring

— CRLB: Cramér-Rao Lower Bound

— RMSE: Root Mean Squared Error

— SNR: Signal-to-Noise Ratio

— CNR: Contrast-to-Noise Ratio

— SSI: Stochastic Subspace Identification

— SVD: Singular Value Decomposition

— LDS: Laser Displacement Sensor

— LDV: Laser Doppler Vibrometer

— LiDAR: Light Detection And Ranging

— GB-SAR: Ground-Based Synthetic Aperture Radar

— DIC: Digital Image Correlation

— MSE: Mean Squared Error

— ROI: Region of Interest

— NDT: Non Destructive Techniques

— GNSS: Global Navigation Satellite System

— IRT: Infrared Thermography

— MAC: Modal Assurance Criterion

List of Symbols

$(\Delta m, \Delta n)$ Pixel-level shift

$(\delta x, \delta y)$ Subpixel shift

(d_x, d_y) Total shift

(N, M) Number of rows and columns for image template

(n, m) Coordinates of row and column

$\boldsymbol{A}$ State transition matrix

$\boldsymbol{C}$ Observation matrix

$I(x, y, t)$ Intensity of image at location (x, y) at time instant t, where x denotes the Cartesian coordinate in the vertical direction, y denotes the Cartesian coordinate in the horizontal direction

Chapter 1

Introduction

Civil engineering structures, e.g., bridges, buildings, road networks, are an integral part of human life, ensuring the economic activity and the mobility of human societies, including medical care, goods, resources and materials transport of any type. Any damage or failure in these structures may result in strong impacts. Damages may take place on infrastructures due to aging, environmental deterioration or natural disasters such as earthquakes, or design and construction weaknesses. As a matter of fact, two recent bridge collapses in Europe, namely, Morandi Bridge in Italy and Mirepoix bridge in France (Wikipedia contributors, 2021), have recalled the requirement for an effective monitoring strategy of civil engineering structures.

The survey of civil engineering structures is not a new task. The oldest and conventional method is based on periodic visual inspection by skilled experts. They look for signs of mechanical damages and chemical corrosion, listen for strange noises, or analyse material samples taken from the structure. Tracking the health of a structure over time with these tedious procedures requires a lot of human labor, yet may not be sufficient to detect damages at their early stage. For obvious practical reasons, such survey procedures are limited or impossible at some locations of the structures, where the inspectors can not easily approach.

Sensors and measurements techniques have been developed for a long time to provide reliable and quantitative information on the structure to the end-users. They contributed to the emergence of Non Destructive Techniques (NDT) and to Structural Health Monitoring (SHM) (« Infrastar Training School », 2021).

Among the existing strategies for SHM (Amafabia et al., 2017), vibration analysis provides important information about the health state of infrastructures. In brief, vibration-based SHM techniques aim to gain knowledge of the integrity of in-service structures on a continuous time basis, by detecting and characterizing changes in the dynamic behavior of the monitored structure from sensor measurements. Traditionally, vibration-based SHM techniques have been based on mechanical sensors installed on the surface of the struc-

ture or embedded in the material, i.e, point-like sensors. The most widespread sensors are accelerometers. The number of sensors and their locations highly affect the performance of SHM techniques.

Within this context, video camera-based vibration analysis has recently emerged as a promising approach to Structural Health Monitoring (SHM), owing to its handiness and low-cost aspects. Based on appropriate image processing techniques, video image pixels act as motion sensors, substituting conventional mechanical sensors on the structure to monitor. In the long term, such image-based SHM technique may monitor any civil engineering structure without having to integrate sensors on it. The preferred structures concerned by this technique may be either specific vertical structures, e.g., the mast of wind turbines, or any element of bridges and buildings.

1.1 Context of the study

This PhD project has been carried out within the Inria-UGE joint research team I4S, which develops vibration analysis techniques based on various sensors for SHM. In parallel, the team is developing image processing techniques for monitoring surface conditions on concrete and pavement surfaces.

Besides, I4S has been a partner of the three-years 2017' EU research project DESDEMONA (« Desdemona project home page », 2021), which aims at integrating sensing and automation technologies for the purpose of survey and monitoring of steel structures. One of the activities of this project aims at developing video-based vibration analysis for the monitoring of some mechanical structures.

Within this context, this thesis has aimed to combine vibration analysis and image processing techniques, in order to perform the vibration analysis of structures by optical remote sensors. Video-based monitoring techniques have emerged in the literature in recent years. A video camera provides a temporal flow of video images of the structure to be monitored. Motion signals of the structure are deduced from selected pixels within the image, which thus act as remote / virtual sensors. The temporal variations of the motions signals allow performing the vibration analysis of the structure.

1.2 Existing methods for vibration-based SHM

Vibration-based SHM techniques usually rely on modal analysis methods. They consist in estimating modal parameters, including natural frequencies and damping factors, from measured vibration data, typically acceleration and displacement measurements (Gürkan et al., 2017; Lin & Lan, 2010).

This section briefly introduces some existing contact and contactless solutions for collecting vibration data over large mechanical structures. Further details are available in the review papers (« Cost action TU1406, part C, Innovation on Technologies », 2021; Feng & Feng, 2018; Spencer et al., 2019; Zona, 2021).

Contact-based methods

Contact-based methods rely on sensors integrated in the monitored structure. The literature mainly reports the use of the following types of sensors.

Accelerometer: The most widespread technique for SHM is based on accelerometer measurements. This is basically the cheapest technology of measurement in SHM. The technology of accelerameter has been improved with time, providing higher precision and larger frequency bandwidth for mechanical structures. As a counterpart, this technique has wiring issues and requires a proper physical placement of the sensor. The full scale characterization of the structure requires a large network of such sensors with synchronous data collection.

Strain gauge: Strain gauge is a device that measures the local lengthening of the monitored structure. The most widely used strain gauge is called "rosette" and relies on highly accurate electrical resistance measurement sensitive to geometrical deformations. One drawback of this method is the addition of mass to the monitored structure, changing its dynamic properties. Two fiber optic technologies allow distributed strain measurements, namely, Fiber Bragg Gratings (Torres et al., 2011) and Rayleigh scattering (Chapeleau et al., 2017).

GNSS (global navigation satellite system): GNSS technology has been experimented for SHM applications in (Yu et al., 2016). The satellite signals and the GNSS internal clock enable estimating the 3D global positioning of the GNSS receiver. Multiple frequency techniques, phase-based computation and some post-processing have been shown to achieve positioning accuracy up to a few millimeters in (« Cost action TU1406, part C, Innovation on Technologies », 2021). Besides, specific technologies enable

acquisition rate up to 100Hz. Time series of 3D coordinates then allow computing the vibration spectrum at the location of the structure where the GNSS receiver is set up.

Non-contact measurement techniques

Non-contact measurement methods have been quickly developing during the last two decades. They rely on radar and optical imaging technologies, namely, laser instruments, visible light and infrared cameras, to collect data in order to investigate the monitored structure conditions. Non-contact methods avoid the drawbacks of contact methods and are being researched intensely for the purposes of SHM.

Laser-based techniques: The laser-based techniques include point-like sensor and laser scanning techniques. A common point-like sensor is laser displacement sensor (LDS), which measures the out-of-plane motion of an object from triangulation principle (Gürkan et al., 2017). Three LDS are used in Chapter 5 to serve as reference displacement measurements when experimentally assessing image-based motion extraction algorithms. Basic Laser Doppler vibrometer (LDV) is another point-like sensor which measures the vibration speed of the surface from the Doppler shift of the reflected laser beam (Nassif et al., 2005).

By extension, laser-based scanning techniques is a powerful alternative imaging technique which allow scanning large objects or scene at daylight and nighttime with high spatial point cloud measurements. They have been reported to be applied to SHM, e.g., (Sohn & Park, 2021), including Laser Doppler vibrometry (LDV) (Sharma, 2008), Light detection and ranging (LiDAR) (Cha et al., 2020) and Terrestrial laser scanner (TLS) (Park et al., 2007).

Ground-Based Synthetic Aperture Radar (GB-SAR): The GB-SAR is a non-contact active electromagnetic technology which has emerged in the past 10 years for the survey of large structures (Monserrat, 2012). The GB-SAR technique is based on an imaging radar-based sensor, which provides high sensitivity to small displacements at the scale of centimetres wavelengths. Vibration information are extracted from the timely phase variations of the signal attached to the centimeter radar resolution cells, provided only one backscatterer is present herein.

Video-based imaging techniques: Most of the aforementioned measurement devices allow collecting vibration signals at limited number of positions along the monitored structure, because of the difficulties for installing a large number of such sensors and cost limitations. Besides, they require skilled people to operate, that may limit the spread of

their use in practical applications.
By contrast, video-based vibration techniques have emerged in the past 10 years as an enticing technology for non-contact methods with a handy use and affordable costs while avoiding tedious sensor installations. They can image a large scene of interest in the visible wavelength to collect motion information on a large structure.
With the use of a video camera, selected pixels in the image can provide vibration information about the corresponding part of the structure. A lot of video-based vibration applications have been reported in the literature (Spencer et al., 2019; Zona, 2021), on either simple mechanical structure in laboratory (Chou & Chang, 2021) and on field experiments (Zhao et al., 2019). Among field experiments, the results on steel bridges, e.g., (Feng & Feng, 2018), are highly demonstrative because the measured vertical displacement of the bridge can reach 20 cm.
To apply such techniques, image processing is the key step to transform pixel brightness time variations from a sequence of images to the vibrating motion signal of the monitored structure. A lot of image processing techniques have been proposed to this aim, but requires to be tailored and tested within both the scope of SHM application and the various experimental configurations. SHM application requires image processing techniques with subpixel accuracy to efficiently monitor vibration behaviors. When observing a structure with a camera at long distance, the movement of the structure is likely to be seen within a single pixel or a few pixels. Subpixel movement of the structure (or a part of it) results in subtle changes in the pixel brightness between two consecutive images. Subpixel motion analysis aims at interpreting the change in pixel brightness in terms of displacement.
Video cameras can range from scientific instruments, with high-frequency acquisition rate, high-resolution video, to cheaper and conventional resolution units such as surveillance video cameras, and inexpensive units such as those on cell phones or action cameras. The universality and low cost of video cameras make them a good candidate for widespread usage in SHM.

Infrared thermography: Within civil engineering field, thermography (IRT) or thermal imaging can be used for two applications. First, the small penetration depth of infrared wavelengths allows probing and detecting defects that are flush with the surface. Within the scope of the thesis, infrared cameras can be used as a complementary device of the latter video cameras to image vibrating structures at night (or for a 24 hour monitoring coverage) with the expense of higher costs (Talai et al., 2016)

1.3 Research objectives

Within the scope of SHM, this thesis aims at developing image-based vibration analysis of mechanical structures. Within this scope, image-based techniques are used to extract motion signals (displacements with subpixel accuracy) from selected pixels of video image flows.

These objectives allow defining an enticing technology for non-contact vibration-based SHM methods. As a result, the SHM technique may be performed without instrumenting the mechanical structure with conventional embedded sensors. Selected pixels in the image, namely, either pixels from specific targets attached to the mechanical structure or pixels at the edges along the mechanical structure, may replace traditional conventional sensors such as accelerometers.

In terms of image processing techniques, the objectives are to select in the state-of-the-art some existing image processing techniques which achieve subpixel accuracy in motion measurements. Besides, two new methods are proposed that afford improved subpixel accuracy. The accuracy of motion extraction methods are tested on simulated and field images on a simple mechanical structure, namely, a vertical steel beam. The accuracy test is carried out on two kinds of pixels: pixels from a few targets attached to the structure and edge pixels from the structure contours. Results are expected to show that the motion extraction algorithm can greatly enhance the time variation of the pixel raw brightness directly from the image to perform the vibration analysis.

For SHM application, the objectives of the thesis is first to demonstrate that video-based motion signals can be used to accurately determine the modal parameters of mechanical structures from subspace identification techniques, i.e., natural frequencies, the associated damping ratios and mode shapes. Second, it is showed that the image-based SHM is sensitive enough to monitor the structure by detecting changes in the modal parameters of the structure.

The experimental objective of the thesis is to test image-based techniques on a laboratory mechanical structure, namely, a vertical steel beam in both indoor and outdoor conditions, which is excited by different sources. This experiment enables to assess the performance of the correlation-based algorithms by comparing the estimated pixel motion to reference laser sensors. The estimated motion signals are then processed by subspace identification techniques to provide the modal parameters of the vertical beam, which are compared to the theoretical modal response of the beam. Finally, a first SHM experiment

is set up, by artificially modifying the free-vibrating beam structure with added magnets (mass) at some places along the beam. The image-based SHM analysis enables to detect changes in the mode shapes.

1.4 Thesis outline

The thesis is organized into the following four chapters after the introduction chapter:

— **Chapter 2** presents first the state-of-the-art in the application field with a review on widespread two-step motion extraction methods, including pixel-level shift estimation methods and subpixel shift estimation methods. Two families of subpixel methods are detailed : ROI-based methods, i.e., Taylor-based method, quadratic surface fitting (QSF), and local phase methods based on the steerable filter (Steer). This chapter is completed with some review on the SHM application.

— **Chapter 3** focuses on the two proposed improved motion extraction methods, the "efficient quadrant interpolation" and its variant version named "efficient octant interpolation". Besides, the third chapter presents the extension of the steerable filter (EX-ST) to pluripixel motion. Finally, some analysis of the quadratic surface fitting (QSF) for subpixel motion extraction from video images is detailed.

— **Chapter 4** presents the performance assessment of the proposed video-based motion extraction methods from simulated synthetic images. First, the simulated image data set is based on two models, a simplified one and physical-based model. We also used an existing dataset for DIC challenge. Each step of the correlation-based motion extraction technique is separately evaluated. The RMSE computation between the estimated and the true motion over the time record is used to evaluate the performance of the algorithms. For convenience, it is expressed in terms of pixel fraction, i.e., subpixel.

— **Chapter 5** presents the laboratory experiment on a vertical steel beam. It first details the experimental set up and materials, namely, a shake-table, a high speed camera, the lighting, targets, three laser sensors which serve as the reference for motion. The shake-table allows to generate either pure sinusoidal excitation or wide bandwidth random excitation. Hammer shocks have been also considered to generate transient vibration source. We first pay attention to the experimental evaluation of the motion extraction algorithms from the ROIs centered on the three targets

attached to the beam. Then, the edge pixels of the beam are used to test the target-free solution. The latter case is finally used for demonstrating the use of video-based motion extraction for SHM application. The mode shapes of the beam are estimated from the estimated motion on the selected edge pixels after performing some stochastic subspace identification algorithm. At last, a defect diagnosis experiment was set up, by attaching magnets at some places along the beam. The image-based SHM analysis achieves fault detection from the changes in the beam natural frequencies while the modal assurance criterion (MAC) enables to detect the changes in the mode shapes.

1.5 Main contributions

The main contributions of this thesis to the SHM literature are expected to be as follows:

Motion-extraction algorithms:

— Review and analysis of some widespread two-step correlation-based techniques. Within this scope, the quadratic surface fitting (QSF) method has been detailed in Section 2.5 and some shortcomings have been identified. In particular, the conditions ensuring the existence of a maximum of the fitted quadratic surface are rigorously analyzed.

— Two new interpolation methods, namely, EQI for Efficient Quadrant Interpolation and octant interpolation (EQI-Octant), have been proposed in Section 2.2 to estimate the subpixel shift. By considering a preliminary step to determine the subpixel shift direction, the accuracy in motion extraction has been improved compared to the conventional first-order Taylor-based method.

— The method of the steerable filter has been extended to estimate the motion larger than one pixel by contrast to the previous work in (J. G. Chen et al., 2015).

Performance assessment:

The proposed correlation-based motion-extraction algorithms have been assessed on both simulated images and field images flows on the basis of the RMS error on displacements. Compared to the literature:

— The performance is detailed at both the pixel and the subpixel steps before reporting on the overall performance. It is clearly demonstrates the importance of the subpixel

step for the overall accuracy of motion extraction.

— The images at hand show a weak texture compared to the speckled images which usually support the evaluation of image registration techniques or DIC challenge. Then, the performance of the selected motion-extraction algorithms do not rank in the same order as in the literature.

— Besides, simulated and experimental results illustrate on the RMS error the influence of the noise model, blurring effect, displacement amplitude, experimental set-up, indoor/outdoor conditions, camera parameters and camera types.

Image-based SHM:

— It is experimentally demonstrated that modal parameters of a mechanical structure can be extracted by directly analyzing raw pixel brightness signals without the step of motion signal extraction, when the vibration amplitude of the structure is sufficiently small.

— The added value of motion signal extraction is confirmed by successfully extracting modal parameters in the case of large vibration amplitude. Moreover, the performance is improved compared to direct raw pixel brightness signals processing in the case of small vibration amplitude.

— The feasibility of video image-based SHM is demonstrated by successfully detecting mass changes of the tested mechanical structure based on motion signals extracted from video images captured by a single camera equivalent to the measurements of hundreds of virtual mechanical sensors.

1.6 List of publications

The work in this thesis has led to the following publications.

Journal paper :

— Bian Xiong, Qinghua Zhang, Vincent Baltazart, *On Quadratic Interpolation of Image Cross-Correlation for Subpixel Motion Extraction*, Sensors 22(3), https://doi.org/10.3390/s22031274, 2022.

Conference papers :

— Bian Xiong, Qinghua Zhang, Vincent Baltazart, *Analysis of Quadratic Surface Fitting for Subpixel Motion Extraction from Video Images*, 28th European Signal Processing Conference, Jan 2021, Amsterdam.

— Boualem Merainani, Bian Xiong, Vincent Baltazart, Jean Dumoulin, Michael Dohler, Qinghua Zhang, *Experimental investigation of structural modal identification using pixels intensity and motion signals from video-based imaging devices: performance, comparison and analysis*, SPIE Multimodal Sensing and Artificial Intelligence: Technologies and Applications, SPIE Optical Metrology conference, june 2021.

Submitted and ongoing journal publications :

— Bian Xiong, Qinghua Zhang, Vincent Baltazart, Boualem Merainani, Ivan Guéguen, Efficient Quadrant Interpolation for Subpixel Motion Extraction from Video Images in the Application of Structural Health Monitoring, To be submitted in Dec. 2021.

— Boualem Mérainani, Michaël Dölher, Bian Xiong, Vincent Baltazart, Qinghua Zhang, Ivan Guéguen, Subspace-based vibration mode identification of a cantilever beam from video image flows, under writing, december 2021

— Boualem Mérainani, Bian Xiong, Ivan Guéguen, Vincent Baltazart, Qinghua Zhang, data paper on "Video images database of vibrating vertical cantilever beam under various laboratory conditions (for assessing video-based vibrating SHM techniques)", Université Gustave Eiffel, under writing, 2021

Miscellaneous:

— B. Merainani, M. Döhler, V. Baltazart, Q. Zhang, B. Xiong, L. Mevel, J. Dumoulin, C. Rinaldi, J. Ciambella, V. Gattulli, Identification techniques by image processing / Structural Modal analysis using computer vision based methods, Desdemona project, Deliverable D5.4, november 2021

— Journée des doctorants de COSYS, on 8th and 9th October 2019, Poster presentation.

— Journées des Doctorants MathSTIC (JDOC), 'Quadrant Interpolation Applied to Subpixel Motion Extraction for Video-Based Structural Health Monitoring', oral presentation, Université de Nantes in May 2021/ Virtual

— Bian Xiong, Vincent Baltazart, Qinghua Zhang, Motion extraction methods using the steerable filter for video-based vibration analysis, accepted to RFIAP conference, Vannes, France, 2020 (Conference cancelled)

Chapter 2

State-of-the-Art on image-based SHM techniques

Image-based SHM applications require image processing techniques with high accuracy to efficiently monitor vibration behaviors. When observing a mechanical structure with a camera at long distance, the vibration of the structure is likely to be seen within a single or a few pixels. Subpixel displacement of the structure results in subtle changes in pixel brightness between consecutive images.

Within this scope, image processing is the key step to transform pixel brightness time variations to the vibrating motion signal of the monitored structure. A lot of existing image processing techniques have been tailored to this application. Most of them originate from other scientific communities and applications, e.g., remote sensing, image registration and super resolution images.

In the first part of this chapter, some widespread image-based motion extraction methods are reviewed. Most of them are based on a two-step coarse-to-fine strategy. The pixel-level shift estimation is presented firstly on the basis of cross-correlation criterion. Then, various subpixel shift estimation methods are presented.

In the second part, some basic theories and literature related to vibration-based SHM are presented. The vibration of mechanical structures is formulated with a discrete time state-space model. Then, subspace-based system identification theory allows performing modal analysis, mode shape identification, and damage detection for SHM purposes.

2.1 Cross-correlation-based pixel-shift estimation

For SHM purpose, vibration analysis can be based on displacement, velocity and/or acceleration measurements collected with various sensors. While acceleration measurements are the most often used, video-based techniques usually make use of displacement or velocity signals extracted from video images. As reported in (Pan et al., 2009; D. Zhang

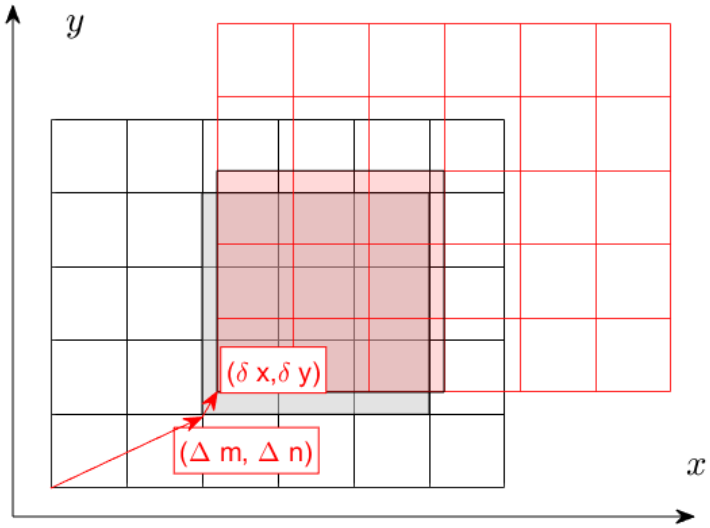


Figure 2.1 – Image gridding of two successive video images at times t_1 in black and t_2 in red. The total shift between the two images is decomposed into pixel-level and subpixel shifts. Each square represents a pixel and the colored filled area is the region of interest(ROI)

et al., 2016; L. Zhang et al., 2015; Z. Zhang et al., 2006), displacement can be estimated with a high accuracy by a two-step coarse-to-fine approach. The pixel-level shift $(\Delta m, \Delta n)$ is estimated from the computation of the cross-correlation (CC), while the subpixel shift $(\delta x, \delta y)$ is accessed by some refinement methods, as illustrated in 2.1. In this Section, we will introduce the cross-correlation principle and some computation variants.

In the following, it is assumed that a locally rigid object is observed through a video image time flow with small displacement amplitude between successive images. Besides, the object is assumed to move in a front plane perpendicular to the camera axis, with

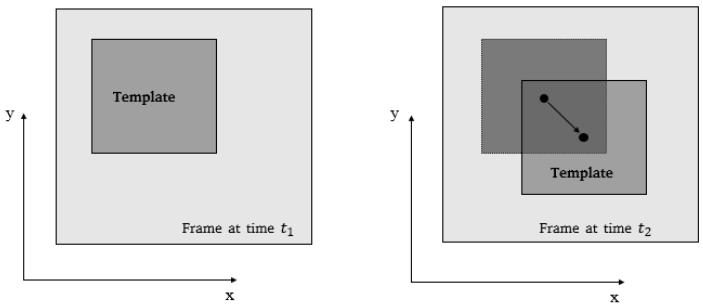


Figure 2.2 – Template shifts between two images at different time

negligible rotational motion.

2.1.1 Cross-correlation basics

Cross-correlation is calculated within a rectangular template as shown in Fig. 2.2, also known as region of interest (ROI), or area of interest (AOI). The image template includes either the whole moving object, some selected part of the object, or targets attached to the object. Given two sub images (templates) of M columns by N rows at two different times $I(m, n, t_1)$ and $I(m, n, t_2)$, the integer number pairs (m, n) are related to the pixel coordinate in the image template plane.

To specify the position of any point in the image plane, consider a Cartesian coordinate system with x and y axes along the horizontal direction and the vertical direction, respectively, as shown in Fig.2.2. In this system, the horizontal and vertical pixel-level shifts are denoted as $(\Delta m, \Delta n)$, respectively corresponding to integer increments in x and y. Usually the first frame is selected as the reference image to measure the displacements of the successive frames.

The pixel-level displacement $(\Delta m, \Delta n)$ has been originally estimated in the literature by locating the peak magnitude in 2D cross-correlation between the two template images $I(m, n, t_1)$ and $I(m, n, t_2)$, as follows:

$$(\Delta m^*, \Delta n^*) = \arg \max_{\Delta m, \Delta n} \mathrm{CC}(\Delta m, \Delta n) \tag{2.1}$$

where CC is the following conventional 2D cross-correlation function:

$$\mathrm{CC}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} I(m, n, t_1) I(m - \Delta m, n - \Delta n, t_2) \tag{2.2}$$

where $\mathbb{T}$ is the set of the pixel indexes of the template image. In practice, we experienced that with the basic 2D cross-correlation criterion defined in Eq. (2.2), the pixel-level shift is sometimes misestimated. So the basic CC will be replaced by other more robust cross-correlation criteria detailed in the following section.

2.1.2 Cross-correlation variants

Cross-correlation is widely used to measure the similarity between signals or images. In image processing, cross-correlation is an efficient tool to match images and therefore

CC criterion	Definition
Cross-correlation (CC)	$\mathrm{CC}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} [I(m,n,t_1) I(m-\Delta m, n-\Delta n, t_2)]$
Normalized cross-correlation (NCC)	$\mathrm{NCC}(\Delta m, \Delta n) = \frac{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_1) I(m-\Delta m, n-\Delta n, t_2)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_1)^2} \sqrt{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_2)^2}}$
Zero-mean cross-correlation (ZCC)	$\mathrm{ZCC}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_1) I_c(m-\Delta m, n-\Delta n, t_2)$
Zero-mean normalized cross-correlation (ZNCC)	$\mathrm{ZNCC}(\Delta m, \Delta n) = \frac{\sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_1) I_c(m-\Delta m, n-\Delta n, t_2)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_1)^2} \sqrt{\sum_{(m,n)\in\mathbb{T}} I_c(m-\Delta m, n-\Delta n, t_2)^2}}$

Table 2.1 – Family of cross-correlation criterion

SSD correlation criterion	Definition
Sum of squared differences (SSD)	$\mathrm{SSD}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} [I(m,n,t_1) - I(m-\Delta m, n-\Delta n, t_2)]^2$
Normalized sum of squared differences (NSSD)	$\mathrm{NSSD}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} [\frac{I(m,n,t_1)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_1)^2}} - \frac{I(m-\Delta m, n-\Delta n, t_2)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_2)^2}}]^2$
Zero-mean sum of squared differences (ZSSD)	$\mathrm{ZSSD}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} \{I_c(m,n,t_1) - I_c(m-\Delta m, n-\Delta n, t_2)\}^2$
Zero-mean normalized sum of squared differences (ZNSSD)	$\mathrm{ZNSSD}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} [\frac{I_c(m,n,t_1)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_1)^2}} - \frac{I_c(m-\Delta m, n-\Delta n, t_2)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_2)^2}}]^2$

Table 2.2 – Family of SSD correlation criterion

is deployed to measure the shift for template matching. Correlation-based similarity criteria can be classified into two categories: cross-correlation (CC) criteria or sum-squared difference (SSD) correlation criteria (P. Liu & Du, 2019; Pan et al., 2009) as reported in Tables 2.1 and 2.2.

In the tables above, $\bar{I}(t)$ and $I_c(m,n,t)$ are the mean value of the pixel brightness and the zero-centered pixel brightness, respectively:

$$\bar{I}(t) = \frac{\sum_{(m,n)\in\mathbb{T}} I(m,n,t)}{MN} \tag{2.3}$$

$$I_c(m,n,t) = I(m,n,t) - \bar{I}(t). \tag{2.4}$$

It is easy to prove that the CC and SSD criteria are related to each other. For example, $\mathrm{ZNSSD}(\Delta m, \Delta n) = 2(1 - \mathrm{ZNCC}(\Delta m, \Delta n))$ and $NSSD(\Delta m, \Delta n) = 2(1 - NCC(\Delta m, \Delta n))$ according to (Pan et al., 2007). In practice, ZNSSD criterion is usually selected owing to its lower computational burden (W. Tong, 2005). It is reported that the correlation values computed by ZNCC or ZNSSD are not changed to a linear transformation of the image gray level, i.e., $I'(m,n) = a \times I(m,n) + b$ (Pan et al., 2007). Thus ZNCC, ZCC, NSSD and ZNSSD are not sensitive to light offset and light linear amplification, while ZCC and ZSSD are insensitive to the light offset only. By contrast, the CC and SSD

are sensitive to all light changes according to (P. Liu & Du, 2019; Pan et al., 2009).

2.1.3 Selected cross-correlation criterion

Different authors have reviewed the previous different cross-correlation criteria (P. Liu & Du, 2019; Pan et al., 2009; Singh & Ganotra, 2012). ZNCC and NCC are usually recommended in the literature for pixel shift estimation, because they have proved robustness against brightness and contrast variations compared to CC.

During our tests in Chapter 4 on simulated images, ZNCC or ZCC criterion were found to lead to the most accurate pixel-level shift estimation, providing appropriate rounding to the nearest integer, e.g., if the preassigned shift is 0.6 pixel, the final estimated pixel-level shift is 1 pixel.

Regarding the computer burden, ZCC or ZSSD would be a better choice. With regards to the pixel-shift accuracy, ZNCC and ZNSSD provide similar performance and are recommended. Finally, ZNCC has been selected in this thesis.

2.2 Subpixel shift estimation methods

Given the estimated pixel-level shift $(\Delta m^*, \Delta n^*)$, the image template after pixel-level shift is denoted as:

$$I_s(m, n, t_1) = I(m + \Delta m^*, n + \Delta n^*, t_1). \tag{2.5}$$

Assume that the pixel-level shift $(\Delta m^*, \Delta n^*)$ has been correctly estimated. After the pixel-level shift compensation, $I_s(m, n, t_1)$ and $I(m, n, t_2)$ are related by a shift smaller than one pixel (subpixel shift) and by some noises affecting the light intensities of each image independently. Then, the relationship between $I_s(m, n, t_1)$ and $I(m, n, t_2)$ can be written as:

$$I(m, n, t_2) \approx I_s(m + \delta x, n + \delta y, t_1) \tag{2.6}$$

where $(\delta x, \delta y)$ represents the subpixel shift and the approximate equality sign $\approx$ accounts for the noises. The total shift is then expressed as:

$$(d_x, d_y) = (\Delta m + \delta x, \Delta n + \delta y), \tag{2.7}$$

as shown in Fig. 2.1.

Now the main problem is how to compute the estimated subpixel shift $(\delta x^*, \delta y^*)$

between $I_s(m, n, t_1)$ and $I(m, n, t_2)$. The estimated total shift will be $(d_x^*, d_y^*) = (\Delta m^* + \delta x^*, \Delta n^* + \delta y^*)$.

The following sections review various approaches to estimate the subpixel shift $(\delta x, \delta y)$.

2.2.1 Optical flow-based subpixel shift estimation method

According to (Burton & Radford, 1978), optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene, which is caused by the relative motion between an observer and the scene. It can be thought as a distribution of apparent velocities of the movement of some brightness pattern in an image. When a pattern moves, it is assumed that the brightness of each point in the pattern is constant. It can be described as:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{2.8}$$

Optical flow is computed with local Taylor series approximations, so the optical flow-based methods are also closely related to differential approaches. To apply the optical flow method, it is assumed that the displacement is small. With the first order Taylor approximation, for each point (x, y), the brightness at $t + \delta t$ is rewritten as follows:

$$I(x + \delta x, y + \delta y, t + \delta t) \approx I(x, y, t) + \delta x \frac{\partial I}{\partial x}(x, y, t) + \delta y \frac{\partial I}{\partial y}(x, y, t) + \delta t \frac{\partial I}{\partial t}(x, y, t) \tag{2.9}$$

and so

$$\frac{I(x + \delta x, y + \delta y, t + \delta t) - I(x, y, t)}{\delta t} = \frac{\delta x}{\delta t} I_x + \frac{\delta y}{\delta t} I_y + I_t \tag{2.10}$$

with

$$I_x = \frac{\partial I}{\partial x}(x, y, t), \quad I_y = \frac{\partial I}{\partial y}(x, y, t), \quad I_t = \frac{\partial I}{\partial t}(x, y, t). \tag{2.11}$$

To the limit $\delta t \to 0$, we obtain:

$$\frac{\delta x}{\delta t} I_x + \frac{\delta y}{\delta t} I_y + I_t = 0 \tag{2.12}$$

Let u, v be the velocity along direction x and y:

$$u = \tfrac{\delta x}{\delta t}, \quad v = \tfrac{\delta y}{\delta t}, \tag{2.13}$$

then

$$I_x u + I_y v + I_t = 0. \tag{2.14}$$

The computation of optical flow based on Eq. (2.14) allows us to estimate a dense field of velocity vectors between two sequential images. There exist two main methods to determine the motion vectors: global method (Horn–Schunck) and local method (Lucas–Kanade).

2.2.1.1 The Horn-Schunck method

(Horn & Schunck, 1981) combined the gradient constraint (2.14) with a global smoothness term to penalize the variation of the velocity field $\mathbf{v}(x, y, t) = (u(x, y, t), v(x, y, t))$, minimizing the following (energy) criterion:

$$E(\mathbf{v}) = \int_{\mathbb{T}} (\mathbf{v} \cdot \nabla I + I_t)^2 dxdy + \alpha \int_{\mathbb{T}} ||\nabla \mathbf{v}||^2 dxdy \tag{2.15}$$

where $||\nabla \mathbf{v}||^2 = ||\nabla u||^2 + ||\nabla v||^2 = (\frac{\partial u}{\partial x})^2 + (\frac{\partial u}{\partial y})^2 + (\frac{\partial v}{\partial x})^2 + (\frac{\partial v}{\partial y})^2$

The first term is the constraint of the optical flow, which is attached to data, thus called external energy. The second term is the regularized stress on $\mathbf{v}$ for regularization, thus called internal energy.

We derive the gradient of E, integrating by part, we find that:

$$\nabla E(\mathbf{v}) = \begin{bmatrix} \frac{\partial E(u,v)}{\partial u} \\ \frac{\partial E(u,v)}{\partial v} \end{bmatrix} = \begin{bmatrix} 2I_x(I_x u + I_y v + I_t) - 2\alpha \triangle u \\ 2I_y(I_x u + I_y v + I_t) - 2\alpha \triangle v \end{bmatrix} \tag{2.16}$$

The problem can be written as:

$$\begin{aligned} I_x(I_x u + I_y v + I_t) - \alpha \triangle u = 0 \\ I_y(I_x u + I_y v + I_t) - \alpha \triangle v = 0 \end{aligned} \tag{2.17}$$

We apply a Gauss-Seidel method, the sequence (u^k, v^k) defined by:

$$\begin{aligned} (\alpha + I_x^2 + I_y^2)(u^{k+1} - \bar{u}^k) = -I_x(I_x \bar{u}^k + I_y \bar{v}^k + I_t) \\ (\alpha + I_x^2 + I_y^2)(v^{k+1} - \bar{v}^k) = -I_y(I_x \bar{u}^k + I_y \bar{v}^k + I_t) \end{aligned} \tag{2.18}$$

where $u^0 = v^0 = 0$

Finally, Horn and Schunck's algorithm obeys the following process:

1. Calculate the spatio-temporal image gradients, namely, (I_x, I_y, I_t)

2. Choose the number of iterations: a large number (slow convergence)
3. Choose the parameter α: it is set arbitrarily (it can be calibrated on test data)
4. Initialize the velocity vector field (u^0, v^0) to zero.
5. Update the motion vector as follows:

$$\begin{aligned} u^{k+1} &= \bar{u}^k - \frac{I_x(I_x\bar{u}^k + I_y\bar{v}^k + I_t)}{\alpha + I_x^2 + I_y^2} \\ v^{k+1} &= \bar{v}^k - \frac{I_x(I_x\bar{u}^k + I_y\bar{v}^k + I_t)}{\alpha + I_x^2 + I_y^2} \end{aligned} \tag{2.19}$$

2.2.1.2 Lucas-Kanade method

(Lucas & Kanade, 1981) implemented a weighted least-square (LS) fit of local first-order constraint (2.9) to a constant model for $\mathbf{v}$ in each small spatial neighborhood by minimizing:

$$\sum_{(x,y)\in\mathbb{T}} W^2(x,y)[\nabla I(x,y,t)\cdot\mathbf{v} + I_t(x,y,t)]^2 \tag{2.20}$$

We can also write in the form of integral where $W(x,y)$ denotes a 5×5 weighting function that gives more influence to constraints at the centre of the neighbourhood than those at the periphery. The solution to (2.20) is given by:

$$A^T W^2 A\mathbf{v} = A^T W^2 \mathbf{b} \tag{2.21}$$

Where, for n points $(x_c, y_c) \in \mathbb{T}$ at a single time t,

$$\begin{aligned} &A = [\nabla I(x_1,y_1), \ldots, \nabla I(x_n,y_n)], \\ &W = diag[W(x_1,y_1), \ldots, W(x_n,y_n)], \\ &\mathbf{b} = -(I_t(x_1,y_1), \ldots, I_t(x_n,y_n)) \end{aligned} \tag{2.22}$$

The solution to equation (2.21) is $\mathbf{v} = [A^T W^2 A]^{-1} A^T W^2 \mathbf{b}$, which is solved in closed form, if $A^T W^2 A$ is a nonsingular 2×2 matrix:

$$A^T W^2 A = \begin{pmatrix} \sum W^2(x,y) I_x^2(x,y) & \sum W^2(x,y) I_x(x,y) I_y(x,y) \\ \sum W^2(x,y) I_x(x,y) I_y(x,y) & W^2(x,y) I_y^2(x,y) \end{pmatrix} \tag{2.23}$$

where all sums are taken over points in the neighbourhood defined by $W(x,y)$.

The drawback of the classic optical flow method is that the computation requires much time (iteration and convergence), and it is not practical to realize the SHM in real time.

Recent researches related applied the convolutional neural networks (CNNs) to compute the optical flow with a reasonable time consumption and accuracy (Fischer et al., 2015; Ilg et al., 2016).

2.2.1.3 Optical flow variant along the gradient direction

Recent works have revisited the mathematical concept and showed that this family of methods can be used to achieve subpixel accuracy for the modal analysis of structures.

(Javh et al., 2017) used a simplified gradient-based optical flow method to estimate the subpixel shift of a steel beam. He applied the optical flow to estimate the displacement in the perpendicular direction of the edge. Equation (2.9) can be rewritten as:

$$\delta x \frac{\partial I}{\partial x}(x,y,t) + \delta y \frac{\partial I}{\partial y}(x,y,t) + \delta t \frac{\partial I}{\partial t}(x,y,t) = 0 \tag{2.24}$$

As the processing is made in a discrete time, $\delta t \frac{\partial I}{\partial t}$ is interpreted as the change in the intensity at a pixel between two sequential images:

$$\delta t \frac{\partial I}{\partial t}(x,y,t) = I(x,y,t+\delta t) - I(x,y,t) \tag{2.25}$$

Equation (2.24) is then developed as:

$$\delta x \frac{\partial I}{\partial x} + \delta y \frac{\partial I}{\partial y} = I(x,y,t) - I(x,y,t+\delta t) \tag{2.26}$$

Equation (2.26) is rewritten in the direction of the gradient (Lucas & Kanade, 1981) in the condition $\frac{\partial I}{\partial x}\delta y - \frac{\partial I}{\partial y}\delta x = 0$:

$$|\nabla I| d_s = I(x,y,t) - I(x,y,t+\delta t) \tag{2.27}$$

where $|\nabla I|$ is known as the scalar value of the intensity gradient and is calculated from the partial derivatives of the reference image:

$$|\nabla I| = \sqrt{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2} \tag{2.28}$$

d_s is the unknown scalar value of the displacement in the direction of the intensity gradient and is equivalent to $d_s = \sqrt{\delta x^2 + \delta y^2}$.

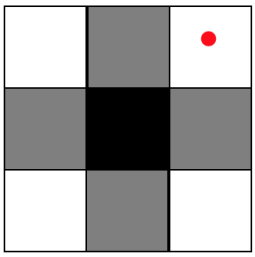


Figure 2.3 – 4-pixels neighborhood which is used for the Taylor-based method; the central pixel in black and its 4 surrounding pixels in gray are used for calculating the value at the red location. Red point represents the position of the central black pixel shifted with the subpixel displacement

2.2.2 Brightness interpolation method: Gradient-based 1st order Taylor approximation

According to (Chan et al., 2010), this approach is a simplified version of the classical optical flow method. It was originally proposed for image registration purposes.

As we only know the intensity of the image at integer coordinates $I_s(m, n, t_1)$, to obtain the intensity after a subpixel shift $I_s(m+\delta x, n+\delta y, t_1)$, a widespread linear approximation method is the 1st order Taylor series approximation (Chan et al., 2010; D. Zhang et al., 2016). Equation (2.6) can be then rewritten as follows:

$$\begin{aligned} I(m, n, t_2) &\approx I_s(m + \delta x, n + \delta y, t_1) \\ &\approx I_s(m, n, t_1) + \delta x \frac{\partial}{\partial x} I_s(m, n, t_1) + \delta y \frac{\partial}{\partial y} I_s(m, n, t_1) \end{aligned} \tag{2.29}$$

where the partial derivatives are computed in practice by the following centered finite difference processing as (2.30) and (2.31)

$$\frac{\hat{\partial}}{\partial x} I_s(m, n, t) = \frac{I_s(m + 1, n, t) - I_s(m - 1, y, t)}{2} \tag{2.30}$$

$$\frac{\hat{\partial}}{\partial y} I_s(m, n, t) = \frac{I_s(m, n + 1, t) - I_s(m, n - 1, t)}{2} \tag{2.31}$$

Then, eq. (2.29) expresses the local tangent plane at the central ongoing pixel, with a 4 pixels connectivity scheme, namely, the Von Neumann pixel neighborhood. The subpixel shift $(\delta x, \delta y)$ is found by minimizing the following sum of squared estimate of errors (SSE)

$\Psi(\delta x, \delta y)$ between data $I(m, n, t_2)$ and the modelled shifted brightness $\hat{I}_s(m+\delta x, n+\delta y, t_1)$ in eq. (2.29):

$$\Psi(\delta x, \delta y) = \sum_{m=1}^{M} \sum_{n=1}^{N} [I(m, n, t_2) - \hat{I}_s(m + \delta x, n + \delta y, t_1)]^2 \tag{2.32}$$

$$(\delta x^*, \delta y^*) = \arg \min_{(\delta x, \delta y)} \Psi(\delta x, \delta y). \tag{2.33}$$

For the problem at hand, the solution to eq. (2.33) can be analytically computed by solving the following linear system:

$$\left[\frac{\partial \Psi}{\partial \delta x}, \frac{\partial \Psi}{\partial \delta y} \right]^T = \mathbf{0} \tag{2.34}$$

The final result can be formulated into the matrix form below:

$$\begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{pmatrix} \delta x^* \\ \delta y^* \end{pmatrix} = \begin{bmatrix} D \\ E \end{bmatrix} \tag{2.35}$$

where:

— $A = \sum_{m=1}^{M} \sum_{n=1}^{N} (\frac{\hat{\partial} I_s(m,n,t_1)}{\partial x})^2$,

— $B = \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{\hat{\partial} I_s(m,n,t_1)}{\partial x} \frac{\hat{\partial} I_s(m,n,t_1)}{\partial y}$,

— $C = \sum_{m=1}^{M} \sum_{n=1}^{N} (\frac{\hat{\partial} I_s(m,n,t_1)}{\partial y})^2$,

— $D = \sum_{m=1}^{M} \sum_{n=1}^{N} (I_s(m, n, t_1) - I(m, n, t_2)) \frac{\hat{\partial} I_s(m,n,t_1)}{\partial x}$,

— $E = \sum_{m=1}^{M} \sum_{n=1}^{N} (I_s(m, n, t_1) - I(m, n, t_2)) \frac{\hat{\partial} I_s(m,n,t_1)}{\partial y}$.

The subpixel shift is finally analytically estimated as follows:

$$[\delta x^*, \delta y^*] = \left[\frac{CD - BE}{AC - B^2}, \frac{BD - AE}{B^2 - AC} \right]. \tag{2.36}$$

2.2.3 Correlation interpolation for subpixel shift estimation

Another widespread family of subpixel shift estimation methods is based on the interpolation of the pixel-level cross-correlation matrix, which is computed in Section 2.1. It

is only required to perform the interpolation within the vicinity of the cross-correlation peak. The selected interpolation function (e.g., quadratic surface or cardinal sinus) is then fitted to the pixel-level cross-correlation matrix around the cross-correlation peak.

The 9 integer pairs (m, n), with $m, n \in \{-1, 0, +1\}$, form a 3×3 grid:

$$\mathbb{G} = \begin{bmatrix} (-1,1) & (0,1) & (1,1) \\ (-1,0) & (0,0) & (1,0) \\ (-1,-1) & (0,-1) & (1,-1) \end{bmatrix}. \tag{2.37}$$

Let $(\Delta m^*, \Delta n^*)$ be resulting from the pixel level maximization (2.2). Accordingly, the 9 cross-correlation values $\rho(\Delta m^* + m, \Delta n^* + n)$, denoted also, *for shorter notations*, by

$$\gamma(m, n) \triangleq \frac{\rho(\Delta m^* + m, \Delta n^* + n)}{\rho(\Delta m^*, \Delta n^*)}, \tag{2.38}$$

form a matrix

$$\Gamma = \begin{bmatrix} \gamma(-1,1) & \gamma(0,1) & \gamma(1,1) \\ \gamma(-1,0) & \gamma(0,0) & \gamma(1,0) \\ \gamma(-1,-1) & \gamma(0,-1) & \gamma(1,-1) \end{bmatrix}. \tag{2.39}$$

2.2.3.1 Quadratic surface fitting (QSF)

A well known method to estimate the subpixel shift is employing the quadratic surface fitting the 3×3 correlation matrix Γ (Pan et al., 2005). The subpixel shift is estimated from the maximum of the fitted analytic equation. The second degree polynomial corresponding to a quadratic surface

$$p_\theta(x, y) = \theta_1 + [\theta_2 \ \theta_3] \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x \ y \end{bmatrix} \begin{bmatrix} \theta_4 & \theta_5/2 \\ \theta_5/2 & \theta_6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \tag{2.40}$$

with the vector $\theta \in \mathbb{R}^6$ collecting the scalar coefficients $\theta_1, \ldots, \theta_6$, is then fitted to the entry values of Γ for $(x, y) \in \mathbb{G}$, by solving the least squares problem

$$\min_{\theta \in \mathbb{R}^6} \sum_{(m,n) \in \mathbb{G}} [\, p_\theta(m, n) - \gamma(m, n)]^2. \tag{2.41}$$

we can easily get the values of $\theta \in \mathbb{R}^6$ minimizing (2.41) and the estimated subpixel shift corresponding to the maximum of this fitted quadratic surface:

$$x^* = \frac{2\theta_2\theta_6 - \theta_3\theta_5}{\theta_5^2 - 4\theta_4\theta_6}, y^* = \frac{2\theta_3\theta_4 - \theta_2\theta_5}{\theta_5^2 - 4\theta_4\theta_6}.$$

In addition to this quadratic surface fitting form, other fitting forms such as cubic surface, Gaussian surface or sinc (sinus cardinal) can also be applied.

2.2.3.2 Upsampling cross-correlation (UCC)

FFT-based upsampling is a widespread method to interpolate brightness data within images. Within the scope of this chapter, this method enables to obtain a refined peak location in the cross-correlation magnitude.

Assuming κ is the upsampling rate, the basic upsampling cross-correlation (UCC) method follows the following steps: (1) computing the discrete Fourier transform (DTF) of two images $I(m, n, t_1)$ and $I(m, n, t_2)$ as $F(u, v)$ and $G(u, v)$, (2) embedding the product $F(u, v)G^*(u, v)$ in a large array of zeros of dimension $(\kappa M, \kappa N)$, where the image resolution is $N \times M$ (known as zero-padding processing), (3) computing the inverse FFT to get the upsampled cross-correlation, (4) locating the peak of the upsampled cross-correlation.

Regarding the UCC method, (Thurman et al., 2008) introduces a special matrix multiplication implementation of the 2D DFT, to achieve the upsampled cross-correlation of dimension $(\kappa M, \kappa N)$ with a reduced computational burden. UCC is implemented by the product of three matrices with dimensions $(1.5\kappa, N)$, (N, M) and $(M, 1.5\kappa)$. The upsampled cross-correlation is thus computed in a very small neighborood of the cross-correlation peak location without the need to zero-pad the $F(u, v)G^*(u, v)$ data matrix. Subpixel displacement estimation is achieved by searching the peak in the output of $(1.5\kappa \times \ 1.5\kappa)$ upsampled array. The algorithm complexity is $O(MN\kappa)$, a obvious improvement compared to the traditional FFT method (Thurman et al., 2008). The limitation of this method is the precision requirement in the initial peak location estimation because the method only performs peak search within the small 1.5×1.5 pixel region of the original image frame.

2.2.3.3 The enhanced cross-correlation (ECC)

(Psarakis & Evangelidis, 2005) proposed an enhanced correlation-based method with subpixel accuracy, which is based on the ZNCC cross correlation computation method

(introduced in Table 2.1). The ECC uses a linear interpolation with subpixel shift in the intensity resulting in constructing the cross-correlation in a nonlinear continuous form. The optimum displacement (d_x, d_y) can be well estimated in an iterative approach for a preset initial displacement with reasonable time consumption.

The ECC is shown to reach better accuracy than the method in (Shimizu & Okutomi, 2002), which relies on some parabola fitting over the NCC cross-correlation variant (as defined in Table 2.1).

2.2.3.4 Phase-Only Correlation (POC) methods

Phase only correlation methods include different variants, which rely on the following normalized phase cross-correlation matrix in the Fourier domain:

$$Q(u,v) = \frac{F(u,v)G^*(u,v)}{|F(u,v)G^*(u,v)|} = e^{i(u\delta x + v\delta y)} \tag{2.42}$$

where $F(u,v)$ and $G(u,v)$ are the DFT of $I_s(m,n,t_1)$ and $I(m,n,t_2)$, respectively. In the Fourier domain, the normalized phase cross-correlation is a pure plane phase shift, which can expressed in the Cartesian coordinates of $u-v$ as:

$$\angle Q = u\delta x + v\delta y \tag{2.43}$$

The inverse Fourier Transform of $Q(u,v)$ leads to a sharper cross-correlation peak magnitude. The magnitude peak location can be computed using the family of methods presented in Section 2.2.3. (Takita et al., 2003) applied a close-form analytic model to fit the correlation peak with a windowing and weighting process to reduce the effect of discontinuity at the image contour and the high frequency components.

Some alternative methods of this family consist in either the plane fitting or the line fitting of the plane phase shift in eq. (2.43). These methods are known to have the best performance in the simulation test for 2D image. However, the phase shift angle is easily contaminated by the image noise. In addition, phase unwrapping can relieve the discontinuity problem, leading the measurement available in the range -0.5 to 0.5 pixels.

To improve the robustness and practicality, we have selected for this thesis the method by (Stone et al., 2001). It estimates the phase plane in the least square sense and allows masking both the high-frequency components and frequency components with small magnitude.

Within this family of techniques, the following two solutions are reported to also achieve good performance in motion extraction :

— (Ye et al., 2018) introduced the image gradient representation, phase filtering operation, a robust model fitting algorithm using higher than minimal subset sampling (HMSS) (Tennakoon et al., 2015) and robustness iteration operation.

— Line fitting computes separably each element of $Q(u, v)$ in eq. (2.42) if the interference term is ignored. Eq. 2.42 can be rewritten as:

$$Q(u, v) = e^{i(u\delta x + v\delta y)} = e^{(iu\delta x)} e^{(iv\delta y)} = q_m(u) q_n(v) \tag{2.44}$$

This implies that the subpixel estimation can be simplified by finding the optimum rank-one approximation of the normalized cross-power spectrum matrix. According to (Hoge, 2003), the best low-rank approximation was achieved by the singular value decomposition (SVD) method by the Eckart-Young-Mirsky theorem. The subpixel displacements in two directions can be separably estimated by line fitting of the phase angle of the left and right domain singular vectors.

2.2.4 Local phase-based subpixel shift estimation

A steerable filter-based method has been recently proposed in (J. G. Chen et al., 2015) and (Wadhwa et al., 2013) to estimate subpixel shift motion of structures from video images. The theory of steerable filter, which was introduced by Freeman in 1990's (Freeman & Adelson, 1991), allows estimating the local spatial phase of images from which the motion of the structure can be computed. This method has given rise to some demonstrative experiments by the MIT research team, e.g., see the videos list in Chen's thesis(J. Chen, 2016), and some further analysis by others, e.g., (Diamond et al., 2017). As a counterpart, the published materials by (J. G. Chen et al., 2015) do not allow to fully understand the practical use and limitations of the method. The theory which is briefly presented hereafter is matched to subpixel motion extraction only. To deal with the motion larger than one pixel, two solutions have been proposed : subsampling preprocessing in (J. G. Chen et al., 2015) to stay within this assumption, a multi-scale analysis of the image in (Wadhwa et al., 2013). In this thesis, another solution is proposed in Section 3.4.

Steerable filters have been originally introduced for contour and edge segmentation in image processing. This method relies on the presence of some strong texture in images. Within the scope of this thesis, the strong texture corresponds to the edges of the structure

to survey. In practice, the method determines the motion in the direction perpendicular to the edges, although the theory of the steerable filter may be generalized to any direction.

The spatial local phase and local magnitude are quantities analogous to the phase and the magnitude of the associated spatial signal. The local phase contains the information on object motion. For a video, with image brightness specified by $I(x, y, t)$ at spatial location (x, y) and time t, with two 2-D real kernels $G_2^\theta(x, y)$ and $H_2^\theta(x, y)$, the spatial filtering is achieved by convolving the image frame with the complex kernel $G_2^\theta(x, y) + iH_2^\theta(x, y)$. The resulting complex value at each location (x, y) is then written in the complex exponential form:

$$A_\theta(x, y, t)e^{i\phi_\theta(x,y,t)} = (G_2^\theta + iH_2^\theta) \bigotimes I(x, y, t) \tag{2.45}$$

where $A_\theta(x, y, t)$ is the local magnitude and $\phi_\theta(x, y, t)$ is the local spatial phase.

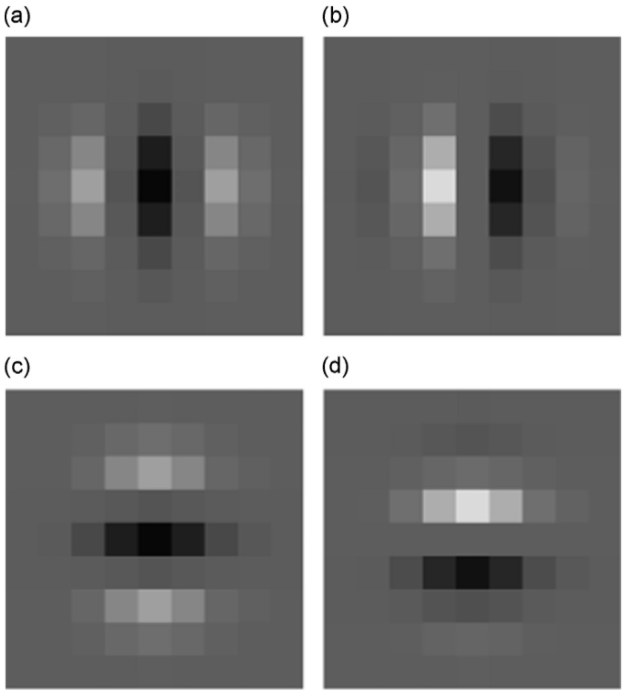


Figure 2.4 – Filter used to compute local phase and local magnitude, (a) real horizontal (G_2^0), (b) imaginary horizontal (H_2^0), (c) real vertical ($G_2^{\pi/2}$), (d) imaginary vertical ($H_2^{\pi/2}$); after (J. G. Chen et al., 2015)

The method for designing the steerable filters is detailed in (Freeman & Adelson, 1991). G_2^θ and H_2^θ are two orthogonal filters. A typical example of G_2^θ is the second derivative of the Gaussian function. Although not clearly stated in the available references, the

Gaussian parameters along the selection for the order of the derivative should be tuned according to the spatial frequency of the image, namely, the spatial variation of the image brightness across the edge. As far as we know, the authors did not report any sensitivity study on this matter.

Then we compute the displacement based on the phase signal. The phase is expected to be more robust against camera's noise compared to brightness amplitude. It has been demonstrated, e.g., (J. G. Chen et al., 2015), that constant contours of the local phase through time correspond to the displacement signal. It can be expressed as:

$$\phi_\theta(x, y, t) = c \tag{2.46}$$

for some constant c. Differentiating with respect to the time t yields

$$\left(\frac{\partial \phi_\theta(x, y, t)}{\partial x}, \frac{\partial \phi_\theta(x, y, t)}{\partial y}, \frac{\partial \phi_\theta(x, y, t)}{\partial t}\right) \cdot (u, v, 1) = 0, \tag{2.47}$$

where "$\cdot$" is the scalar product, u and v are the velocity in the xa and y directions respectively. To estimate u and v separately, the latter equation is projected onto the two perpendicular x and y axis, for which it is assumed that $\partial\phi_0(x, y, t)/\partial y \approx 0$ and $\partial\phi_{\pi/2}(x, y, t)/\partial x \approx 0$. The velocities in units of pixel are then computed as:

$$\begin{aligned} u &= -\left(\frac{\partial \phi_0(x,y,t)}{\partial x}\right)^{-1} \frac{\partial \phi_0(x,y,t)}{\partial t} \\ v &= -\left(\frac{\partial \phi_{\pi/2}(x,y,t)}{\partial y}\right)^{-1} \frac{\partial \phi_{\pi/2}(x,y,t)}{\partial t} \end{aligned} \tag{2.48}$$

(J. G. Chen et al., 2015) computes also the displacements d_x, d_y at time t by integrating the velocity, which give the local displacement of the image with subpixel accuracy at that pixel relative to the first frame $I(x, y, t_0)$.

2.3 Review on the SHM application

This section recalls some basic elements to formulate the dynamics of mechanical systems for the purpose of SHM based on sensor data. Further mathematical details are available in Appendix B.1.5.

Modal analysis is the process of determining the inherent vibration characteristics of a mechanical structure (natural frequencies, damping factors and mode shapes). The mathematical model from which the dynamic behavior of the system is extracted is called

the modal model of the system. In the experimental modal analysis, the aim is to achieve the dynamic characteristics of the system using the mathematical relations contained in the modal model. According to structural dynamics, the dynamic behavior of a system with multiple degree of freedom can be described by a linear time-invariant (LTI) equation:

$$\boldsymbol{M}\,\ddot{z}(t) + \boldsymbol{D}\,\dot{z}(t) + \boldsymbol{K}\,z(t) = \boldsymbol{u}(t) \tag{2.49}$$

where t denotes continuous time, and matrices $\boldsymbol{M}, \boldsymbol{D}, \boldsymbol{K} \in \mathbb{R}^{s\times s}$ denote mass, damping and stiffness matrices, respectively. Vectors $z(t) \in \mathbb{R}^s$ and $\boldsymbol{u(t)} \in \mathbb{R}^s$ denote the continuous-time displacements and the unknown external forces, respectively.

In practice, the measurements are realized at discrete time instants. According to the detailed derivation in Appendix B.1.5 and (Peeters & De Roeck, 1999), the discrete time state-space model based on the previous LTI system can be written as:

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \omega_k \\ \boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \nu_k \end{cases} \tag{2.50}$$

where the state vector $\boldsymbol{x}_k \in \mathbb{R}^l$ is composed of the position $z(k)$ and the speed $\dot{z}(k)$ variables, $\boldsymbol{y}_k \in \mathbb{R}^{n_s}$ is a vector collecting the outputs of sensors, ω_k and ν_k represent input and output noises respectively. The noise vectors are both unmeasurable vector signals, which are assumed to be zero mean white noise. The matrix $\boldsymbol{A} \in \mathbb{R}^{l\times l}$ is the state transition matrix. The matrix $\boldsymbol{C} \in \mathbb{R}^{n_s\times l}$ is the output matrix which specifies the relation between the internal states and the sensors.

In real applications, the inputs are not observed and only the outputs of sensors are available. Thus stochastic identification methods are deployed to estimate the modal parameters for this output-only modal analysis. This time domain state-space model can be identified by subspace identification methods. Let λ_j, ϕ_j be respectively the j-th eigenvalue of $\boldsymbol{A}$ and the associated eigenvector, with $j = 1, 2, \ldots, n$. The *mode shape* is defined as

$$\psi_j = \boldsymbol{C}\phi_j. \tag{2.51}$$

Let τ be the sampling period. The eigenvalues λ_j of $\boldsymbol{A}$ and the eigenvalues μ_j of the continuous time model are related by

$$\lambda_j = e^{\mu_j \tau}. \tag{2.52}$$

In vibration monitoring, two important parameters, the natural frequency f_j and the damping ratio ξ_j, are related to μ_j through

$$\mu_j = -2\pi f_j \xi_j \pm i2\pi f_j \sqrt{1-\xi_j^2} \tag{2.53}$$

The triples (f_j, ξ_j, ψ_j), for $j = 1, 2, \ldots, n_s$ are called modal parameters.

Several wide-spread methods to estimate the modal parameters from output sensor signals are reviewed in this thesis. They can be classified into 3 categories: 1. frequency-domain spectrum-driven methods; 2. time-domain covariance-driven methods; 3. time-domain data-driven methods.

2.3.1 Frequency-domain spectrum-driven methods

The simplest way to estimate the modal parameters is the peak-picking method (Bendat & Piersol, 1993) based on the spectral analysis of the displacement signals, where the eigenfrequencies are determined by identifying the peaks of the spectrum plot. The damping ratios are estimated by the half-power bandwidth method, which is generally considered not accurate. This method is popular due to its implementation simplicity with the only use of the Fast Fourier Transform (FFT) to transform time data to frequency spectra. A more advanced method, complex mode indication function (CMIF), uses the Singular Value Decomposition (SVD) of the frequency response functions (Shih et al., 1988). The maximum likelihood (ML) identification estimates the parameters of the model by minimizing the error in the least-squares sense for the measured frequency response functions(Pintelon et al., 1994).

2.3.2 Time-domain covariance-driven methods (SSI-COV)

Considering the similarity of mathematical expression for impulse responses and output covariance of the system excited by white noise, this observation was applied to classical impulse response based modal parameter estimation methods with output covariance instead (James et al., 1993). Two main methods, the instrumental variable method (IV) and covariance-driven stochastic subspace identification (SSI-COV), are presented in (Benveniste & Fuchs, 1985; Hermans & Van Der Auweraer, 1999).

2.3.3 Time-domain data-driven methods (SSI-DATA)

Different from SSI-COV, the time-domain data-driven method does not compute the covariance of the outputs. The idea is projecting the row space of future outputs into the row space of past outputs. What the SSI-DATA and SSI-COV have in common is that they use covariance or projection to cancel out the (uncorrelated) noise. The first SSI-DATA algorithms were presented in (Viberg, 1995).

2.3.4 Damage detection and identification

Based on the analysis of the modal parameters from a monitored structure, structural health monitoring can be achieved by detecting changes in modal parameters. According to (Rytter, 1993), vibration-based damage identification typically involves tasks of 4 levels:

1. Damage detection
2. Damage localization
3. Damage quantification
4. Prediction of the remaining service life of the monitored structure

In this thesis, we focus on damage detection. To our knowledge, the simplest way to detecting damages is by evaluating changes in frequencies compared with a previously registered reference. (O. Salawu, 1997) has reviewed the application of damage diagnosis using frequency changes. It has been noticed that the frequency shifts are limited to the measurements precision and sometimes frequency shifts are not sensitive enough for some special structure (Doebling et al., 1996; Farrar et al., 1994), e.g., offshore platforms. Thus damage identification based on mode shape changes can be applied to enhance the damage detection by frequency shifts. The Modal Assurance Criterion (MAC) is utilized to estimate the correlation level between the tested structure and the undamaged reference. Various studies (Ko et al., 1994; Ratcliffe, 1997; O. S. Salawu & Williams, 1994) have been conducted to examine changes in mode shapes. A recent related work (Greś et al., 2021) in our team relying on the MAC with uncertainty quantification is implemented to detect damages. More details are presented in chapter 5.

2.4 Conclusion

In this chapter, a variety of motion extraction methods from the literature were reviewed. A two-step coarse-to-fine strategy was introduced. We reviewed the cross-correlation criteria to estimate the pixel-level shift. In terms of subpixel shift estimation, the motion extraction algorithms are classified into 3 categories: 1. gradient-based approach; 2. correlation-based approach; 3. local phase-based approach. For the purpose of SHM, stochastic subspace identification methods were introduced, as well as their application to damage detection.

Chapter 3

Specific contributions to subpixel motion-extraction methods

This chapter is mainly devoted to our proposed new method for subpixel shift estimation, called *efficient quadrant interpolation* (EQI). To well explain the idea behind the EQI method, we will first present a method based on image interpolation from 5 pixels or 9 pixels. This method is quite similar to the Taylor method: it interpolates the image light intensity with a plane locally fitted to 5 or 9 neighboring pixels, instead of the tangent plane used by the Taylor method. To reduce the extent of the involved pixels, a natural idea is to fit a plane to 4 adjacent pixels at the corners of a square. However, before choosing the 4 pixels for fitting a plane, it is necessary to know the quadrant in the direction of subpixel displacement. Due to the fact that displacement in the horizontal or the vertical direction is frequently encountered in practice, the EQI method is then extended with interpolations in 8 directions. It is expected that the plane fitted to local pixels in the least square sense will produce more accurate interpolations than the tangent plane used by the Taylor method, yet preserving the simplicity of a quadratic optimisation problem for subpixel displacement estimation.

Then we present an extension of the steerable filter-based method to overcome the limitation to small displacements.

Finally,, we present a theoretic analysis of the quadratic surface fitting (QSF) method for subpixel displacement extraction. By fitting a quadratic surface the 3×3 correlation matrix around the correlation peak, the maximum value of the fitted surface provides the estimated subpixel shift. However, this method neglects the facts that the fitted quadratic surface may not have a maximum value, or the maximum value may lead to an estimated subpixel shift larger than 1 pixel. To avoid the occurrence of these two unexpected situations, we propose two sufficient conditions of the 9 values in the 3×3 cross-correlation matrix to ensure that the maximum exists and is within the one pixel vicinity of the cross-correlation peak.

3.1 Plane interpolation from 5 pixels or 9 pixels

The 1st order Taylor approximation method reviewed in Section 2.2.2 is based on estimated brightness $I_s(m+\delta x, n+\delta y, t_1)$ for small values of $(\delta x, \delta y)$. In Fig. 2.3, consider the first order Taylor expansion around the central point of the represented 3×3 pixels. In principle, the Taylor 1st order approximation is based on the tangent plane of $I_s(x, y, t_1)$ at the central point of this area of 3×3 pixels. In practice, the partial derivatives of $I_s(x, y, t_1)$ are estimated by finite difference as in (2.30) and (2.31), then this tangent plane is approximately determined from the 5 pixels (the central black pixel and its 4 neighbors gray pixels) represented in Fig. 2.3. Instead of this tangent plane determined by finite difference approximation, we can interpolate $I_s(m + \delta x, n + \delta y, t_1)$ with a plane that best matches the same 5 pixels in the least squares sense as shown in Fig. 3.1. This modification then leads to a method similar to the 1st order Taylor approximation method, with the only difference consisting in replacing, in the 1st order Taylor approximation method, Eq. (2.29) by Eq. (3.1)

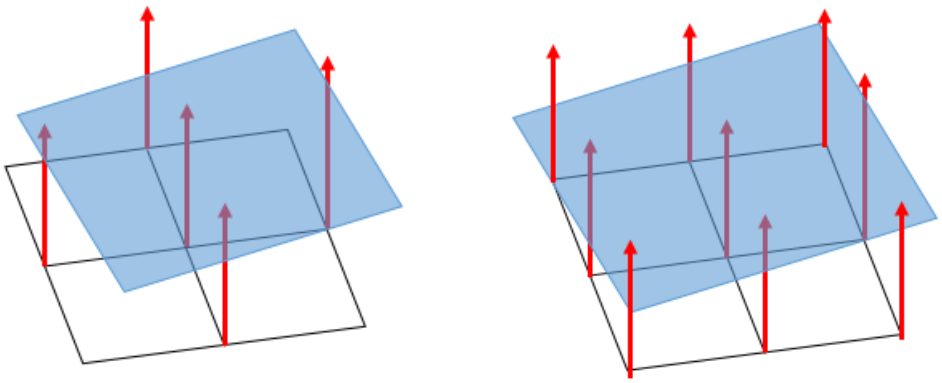
$$I_s(m + \delta x, n + \delta y, t_1) \approx \begin{bmatrix} 1 & m + \delta x & n + \delta y \end{bmatrix} \hat{\mathbf{a}}_{mn} \tag{3.1}$$

where the coefficient vector $\hat{\mathbf{a}}_{mn} = [\hat{a}_{0mn}, \hat{a}_{1mn}, \hat{a}_{2mn}]^T$ that best fits the image brightness within the selected 5 pixels around the central pixel located at (m, n) minimizes the following least squares criterion:

$$(\hat{a}_{0mn}, \hat{a}_{1mn}, \hat{a}_{2mn}) = \arg\min_{\mathbf{a}_{mn}} (\boldsymbol{X}_5 \mathbf{a}_{mn} - \boldsymbol{I}_5)^T (\boldsymbol{X}_5 \mathbf{a}_{mn} - \boldsymbol{I}_5). \tag{3.2}$$

where $\boldsymbol{X}_5 = \begin{bmatrix} 1 & m & n \\ 1 & m & n+1 \\ 1 & m & n-1 \\ 1 & m+1 & n \\ 1 & m-1 & n \end{bmatrix}$ and $\boldsymbol{I}_5 = \begin{bmatrix} I_s(m, n, t_1) \\ I_s(m, n+1m, t_1) \\ I_s(m, n-1, t_1) \\ I_s(m+1, n, t_1) \\ I_s(m-1, n, t_1) \end{bmatrix}$.

This proposed method, referred to as "Linear5" or "Linear interpolation with 5 pixels", interpolates $I_s(m + \delta x, n + \delta y, t_1)$ in the neighborhood of (m, n) from 5 pixels. A variant method would consist in fitting a plane, in the least squares sense, to the 9 pixels illustrated



(a) Plane interpolation with 5 pixels (b) Plane interpolation with 9 pixels

Figure 3.1 – The fitted local plane by interpolation with 5 pixels or 9 pixels. Each arrow represents a pixel, and the height of the arrow refers to the pixel intensity. The blue plan covering the selected 5 or 9 pixels is the local plane fitted with its pixel intensity

in Fig. 3.1b, instead of the 5 pixels, by replacing $\boldsymbol{X}_5$ and $\boldsymbol{I}_5$ with $\boldsymbol{X}_9$ and $\boldsymbol{I}_9$ defined as:

$$\boldsymbol{X}_9 = \begin{bmatrix} 1 & m & n \\ 1 & m & n+1 \\ 1 & m & n-1 \\ 1 & m+1 & n \\ 1 & m-1 & n \\ 1 & m+1 & n+1 \\ 1 & m+1 & n-1 \\ 1 & m-1 & n+1 \\ 1 & m+1 & n-1 \end{bmatrix}, \quad \boldsymbol{I}_9 = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n+1,t_1) \\ I_s(m,n-1,t_1) \\ I_s(m+1,n,t_1) \\ I_s(m-1,n,t_1) \\ I_s(m+1,n+1,t_1) \\ I_s(m+1,n-1,t_1) \\ I_s(m-1,n+1,t_1) \\ I_s(m+1,n-1,t_1) \end{bmatrix}. \tag{3.3}$$

Increasing the number of pixels would decrease the error variance when fitting the plane and increase the robustness to the noise, but also increase the bias due to the larger area covered by the 9 pixels.

To quantify the performance of various methods, the following root mean square error (RMSE) is computed for each method:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_s}[(d_x(i) - \hat{d}_x(i))^2 + (d_y(i) - \hat{d}_y(i))^2]}{N_s}} \tag{3.4}$$

RMSE (pixel) Noise condition / Method	Noiseless	High-level Noise
Taylor	0.0175	0.0570
Linear5	0.0175	0.0514
Linear9	0.0296	0.0247

Table 3.1 – Comparison of RMSE for methods Taylor, interpolation with 5 pixel (Linear5) and interpolation with 9 pixel (Linear 9) in the simulation. For the high-level noise condition, the intensity of the circular target is 215 and the intensity of the uniform background is 40, white Gaussian noise is added with variance 8.

where N_s is the frame number and $\hat{d}_x, \hat{d}_y$ represent the measured displacement in the horizontal and vertical directions, while d_x, d_y represent the true displacement. The RMSE will be used as the criterion to compare the subpixel shift estimation accuracy for all the applied methods.

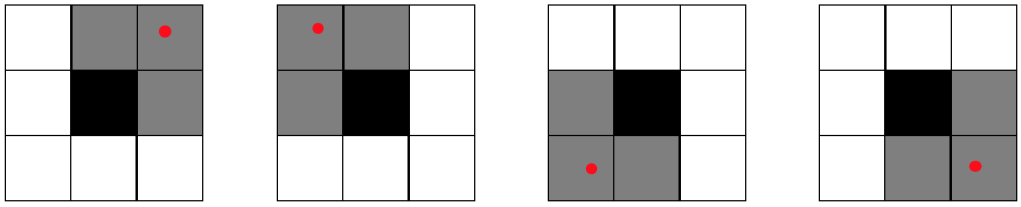
Tab. 3.1 shows a performance (RMSE) comparison between the Taylor method and the linear interpolation with 5 or 9 pixels based on numerical simulations. In the noiseless and noisy condition, the performance of the method Taylor is similar to that of the interpolation with 5 pixels. The interpolation with 9 pixels is more robust to noise compared with the Taylor method and the interpolation with 9 pixels, sacrificing the accuracy in the noiseless condition.

To favor small bias errors, a plane can be fitted to 4 neighboring pixels forming a 2×2 square. Though it is possible to determine a plane with 3 pixels, such a configuration would lack symmetry. Therefore, a new method will be proposed in the next section based on the interpolation of $f_s(m + \delta x, n + \delta y)$ from 2×2 pixels.

3.2 Efficient quadrant interpolation

Before introducing our proposed method, we recall the schema of the Taylor approximation in Fig. 2.3. Based on the central pixel located at (m, n) in the shifted image $I_s(m, n, t_1)$ as in Eq. (2.5) with pixel-level shift $(\Delta m, \Delta n)$, it remains the subpixel shift to estimate. The central pixel located at (m, n) changes its position after the subpixel shift and the new position can be in any direction around the central pixel.

We can divide the example in Fig. 2.3 into 4 square areas (or 4 quadrants which will be explained afterwards), each square area consists of 4 pixels as shown in Fig. 3.2. The



(a) Subpixel moving direction in Quadrant I

(b) Subpixel moving direction in Quadrant II

(c) Subpixel moving direction in Quadrant III

(d) Subpixel moving direction in Quadrant IV

Figure 3.2 – Quadrant-dependent 3-pixels neighborhood which is used for the EQI method; the central pixel in black and its 3 surrounding pixels in gray are used for interpolating the value at the red location.

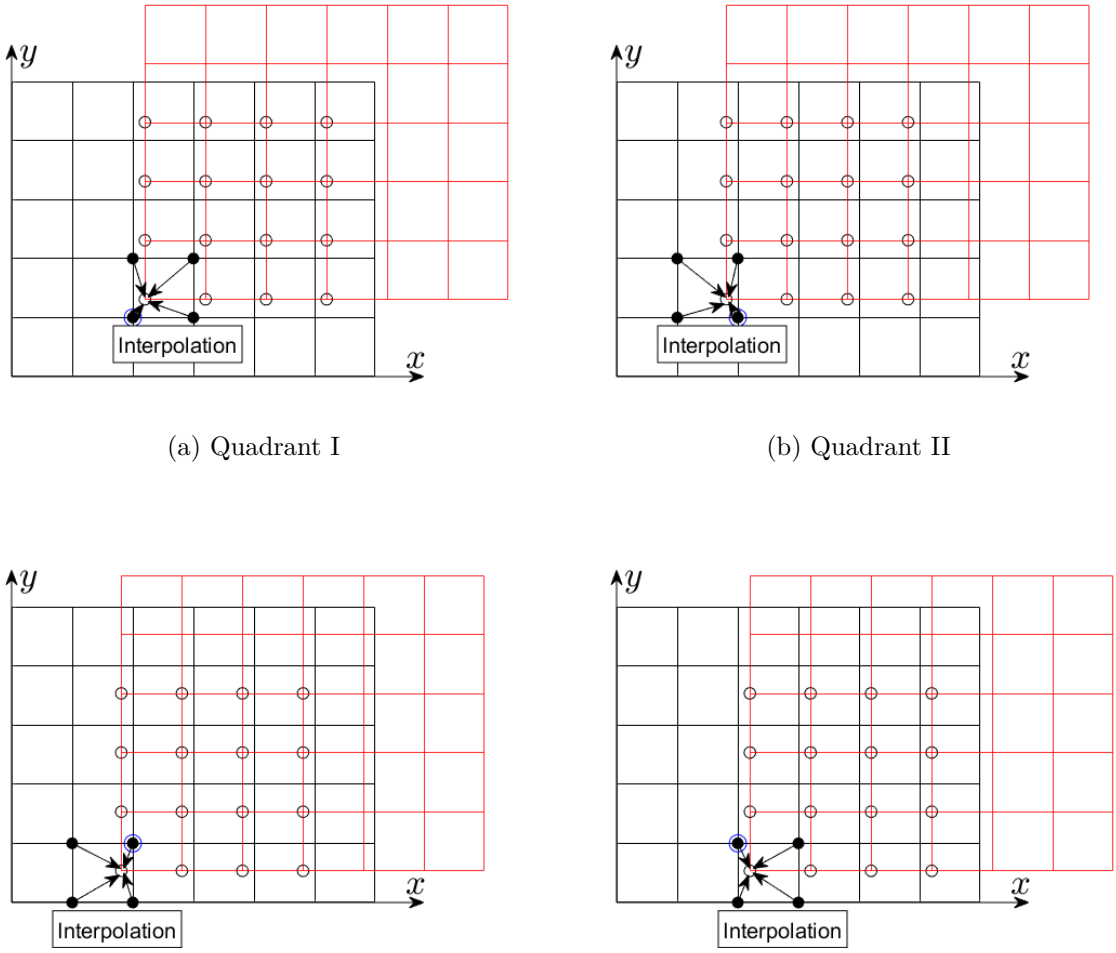
central pixel after the subpixel shift can be located at one of the 4 square areas. Consider the Cartesian coordinate system with its origin coinciding with the central pixel before its subpixel shift, the four square areas correspond to the 4 quadrants in this Cartesian system, i.e., the top-right area corresponds to the quadrant I, the top-left area corresponds to the quadrant II, the bottom-left area corresponds to the quadrant III and the bottom-right area corresponds to the quadrant IV, as shown in Fig. 3.2.

Different from the 1st order Taylor approximation, the method proposed in this section, called *efficient quadrant interpolation* (EQI), intends to interpolate the brightness $I_s(m+\delta x, n+\delta y, t_1)$ in a smaller pixel neighborhood: the square area surrounded by 4 pixels including the central pixel and 3 other pixels, corresponding to one of the 4 quadrants as shown on Fig. 3.2.

As $I_s(m, n, t_1)$ is the shifted image $I(m, n, t_1)$ based on the pixel-level cross-correlation maximisation, we assume that the shift between $I_s(m, n, t_1)$ and $I(x, y, t_2)$ is not larger than one pixel. Then subpixel-level estimation will be computed between $I_s(m+\delta x, n+\delta y, t_1)$ and $I(m, n, t_2)$ for $\delta x \in (-1, 1)$ and $\delta y \in (-1, 1)$. Fig. 3.3 shows the 4 possible subpixel moving directions resulting in different situations of interpolation.

In reality, the subpixel moving direction may be in any of the 4 quadrants. If we do not know the direction of the subpixel shift, we may try all the 4 possible directions, leading to an exhaustive computation. For computational efficiency, the direction of the subpixel shift should be "guessed" somehow.

Now the key problem is to estimate the direction of the subpixel shift before applying the linear interpolation in the corresponding quadrant. The pixel neighborhood where to



(a) Quadrant I

(b) Quadrant II

(c) Quadrant III

(d) Quadrant IV

Figure 3.3 – 4 cases of interpolation in the 4 subpixel moving directions

apply the linear interpolation is selected from further data analysis of the cross-correlation matrix. We will give more details in the following parts to explain how to select the subpixel moving direction.

After determining the subpixel moving direction, we apply linear interpolation so that the subpixel shift estimation remains a simple least squares minimisation, similar to (2.32). The new proposed method is expected to achieve better performance in motion extraction.

The proposed quadrant interpolation consists of the following three steps:

1. Constructing a 2×2 matrix based on the 3×3 cross-correlation matrix ZNCC to determine the most likely quadrant enclosing the subpixel moving direction.

2. Determining the local plane that fits the brightness of the 4 pixels in the selected quadrant.
3. Analytically computing the subpixel shift.

3.2.1 Quadrant selection

The 8-connected pixels neighborhood around the ongoing central pixel is divided into 4 subsets of 4−connected pixels sharing the common central pixel, as shown on Fig. 3.2. The first step of the proposed method consists in determining the most likely subpixel moving direction.

To this aim, a simple test relies on some further analysis of the 3×3 cross-correlation matrix ZNCC, which was introduced in chapter 2 and can be expressed as follows:

$$\text{ZNCC} = \begin{bmatrix} \alpha_{-1,1} & \alpha_{0,1} & \alpha_{1,1} \\ \alpha_{-1,0} & \alpha_{0,0} & \alpha_{1,0} \\ \alpha_{-1,-1} & \alpha_{0,-1} & \alpha_{1,-1} \end{bmatrix} \tag{3.5}$$

where $\alpha_{0,0}$ is the normalized peak value in cross-correlation, (i, j) is the relative position to the correlation peak. To determine the most likely moving displacement, the following 2×2 matrix H is introduced:

$$\mathbf{H} = \begin{bmatrix} h_{II} & h_{I} \\ h_{III} & h_{IV} \end{bmatrix} \tag{3.6}$$

where each element in $\mathbf{H}$ is defined as:

$$h_I = \alpha_{1,0} + \alpha_{1,1} + \alpha_{0,1} \tag{3.7a}$$

$$h_{II} = \alpha_{-1,0} + \alpha_{-1,1} + \alpha_{0,1} \tag{3.7b}$$

$$h_{III} = \alpha_{-1,-1} + \alpha_{0,-1} + \alpha_{-1,0} \tag{3.7c}$$

$$h_{IV} = \alpha_{1,-1} + \alpha_{0,-1} + \alpha_{1,0} \tag{3.7d}$$

Most of the time, the largest element in $\mathbf{H}$ indicates the subpixel moving direction among the first 4 quadrants in Fig. 3.2. For example, if the top-right element h_I is the largest in matrix $\mathbf{H}$, the moving direction is towards top-right, i.e., in quadrant I. Then the local plane will be fitted to the 4 pixels in the selected quadrant.

During simulation tests, we add additive white Gaussian noise to verify the stabil-

ity of estimating the direction of quadrant with our proposed technique compared with quadratic surface fitting (QSF).

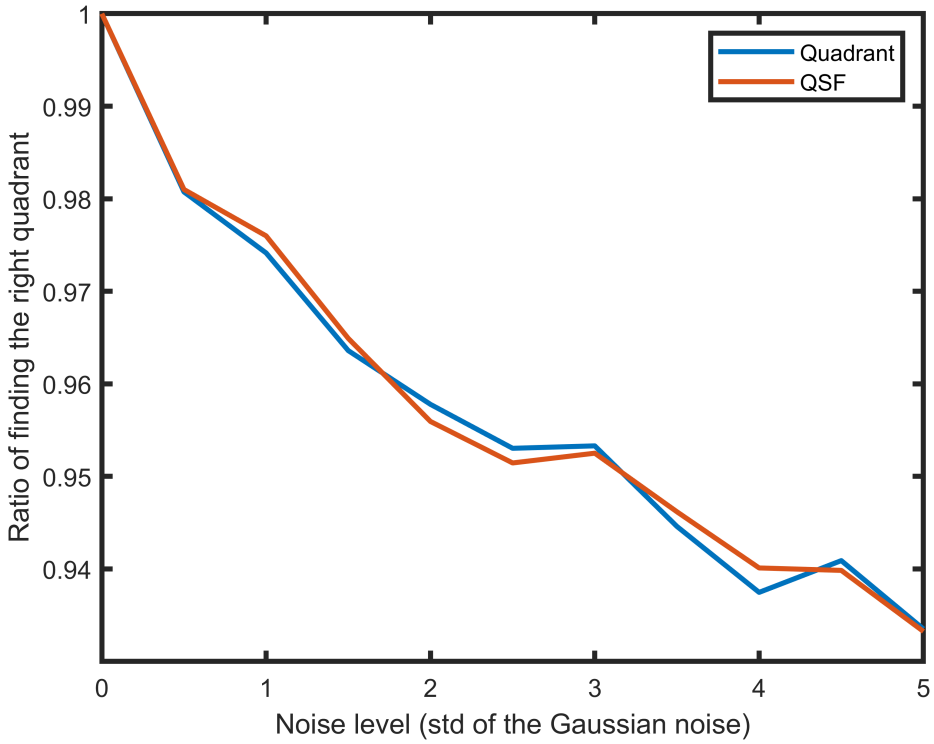


Figure 3.4 – The ratio of finding the right quadrant by our proposed method compared to QSF method for different levels of noise in the simulation test. The noise level is the standard deviation of the Gaussian noise. The details of the simulation will be given in Chapter 4.

As shown in Fig. 3.4, without the quadratic surface fitting and the computation of the subpixel shift, our proposed method to select the direction of the quadrant can reach a performance similar to QSF, which implies that our quadrant selection is a reliable technique.

3.2.2 Local plane fitting

Instead of the first order Taylor approximation in Eq. (2.29), the image brightness is now modeled by the following plane equation:

$$\hat{I}_s(m+\delta x, n+\delta y, t_1) \approx a_{0mn} + a_{1mn}(m+\delta x) + a_{2mn}(n+\delta y). \tag{3.8}$$

Let us first consider the case of plane fitting in quadrant I. The coefficient vector $\mathbf{a}_{mn} = [a_{0mn}, a_{1mn}, a_{2mn}]^T$ that fits the image brightness within the selected 4 pixels in quadrant I minimizes the following MSE error:

$$\min_{\mathbf{a}_{mn}} \psi(\mathbf{a}_{mn}) = \min_{\mathbf{a}_{mn}} \sum_{m_0=m}^{m+1} \sum_{n_0=n}^{n+1} [a_{0mn} + a_{1mn} m_0 + a_{2mn} n_0 - I_s(m_0, n_0, t_1)]^2, \tag{3.9}$$

or expressed in matrix form as follows :

$$\min_{\mathbf{a}_{mn}} \psi(\mathbf{a}_{mn}) = \min_{\mathbf{a_{mn}}} ||\mathbf{X}_I \mathbf{a}_{mn} - \mathbf{F}_I||^2 \tag{3.10}$$

where $\mathbf{X}_I = \begin{bmatrix} 1 & m & n \\ 1 & m & n+1 \\ 1 & m+1 & n \\ 1 & m+1 & n+1 \end{bmatrix}$ and $\mathbf{F}_I = \begin{bmatrix} I_s(m, n, t_1) \\ I_s(m, n+1, t_1) \\ I_s(m+1, n, t_1) \\ I_s(m+1, n+1, t_1) \end{bmatrix}$.

The LS solution for quadrant I is then given by:

$$\hat{\mathbf{a}}_{mn} = (\mathbf{X}_I^T \mathbf{X}_I)^{-1} \mathbf{X}_I^T \mathbf{F}_I. \tag{3.11}$$

For completeness, the expression for $\mathbf{X}$ and $\mathbf{F}$ in quadrants II to IV is available in appendix A.3. Thus, this processing step provides a collection of $M \times N$ coefficient vectors $\hat{\mathbf{a}}_{mn}$ of fitted local planes by treating every pixel of the whole ROI as the central pixel in Fig. 3.2. The next step is then to use the local linear model to compute the subpixel shift.

3.2.3 Subpixel shift computation

The subpixel shift $(\delta x, \delta y)$ is found by minimizing the least squares criterion between the data $I(m, n, t_2)$ and the modeled shifted brightness $\hat{I}_s(m + \delta x, n + \delta y, t_1)$ in eq (3.8):

$$(\delta x^*, \delta y^*) = \arg \min_{(\delta x, \delta y)} \Psi(\delta x, \delta y) \tag{3.12}$$

where $\Psi(\delta x, \delta y)$ is the sum of squared errors over the $M \times N$ image pixels, defined as:

$$\Psi(\delta x, \delta y) = \sum_{m=1}^{M} \sum_{n=1}^{N} [I(m, n, t_2) - \hat{I}_s(m + \delta x, n + \delta y, t_1)]^2. \tag{3.13}$$

The solution of Eq. (3.12) is obtained the same way as in eq. (2.34). The solution of the corresponding linear system in eq. (2.35) is given by Eq. (2.36). The elements A to E of the linear system are now defined as follows for the EQI method:

— $A = \sum_{m=1}^{M} \sum_{n=1}^{N} \hat{a}_{1mn}^2$,

— $B = \sum_{m=1}^{M} \sum_{n=1}^{N} \hat{a}_{1mn}\hat{a}_{2mn}$,

— $C = \sum_{m=1}^{M} \sum_{n=1}^{N} \hat{a}_{2mn}^2$,

— $D = \sum_{m=1}^{M} \sum_{n=1}^{N} [\hat{a}_{1mn} I(m,n,t_2) - (\hat{a}_{0mn}\hat{a}_{1mn} + \hat{a}_{1mn}^2 m + \hat{a}_{1mn}\hat{a}_{2mn} n)]$,

— $E = \sum_{m=1}^{M} \sum_{n=1}^{N} [\hat{a}_{2mn} I(m,n,t_2) - (\hat{a}_{0mn}\hat{a}_{2mn} + \hat{a}_{1mn}\hat{a}_{2mn} m + \hat{a}_{2mn}^2 n)]$.

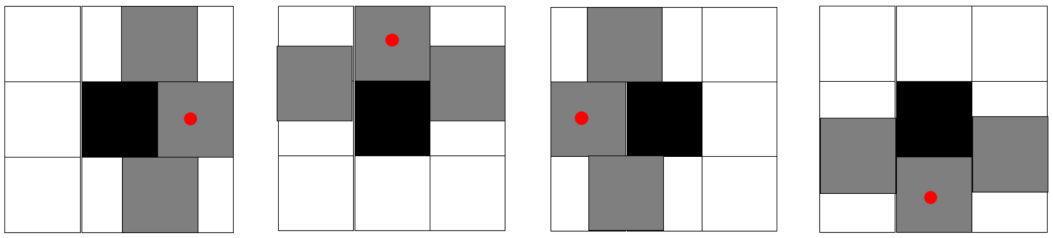
3.3 Efficient octant interpolation

The EQI method assumes that the subpixel displacement is in one of the 4 quadrants, excluding the cases of displacements in the horizontal or vertical direction. If the true displacement was randomly distributed in all the possible directions, then the probability of displacements in the horizontal or vertical direction would be negligible. However, in practice, horizontal and vertical displacements are frequently encountered. In such situations, the EQI method is not suitable. For this reason, we extend the EQI method by considering 8 octants instead of 4 quadrants, as shown in Fig. 3.5.

Based on the method EQI, after finding the maximum element in **H**, we will compare with its adjacent element. If the difference between the largest element and its adjacent element is below a certain threshold, the row/column location of the two largest matrix elements points out the moving direction among the directions V to VIII in Fig. 3.5. For example, if (h_I, h_{II}) are the two largest elements with $h_I - h_{II} < 0.1\%$, the moving direction is upward within the direction V, namely, in the x positive, and then we apply the same process as implemented in the EQI.

3.4 Extension of the streerable-based local phase approach

A subpixel motion extraction method based on spatial local phase extraction with steerable filters has been shortly introduced in Section 2.2.4. Steerable filters have been



(a) Subpixel moving direction in the direction V, namely, in the x positive direction

(b) Subpixel moving direction in the direction VI,namely, in the y positive direction

(c) Subpixel moving direction in the direction VII, namely, in the x negative direction

(d) Subpixel moving direction in the direction VIII,namely, in the y negative direction

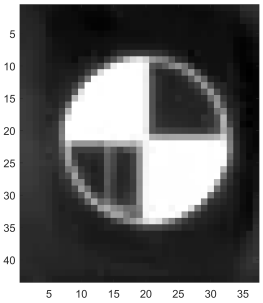
Figure 3.5 – Octant-dependent 3-pixels neighborhood which is used for the octant method; the central pixel in black and its 3 surrounding pixels in gray are used for interpolating the value at the red location. The gray square covering the 2 white squares takes the mean gray level value of the two white squares

used for motion extraction in (J. G. Chen et al., 2015). In this approach, the local phases of active pixels in two images at different sampling time instants are used to estimate the velocity of the moving object captured in the images. This computation is focused on "active pixels", which are selected from the magnitude of the steerable filter and usually match to the edge of the moving object. For example, (J. G. Chen et al., 2015) have selected a threshold from the 30 pixels with the largest magnitude of steerable filters. The final estimated displacement is the average of the displacements extracted from "active pixels".

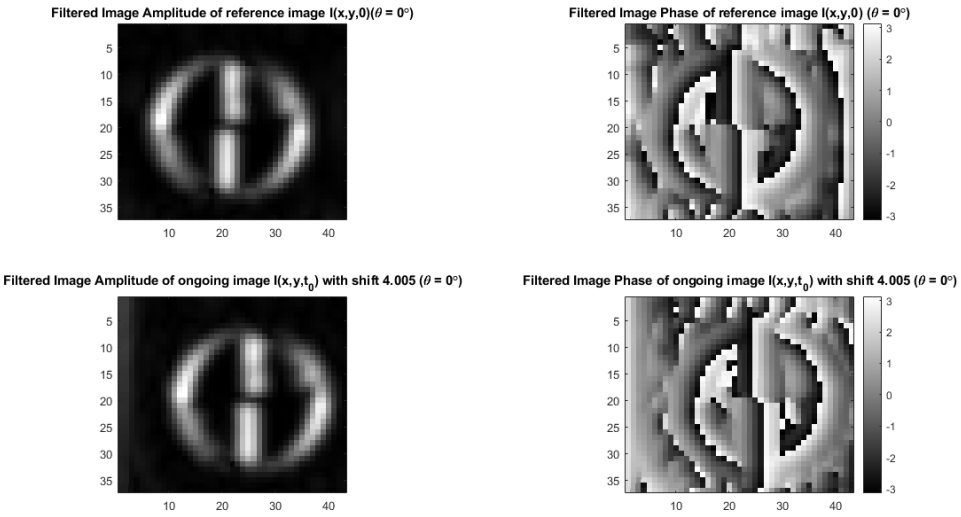
(J. G. Chen et al., 2015) assumed subpixel motions between the ongoing image $I(x, y, t)$ at instant t and the reference image $I(x, y, t_0)$ at time t_0. In real world applications, this condition is difficult to check, since the displacement amplitude is not known beforehand. For motions of larger amplitude, the solution proposed in (J. G. Chen et al., 2015) consists in subsampling images by some factor in order to recover the subpixel condition limit. The subsampling can be eventually iteratively performed. However, subsampling reduces the image contrast, smooths the image texture, and reduces the amount of active pixels.

In particular, according to the results in Fig. 3.6 , the phase coherence between two images extends to a neighbourhood larger than one pixel. Thus, the one pixel condition limit for the local phase using steerable filters is likely to be restrictive.

In this thesis, three solutions were explored to generalize the steerable filter-based method to large image motion (more than 1 pixel). We first tested a cumulative ap-



(a) Experimental image



(b) Output of steerable filter in the horizontal direction

Figure 3.6 – The extracted phase and amplitude of the reference experimental image $I(x, y, t_0)$ and the ongoing image $I(x, y, t)$ with shift 4.005 pixel by application of steerable filters in the horizontal direction

proach by considering the ongoing image at time t_n as the reference image to estimate the motion at time t_{n+1}. The subpixel motion condition is then much easier to satisfy. The motion is obtained by accumulating the estimated subpixel motions from successive images. Unfortunately, this solution provided uncontrolled trends (accumulated error) in motion estimation.

The second approach relies on the two-step processing strategy. It takes advantage of the estimated pixel level shift between the reference and current image frame by using the cross-correlation technique. Then, the image mask containing the selected active pixels on the reference image frame can be accordingly shifted to stay in the range of subpixel motions with the ongoing image. This solution has been found sensitive to some unexpected gaps (outliers for large displacement) sometimes, which reduces the overall performance for motion extraction.

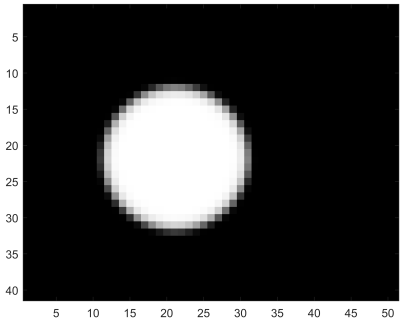
Finally, the third approach entitled "extended local phase using steerable filters" combines the two previous solutions, by introducing a moving reference image. The first solution is performed when the pixel-level shift is null compared to the reference image. Otherwise, the second solution is applied and a new reference image at that ongoing time sampling is changed accordingly.

The following sections will give the details of our proposition of selection of active pixels and the improved version of local phase using steerable filters titled "extended local phase using steerable filters" (EX-ST).

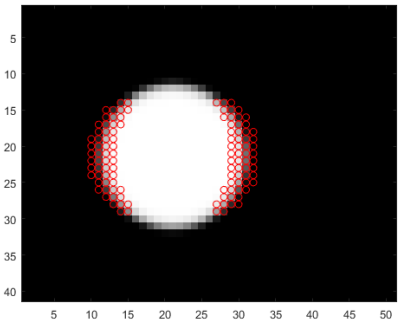
3.4.1 Chen's selection of active pixels

The proposition of (J. G. Chen et al., 2015) consists in selecting active pixels based on the local amplitudes in the results of the steerable filters as in e.q (2.45). The selected active pixels have their amplitudes larger than a threshold, which is calculated as the half of the median of the 30 pixels with the largest amplitudes. The active pixels selected following this proposition are represented for our simulated image in Fig. 3.7(a).

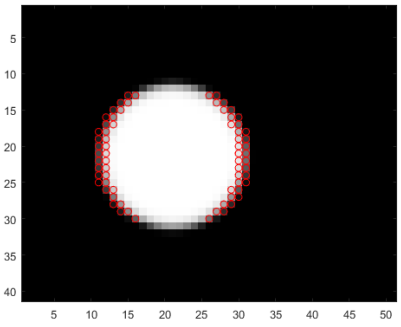
We have noticed that, with this proposition, some pixels selected represent badly the simulated displacements as shown in Fig. 3.8. By averaging the estimated displacements for all the active pixels could obviously reduce the estimation error, but the result is still not satisfactory. For the large displacement (more than 1 pixel), (J. G. Chen et al., 2015) proposed to downsample the original image until the displacement is less than 1 pixel. A drawback is that if the size of our video image is not large enough, after the process of downsampling, the final image cannot provide sufficient information. For example, in



(a) Simulated image



(b) Active pixels selected in the simulated image with the proposition of (J. G. Chen et al., 2015)



(c) Our proposed technique to select the active pixels in the simulated image

Figure 3.7 – Different techniques to select the active pixels (red circles) in the simulation: (a) Simulated image (b) Proposition by (J. G. Chen et al., 2015), (c) Our technique

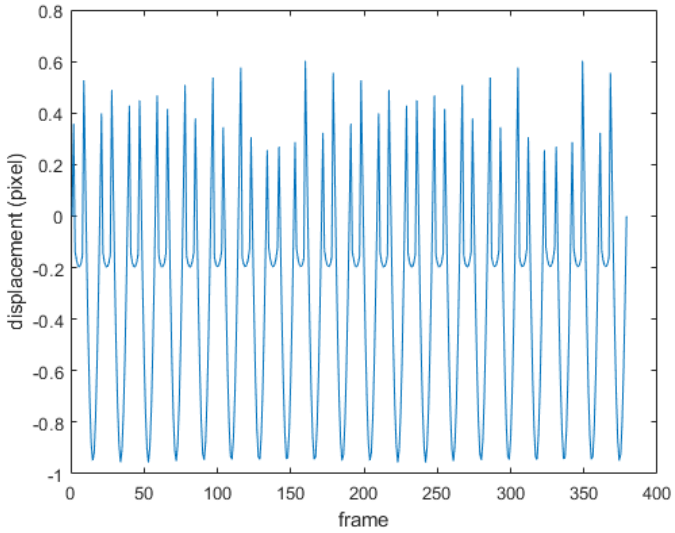


Figure 3.8 – The active pixel selected with the proposition (J. G. Chen et al., 2015) represent badly the displacement for the real displacement signal monotonic sinusoidal of amplitude 1 pixel for the simulation presented in Fig. 3.7

our simulation, the simulated image template is of size 50×50 with maximum shift amplitude 4 pixels, thus we need to downsample the image template with a factor of 4 for each dimension to a size of 15×15. With this small dimension, the method "local phase using steerable filters" behaves poorly. Another limitation is that we need to know the maximum displacement amplitude in advance in order to apply suitably downsampling.

3.4.2 Vibration-based selection of active pixels

We propose a technique to select active pixels based on the energy of the light intensity. For each pixel at (x, y), we construct a time-series vector $\mathbf{L}(x, y)$ recording the gray level for all the frames. We apply a band-pass filter to filter the time-series gray level vector $\mathbf{L}(x, y)$ to obtain the filtered time-series $\tilde{\mathbf{L}}(x, y)$. Then its l^2-norm $\|\tilde{\mathbf{L}}(x, y)\|$ is calculated for each pixel from $\tilde{\mathbf{L}}(x, y)$. The threshold is chosen as $1/4$ of the maximum l^2-norm among all the processed pixels, and the pixels whose l^2-norm is larger than this threshold will be considered as active pixels. The whole process can be summarized as:

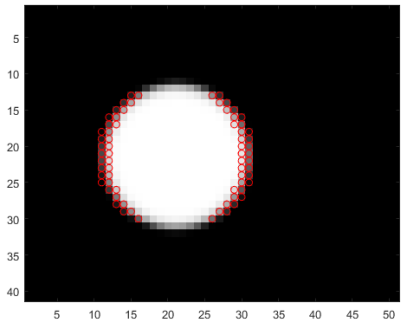
1. Constructing a time-series vector $\mathbf{L}(x, y)$ recording the gray level for all the frames at each position (x, y)
2. Filtering the time-series gray level vector $\mathbf{L}(x, y)$ with a band-pass filter to obtain

the filtered gray level time-series vector $\tilde{\mathbf{L}}(x, y)$

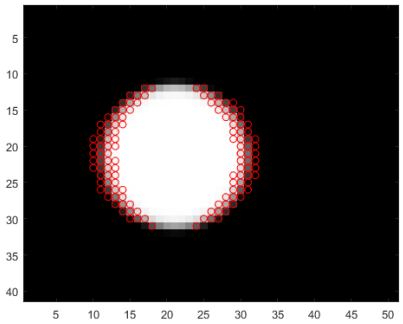
3. Calculating the l^2-norm $\|\tilde{\mathbf{L}}(x, y)\|$

4. The threshold is selected as 1/4 of the maximum l^2-norm, and the pixels whose l^2-norm is larger than this threshold will be considered as active pixels.

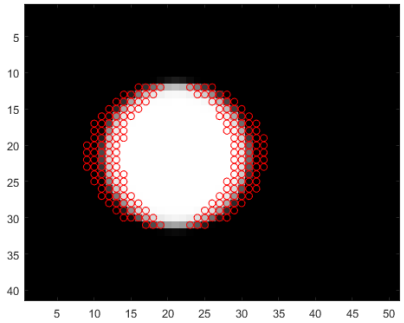
With this process, the active pixels selected are shown in Fig. 3.7(c). With our proposed technique, we have less active pixels which badly represent the displacements. When tested with experiments of different vibration amplitudes, this process is capable of automatically adapting the number and extend of selected pixels to the amplitudes as shown in Fig. 3.9.



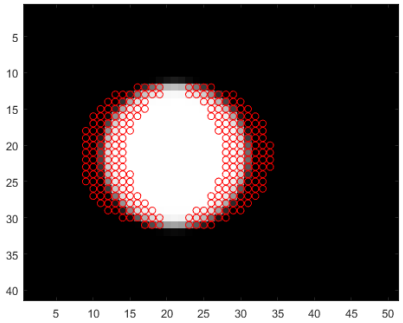
(a) Active pixels selected for maximum amplitude 1 pixel



(b) Active pixels selected for maximum amplitude 2 pixels



(c) Active pixels selected for maximum amplitude 3 pixels



(d) Active pixels selected for maximum amplitude 4 pixels

Figure 3.9 – Our proposed technique to select the active pixels for different maximum amplitudes of the displacement for the simulated image represented in Fig. 3.7(a)

3.4.3 Extended local phase using steerable filters (EX-ST)

The local phase computed with steerable filters can only estimate displacements less than 1 pixel for the reason of phase continuity. For larger displacements, it requires a pre-processing of downsampling to ensure that the displacement is smaller than 1 pixel. An issue of downsampling is that we may lose the texture information if the size of the original image is not large enough. To avoid this problem, we propose a method titled "Extended local phase using steerable filters" (EX-ST) taking advantages of the pixel-level shift estimated by cross-correlation techniques. With the obtained pixel-level shift, the selected active pixels can be accordingly shifted. The algorithm to estimate the displacement in horizontal direction is presented and the flowchart of the proposed extended local phase using steerable filters are presented below:

Algorithm 1: Extended local phase using steerable filters

input : A sequence of images: $I(i)$, $i \in 1 \dots N$
output: Horizontal displacement vector $d(i)$, $i \in 1 \dots N$
initialization;
$Ir \leftarrow I(1)$;
$activePixels \leftarrow determineMask(Ir)$ Determine the active pixels from the reference image;
for $i \leftarrow 1$ **to** N **do**
  $\Delta m = findPixelshift(Ir, I(i))$ Estimate the pixel-level shift;
  if $\Delta m == 0$ **then**
    $\delta x \leftarrow applyLocalPhaseSteerableFilter(Ir, I(i), activePixels)$ Use activePixels to calculate the subpixel displacement between the reference image and actual image;
    $d(i) \leftarrow \delta x$;
  else
    $activePixelNew = shiftMask(activePixels, \Delta m)$ Shift the mask with pixel-level shift;
    $\delta x \leftarrow applyLocalPhaseSteerableFilter(I(i-1), I(i), activePixelsNew)$ Estimate the subpixel shift from the two sequential images;
    $d(i) \leftarrow d(i-1) + \delta x$;
  end
end
return d

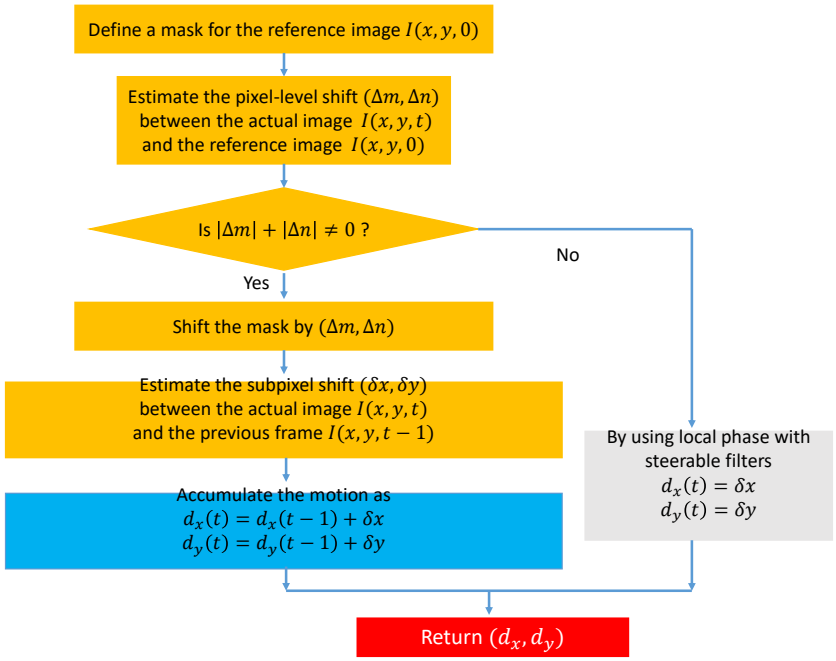


Figure 3.10 – Flowchart of our proposed extended local phase using steerable filters to estimate multi-pixel displacements

Table 3.2 – RMSE comparison between Steer (J. G. Chen et al., 2015) and our proposed EX-ST. N/A means that the method Steer is not capable to measure large displacements (4 pixels)

Algorithm	**Small displacements (1 pixel)**	**Large displacements (4 pixels)**
Steer	0.0146	N/A
EX-ST	0.0122	0.0449

With this extended version using cross-correlation technique, we can well estimate the displacement even though the displacement is larger than 1 pixel. We can also avoid the process of downsampling which reduces the image quality.

3.4.4 Performance assessments

Tab. 3.2 shows the RMSE obtained with the original local phase using steerable filters (Steer) and with our extend local phase using steerable filters (EX-ST) for small displacements and large displacements.

Our EX-ST has a better accuracy than the Steer (J. G. Chen et al., 2015) for the subpixel shift estimation and can reach a good estimation for the large displacements.

3.5 Analysis of the QSF method for subpixel motion extraction

In the quadratic surface fitting (QSF) method introduced in section 2.2.3.1, in order to gain subpixel accuracy at the second step, we apply a second degree polynomial (or, geometrically, a quadratic surface) $p_\theta(x, y)$ fitted to the 9 neighboring cross-correlation values around the cross-correlation peak, namely $\gamma(m, n)$ with $m, n \in \{-1, 0, +1\}$ in Eq. (2.39). Then the maximum of the fitted polynomial yields the estimated subpixel shift between the two templates (D. Zhang et al., 2016).

Satisfactory experimental results of this method have been reported by different authors, e.g., (Alba et al., 2015; X. Tong et al., 2019; D. Zhang et al., 2016). Our own experiments confirm also its good performance compared to other existing methods for subpixel shift extraction. The main purpose of this section is to consolidate the theoretical basis of this method.

More specifically, the quadratic surface fitting method, as described in the aforementioned references, assumes *implicitly* that the second degree polynomial $p_\theta(x, y)$ fitted to the 9 correlation values $\gamma(m, n)$, with $m, n \in \{-1, 0, +1\}$, *always* has a unique global maximum, corresponding to (x^*, y^*) located in the close vicinity of $(x, y) = (0, 0)$, so that the total shifts as expressed in (2.7) do not fall too far from the pixel-level optimal shifts $(\Delta m^*, \Delta n^*)$. The following issues will be investigated.

1. Does the second degree polynomial (or quadratic surface) $p_\theta(x, y)$ fitted to the 9 cross-correlation values around the cross-correlation peak *always* have a unique global maximum?
2. If this fitted polynomial *does not always* have a unique global maximum, what are the conditions ensuring its existence?
3. What should the algorithm do if ever this polynomial has no maximum?

Let us recall the process of the QSF method. The 9 cross-correlation values form a matrix as in Eq. (2.39). The central entry of Γ

$$\gamma(0, 0) = \frac{\rho(\Delta m^* + 0, \Delta n^* + 0)}{\rho(\Delta m^*, \Delta n^*)} = 1 \tag{3.14}$$

The second degree polynomial (2.40) is then fitted to the entry values of Γ by solving the least squares problem (2.41).

Does this fitted second degree polynomial $p_\theta(x, y)$ *always* have a unique global maximum?

As explained above, the central entry of the matrix Γ, namely $\gamma(0, 0) = 1$, is the maximum value among all the 9 entries of Γ. It then seems reasonable to expect that the second degree polynomial (or quadratic surface) $p_\theta(x, y)$ fitted to the 9 entries of Γ has a maximum somehow close to the (maximum) central entry of the matrix Γ corresponding to the origin $(x, y) = (0, 0)$.

Unfortunately, the fact that the central entry $\gamma(0, 0)$ is the maximum value among the 9 entries of Γ does not really ensure that the fitted second degree polynomial $p_\theta(x, y)$ always has a global maximum, as demonstrated by the following counterexample.

Counterexample. The counterexample presented below has been chosen for its simplicity so that it can be easily reproduced. The Matlab code for generating the presented result is available for download (Xiong et al., 2021). To show the robustness of this counterexample, the Matlab code can be optionally run with some random noise added to the generated synthetic images, though the result presented below is noise-free.

Consider two binary images of 18×18 pixels as shown in Fig. 3.11, with the template (ROI) chosen as the red square window of 8×8 pixels in each image. The template in the first image contains a diagonal pattern. In the second image, this diagonal pattern is shifted by 1 pixel toward the right and also by 1 pixel toward the bottom. In these binary images, the intensity is 1 at the darker pixels and 10 at the brighter pixels.

The normalized cross-correlation notation $\gamma(m, n)$ has been defined in Eq. (2.38) and the normalized 3×3 cross-correlation matrix around the peak as defined in Eq. (2.39) is

$$\Gamma = \begin{bmatrix} 0.2236 & 0.2236 & 0.8059 \\ 0.2236 & 1 & 0.2236 \\ 0.8059 & 0.2236 & 0.2236 \end{bmatrix}.$$

Fitting the second degree polynomial $p_\theta(x, y)$ to Γ for $(x, y) \in \mathbb{G}$ yields

$$\theta^* = \begin{bmatrix} 0.5256 \\ 0 \\ 0 \\ -0.0647 \\ 0.2911 \\ -0.0647 \end{bmatrix}.$$

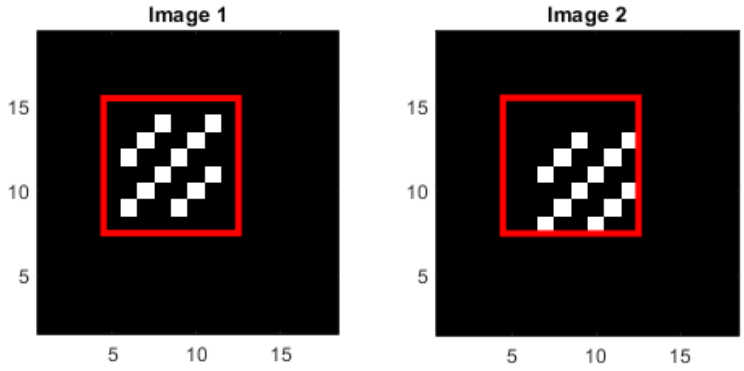


Figure 3.11 – Counterexample of the QSF method: two binary images of 18×18 pixels. The intensity is 1 at the darker pixels and 10 at the brighter pixels, The template (ROI) is selected as the red square window of 8×8 pixels.

The corresponding quadratic surface $p_\theta(x, y)$ exhibiting a saddle point, as illustrated in Fig. 3.12, has no global maximum, despite the fact that the central entry of the matrix γ is its largest entry.

This counterexample clearly invalidates the widespread intuition that the fitted polynomial $p_\theta(x, y)$ in the QSF method always has a global maximum.

After the answer to the first question raised before, the following sections will then answer the other two remaining questions.

3.5.1 Conditions for the existence of a maximum

In elementary algebra, it is well known that the second degree polynomial $p_\theta(x, y)$ as expressed in (2.40) has a unique global maximum if its Hessian matrix

$$H \triangleq \begin{bmatrix} \theta_4 & \theta_5/2 \\ \theta_5/2 & \theta_6 \end{bmatrix} \tag{3.15}$$

is negative definite (O'Meara, 1973). However, this simply stated fact does not directly

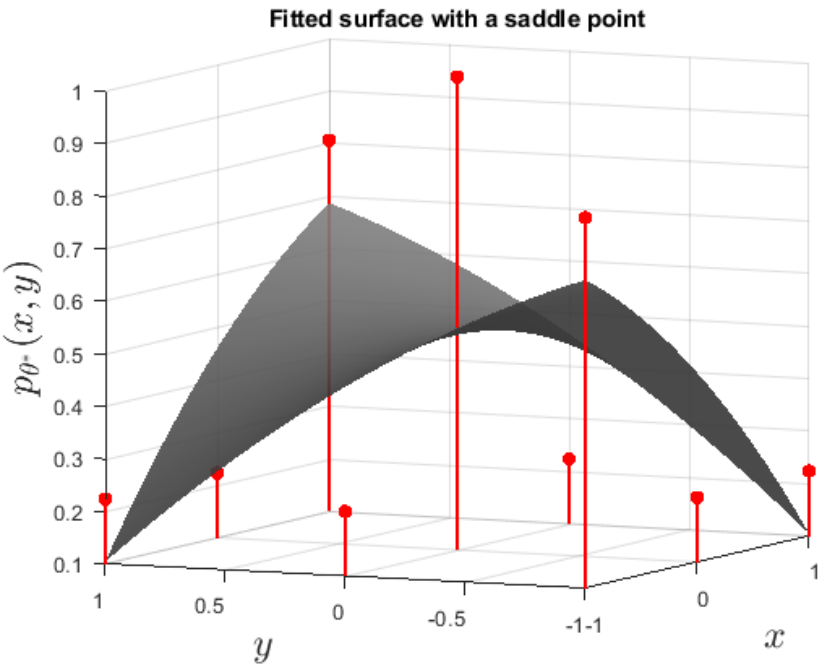


Figure 3.12 – Fitted quadratic surface exhibiting a saddle point. The vertical red line segments represent the entries of Γ.

help to understand how the normalized cross-correlation values $\gamma(m, n)$ (those filling up the Γ matrix in Eq. (2.39)) should be, so that the fitted $p_\theta(x, y)$ has a global maximum. Because the polynomial coefficients $\theta_1, \ldots, \theta_6$ are determined from the values of $\gamma(m, n)$ by solving the least squares problem Eq. (2.41), it is straightforward to express the negative definiteness condition of H in terms of $\gamma(m, n)$. Then, in principle, the condition for the existence of a global maximum of $p_\theta(x, y)$ will be formulated in terms of the normalized cross-correlation values $\gamma(m, n)$. Nevertheless, this approach results in a sophisticated condition, notably an inequality involving the determinant of H expressed in terms of $\gamma(m, n)$. For a better understanding, the result presented below will be formulated with simple and easily interpretable inequalities about the values $\gamma(m, n)$ filling up Γ For instance, one of these simple inequalities states that the central entry $\gamma(0, 0)$ of Γ is its largest entry. As shown by the previously presented counterexample, this condition alone is not sufficient. It is then completed by similar simple inequalities.

Theorem 1 *If the cross-correlation values $\gamma(m,n)$ filling up the matrix Γ satisfy*

$$\gamma(0,0) \geq \gamma(m,n) \quad \textit{for all } m,n \in \{-1,0,+1\} \tag{3.16}$$

$$\gamma(m,0) > \gamma(m,n) \quad \textit{for all } m,n \in \{-1,+1\} \tag{3.17}$$

$$\gamma(0,n) > \gamma(m,n) \quad \textit{for all } m,n \in \{-1,+1\}, \tag{3.18}$$

then the second degree polynomial $p_\theta(x,y)$ fitted to the entries of Γ on the grid $\mathbb{G}$ defined in (2.37)*, by solving the least squares problem* (2.41)*, has a unique global maximum.*

Interpretation of the conditions of Theorem 1.

— Inequalities (3.16): the central entry $\gamma(0,0)$ has the largest value among all the 9 entries of Γ.

— Inequalities (3.17): the middle entry $\gamma(0,\pm 1)$ is the largest entry of the top or the bottom row of Γ.

— Inequalities (3.18): the middle entry $\gamma(\pm 1,0)$ is the largest entry of the right or the left column of Γ.

Proof of Theorem 1.

In order to shorten lengthy equations and inequalities, let us introduce more compact notations for the cross-correlation values $\gamma(m,n)$ filling up the matrix Γ defined in (2.39), so that Γ is rewritten as

$$\Gamma = \begin{bmatrix} a & e & b \\ f & i & g \\ c & h & d \end{bmatrix}. \tag{3.19}$$

Remark that the letters $a, b, d, \ldots, i$ fill Γ first at the 4 corners, then the middles of side rows and columns, before finishing at the central entry.

With these compact notations, the least squares solution (2.41) leads to

$$9\theta_1 = 2(e+f+g+h) + 5i - (a+b+c+d) \tag{3.20a}$$
$$6\theta_2 = (b-a) + (g-f) + (d-c) \tag{3.20b}$$
$$6\theta_3 = (a-c) + (e-h) + (b-d) \tag{3.20c}$$
$$-6\theta_4 = (e-a) + (e-b) + (h-c) + (h-d) + (i-f) + (i-g) \tag{3.20d}$$
$$4\theta_5 = (b-a+c-d) \tag{3.20e}$$
$$-6\theta_6 = (f-a) + (f-c) + (g-b) + (g-d) + (i-e) + (i-h). \tag{3.20f}$$

As already mentioned in this section, the negative definiteness of the Hessian matrix H defined in (3.15) ensures that the polynomial $p_\theta(x,y)$ has a global maximum. Based on Sylvester's criterion [1], this negative definiteness will be checked through:

$$\theta_4 < 0 \tag{3.21}$$
$$\det(H) > 0. \tag{3.22}$$

According to the inequalities assumed in (3.17), e (or h, resp.) is the largest entry of the top (or bottom, resp.) row of Γ, then

$$e-a>0, \quad e-b>0 \tag{3.23}$$
$$h-c>0, \quad h-d>0 \tag{3.24}$$

and according to (3.16), i is the largest entry of Γ, then

$$i-f \geq 0, \quad i-g \geq 0. \tag{3.25}$$

These inequalities together with (3.20d) imply immediately (3.21).

It is more involved to check (3.22). The inequalities in (3.23) imply

$$\left(1+\frac{3}{4}\right)(e-a) > 0 > \left(-1+\frac{3}{4}\right)(e-b), \tag{3.26}$$

1. Usually Sylvester's criterion (Meyer, 2000) is about the *positive* definiteness of a real symmetric (or complex Hermitian) matrix. It is trivial to translate this criterion to the case of *negative* definiteness.

then

$$(e-a)+(e-b) > -\frac{3}{4}(e-a)+\frac{3}{4}(e-b) \tag{3.27}$$

$$= -\frac{3}{4}(b-a). \tag{3.28}$$

Repeat the reasoning from (3.26) to (3.28) while interchanging the positions of $(e-a)$ and $(e-b)$:

$$\left(1+\frac{3}{4}\right)(e-b) > 0 > \left(-1+\frac{3}{4}\right)(e-a), \tag{3.29}$$

leading to

$$(e-a)+(e-b) > \frac{3}{4}(b-a). \tag{3.30}$$

Combining (3.28) and (3.30) yields

$$(e-a)+(e-b) > \frac{3}{4}|b-a|. \tag{3.31}$$

This result expresses a relationship between the entries in the top row of the matrix Γ. A similar reasoning then leads to the following relationship between the entries in the bottom row of Γ:

$$(h-c)+(h-d) > \frac{3}{4}|c-d|. \tag{3.32}$$

According to the inequalities assumed in (3.16), i is the largest entry of Γ.

Take the sums of the respective sides of the 4 inequalities in (3.31), (3.32) and (3.25), then

$$(e-a)+(e-b)+(h-c)+(h-d)+(i-f)+(i-g)$$

$$> \frac{3}{4}|b-a|+\frac{3}{4}|c-d| \tag{3.33}$$

$$\geq \frac{3}{4}|b-a+c-d|. \tag{3.34}$$

This result then implies that $-6\theta_4$ and $4\theta_5$, as expressed respectively in (3.20d) and

(3.20e), satisfy

$$-6\theta_4 > \frac{3}{4}|4\theta_5|, \tag{3.35}$$

hence

$$-\theta_4 > \frac{1}{2}|\theta_5| \geq 0. \tag{3.36}$$

Following the same approach, it is then similarly shown that

$$-\theta_6 > \frac{1}{2}|\theta_5| \geq 0. \tag{3.37}$$

This last inequality can also be deduced from a certain "symmetry" between the formulae expressing θ_4 and θ_6 in (3.20d) and (3.20f).

It then follows from (3.15), (3.36) and (3.37) that

$$\det(H) = \theta_4\theta_6 - \frac{1}{4}\theta_5^2 > 0. \tag{3.38}$$

Therefore, the two inequalities (3.21) and (3.22) ensuring the negative definiteness of H are successfully checked. It is then established that the second degree polynomial $p_\theta(x, y)$ has a unique global maximum.

3.5.2 Conditions for the estimated subpixel shift to stay inside the one pixel vicinity with $|x^*| < 1, |y^*| < 1$

When the fitted polynomial $p_\theta(x, y)$ has a unique global maximum, it may happen that this maximum is far away from the origin $(x, y) = (0, 0)$ corresponding to the optimized integer shifts $(\Delta m^*, \Delta n^*)$, thus outside the square area corresponding to the one pixel vicinity of the cross-correlation peak. The following result ensures that the maximum of $p_\theta(x, y)$ stays inside this square area, under easily interpretable conditions.

Theorem 2 *If, in addition to the conditions of Theorem 1, the cross-correlation values*

$\gamma(m,n)$ filling up the matrix Γ satisfy, for all $m,n \in \{-1,+1\}$,

$$\gamma(m,0) - \gamma(m,n) > \frac{1}{5}\left[\gamma(m,0) - \gamma(m,-n)\right] \tag{3.39}$$

$$\gamma(0,n) - \gamma(m,n) > \frac{1}{5}\left[\gamma(0,n) - \gamma(-m,n)\right], \tag{3.40}$$

then the maximum of the fitted polynomial $p_\theta(x,y)$ is located at (x^*,y^*) such that $|x^*|<1$ and $|y^*|<1$.

Interpretation of the conditions of Theorem 2.

The conditions inherited from Theorem 1 ensure that the middle entry in each row or column of Γ is the largest entry of the row or column, without imposing any "degree of symmetry". For example, among the top row of Γ as expressed in (2.39), inequalities formulated in (3.17) ensure that $\gamma(0,1)$ is the largest entry, but the ratio $[\gamma(0,1)-\gamma(-1,1)]/[\gamma(0,1)-\gamma(1,1)]$ can be any positive number. Two of the inequalities in the extra condition (3.39) of Theorem 2 constrain this ratio between 1/5 and 5, thus limiting the dissymmetry between $\gamma(-1,1)$ and $\gamma(1,1)$.

Proof of Theorem 2.

Based on Theorem 1 (its conditions are inherited here), the fitted polynomial $p_\theta(x,y)$ has a unique global maximum in $\mathbb{R}^2$, which is located at

$$x^* = \frac{2\theta_2\theta_6 - \theta_3\theta_5}{\theta_5^2 - 4\theta_4\theta_6} \tag{3.41}$$

$$y^* = \frac{2\theta_3\theta_4 - \theta_2\theta_5}{\theta_5^2 - 4\theta_4\theta_6}. \tag{3.42}$$

In order to prove $|x^*|<1$, it will be shown that the numerator of $|x^*|$ is smaller than its denominator. The proof for proving $|y^*|<1$ will be made similarly.

With the compact notations filling up the matrix Γ introduced in (3.19) for the correlation values $\gamma(n,m)$, one of the inequalities contained in (3.39), namely

$$\gamma(0,1) - \gamma(-1,1) > \frac{1}{5}\left[\gamma(0,1) - \gamma(1,1)\right], \tag{3.43}$$

is translated into

$$e - a > \frac{1}{5}(e-b), \tag{3.44}$$

which is rewritten as

$$\frac{5}{2}(e-a) > \frac{1}{2}(e-b), \tag{3.45}$$

or in a slightly different form

$$\left(1+\frac{3}{2}\right)(e-a) > \left(-1+\frac{3}{2}\right)(e-b). \tag{3.46}$$

Then

$$(e-a)+(e-b) > \frac{3}{2}(e-b) - \frac{3}{2}(e-a) \tag{3.47}$$

$$= \frac{3}{2}(a-b). \tag{3.48}$$

A similar reasoning (interchanging the positions of $(e-a)$ and $(e-b)$) leads to

$$(e-b)+(e-a) > \frac{3}{2}(e-a) - \frac{3}{2}(e-b) \tag{3.49}$$

$$= -\frac{3}{2}(a-b). \tag{3.50}$$

Combining (3.48) and (3.50) then amounts to

$$(e-a)+(e-b) > \frac{3}{2}|b-a|. \tag{3.51}$$

This last inequality concerns the entries in the top row of Γ as in (3.19). Similar reasonings about the bottom row, the left and right columns of Γ then lead to

$$(h-c)+(h-d) > \frac{3}{2}|c-d|. \tag{3.52}$$

$$(f-a)+(f-c) > \frac{3}{2}|a-c|. \tag{3.53}$$

$$(g-b)+(g-d) > \frac{3}{2}|b-d|. \tag{3.54}$$

Because i is the largest entry of Γ (condition inherited from Theorem 1),

$$i-f \geq 0, \quad i-d \geq 0. \tag{3.55}$$

Summarizing (3.51), (3.52) and (3.55) then implies that θ_4 as expressed in (3.20d)

satisfies

$$\begin{aligned} -6\theta_4 &= (e-a)+(e-b)+(h-c)+(h-d) \\ &\quad +(i-f)+(i-g) && (3.56)\\ &> \frac{3}{2}|b-a|+\frac{3}{2}|c-d| && (3.57)\\ &\geq \frac{3}{2}|b-a+c-d|. && (3.58) \end{aligned}$$

This result, together with (3.20e), then yields to

$$-\theta_4 > |\theta_5| \geq 0. \tag{3.59}$$

Some similar reasonings (due to a certain "symmetry" between θ_4 and θ_6 in (3.20)) then lead to

$$-\theta_6 > |\theta_5| \geq 0. \tag{3.60}$$

Add together the respective sides of (3.20b) and (3.20d), then

$$6(\theta_2-\theta_4) = 2\left[(e-a)+(h-c)+(i-f)\right]. \tag{3.61}$$

Every parenthesis at the right hand side is positive, due to conditions inherited from Theorem 1. Then

$$\theta_2-\theta_4 > 0. \tag{3.62}$$

Similarly,

$$6(\theta_2+\theta_4) = -2\left[(e-b)+(h-d)+(i-g)\right] < 0. \tag{3.63}$$

Based on the last 2 inequalities, $-\theta_4$ satisfies

$$-\theta_4 > \pm\theta_2, \tag{3.64}$$

hence

$$-\theta_4 > |\theta_2| \geq 0. \tag{3.65}$$

In the same manner, combining (3.20c) and (3.20f) yields

$$6(\theta_3 - \theta_6) = 2\left[(f-c) + (g-d) + (i-h)\right] > 0 \tag{3.66}$$

$$6(\theta_3 + \theta_6) = -2\left[(f-a) + (g-b) + (i-e)\right] < 0, \tag{3.67}$$

then

$$-\theta_6 > |\theta_3| \geq 0. \tag{3.68}$$

The following steps of the proof will be essentially based on

$$|\theta_4| > |\theta_5| \tag{3.69}$$
$$|\theta_6| > |\theta_5| \tag{3.70}$$
$$|\theta_4| > |\theta_2| \tag{3.71}$$
$$|\theta_6| > |\theta_3|, \tag{3.72}$$

respectively due to (3.59), (3.60), (3.65) and (3.68).

On the one hand, (3.71), (3.72) and (3.69) imply (the step (3.74) below)

$$|2\theta_2\theta_6 - \theta_3\theta_5| \leq 2|\theta_2||\theta_6| + |\theta_3||\theta_5| \tag{3.73}$$
$$< 2|\theta_4||\theta_6| + |\theta_6||\theta_4| \tag{3.74}$$
$$= 3|\theta_4||\theta_6|; \tag{3.75}$$

and on the other hand, (3.69) and (3.70) lead to (the step (3.78) below)

$$|\theta_5^2 - 4\theta_4\theta_6| \geq |4\theta_4\theta_6| - |\theta_5^2| \tag{3.76}$$
$$= 4|\theta_4||\theta_6| - |\theta_5||\theta_5| \tag{3.77}$$
$$> 4|\theta_4||\theta_6| - |\theta_4||\theta_6| \tag{3.78}$$
$$= 3|\theta_4||\theta_6|. \tag{3.79}$$

Therefore

$$|2\theta_2\theta_6 - \theta_3\theta_5| < |\theta_5^2 - 4\theta_4\theta_6|. \tag{3.80}$$

It is then concluded that x^* as expressed in (3.41) satisfies

$$|x^*| < 1. \tag{3.81}$$

In the same way, it is also proved that

$$|y^*| < 1. \tag{3.82}$$

The proof of Theorem 2 is thus completed. □

3.5.3 Handling unexpected cases

There are two likely unexpected cases:

— Case 1, the fitted polynomial $p_\theta(x, y)$ has no global maximum;

— Case 2, $p_\theta(x, y)$ has a global maximum, reached at (x^*, y^*), but $|x^*| \geq 1$ and/or $|y^*| \geq 1$.

In the first case, the only reasonable proposition is to retain the optimized integer shifts $(\Delta m^*, \Delta n^*)$ as the estimated total shifts.

In the second case, solutions on the boundary satisfying $|x^*| \leq 1$ and $|y^*| \leq 1$ are accepted. If $|x^*| > 1$ and/or $|y^*| > 1$, then the subpixel shifts are estimated by solving the constrained optimization problem

$$(x^\star, y^\star) = \arg \max_{|x|\leq 1, |y|\leq 1} p_\theta(x, y), \tag{3.83}$$

then the estimated total shifts amount to

$$(\Delta m^* + x^\star, \Delta n^* + y^\star). \tag{3.84}$$

As $p_\theta(x, y)$ is a second degree polynomial, the constrained optimization problem (3.83) can be solved by quadratic programming algorithms. In this considered Case 2, the unique unconstrained maximum of $p_\theta(x, y)$ is outside the square area constrained by $|x| \leq 1$ and $|y| \leq 1$, the constrained solution is certainly on the boundary of the constrained square

area, on one of its sides or of its corners. Given the simplicity of this quadratic problem, instead of applying a general quadratic programming tool, the constrained optimization problem (3.83) can be solved as follows.

— Compute the values of $p_\theta(x, y)$ at the four corners of the square, namely $p_\theta(\pm 1, \pm 1)$.

— Find the maximums of $p_\theta(x, \pm 1)$ in x and $p_\theta(\pm 1, y)$ in y:

$$x_\pm^\star = \arg\max_{x \in \mathbb{R}} p_\theta(x, \pm 1) = -\frac{\theta_2 \pm \theta_5}{2\theta_4} \tag{3.85}$$

$$y_\pm^\star = \arg\max_{y \in \mathbb{R}} p_\theta(\pm 1, y) = -\frac{\theta_3 \pm \theta_5}{2\theta_6}. \tag{3.86}$$

Eliminate any results(s) not satisfying $|x_\pm^\star| \leq 1$ or $|y_\pm^\star| \leq 1$.

— Find the maximum value among the 4 corner values $p_\theta(\pm 1, \pm 1)$ and the non eliminated values $p_\theta(x_\pm^\star, \pm 1)$ and/or $p_\theta(\pm 1, y_\pm^\star)$, if any. The solution of the constrained optimization problem (3.83) is then given by the location of this maximum.

3.6 Conclusion

This chapter presented our proposed motion extraction method "Efficient Quadrant Interpolation" and its extended version "Efficient Octant Interpolation". We first introduced a linear interpolation-based method similar to the Taylor method, then the EQI method is presented aiming at gaining the accuracy of measurements compared to the Taylor method. Its extended version "Efficient Octant Interpolation" enhances the measurements in the vertical and horizontal directions. An extended method based on local phase using steerable filters are introduced to break the limitation to small displacement. In the end, two theorems are proposed to ensure the effectiveness of quadratic surface fitting for estimating the subpixel shift.

Chapter 4

Performance assessment of motion extraction methods from simulated images

In this chapter, the motion extraction methods which have been presented in Chapters 2 and 3 are evaluated on two types of simulated image flows. Both rely on some image pattern which are numerically moved with subpixel accuracy to simulate the motion. Most of the performance assessment of algorithms is performed on the first type.

Chapter 4 introduces at first the texture-free pattern which represents a simple bright disc target on a dark background. Two techniques to simulate the image formation are introduced, providing anti-aliased and parameterized gray-scale images. They represent the simplified version of the existing physical-based simulation models in Section 4.6, which are known to achieve realistic synthetic images on either CCD (Charge-Coupled Device) and CMOS (Complementary Metal Oxide Semi-conductor) cameras.

Four methods are compared in section 4.2 to generate motion in one direction with subpixel shifts, namely, analytic approach, Fourier transform, downsampling high-resolution image and spatial interpolation. To complete the simulation, the additive noise (AWGN) and the multiplicative noise models are considered to simulate noisy images. The latter model is known to provide more realistic images on either CCD or CMOS photosensor.

Then, the performance of the proposed EQI method (Efficient Quadrant Interpolation in Chapter 3) is compared to the existing motion extraction methods, which were introduced in Chapter 2, namely the 1st order Taylor approximation (Taylor), the quadratic surface fitting (QSF), the upsampling cross-correlation (UCC) (Thurman et al., 2008), and the enhanced cross-correlation (ECC) (Psarakis & Evangelidis, 2005) methods. The performance assessment for all motion extraction methods mostly rely on the computation of the RMSE on the estimated motion. Simulations are flexible for generating images with different parameterizations. Then, the RMSE behaviour enables to conduct sensitiv-

ity analysis of the motion extraction algorithms in relation with the motion magnitude, the noise level, the slope rate of the contour, the image contrast and the computing time.

Finally, the second type of image flows is presented at the end of Chapter 4. SEM (society for experimental mechanics) provides some image data sets that represent strongly textured speckle images. Data sets were created by numerically binning a 14-megapixel image to supply subpixel shift by Fourier transform. Performance assessment is performed on both the high contrast and the low contrast image data sets.

4.1 Simulating a simple target pattern

In this section, two solutions are proposed to simulate the image of a perfect disk target using simplified physical assumptions. This target will be then mathematically moved in section 4.2 to provide a first synthetic image data flow for assessing the performance of the motion extraction algorithms in section 4.3. By contrast, a strong textured simulated image data base is used in Section 4.7.

The two proposed solutions hereafter can be viewed as the simplified version of the more complex simulation process in Section 4.6 to simulate CCD and CMOS images. In both cases, the simulated images represent a white disk over a black uniform background and the size is 100×100 pixels.

The pixels edge of the target are processed in two ways:

1. Anti-aliased image: The intensity of the pixels along the target contour is related to the surface ratio of the pixel inside the boundary of the target.
2. Parameterized gray-scale image: the disk target is defined as a 2D Gaussian function, whose parameters control the sharpness of the intensity of contour pixels.

4.1.1 Anti-aliasing images

It is assumed that the disk target is viewed at oblique incidence and then looks like an ellipse. Assuming the principal axes to be aligned with the Cartesian coordinate system, the ellipse is specified by the following equation:

$$\left(\frac{x}{A}\right)^2 + \left(\frac{y}{B}\right)^2 = 1, \tag{4.1}$$

where A and B are the semi-major and semi-minor axes.

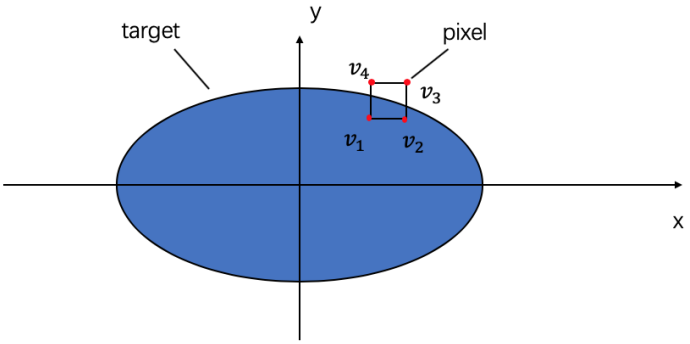


Figure 4.1 – Intersection case between the elliptic target and the square pixel in quadrant I.

The method by (Chojnacki & Szpak, 2019) has been used to provide simulated images with limited spatial resolution without aliasing effect. It takes account of the spatial discretization of the digital image.

It is assumed that the pixels within the ellipse region are set to a light gray level, whereas the background pixels are set to a dark gray level. The brightness of pixels is determined according to the surface ratio between the bright and dark areas within each pixel. The pixels along the ellipse border then require a specific computation.

Assuming the ellipse center is matched to the Cartesian axes system, we consider the brightness computation of the pixel corresponding to the rectangular area in Fig. 4.1. According to the location of the pixel along the border, (Chojnacki & Szpak, 2019) distinguish between six intersection cases (for each of the four quadrants) to analytically compute the pixel brightness. The pixel brightness in Fig. 4.1 (quadrant I) is then determined as the average of the two extreme gray levels weighted by the surface areas of the two parts within the pixel boundary. Further details are available in appendix A.2)

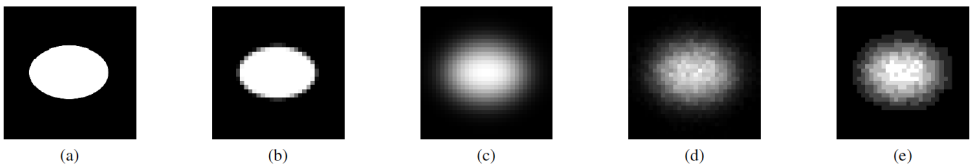


Figure 4.2 – Illustration of the six different steps to simulate the synthetic images in (Chojnacki & Szpak, 2019): (a) ideal binary image, (b) after accounting for the spatial discretization described in this section, (c) after accounting for the point spread function of the optical system, (d) after integrating Poisson noise, (e) after numerical quantization.

The simulation process in (Chojnacki & Szpak, 2019) is illustrated in Fig 4.2. The simplified synthetic image which has been simulated in this section corresponds to the two first stages, as displayed in Fig 4.2b, with a 8-bits quantization.

4.1.2 Parameterized 2D Gaussian distribution

To simulate a disk target which the gray level spatially decreases from its center to its periphery, a simple mathematical model would be the 2D Gaussian function. However, the conventional Gaussian function is not flexible enough for achieving a sharp brightness transition at the edge. In this section, it is proposed to use the 2D generalized Gaussian function instead.

The 1D Generalized Gaussian function has been introduced by (Varanasi & Aazhang, 1989) and is expressed as follows:

$$f(x, \rho, \mu, r) = e^{-(\frac{|x-\mu|}{r})^{\rho}} \tag{4.2}$$

where ρ is the "shape parameter", μ is the center position, $r > 0$ is a scale parameter. When $\rho = 2$, the generalized Gaussian function matches to the conventional Gaussian function. When $\rho = 20$, the function has a sharper edge, which the slope is similar to the experimental images to be shown in the next chapter. Fig. 4.3 illustrates the large variety of Gaussian shapes which can be achieved with different ρ values:

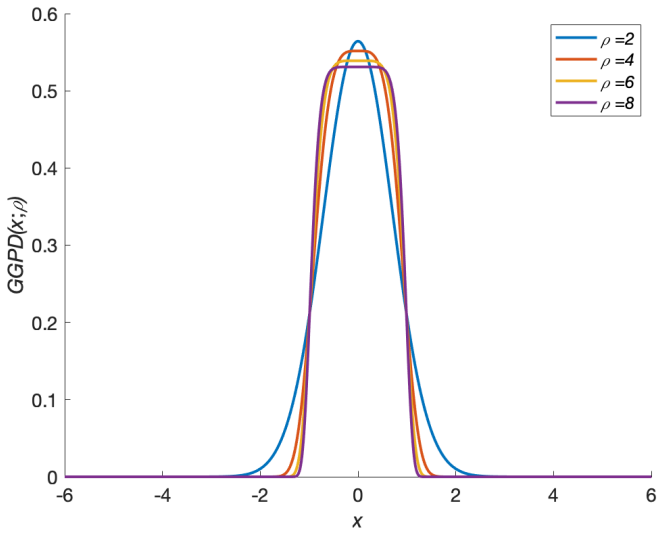


Figure 4.3 – Generalized Gaussian function for different ρ values

To simulate the disk target, the following 2D generalized Gaussian function has been inspired by (Xu et al., 2009) and is expressed as:

$$f(x, y, \rho, \mu_x, \mu_y, r) = e^{-((\frac{x-\mu_x}{r})^2+(\frac{y-\mu_y}{r})^2)^{\rho/2}} \tag{4.3}$$

To match the 2D generalized Gaussian function $f(x, y)$ to the gray level in the image, the following linear transformation is applied:

$$I(x, y) = \eta f(x, y) + \zeta \tag{4.4}$$

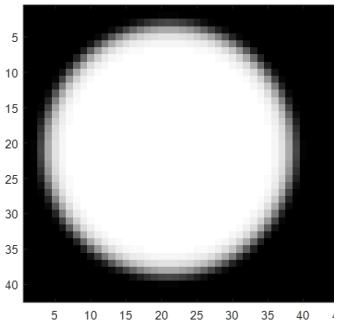
where η and ζ are two constants to be tuned to data.

Then, the two following approaches may be deployed to assign the gray level of each pixel and thus accounting for the spatial discretization of the image:

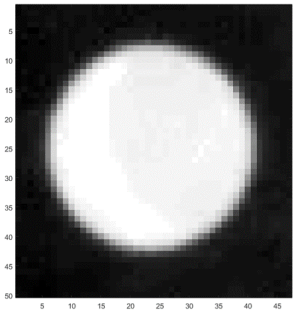
1. The gray level of pixels $I(m, n)$ is assigned to the sampled value of the 2D function at the pixel location (m, n);

2. In addition, the gray level of the pixel at (m, n) (pixel center at $m - 0.5, n - 0.5$)) is averaged over the pixel size as follows: $I(m, n) = \int_{m-1}^{m} \int_{n-1}^{n} I(x, y)\, dx\, dy$.

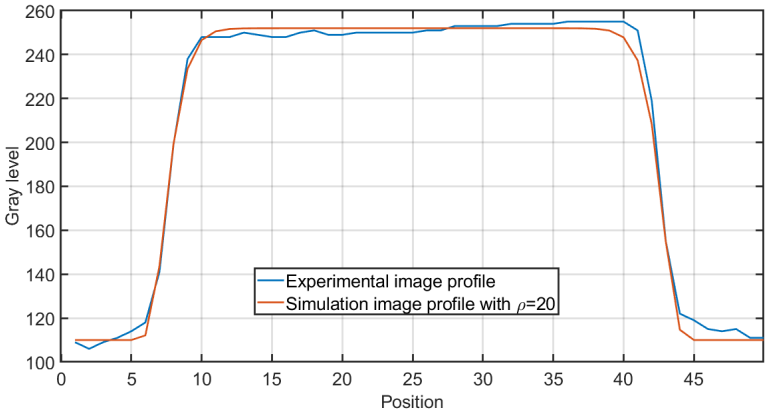
The second approach provides similar images to the one shown in Fig 4.2. However, due to its lower computational burden, the first approach has been selected in the thesis and some result is illustrated in Fig 4.4.



(a) Simulated image



(b) Experimental image



(c) Gray level profile along axis X through the disk center

Figure 4.4 – Simulated synthetic image by the 2D generalised Gaussian function compared to experimental image: a) 2D synthetic image, a) 2D experimental image (see chapter 5), c) Synthetic and experimental 1D image profiles superimposed.

4.2 Simulating image flows

The target image which was previously simulated is assumed to be moved along the θ direction with the following oscillating motion:

$$d_x(t) = P_x \cos(2\pi f_0 t) \cos\theta \tag{4.5}$$

$$d_y(t) = P_y \cos(2\pi f_0 t) \sin\theta \tag{4.6}$$

where $d_x(t)$ and $d_y(t)$ are respectively the horizontal and vertical displacements at time t, and $P_{x,y}$ is the motion amplitude along (x, y) directions. The oscillation frequency and the sampling frequency are set to $f_0 = 10\ Hz$ and $f_s = 89\ Hz$, respectively. The 50×50 pixels image template (ROI) entirely includes the target disk.

To generate synthetic image flows, it is required to compute the shifted image at each sampling time with subpixel accuracy. The four main existing techniques to this aim are recalled in the next subsection. To help in the selection among the latter techniques, the performance is compared then on the basis of the subpixel motion error. Adding noise to images is discussed in Section 4.4.

4.2.1 Computing shifted images

4.2.1.1 Analytical approach

This method can be easily applied to any image target characterized by an analytic model. For example, the simulated ellipse disk target in the previous subsection is determined by the minor and major axes and the coordinates of its center, namely (μ_x, μ_y).

To simulate shifted images by $(d_x(t), d_y(t))$, the target image is computed by varying the position of the center $(\mu_x + d_x(t), \mu_y + d_y(t))$ with subpixel values $d_x(t)$ and $d_y(t)$.

4.2.1.2 Phase shifting in the Fourier domain

A subpixel shift can be simulated by using the shift theorem of the Fourier transform. In the frequency domain, a spatial shift corresponds to a multiplication of the Fourier transform by a complex phase plane. The inverse Fourier transform of the latter result then leads to the synthetic shifted image. The steps of this method are summarized as follows.

— $I(m,n) \xrightarrow{\text{DFT 2D}} I_F(k,l)$

— $I_F(k,l) \xrightarrow{\text{applying a phase difference } (d_x, d_y)} I_F(k,l) * e^{-j2\pi(\frac{d_x \times k}{M} + \frac{d_y \times l}{N})}$

— $I_F(k,l) * e^{-j2\pi(\frac{d_x \times k}{M} + \frac{d_y \times l}{N})} \xrightarrow{DFT^{-1}}$ equivalent shifted image $I^{\text{shifted}}(m,n)$ with the shift (d_x, d_y)

4.2.1.3 Downsampling high resolution images

It is assumed that the disk target is some high-resolution image, L times larger than the wanted final resolution image, e.g. 100×100 pixels, with Δ as the pixel size. Then, a

pixel-level shift is performed on the high resolution image by $[Ld_{x,y}(t)/\Delta]$.

Then the high resolution shifted images is down-sampled, i.e., numerically binned by $L \times L$ pixels, yielding the final 100×100 resolution image with subpixel shift by $1/L$ pixel. According to (Bornert et al., 2012), this method provides a limited number of shifted positions, namely, L.

4.2.1.4 Brightness interpolation

The gray levels of the original images can be interpolated on a sub-grid to obtain a virtual image with arbitrary higher resolution. Then, the aforementioned method in the previous Section can be applied to simulate subpixel shifts.

This method does not require any analytical model of the simulated target and can simulate arbitrary subpixel shifts. However, the simulated results strongly depend on the interpolation method. In Section 4.2.3, we have compared two interpolation methods, namely, cubic (Keys, 1981) and bi-linear (Smith, 1981) methods.

4.2.2 Noise models

This section briefly recalls the two noise models which have been used in the literature to achieve more realistic simulated images. The noise measured on the high speed camera to be used in experiments is illustrated in Section 5.3.1. The sensitivity of the motion extraction algorithms to the noise is illustrated in Section 4.5.

4.2.2.1 Additive white gaussian noise (AWGN) model

The conventional additive noise model consists in adding an ideal white Gaussian noise $N_1(\mu_N, \sigma_N)$ to the noiseless simulated gray-level images as follows:

$$I_{\text{noisy}}(m,n) = I(m,n) + N_1(\mu_{N_1}, \sigma_{N_1}) \tag{4.7}$$

where μ_{N_1} and σ_{N_1} are the mean value and standard deviation of the noise respectively. Usually a zero-mean AWGN is considered in simulations, i.e., $\mu_{N_1} = 0$.

4.2.2.2 General noise model

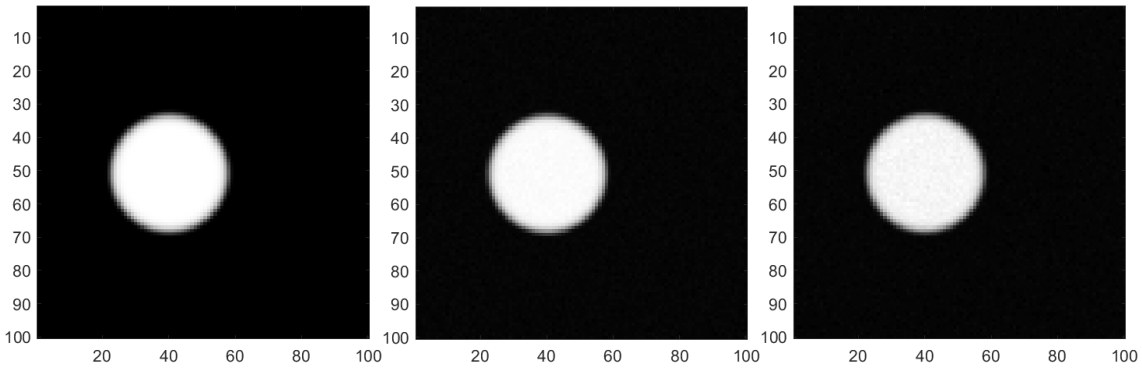
Physics-based image model (see Section 4.6 for example) and experimental evidence (see Section 5.3.1 for illustration) show that the two first order statistics of the image,

i.e., the standard deviation and the average brightness level, can be related to each other.

The following general noise model which combines the previous additive model with a multiplicative one, has been used in the literature to account for this dependency :

$$I_{\text{noisy}}(m,n) = I(m,n)\,(1 + N_2(0,\sigma_{N_2})) + N_1(\mu_{N_1},\sigma_{N_1}) \tag{4.8}$$

where $N_2(0,\sigma_{N_2})$ is a zero-mean Gaussian noise with variance σ_{N_2}.



(a) Noiseless (b) Additive noise (c) General noise

Figure 4.5 – Illustration of the different simulated images with the two proposed noise models. (a) Noiseless; (b) Additive noise with $\sigma_{N1} = 1$ (SNR = 48 dB) ; (c) Additive and multiplicative noise with $\sigma_{N2} = 0.01$ (SNR = 40 dB)

4.2.2.3 Signal-to-Noise Ratio (SNR)

For the small textured target image which is simulated in Section 4.1, the Signal-to-noise-ratio (SNR) is related to the contrast between the bright target pixels and the darker background pixels as follows:

$$\text{SNR}_{dB} = 20\log_{10}\frac{max(I(m,n)) - min(I(m,n))}{\sigma_N} \tag{4.9}$$

where the standard deviation σ_N is estimated from the image.

4.2.3 Comparison between the computation methods

The latter techniques have been evaluated in (Bornert et al., 2012) on images with strong texture. Spline interpolation and Fourier-based techniques were found to achieve

the best performance in that case. However, it is required to further evaluate these techniques on the texture-free target images, which are simulated in this chapter.

To simplify the comparison, two methods only are selected to extract the motion from the image data flow, namely, the gradient-based method (labelled as Taylor in the following) introduced in Section 2.2 and the proposed EQI in Section 3.2.

The displacement (or shift) error is computed for each image as the difference between the estimated motion $\hat{d}(t)$ and the simulated motion $d(t)$: $\Delta d(t) = \hat{d}(t) - d(t)$.

Bias and standard deviation on the latter shift error are introduced to compare the different methods to generate subpixel shift. To achieve reliable result on noisy images (see Section 4.4 for noise models), a Monte Carlo process is used. For each simulated displacement $d(t)$, we generate N_{MC} noisy synthetic images (N_{MC} is set to 50 in the following simulations).

The shift error for the k^{th} Monte Carlo iteration is expressed as follows:

$$\Delta d_k(t) = \hat{d}_k(t) - d_k(t)$$

The averaged motion error $\overline{\Delta d}$ and the standard deviation σ_d are calculated at each time as follows:

$$\overline{\Delta d}(t) = \frac{\sum_{k=1}^{N_{\mathrm{MC}}} \Delta d_k(t)}{N_{\mathrm{MC}}} \tag{4.10}$$

$$\sigma_d(t) = \sqrt{\frac{\sum_{k=1}^{N_{\mathrm{MC}}} (\Delta d_k(t) - \overline{\Delta d}(t))^2}{(N_{\mathrm{MC}} - 1)}} \tag{4.11}$$

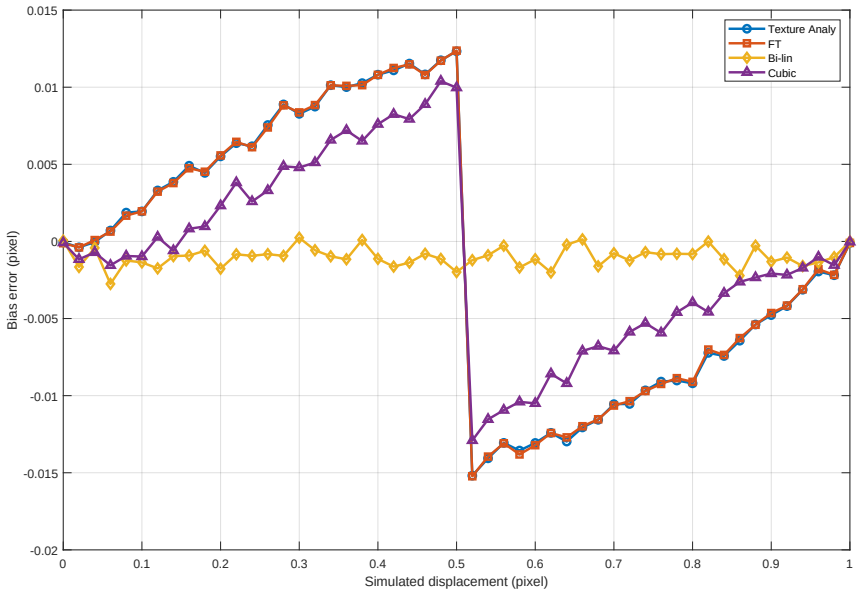


Figure 4.6 – Taylor-based shift error $\overline{\Delta d}$ variations w.r.t the simulated motion amplitude for all the four computing methods in section 4.2.1. "Texture Analy" refers to the analytic approach, "FT" refers to Fourier Transform, "Bi-lin" and "Cubic" refer to spatial interpolation with the bi-linear and the cubic methods, respectively.

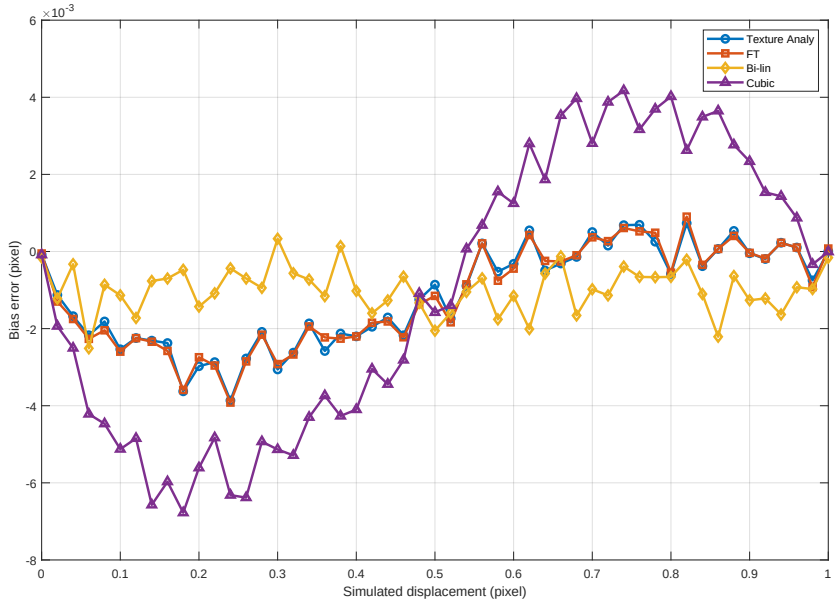


Figure 4.7 – EQI's shift error $\overline{\Delta d}$ variations w.r.t the simulated motion amplitude for all the four computing methods in section 4.2.1.

The variations of the error bias $\overline{\Delta d}$ (resp. the standard deviation σ_d) are plotted in Fig. 4.6 and Fig. 4.7 with regards to the motion amplitude (resp. Fig. 4.8 and Fig. 4.9). For convenience, these values are converted in pixel units. The following comments on the results of the error bias (Fig. 4.6 and Fig. 4.7) can be made:

1. The two curves are symmetric with respect to zero motion; the average error bias is about -10^{-3} pixels.
2. The error bias varies with respect to the motion amplitude and the variation shape depends on the motion extraction method. For the Taylor method, the error bias is seemingly linear within the motion range $[-0.5, 0.5]$; it is then maximum when the shift is 0.5 pixel. For the EQI method, the error bias variation shows an S-shaped, as reported in Bornert et al., 2012.
3. Comparing the methods to shift the texture-free images at hand, bi-linear interpolation ("Bi-lin") presents the smallest error bias for both the motion extraction methods. The cubic interpolation ("Cubic") reports the worst performance for the EQI method. The analytic approach ("Texture Analy") and Fourier Transform ("FT") show similar performance for both the motion extraction methods.

4. The proposed EQI method shows smaller error bias than the Taylor method, i.e., 30 times smaller at least for "FT" and "cubic" methods.

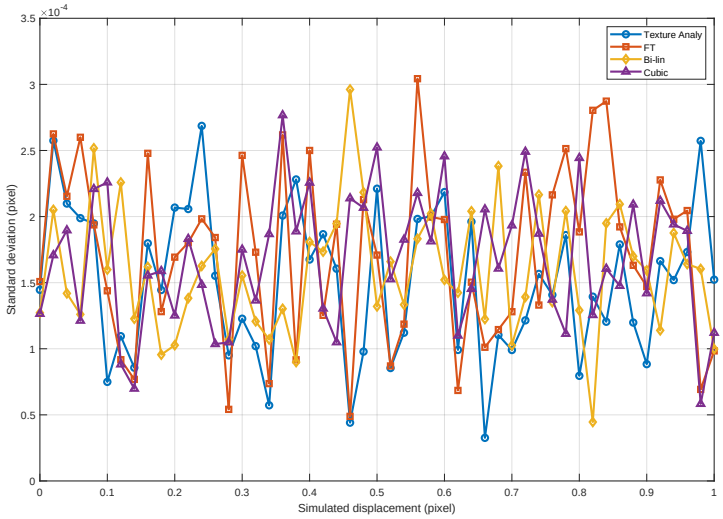


Figure 4.8 – Taylor-based standard deviation σ_d variations w.r.t the simulated motion amplitude for all the four computing methods in section 4.2.1. "Texture Analy" refers to the analytic approach, "FT" refers to Fourier Transform, "Bi-lin" and "Cubic" refer to spatial interpolation with the bi-linear and the cubic methods, respectively.

Fig. 4.8 and Fig. 4.9 provide the standard deviation variation with regards to the motion amplitude. The following comments can be made:

1. All the standard deviation variations shows random variation w.r.t the motion amplitude; then, the standard deviation does not depend on either the motion amplitude, the computing shifted methods or the motion extraction methods.

2. The two motion extraction methods, i.e., Taylor and EQI method, show similar standard deviation on the shift error; it is about 1.6×10^{-4} pixels in average.

Finally, following (Bornert et al., 2012; Su et al., 2018), the Fourier Transform was selected among the four computing shifted methods for generating subpixel shifted images, as a good trade-off between computing burden and accuracy.

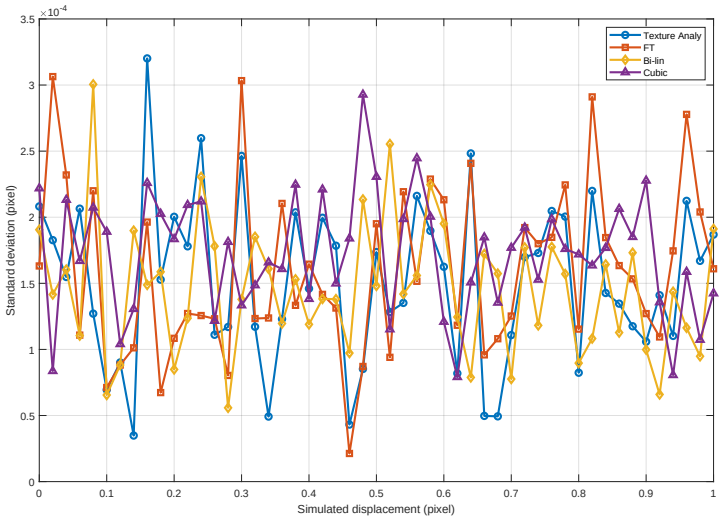


Figure 4.9 – EQI's standard deviation σ_d variations w.r.t the simulated motion amplitude for all the four computing methods in section 4.2.1.

4.3 Performance analysis of the selected motion extraction algorithms

In this section, the motion extraction algorithms which were presented in Chapters 2 and 3 are evaluated on the simulated image data flows introduced in Section 4.2.

Because the algorithms rely on a two-step correlation-based technique, we first evaluate the accuracy of the pixel-level shift motion estimation, before analysing the global performance of the algorithms. Then, some sensitivity analysis is performed on the motion extraction algorithms with regards to the slope rate of the target contour and the noise model. Finally, the computer burden of algorithms are compared.

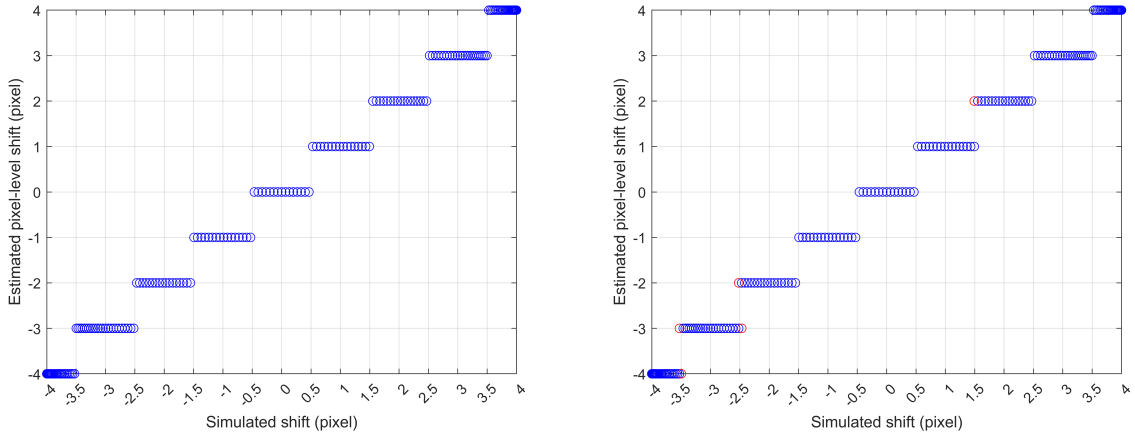
The performance evaluation of the motion extraction methods is based on the RMSE (root mean squared error) (3.4) computation of the error between the estimated and the true displacement values, which is defined in Section 3.4.4.

4.3.1 Pixel-level shift evaluation

The first step of the correlation-based techniques presented in Chapters 2 and 3, namely, the pixel-level shit estimation, is evaluated on a multi-pixel shift image data flow.

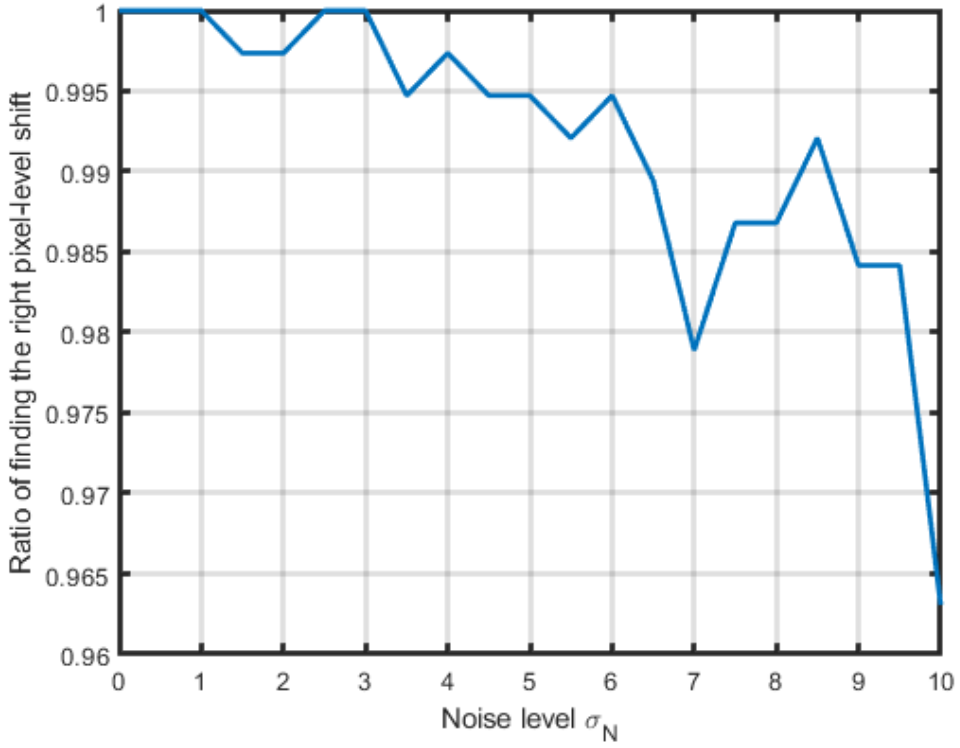
The cross-correlation technique which is selected in Section 2.1.3 is the ZNCC method. It is used to estimate the pixel-level shift of the simulated motion. It is compared to the theoretical pixel-level shift which is defined by rounding off the preassigned shift. The estimated (in red) and the theoretical pixel-shift (in blue) are superimposed in Fig. 4.10. ZNCC perfectly performs in the noiseless condition with 100% accuracy as shown in Fig. 4.10(a).

As the noise level increases, ZNCC provides some errors in the pixel-level shift estimation as shown in Fig. 4.10(b) by the red circles. In practice, this error may be corrected at some extent at the second step of the motion extraction algorithms. Fig. 4.11 shows the ZNCC performance with increasing noise level. The ratio of finding the right pixel-level shift, namely, the true positive ratio reduces to 96% at $\sigma_N = 10$ (SNR = 28 dB).



(a) Noiseless condition (b) Noisy condition (SNR = 28 dB)

Figure 4.10 – Comparing pixel-level shift ZNCC estimation along the horizontal direction in noiseless and noisy conditions. The blue circles refer to the right estimation of the pixel-level shift (true positive result), and the red circles refer to erroneous results (false negative result)



]

Figure 4.11 – True positive ratio of the pixel-level shift estimation

4.3.2 Estimation error vs. motion amplitude

In this section, a monotonic motion of the target image is simulated in the x direction with a 4 pixels magnitude, and the total shift is estimated by the seven selected motion extraction algorithms, namely, Taylor, EQI, EQI-octant, QSF, POC, UCC and ECC methods.

The performance of methods are compared at first by plotting the motion error in eq. 4.10 against the motion amplitude. For the sake of clarity, only the three results, namely, Taylor, EQI and QSF methods, are displayed on Fig. 4.12-a. The results show a periodic error signal which the period is one pixel. For each period, the error bias is S-shaped as it was observed for the subpixel shift error in section 4.2.3. As observed in Section 4.2.3, the magnitude of the motion error is the smallest for the EQI method among the 3 methods. QSF shows close performance to EQI. The Taylor method shows the largest error magnitude.

The periodic variations of the error bias confirms the conclusion drawn in Section

4.3.1, i.e., the pixel-level shift estimation well performs whatever the motion magnitude. In addition, the motion error on the 7 periods is averaged to provide the motion error vs. the subpixel motion amplitude on Fig. 4.12-b.

All the methods show a small bias of about 2/1000 to 4/1000 pixels in average. They can have different phase polarity vs. the motion amplitude. Among the 7 selected methods, only POC and UCC methods depict the motion error to be independent on the motion amplitude. For this simulation condition, EQI, UCC and POC seem to show the smallest magnitude error. EQI, QSF and ECC achieve the smallest standard deviations.

For a more accurate comparison between the 7 methods, the RMSE on the motion error has been computed in Table 4.1 for this simulation scenario. UCC and the proposed EQI methods are found to achieve the best accuracy among all the motion extraction methods at high SNR.

4.3.3 Sensitivity to the slope rate of the target contour

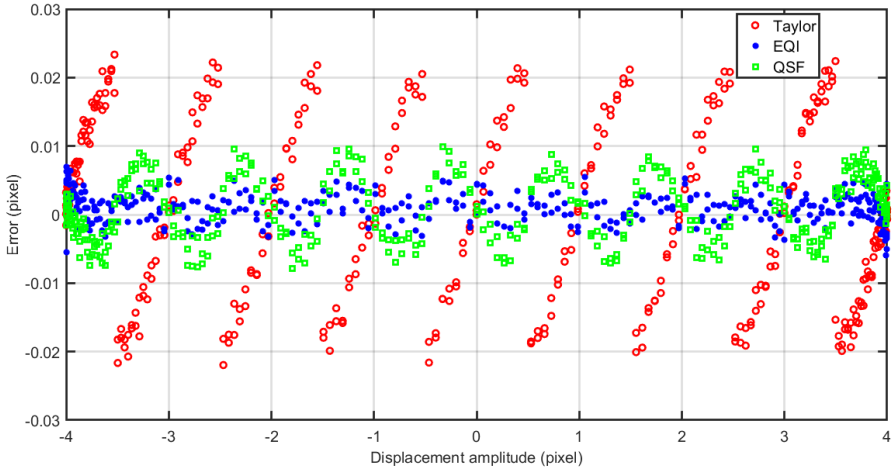
Some preliminary tests have shown that the performance of the Taylor method is degraded when the edge is sharp, because the estimated gradient relies too heavily on the variance of the brightness on the edge.

In this section, we then study the sensitivity of the motion extraction methods to the edge of the target contour. To this aim, the shape parameter ρ of the generalized Gaussian function in section 4.1.2 is varied within the range $[10, 20]$ to simulate smooth to sharp edges. When $\rho = 20$, we obtain the ideal binary image.

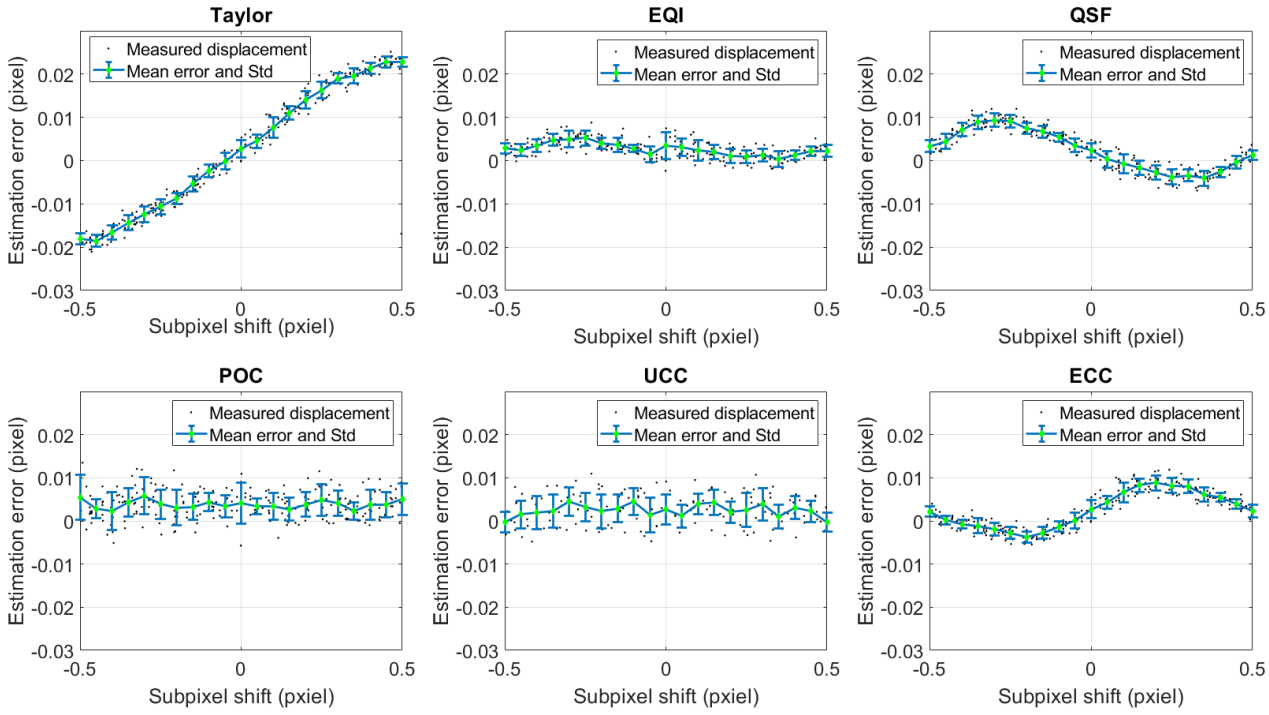
The sensitivity to the shape parameter is evaluated according to Section 3.4.4, by the RMSE computation on the motion error between the estimated and the true displacement values as defined in eq. 4.10.

In the following illustrations, Taylor refers to the gradient-based method ((Chan et al., 2010)), EQI-Octant to the proposed octant method in Section 3.3, EQI refers to the proposed efficient quadrant interpolation in Section 3.2, UCC refers to upsampling cross-correlation (Thurman et al., 2008), POC refers to phase-only correlation (Takita et al., 2003), QSF refers to quadratic surface fitting (Pan et al., 2005; D. Zhang et al., 2016), and ECC refers to the enhanced cross-correlation (Psarakis & Evangelidis, 2005).

Fig. 4.13 confirms that Taylor method is heavily affected by the edge slope. By contrast, the other methods are more robust and reach a good precision for ρ within the range $[10, 20]$. POC is the least sensitive to the shape parameter and provides the best performance at the same time.



(a) Error on motion vs. the full range motion amplitude



(b) Averaged motion error vs. the subpixel motion range

Figure 4.12 – Estimated motion error vs. the motion amplitude for the selected motion extraction methods; noisy image data flow with AWGN ($\sigma_N = 1$) for 4 pixels magnitude monotonic motion along the x direction. a) Error on motion vs. the full range motion amplitude; only 3 methods are represented for clarity; b) Averaged motion error vs. the subpixel motion range; the error bars represent the standard deviation on the motion error.

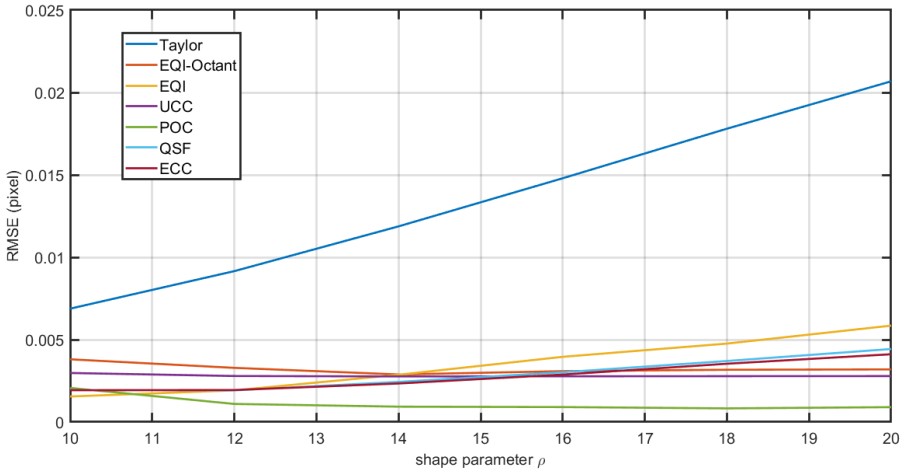


Figure 4.13 – RMSE vs. shape parameter ρ for horizontal shift of the image without noise

For target images, there is a strong relationship between the sharpness of the contour edge and the focus tuning of the camera. In practice, for a perfect focused image, EQI and other methods would be more accurate to perform the motion extraction. By contrast, the Taylor method would require some image preprocessing beforehand, namely, some smoothing of the image, as it is recommended in the literature.

4.3.4 Sensitivity to noise

To study the sensitivity of noise on the performance of motion extraction algorithms, a simulated image data flow is generated for each noise level and noise model according to Section 4.2.2. The RMSE on the motion error (eq. (3.4)) on each image data flows is then computed and plotted against increasing noise level.

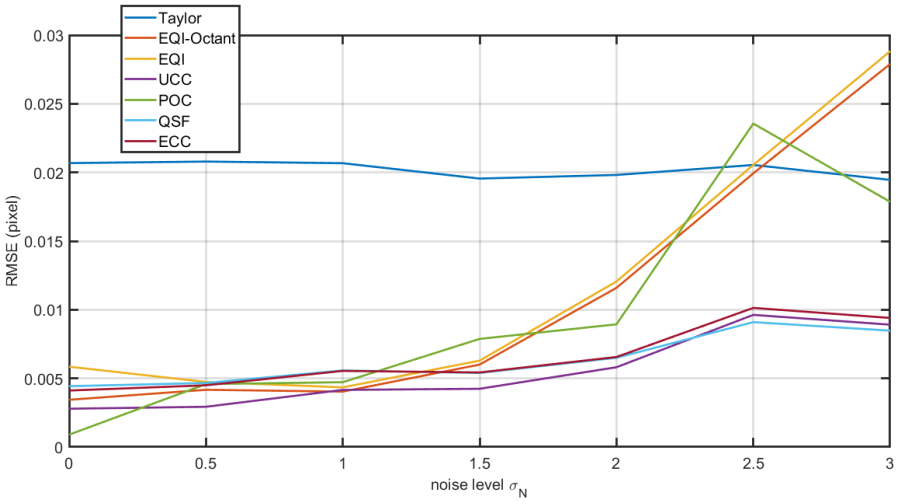


Figure 4.14 – RMSE vs. noise level σ_{N_1} for horizontal motion ($P_x = 4$ pixels) providing AWGN noise model; the shape parameter is $\rho = 20$.

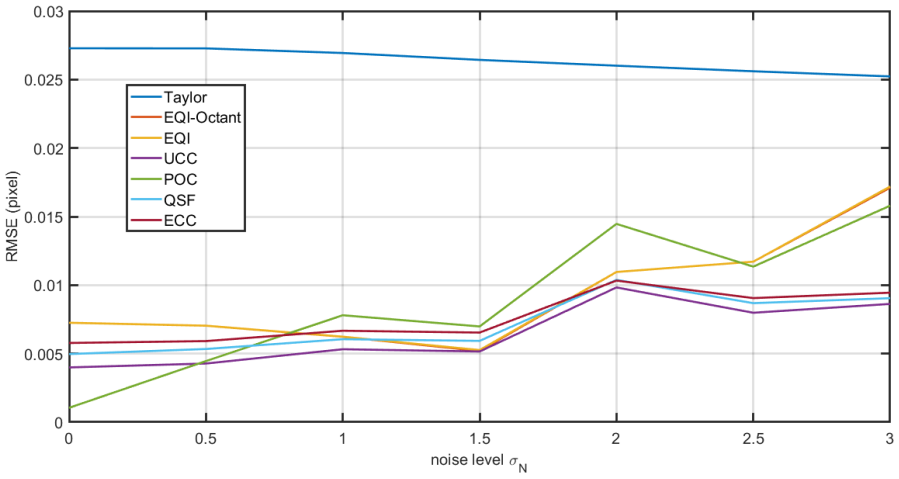


Figure 4.15 – RMSE vs. noise level σ_{N_1} for oblique motion ($\theta = 53°$, $P_x = 4$ pixels, $P_y = 3$ pixels) providing AWGN noise model and $\rho = 20$.

Fig. 4.14 and 4.15 show the result for the additive noise model, namely, AWGN in Section 4.2.2, for the simulated horizontal and the oblique motions in eq. 4.6, respectively. The following comments on the results can be made:

1. Taylor method provides the largest RMSE for the selected shape parameter ρ.
2. For the horizontal motion in Fig. 4.14, we can verify the slight improvement of the EQI-octant compared to EQI. The two algorithms provides equivalent performance

for the oblique motion in Fig. 4.15.

3. EQI, EQI-octant and POC are more sensitive to noise compared to UCC, ECC and QSF methods; the associated performance quickly drops beyond $\sigma_{N_1} > 1.5$.

4. For low noise level ($\sigma_{N_1} \leq 1.5$), the methods shows similar RMSE, except for the Taylor method.

The result for the additive and multiplicative noise model is shown in Fig. 4.16. UCC is found the most accurate method over the noise level range. When $\sigma_N \leq 1$, the proposed EQI and EQI-octant methods achieve the same performance as UCC. Compared to Fig. 4.14, POC shows degraded performance while EQI methods are found more sensitive to σ_{N_1}. UCC, ECC and QSF methods are found more robust to noise than others as for the AWGN noise model.

In conclusion, the performance of the proposed EQI and EQI-Octant methods is found equivalent to some reference motion extraction methods at low noise level, namely, UCC, ECC, QSF, and more sensitive to noise beyond $\sigma_{N_1} \geq 1.5$. Then, to maintain high performance for the proposed EQI methods, it is recommended to use beforehand some filtering techniques that are known to preserve the image texture and the contour edges while smoothing the noise, e.g., bilateral or non-local mean filters.

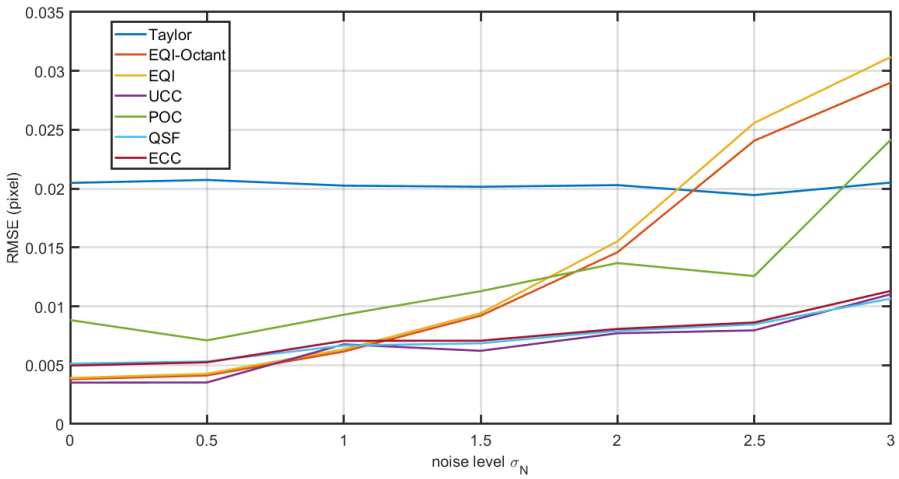


Figure 4.16 – RMSE vs. noise level σ_{N_1} for horizontal shift ($P_x = 4$ pixels) providing the additive noise and multiplicative noise model with $\sigma_{N_2} = 0.1$; the shape parameter is $\rho = 20$.

4.3.5 Time computation and overall performance

At the end, the time computation of the selected motion extraction algorithms are compared in Table 4.1 and related to the measured RMSE, in order to draw some possible trade-off between computation burden and accuracy. AWGN noisy images were used for the test, with $\sigma_{N_1} = 1$. The shape parameter in Eq. 4.3 is set to $\rho = 20$.

Table 4.1 – RMSE (in both x and (x, y) directions) and time computation associated to the data in Fig. 4.12 for the 7 selected motion extraction algorithms.

Algorithm	**RMSE**$_x$ (10^{-3} pixel)	**RMSE**$_{xy}$ (10^{-3} pixel)	**Time per frame** (ms)
Taylor	13.4	13.6	4.9
EQI	**3.5**	**4.6**	9.6
EQI-Octant	**4**	**4.6**	9.0
QSF	5.3	5.8	4.7
ECC	5	5.6	18.2
POC	5.1	6.2	**1.1**
UCC	**4**	**4.3**	3.4

According to Table 4.1, UCC and the proposed EQI and EQI-Octant methods have the best accuracy among all the motion extraction methods at high SNR.

However, this results is achieved at the expense of a larger computer burden. POC achieves the shortest computing time, followed by UCC and Taylor methods. The computing time of the proposed EQI and EQI-octant methods is faster than ECC (recalling ECC uses an iterative process).

Then, the proposed issue would be to reduce the computer burden of the EQI methods by some coding optimization, and then to make the computer burden closer to that of the Taylor method.

On the basis of the latter Table, UCC method may appear as the best trade-off between accuracy and computer burden on weak textured image flows. Besides, it shows more robustness to noise level than other methods according to Section 4.3.4.

4.4 Tests on the photo-sensor simulation model

In the literature, a few authors have developed simulation models which rigorously take account of both the physics and electronics in the image formation within either CCD or CMOS cameras (Chojnacki & Szpak, 2019; Konnik & Welsh, 2014).

The image formation model proposed by (Chojnacki & Szpak, 2019) is quite theoretic and mathematical. This model has inspired the approach in Section 4.1.1. Further details on this modeling are available in appendix A.2.

By contrast, (Konnik & Welsh, 2014) proposed a more realistic modeling, which is based on a high-level numerical simulation of the image formation in either CCD or CMOS photosensor in three stages: 1. from photon to charge, 2. from charge to voltage, 3. from voltage to digital numbers. The main sources of noise are the shot noise (photon shot noise, dark current shot noise) which is of type Poisson noise, and the fixed pattern noise (dark current fixed pattern noise, offset fixed pattern noise) which is of type Gaussian noise. By accounting for all the noise sources, this modeling is somehow equivalent to the additive and multiplicative noise in Section 4.2.2, but in a more complex form.

Within the scope of this thesis, we used the Matlab coding developed in (Toullier et al., 2020) on the basis of the photosensor model in (Konnik & Welsh, 2014) to generate more realistic synthetic image (as shown in Fig. 4.18) data set. To this aim, the grey-level image data flow in Section 4.2 were converted to radiosity images before being processed by the latter coding, and then finally converted back to 8-bits grey-level images.

Method / RMSE	Taylor	EQI	EQI-Octant	QSF	POC	UCC	ECC
CCD model	14.3	4.9	4.8	5	**4**	5.1	5.6

Table 4.2 – RMSE results (in 10^{-3} **pixel units**) by the selected motion extraction algorithms on the simulated images by the photosensor model in (Konnik & Welsh, 2014)

In Table 4.2, EQI, QSF, ECC, and UCC achieved very close performance (RMSE is about 5/1000 pixels) while POC performs the best (RMSE $\approx$ 4/1000 pixels) and Taylor method shows three times larger RMSE. Compared to the results in Table 4.1 (where the data set was obtained with the simplified image model in Section 4.1.2), EQI methods show consistent performance while the other methods achieve improved RMSE.

4.5 Tests on the DIC challenge data set

SEM (Society for Experimental Mechanics) supplies different sets of image to scientific people for assessing the performance of various Digital Image Correlation (DIC) techniques. The synthetic images are digital speckle images which were created by numerically binning shifted 14 MPx image.

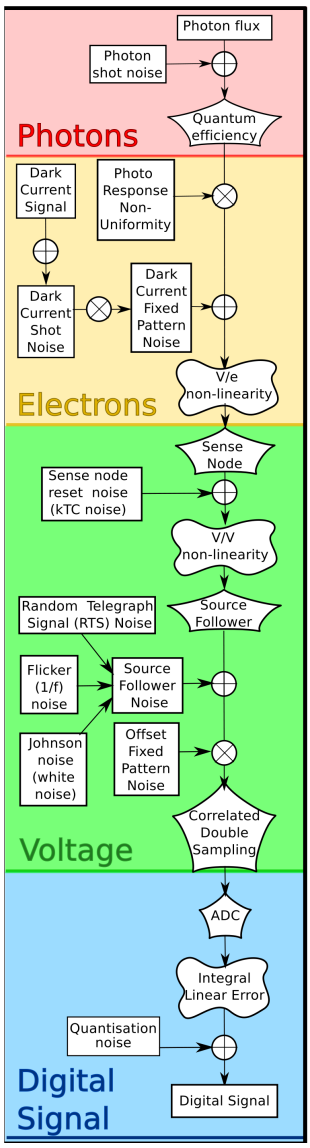
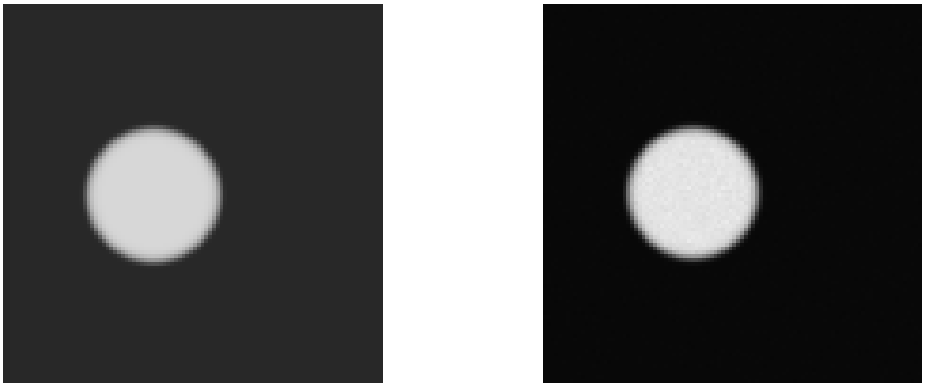


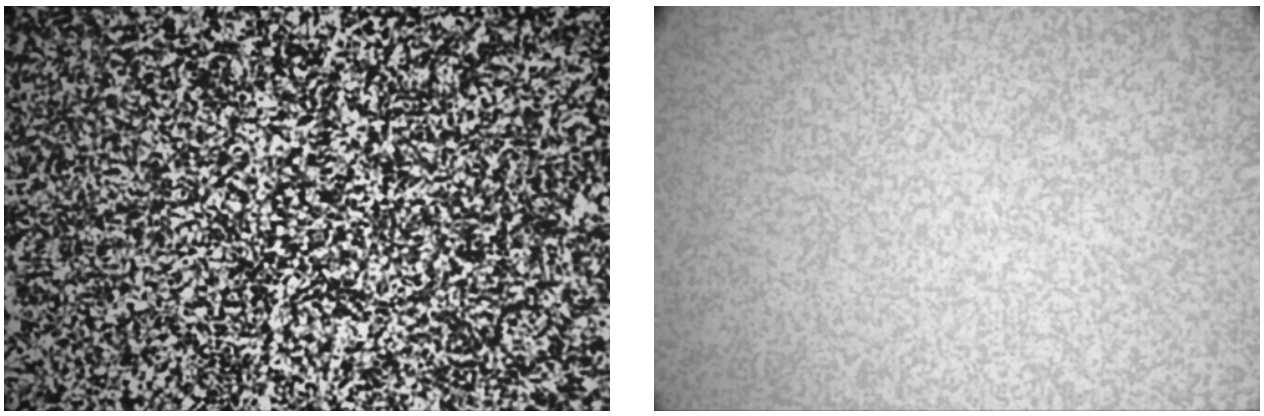
Figure 4.17 – Synoptic of the photosensor model in (Konnik & Welsh, 2014)



(a) Noiseless (b) Photosensor model

Figure 4.18 – Illustration of the photosensor model to generate noise. (a) Noiseless; (b) Photosensor model.

The data set includes 122 325×487 pixels images with 0.1 pixel shift from 0 to 1 pixel in both the x and y directions. Each shifted image was computed according to the method described in Section 4.2.1.3 (Reu, 2011). The data base includes a high contrasted and low contrasted image data sets as shown in Fig. 4.19.



(a) High contrast image (b) Low contrast image

Figure 4.19 – Sample synthetic images from the *DIC challenge* data base.

The DIC challenge data set enables to test the sensitivity of the selected motion extraction algorithms to the image texture. Algorithms are then performed on both the high and low contrast image data sets, and the RMSE on the 122 estimated displacement is shown in Table 4.3. It is noticed that ECC achieves the best accuracy on low contrast images and is the most consistent method. The proposed EQI methods show close per-

RMSE Method / Images	Taylor	EQI	EQI-Octant	QSF	POC	UCC	ECC
High contrast	39.6	16.1	16.1	12.3	**9.6**	17.4	12.2
Low contrast	43	23.2	23	20.2	353.9	69.9	**13.3**

Table 4.3 – RMSE results (in 10^{-3} **pixel units**) by the different selected motion extraction methods on the DIC challenge images.

formance to QSF. The latter three methods are less sensitive to the image contrast than POC and UCC. For example, POC shows the best accuracy on high contrast images while bringing the worst performance on low contrast images.

Compared to the results in Table 4.1 with the texture-free image data set at large SNR, we observed that RMSE values increases with image texture for all the methods. Besides, the methods do not order the same way on both data sets. ECC, POC and QSF are the best methods on speckled images, while EQI and UCC are the best on texture-free images.

4.6 Conclusion

In this chapter, two types of synthetic images have been presented and provide ad hoc image data sets to assess the performance of the selected motion extraction methods in Chapters 2 and 3 in different conditions. The proposed Efficient Quadrant Interpolation (EQI) in Chapter 3 is then compared with the existing methods introduced in Chapter 2, namely Taylor, QSF, POC, UCC, and ECC methods.

The first type of synthetic image data flow simulates some monotonic motion of a texture-free disk target over a darker background. This is a simplified solution of the physics-based simulation models which are briefly described in Section 4.6. Among the two proposed methods that are detailed to simulate the synthetic target images, we mostly used the analytical approach which relies on the 2D generalized Gaussian function; among other, this function allows simulating target images with a versatile sharpness of edges. Additive noise or/and multiplicative noise models are finally deployed to improve the image realism. Some existing dataset has been used as the second type of synthetic images. Compared to the first one, it relies on a strong textured speckled images.

The two types of image data flows allowed us to evaluate the performance of the selected motion extraction algorithms and to conduct some sensitivity analysis. The RSME

on the motion error has been selected to assess the performance of algorithms. On weakly textured image flows, UCC method has appeared as the best trade-off between accuracy and computer burden (Table 4.1). Besides, it shows more robustness to noise level than other methods according to Section 4.3.4.

The proposed EQI methods have shown consistent results with regards to the various image data sets introduced in this chapter. It is then found as accurate asthe UCC method in the measurement of subpixel shift at high SNR, i.e., noise level $\sigma_{N_1} < 2$, robust to the motion amplitude, the slope rate of the contour, the image texture and the image contrast. Some coding optimization would be required to reduce the computer burden.

Chapter 5

Laboratory experiments

At beginning of this chapter, I repeat my heartfelt thanks to Boualem MERAINANI for his help in the experimental tests. His experience (Merainani, Döhler, et al., 2021; Merainani, Xiong, et al., 2021) helped me to complete the experimental part and SHM tests.

After the simulations presented in the previous chapter, this chapter will present laboratory experiments to further validate the results. We have installed a video-camera to capture images of a vibrating cantilever beam on a shake table. As showed in Fig 5.1(a), the beam was excited only in the horizontal direction by the shake table with a random white noise. Laser sensors were installed on a special structure which has no contact with the vibration system. Three laser sensors were installed to measure displacements at the positions of the targets. Sensors characteristics are given in table. 5.1. Three artificial circle patterns were attached to the beam at the different positions (bottom 1/4, middle 1/2, high 1/1) of the beam, while a target with a different pattern was installed directly on the shake table.

A PHANTOM MIRO 320S video camera was equipped with CMOS sensor and mounted with NIKKOR lens with a fixed $50mm$ focal length. The maximum resolution is 1920×1200 with $10\mu m$ pixel size. The camera was placed at the distance 5.30 m from the shake table and the registration rate is 512 fps with a 300 microseconds exposure time. The scaling factor was about 1.007 $mm/pixel$ indicating the pixel size in the image corresponding to the real world. The scaling factor is calculated based on a known object with its known size and the covering numbers of pixels recorded in the image. The artificial circle target recorded in the image had a diameter about 30 pixels. The ROI selected in Fig 5.1(b) was of size 43×37 pixels. The displacements measured by laser sensors were used as reference to compare the performance of the motion extraction methods by image processing. Because our video camera system and laser sensors system do not have the same trigger, an off-line processing was then applied after the processing of motion extraction from the video frames in order to synchronize the motion signals obtained by

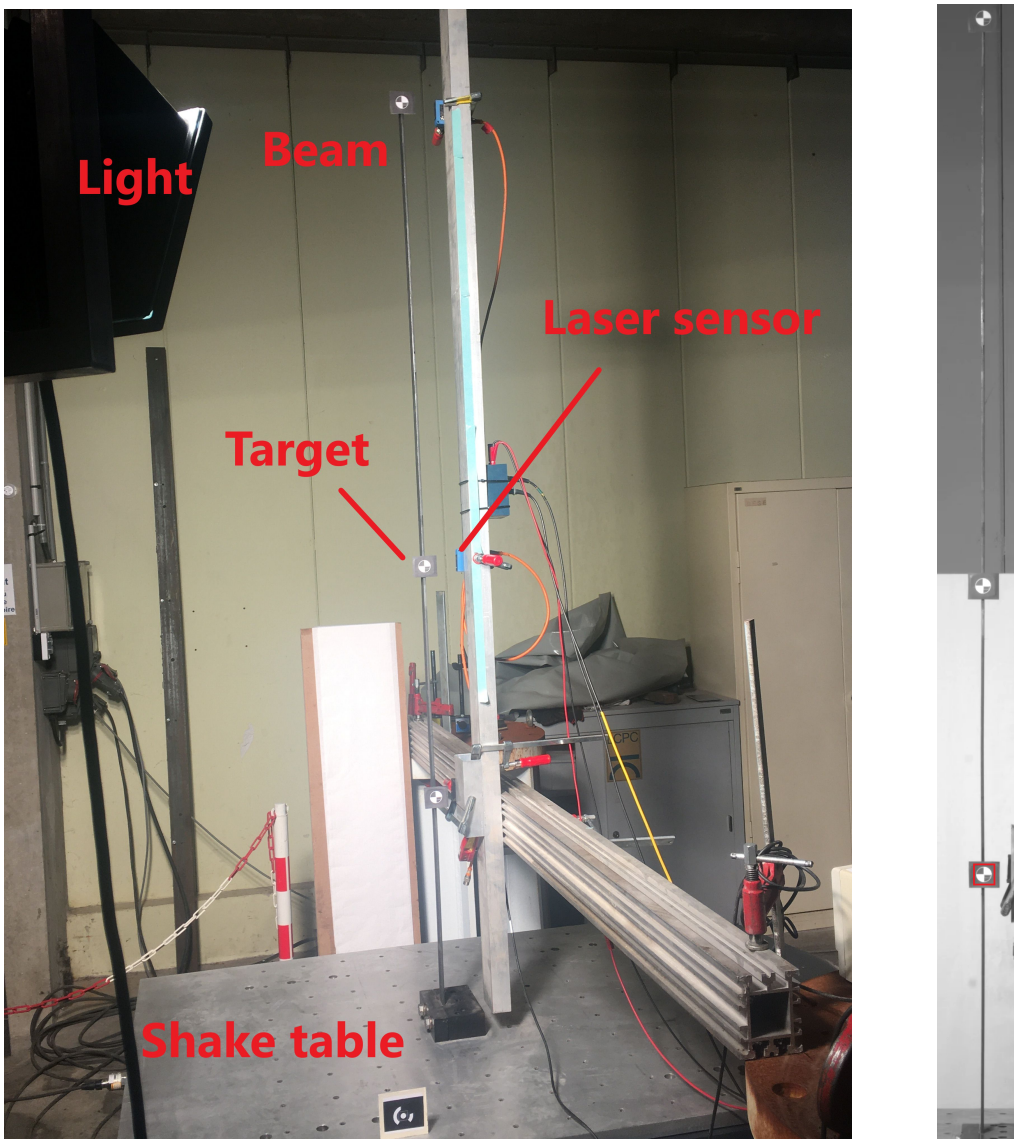
different means. It is done by maximizing the cross-correlation between the displacement signal measured by LDS (laser displacement sensor) and the displacement signal extracted by image processing.

Table 5.1 – characteristics of the laser displacement sensors displacements

Features	Model	
	Keyence IL-030	Wenglor YP06MGV80
— Measuring range :	20 to 45 mm	50 to 100 mm
— Reference distance :	30 mm	50 mm
— Wavelength of light source :	655 nm	655 nm
— Resolution :	$1\mu m$	$< 100\mu m$

5.1 Processing workflow

A general overview of the processing workflow is shown in Fig. 5.2. The input is a sequence of images from the source video of a vibrating structure. The workflow consists of three parts: 1. Preprocessing, 2. Processing, 3.Post-processing. In the preprocessing, we manually select the ROI covering the vibrating structure or object, and for the application of local phase using steerable filters, an additional selection of active pixels is required. In the processing part, we apply the two-steps strategies: coarse to fine approaches. The first step is using ZNCC to estimate the pixel-level shift, and we move the reference image (usually the first the frame of the video) with this estimated pixel-level shift. In the second step, we apply the subpixel shift estimation method between the shifted reference image and the target image to obtain the subpixel shift. In the end, based on the extracted displacement from the video, a time synchronisation is employed to align the displacement signal extracted from the video and the displacement measured by the LDS (laser displacement sensor). The displacement signal is then analyzed for the purpose of SHM: natural frequencies estimation, mode shape identification, deformed model analysis, stabilization diagram and MAC (modal assurance criterion) analysis.



(a) Experimental setup

(b) Raw frame

Figure 5.1 – Laboratory experiment setup

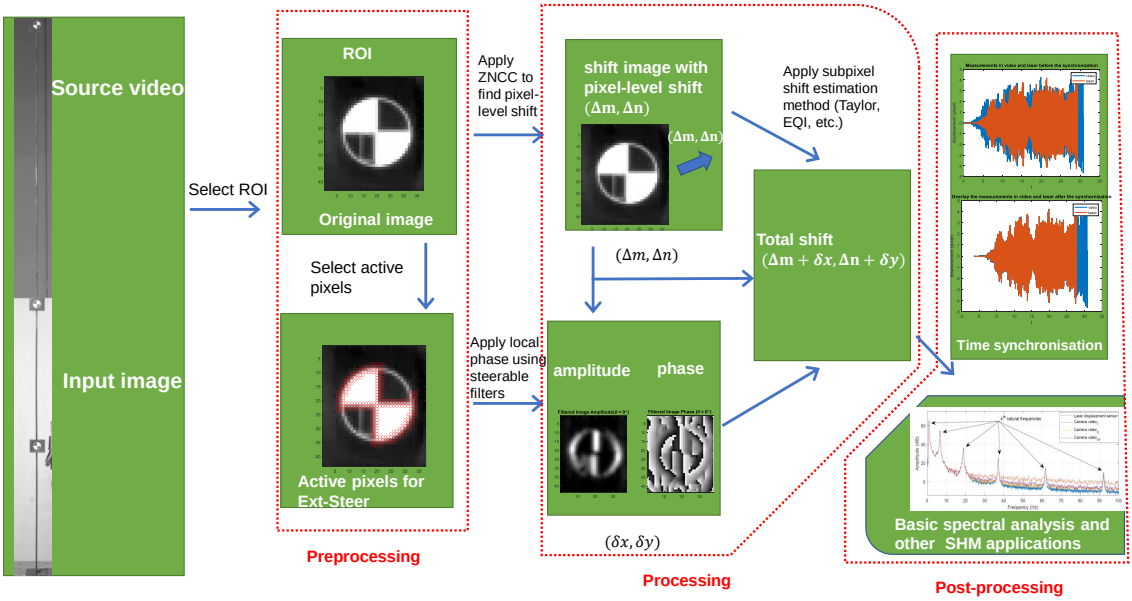


Figure 5.2 – The general overview of the processing workflow, from the input image of the source video, preprocessing including the selection of ROI and active pixels, and outputs of displacements extracted from the video with different methods, in the end, a post-processing including a basic spectral analysis and other SHM applications.

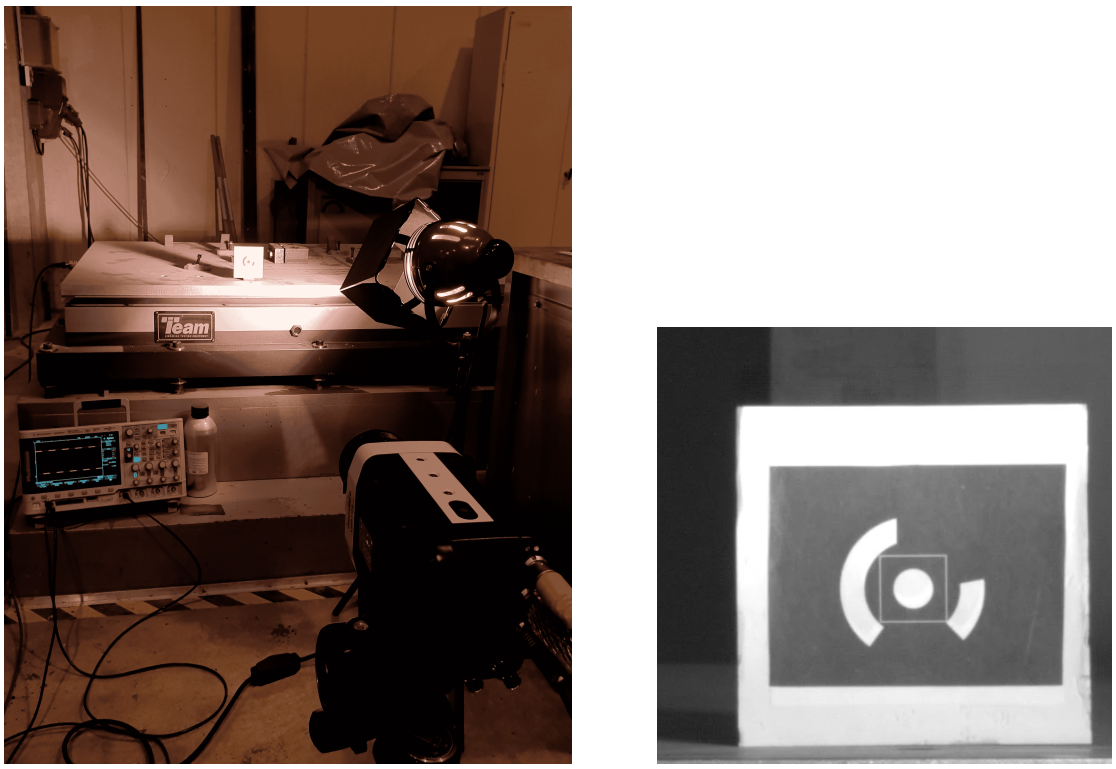
5.2 Two laboratory experiments: shake table test with sinus signal, cantilever beam test with white noise

In this section, we will give more details of the experiments on the two simple structural configurations: a shake table and a cantilever beam. Measurements were made to identify the excitation frequency, their natural frequencies and mode shapes. The study of the deformed model was also conducted. A quasi real-time motion extraction process was verified during the shake table test and cantilever beam test. The performance assessment of different methods has been made in the field experiments.

5.2.1 Shake-table test

The first part of the experiments consisted of installing a target directly on the shake table, as seen in Fig. 5.3(a). A laser displacement sensor was also installed and used to collect the reference displacement of the shake-table. The camera is controlled from the

computer through Phantom Camera Control software. In the cine setting options, the camera resolution, sample rate, exposure time and the extreme dynamic range (EDR) can be adjusted. The exposure time or shutter speed is the length of time when the film or digital sensor inside the camera is exposed to light. Exposure time and frame rate are two closely related. According to the Phantom manual, the exposure should be less than 1/sample rate. In the experiments, resolution for the videos was chosen to give as much spatial data as possible and a sample rate was chosen to be sufficient for the presumed vibration frequencies of the structures while still leaving enough exposure time so that the camera collects enough light.



(a) Experimental setup (b) Raw frame

Figure 5.3 – Setup for motion extraction of a target mounted on the shake-table.

The unidirectional shake table facility at the SII laboratory (Université Gustave Eiffel, Nantes) can move in the horizontal direction, driven by a hydraulic actuator. The later was excited with two types of signals: varying frequency sinusoidal signal with frequency changing from 20 Hz to 10 Hz and a limited band Gaussian noise. The first one supplies the displacements information in the similar condition during our simulation test, the sinusoidal signal remained the fixed frequency in several periods and then it changed

its frequency step by step. The sinusoidal signals can provide data for comparing the performance in the simulation condition and experimental condition, while the other, the white Gaussian noise supplies more complicated displacement signals with varieties of frequency information. Both the two types of signals were generated in two levels: 1. Small displacement (maximum displacement amplitude less than 1 pixel); 2. Large displacement (maximum displacement amplitude more than 1 pixel). The video was recorded with 500 fps and a resolution of 704×496 pixels. The scaling factor for the camera in the testing was 0.27586 mm/pixel. The images extracted from the video source are saved in 8 bits depth.

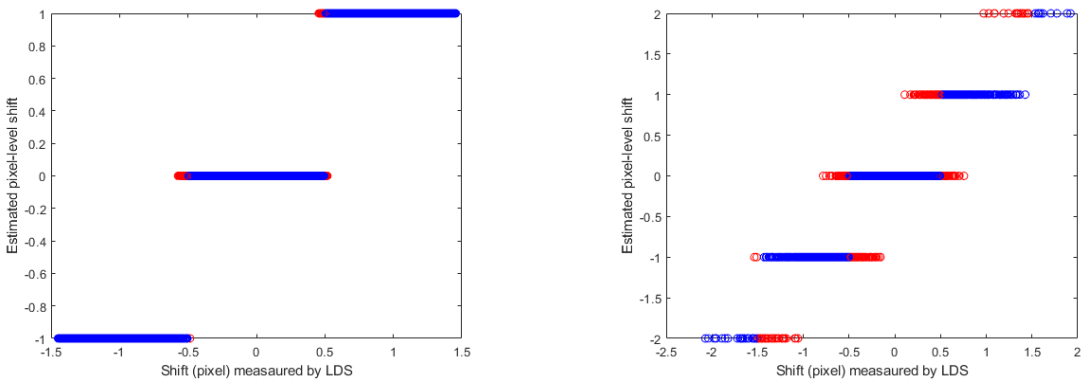
5.2.1.1 Pixel-level shift estimation validation

With the same process that we have followed in the simulation studies presented in the previous chapter, the first step is to validate the estimation of pixel-level shift by Zero-mean Normalized Cross-Correlation (ZNCC). The results are shown in the Tab. 5.2.

Signals of the shake-table	Sinusoidal signal		Gaussian noise	
	Small displacements	Large displacements	Small displacements	Large displacements
Accuracy	0.9231	0.9915	0.9988	0.9064

Table 5.2 – The accuracy of ZNCC to find the right pixel-level shift in the shake-table test

In these results, the pixel-level shift found by cross-correlation is considered "correct" if it is in agreement with the shift indicated by Laser Displacement Sensor (LDS). It is noticed that in the experiment test, the ratio of correct pixel-level shifts found by cross-correlation is always above 90%. When the excitation of the shake-table is a sinusoidal signal, the recording frequencies of the video camera and the LDS are both 512 Hz, however when the excitation of the shake-table is the Gaussian noise, the recording frequency of the video camera is 500 Hz and the recording frequency pf the LDS is 1024 Hz, thus for aligning the measurement of the two systems, a resample process with linear interpolation is applied on the LDS measurement data. When the amplitude of the displacement is small (less than 1 pixel), this resampling process with linear interpolation brings little error, but as the amplitude of the displacement is large, this process produces large error, which can be observed in the Fig. 5.4(b). For the Gaussian noise excitation, we have more outliers (red circles) compared with the previous simulation tests and the sinusoidal signal excitation.



(a) Sinusoidal signal (b) Gaussian noise

Figure 5.4 – Pixel-level shift estimation compared with the preassigned shift. The blue circles refer to the right estimation of the pixel-level shift, and the red circles refer to the wrong estimation of the pixel-level shift (outliers)

5.2.1.2 Subpixel shift estimation validation

After finding the pixel-level shift, we estimate the subpixel shift with the subpixel shift estimation methods Taylor-based method, QSF, POC, UCC, ECC and our proposed method EQI and EQI-Octant. Then the total shift is compared with the displacement measured by LDS. In the Tab. 5.3, we list the RMSE of all the methods for the shake-tale test. The performances of EQI and EQI-Octant are quite similar. The method Taylor have the worse RMSE among all the tested methods. Our two proposed methods have good performance for small displacements: RMSE = 0.0206 and 0.0220 for the sinusoidal signal excitation and Gaussian noise excitation. For large displacements, the estimation of our proposed methods is degraded but still maintain good accuracy. In the Fig, 5.5, we present the results of the displacements measured by our proposed EQI method compared to the displacements measured by LDS

Methods	Sinusoidal signal excitation		Random noise excitation	
	Small displacements	Large displacements	Small displacements	Large displacements
Taylor	0.0270	0.0218	0.0380	0.1363
EQI	0.0208	0.0217	**0.0220**	0.1329
EQI-Octant	0.0206	0.0217	0.0293	0.1337
QSF	**0.0195**	**0.0192**	0.0234	**0.1285**
POC	0.0289	0.0266	0.0545	0.1379
UCC	0.0206	**0.0194**	**0.0219**	0.1312
ECC	0.0216	0.0203	0.0273	0.1328

Table 5.3 – The RMSE of different methods in the shake-table test

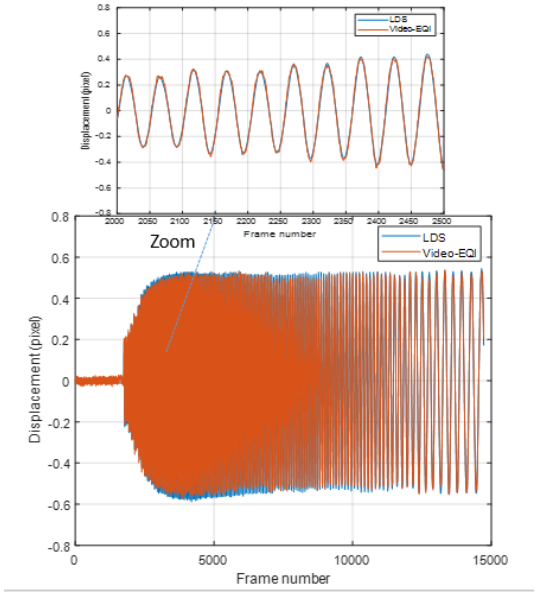
5.2.2 Cantilever beam test

The second part of the experiment consisted in testing a beam on the shake-table with 3 targets fixed at specific positions of the beam. Before this experiment, we have simulated the 3D cantilever Timoshenko beam and have studied its physical property: natural frequencies and mode shapes. Then based on the simulated result, we conducted the experimental test to evaluate performance of different subpixel shift estimation methods.

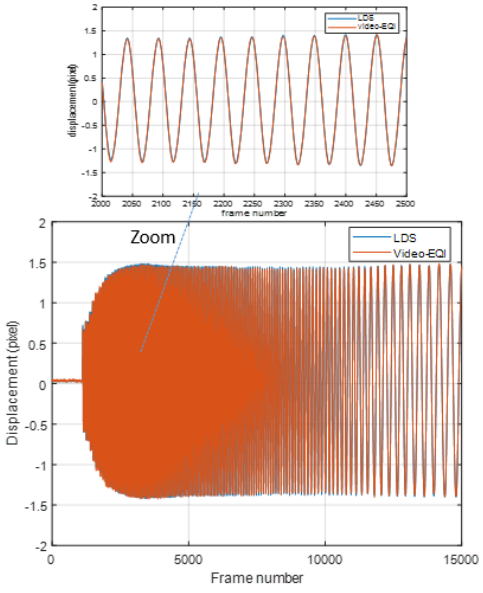
5.2.2.1 Finite element model of a 3D cantilever Timoshenko beam

We build a finite element model of a 3D cantilever Timoshenko beam to simulate and analyse the mode shapes of the cantilever beam. The beam element will be assumed to be a straight bar of uniform cross section capable of resisting axial forces, bending moments about the two principal axes in the plane of its cross section, and twisting moments about its centroidal axis.

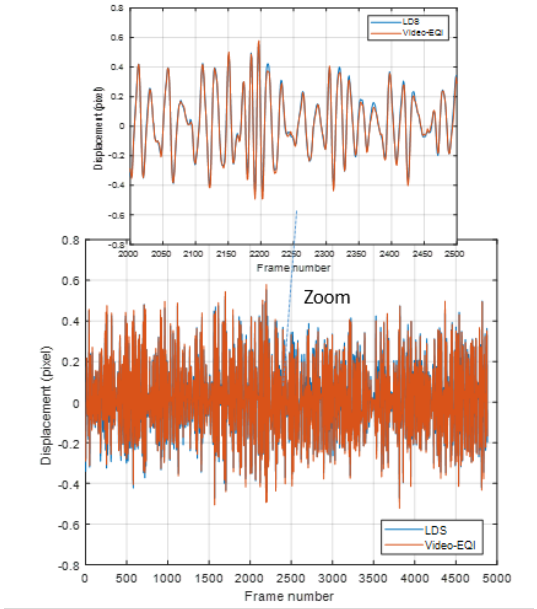
Table 5.4 gives the parameters used in the model. The obtained natural frequencies corresponding to the first six mode shapes shown in Fig. 5.6 are given in Tab. 5.5.



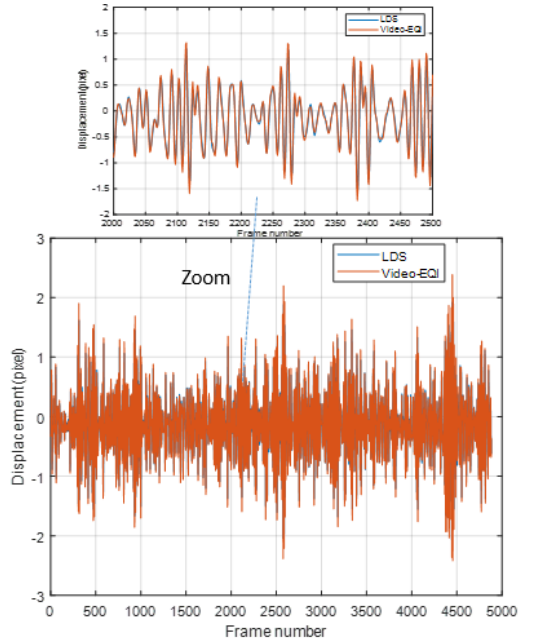
(a) Sinusoidal signal (small)



(b) Sinusoidal signal (large)



(c) Gaussian noise (small)



(d) Gaussian noise (large)

Figure 5.5 – Measurements of displacements by our proposed EQI method compared with the displacement measured by LDS in the shake-table test, (a) sinusoidal signal with small displacements. (b) sinusoidal signal with large displacements. (c) Gaussian noise with small displacements. (d) Gaussian noise with large displacements

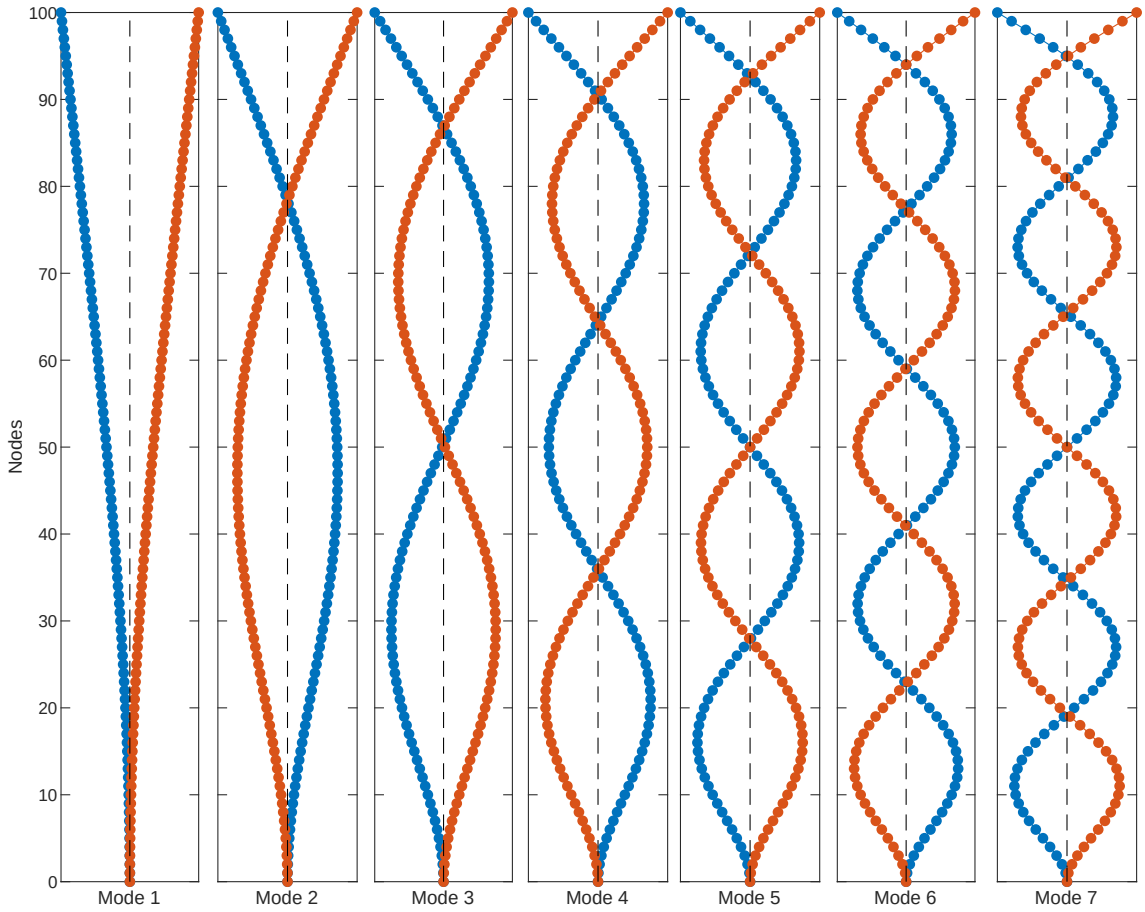


Figure 5.6 – Theoretical first seven mode shapes.

Table 5.4 – Beam parameters

Index	Value
Total beam length(m)	2
Number of elements	100
Beam thickness (m)	0.005
Beam height (m)	0.03
Density of steel $[kg/m^3]$	7700
Young's modulus of steel $[N/m^2]$	2.1e11
Poisson's ratio	0.28
Damping ratio	$\alpha = 8e-6$, $\beta = 8e-2$

The free vibration of the beam is first initiated through an impulse at $t = 0$. Three displacement signals have been extracted from nodes 100, 50 and 25 respectively, corre-

Table 5.5 – Natural frequencies

Mode	Natural frequencies (Hz)
1	1.04
2	6.54
3	18.32
4	35.90
5	59.35
6	88.65

sponding to the beam positions top (1/1), middle (1/2) and (1/4) of the total length from the base. They are shown in Fig. 5.8.

With the simulation of the 3D cantilever Timoshenko beam, the natural frequencies can be identified by a spectral analysis based on the displacement signals. As in the future experimental test with a cantilever beam, we identify the natural frequencies from the displacement signals, this simulation can provide a strong reliance. To perform spectral analysis, fast Fourier transform (FFT) is the simplest tool. However, it is clear that the measured signals are non-stationary, while FFT is mainly used for stationary signal analysis. To overcome this problem, a time-frequency analysis technique, short time Fourier transform (STFT)(Allen, 1977), was used.

The STFT algorithm, uses a displacement signal divided into overlapping blocks of length L with a delay T_s as shown in Fig. 5.7. Finally, the adopted spectra will be the average of the spectra of all blocks.

In our case, L was chosen equal to 2^{10}. The overlap was 80 %. Finally, the adopted spectra will be the average of the spectra of all blocks.

The 3D cantilever Timoshenko beam is first simulated in Matlab with a precise model and then simulated in SolidWorks.

In the simulation Matlab, according to Fig. 5.6, we will show that the dominant resonant frequency of the first (resp. second, third) position will be related to the 1^{st} (resp. 2^{nd}, 3^{rd}) mode shape. The frequency contents in Fig. 5.8 show the resonant frequencies corresponding to the modes, 1 to 6 respectively where the dominant ones were as the expected sequences. In position (1/2), the resonant frequency of the 5^{th} mode shape was not excited.

As our laboratory (Structures and Integrated Instrumentation laboratory (SII)) is equipped with a shake table, it seems judicious to simulate the beam behaviour under base

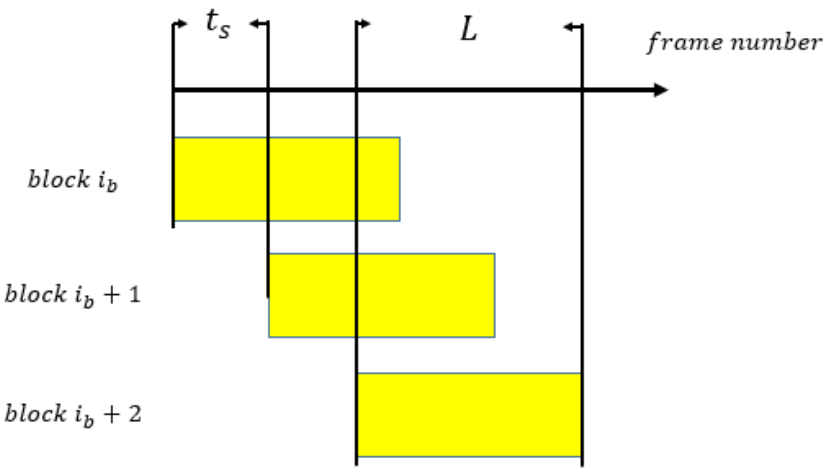


Figure 5.7 – Schematic visualization of overlapping block processing in STFT.

motion excitation. Then, a finite element model of the beam is simulated using SolidWorks software. In simulation module, a modal analysis option is chosen to determine the natural modes of vibration and afterwards the beam displacements. Analysis parameters are set as follows:

- Mesh parameters
  - Jacobian points : 4 Points
  - Element Size : 7.49133 mm
  - Tolerance : 0.374566 mm
  - Total Nodes : 14639
  - Total Elements : 6564
  - Maximum Aspect Ratio : 6.2094
- Load parameters
  - Impulse force with an amplitude of 40 N
  - Uniform base excitation : white noise with standard deviation equal to 0.2 g

Figure 5.9 shows the simulated beam in SolidWorks module. To simulate the effect of sensor placement. Three positions were chosen as on the previous simulation (top (1/1), middle (1/2) and 50 cm from the base (1/4)). Before simulating beam vibration on shake table, we excite the beam by an impulse force near the base. The displacements and their

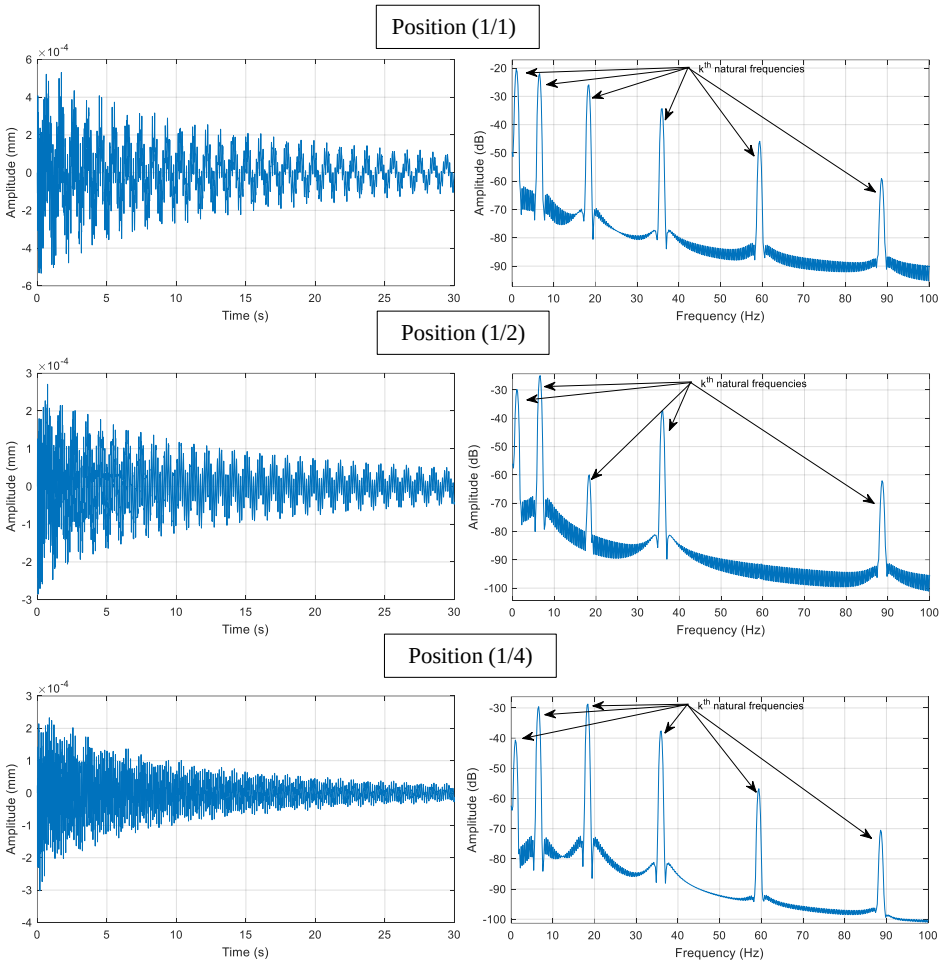


Figure 5.8 – Vibration response of cantilever beam due to impulse excitation in the simulation with Matlab. Left: time domain signals, right: spectra amplitudes.

spectra are shown in Fig. 5.10. One can see that resonant frequencies corresponding to the modes, 1 to 6 appeared, and the dominant ones are in relation with the node position.

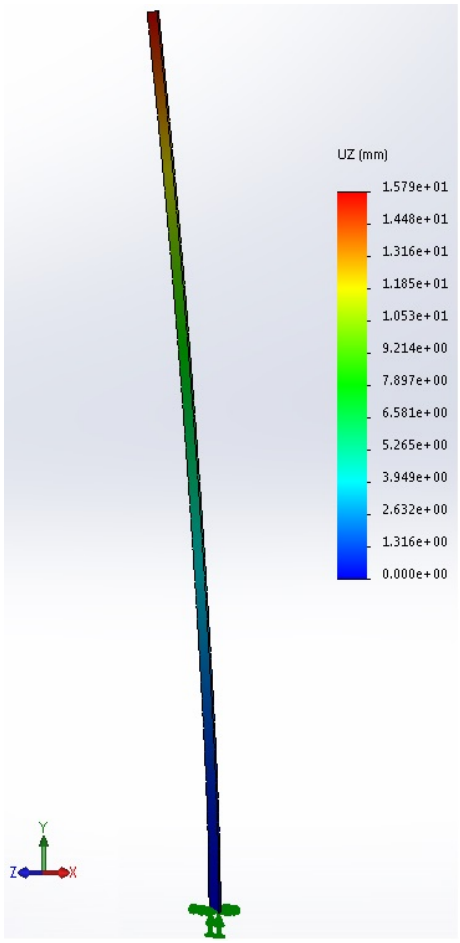


Figure 5.9 – Screenshot of the simulated beam in SolidWorks.

In the second part, we excite the base by a white noise (std =0.03 g). The displacements signals are given in Fig. 5.11. As in the previous simulations, the natural frequencies corresponding to the first six modes are identified in the spectra.

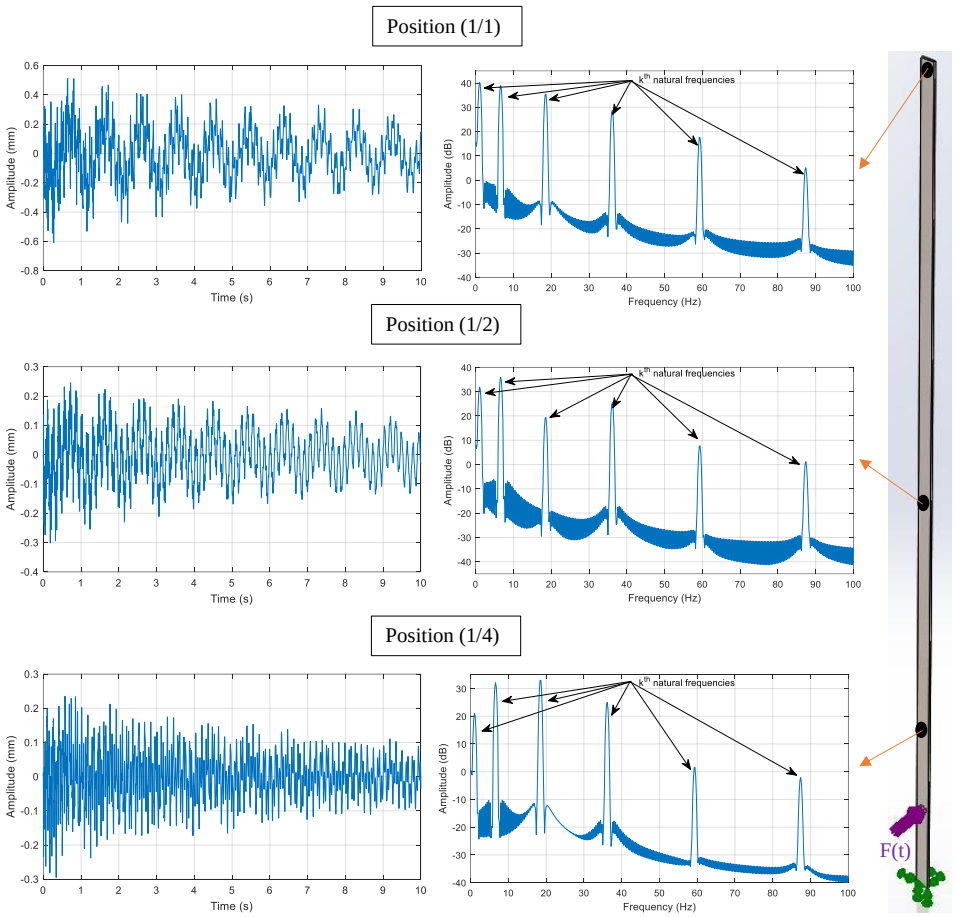


Figure 5.10 – Vibration response of cantilever beam due to impulse excitation in the simulation SolidWorks

5.2.2.2 Experimental verification

After the simulation of the 3D beam structure, in the experimental test, we used a cantilever beam having the same dimension as in simulation. The targets location were not chosen arbitrarily. In fact, according to Fig. 5.6, each mode shape presents a maximum displacement in a specific beam position, namely, vibrating loop. For example, at the top of the beam, the dominating frequency will be the resonant frequency of the first mode. Three target positions were chosen. The first one is located in the top (1/1), the second, in the middle (1/2) and the third (1/4) from the base. By doing so, we then expect to emphasize the detection of the first six vibrating modes. In what follows, the beam was excited either by a hammer near the base or by a random noise.

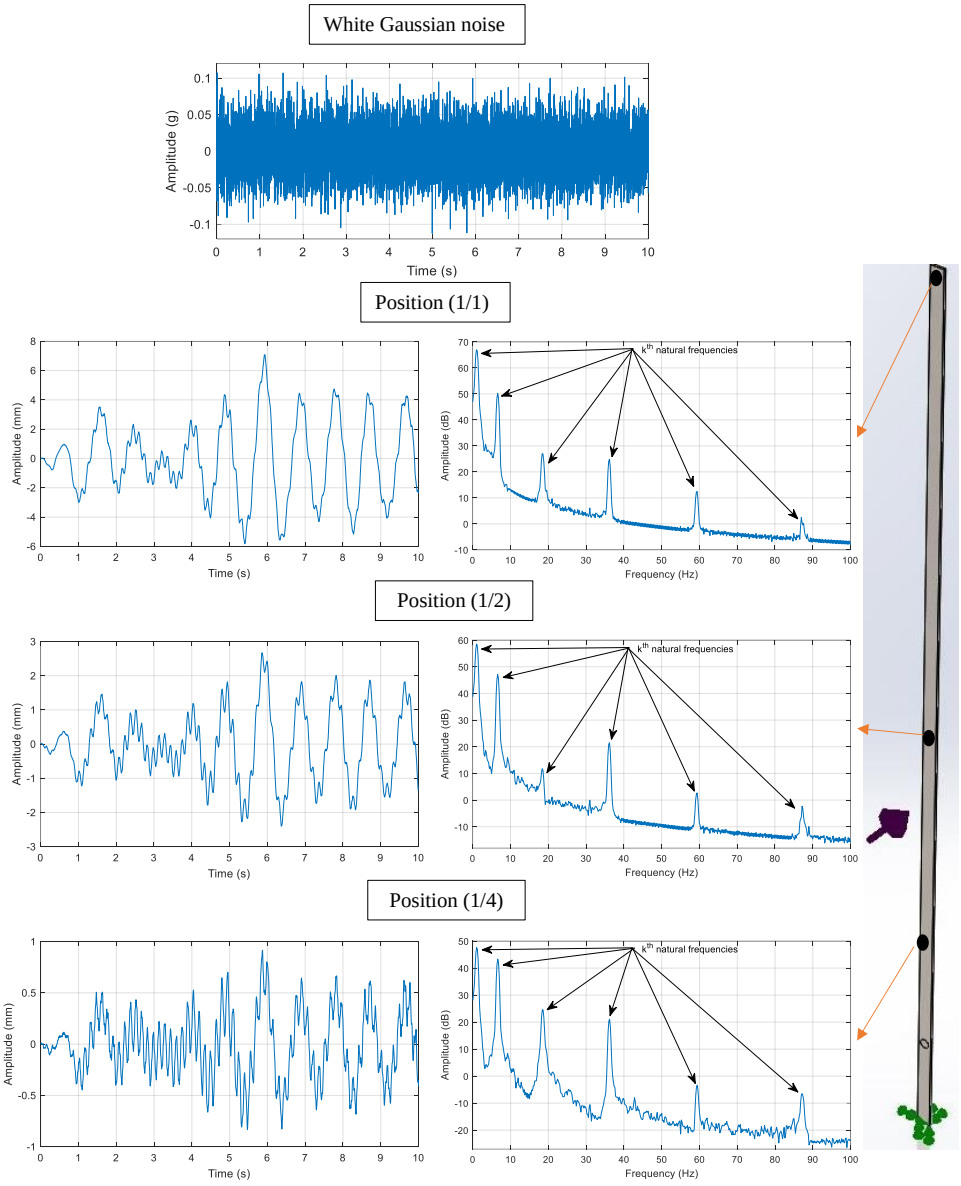
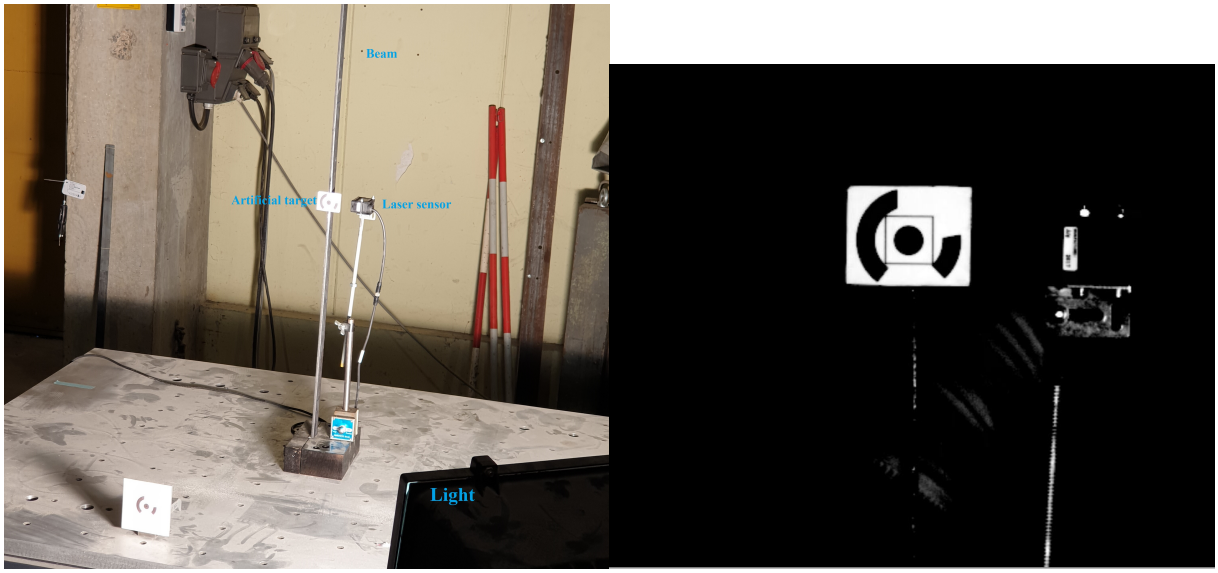


Figure 5.11 – Vibration response of cantilever beam due to noise excitation in the simulation SolidWorks

5.2.2.3 Hammer strike test

In the first part, the cantilever beam was excited by a hammer strike. Two strike intensities scenarios with the hammer have been tried, referred to as "weak strike" and "strong strike". Strong strike is aimed at generating strong displacement amplitude larger than 1 pixel and weak strike is aimed at generating small displacement amplitude less than 1 pixel. In this hammer strike test, the camera video system was recorded with 500 Hz and our LDS system was recorded with 512 Hz. A resampling process was applied to LDS measurement data to align the two signals. This experiment setup is shown in Fig. 5.12. Consequently, only one target located at 1/4 position of the beam appears in the scene as seen in Fig. 5.12b.



(a) Experimental Setup (b) Raw Frame from camera

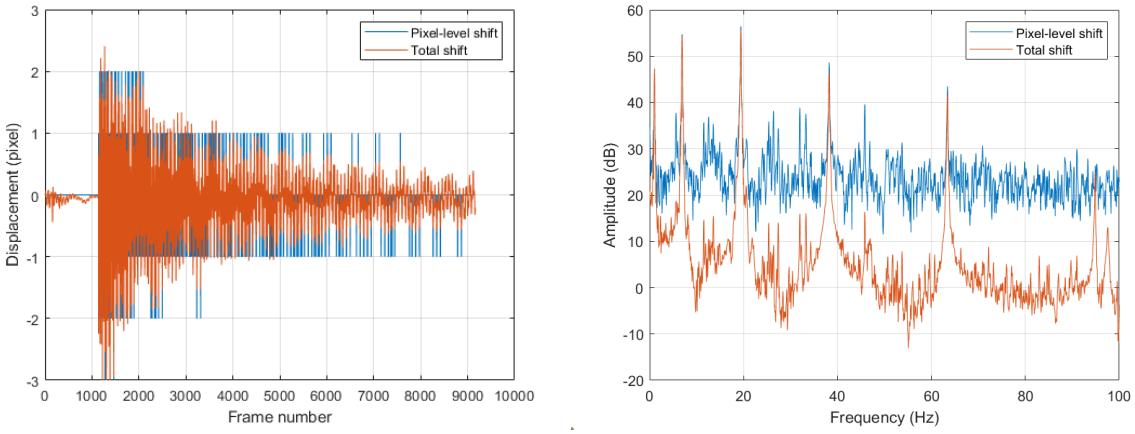
Figure 5.12 – Setup for motion extraction of a target located at 1/4 position of the beam.

A) Strong strike :

The extracted displacement using the vision system based on the method EQI and the measured one using the LDS are given in Fig. 5.14(a). The accuracy of finding the right pixel-level shift is presented in Tab. 5.6.

The natural frequencies, in Fig. 5.14(b), obtained from both systems are compared in table. 5.7. Figure. 5.13 compares the extracted pixel-level shift with total shift measured by the method EQI. Since the strike is strong, displacements over one pixel detected by the cross correlation are significant. Total shift with subpixel accuracy in Fig. 5.13a gives

more and meaningful details to motion signal. The spectrum based STFT of the pixel-level shift does not show the natural frequency of six modes, with low amplitude, but overwhelmed by other spectral components because of the square form of the pixel-level shift signal. This implies that in the application of SHM, it is meaningful to realize a measurement of the displacement with subpixel accuracy.



(a) (b)

Figure 5.13 – Comparison of the beam test with hammer strike (a) pixel-level shift and total shift measured by vision system based on EQI. (b) spectra based STFT

As the sensor and target are in the position (1/4), we expect that the dominant resonant frequency will be related to the 2nd and 3rd mode shape (see Figs. 5.14(b)). This result is coherent with the simulation of the beam in Fig. 5.8.

B) Weak strike :

We followed the same procedure as the last case, the extracted displacement using the vision system and the measured one are given in Fig. 5.14(a)(b).

The spectrum of video cameras signal looks more noisy than the ones of LDS. In fact, there are some parameters that influence the video cameras signals. In the work of (J. G. Chen et al., 2015), the author investigates the effect of the exposure time, rate frame and focal length on the noise floor. He concluded that :

— The larger the exposure time, the less the noise floor will be.

— Lower noise floors can be achieved at lower frame rates, because longer exposure times can be used, potentially increasing the contrast between dark and bright objects. But the difference was not obvious.

— In case of focal length impact on noise floor, the author made a measurements by changing the focal length from 85mm, 50mm, and 24mm. It had an f-stop of f/3.5-4.5, which means that at 24mm the lens has an f-stop of f/3.5 while at 85mm the lens has an f-stop of f/4.5. That is to say that the lens lets in more light at 24mm than at 85mm. The difference in noise floor is thus explained by the change of brightness in the video.

In Tab. 5.8, we compared the RMSE of different methods tested on the beam test with hammer strike. Our proposed method EQI and EQI-Octant have the best accuracy for the beam test with strong strike. For the weak strike, our proposed method can still reach good accuracy (0.0437 and 0.0475), a little less accurate than the method ECC (0.0399). By contrast, for strong strike, our proposed methods EQI and EQI-Octant reach a similar precision compared with the method ECC. The method ECC is limited by its initial condition to set previously and the number of iteration, and the computation cost of ECC is relatively higher than EQI as shown in Tab. 4.1. For several frames, with a fixed number of iteration (in our case, 10) and the non-ideal initial value, we had not found the optimal solution for the subpixel shift.

Table 5.6 – The accuracy of ZNCC to find the right pixel-level shift in the beam test with hammer strike

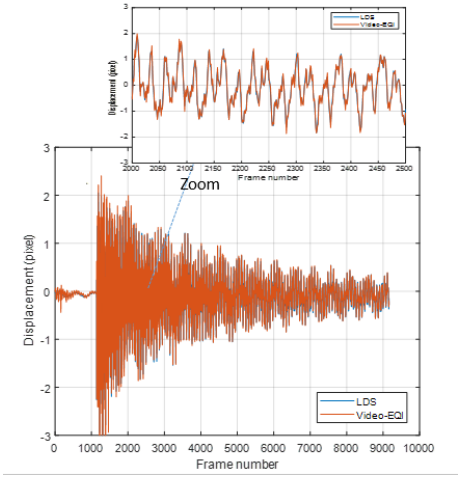
Hammer strike	Strong strike	Weak Strike
Accuracy	0.9596	0.9879

Table 5.7 – Comparison results of natural frequencies obtained by vision and laser systems

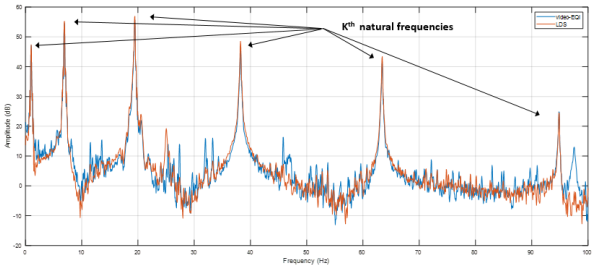
	System	Modes				
		1	2	3	4	5
Natural frequencies	Vision system EQI (Hz)	0.9966	6.777	18.99	37.32	61.94
	Laser system (Hz)	0.9972	6.731	19.01	37.33	61.95
	Absolute error (Hz)	0.0006	0.046	0.02	0.01	0.01

5.2.2.4 Random signal excitation test

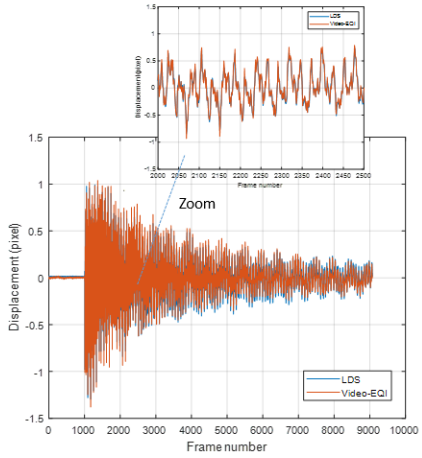
In this part, the beam was excited by the shake table with random Gaussian signal. But for that, we had to isolate the laser sensors from the shake table. A structure has been



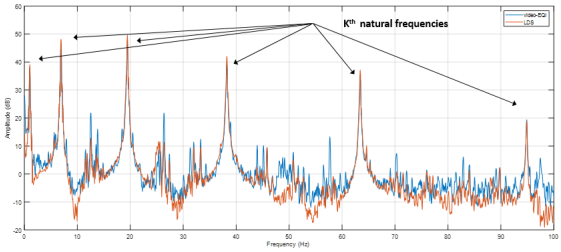
(a) Hammer strike (strong)



(b) Hammer strike (strong)



(c) Hammer strike (weak)



(d) Hammer strike (weak)

Figure 5.14 – Measurements of displacements by our proposed EQI method compared with the displacement measured by LDS in the beam test with hammer strike, (a) measurements of displacements with strong strike. (b) spectra based STFT with strong strike. (c) measurements of displacements with weak strike. (d) spectra based STFT with weak strike

mounted as seen in Fig. 5.15a. In order to compare the results of artificial and natural targets (the edges) of the beam, a white background was added. The distance between the camera and the beam was also increased so that the full beam length will appear in the image (see Fig 5.15b).

The shake table was piloted in random mode with a hydraulic cylinder. Random vibration can be represented in the frequency domain by a power spectral density function (PSDF) ($[G^2/Hz]$ versus frequency $[Hz]$). Note that the amplitude is actually $[GRMS^2/Hz]$. Five PSDFs were generated and tested. Their amplitudes were modified one time (denoted *Low* and *high*) in such a way to generate displacements, in one, two or all targets, lower/higher than one pixel size. This allows to highlight the importance of subpixel displacement estimation. In this report, we only present two PSDFs presented in Fig. 5.16. Note that, the generated noise was band limited between $1Hz$ and $200Hz$. Extra lighting was used so that the high-speed camera would be able to see the vibrating structures.

The camera nominally measured the motions of the three targets (ie. all beam length), on one recording, mounted to the beam. Measurements were made at 512 fps with a 300 microsecond exposure time, at a resolution of 1920×120 pixels, and 8-bit depth or 12-bit depth. The distance between the camera and the beam was $5.30m$. The scaling factor was, thus, estimated to be 1.007 $mm/pixel$. To obtain more precise value, A4 checkerboard with $55mm$ squares was printed and pasted on the beam, before doing the tests. The scaling factor was 1.010 $mm/pixel$. The later value was adopted. The displacements measured and spectra based STFT are shown in Fig. 5.17.

Table 5.8 – The RMSE of different methods in the beam test with hammer strike

Methods	Hammer strike excitation	
	Strong	Weak
Taylor	0.0925	0.0457
EQI	**0.0890**	0.0437
EQI-Octant	**0.0890**	0.0475
QSF	0.0902	0.0450
POC	0.0944	0.0478
UCC	0.0915	0.0479
ECC	**0.0892**	**0.0399**

(a) Experimental Setup

(b) Raw Frame from camera

Figure 5.15 – Setup for motion extraction of three targets pasted on the beam with white background of beam test oscillated by random Gaussian signal

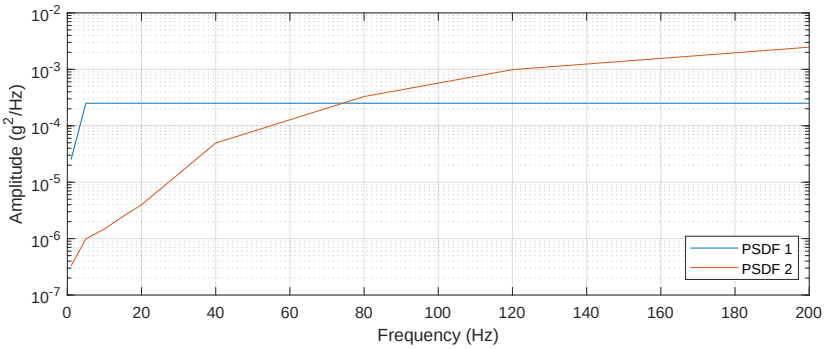


Figure 5.16 – Adopted power spectral density functions

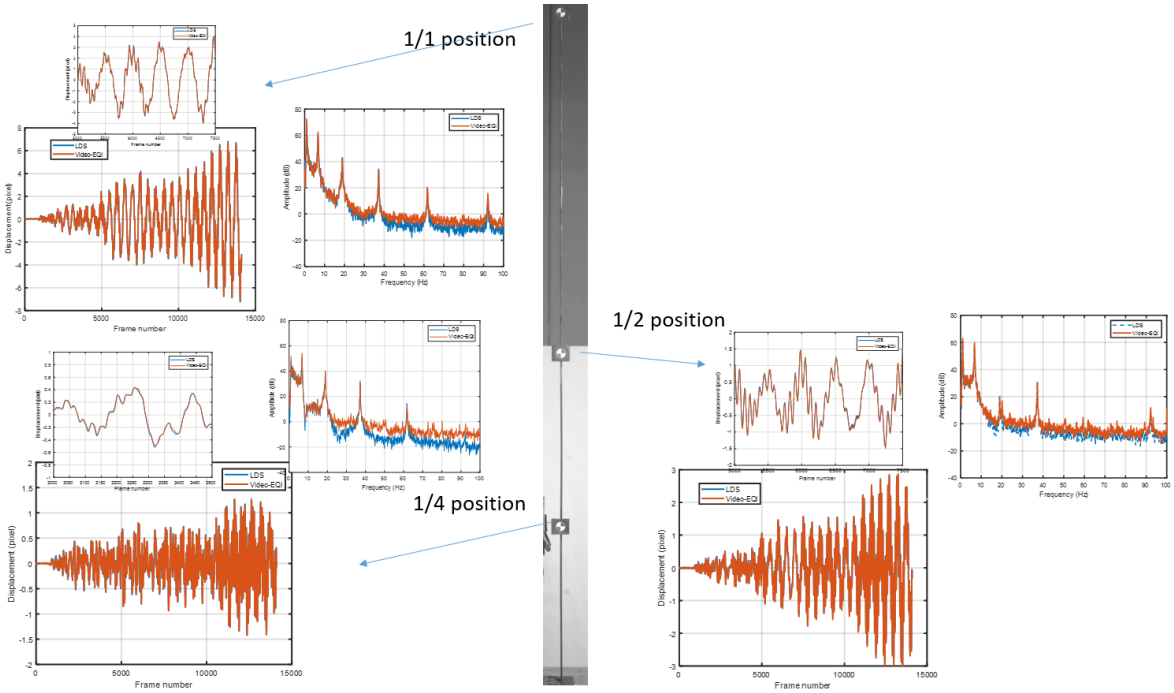


Figure 5.17 – Comparison of the beam displacements extracted from targets at positions (1/1), (1/2) and (1/4), measured by LDS and vision systems based on our proposed method EQI, the spectra based STFT is presented correspondingly for each position.

Table 5.9 shows the accuracy of the different methods in the beam test excited by random Gaussian for all targets located at different positions of the beam. Our proposed method EQI and EQI-Octant have almost the best accuracy in all the tested methods (0.0396 at position 1/1 , 0.0344 at position 1/2, 0.0188 and 0.0532 at position 1/4), It is noticed that at higher position the target is located, the RMSE of the motion extraction is larger. The reason is that in the bottom of the beam, the light condition is better than for

the high position of the beam. The contrast is different: the ROI located at 1/4 position has larger contrast than the one located at 1/1 position. The laser displacement sensor installed at the bottom 1/4 position has best accuracy than the other 2 LDS installed at the upper position, resulting in the larger RMSE of the motion extraction at 1/2 and 1/1 position. The spectra presented in Fig. 5.17 shows a good natural frequency identification compared with the signal measured by the LDS. We have observed that the noise floor of video extraction signal is quite similar to the displacement signals measured by the LDS system.

A summary of the estimation accuracy (RMSE) of all the motion extraction algorithms is available in Tab. A.1. In the various conditions of measurements, our proposed method EQI acquires the best performance (smallest mean RMSE and highest average rank).

Table 5.9 – RMSE comparison for different methods in the beam test excited with random Gaussian signal in the different positions of the beam

Positions	Motion extraction algorithms						
	Taylor	EQI	EQI-Octant	QSF	POC	UCC	ECC
Position (1/1)	0.0506	**0.0396**	**0.0396**	0.0413	0.0480	0.0406	0.0405
Position (1/2)	0.0530	0.0344	0.0344	**0.0302**	0.0536	0.0424	0.0361
Position (1/4)	0.0566	**0.0188**	**0.0188**	0.0204	0.0251	0.0197	0.0275

5.2.2.5 Motion extraction with targets and natural contours

The previous experimental tests have used three targets (with some artificial texture on it) to extract the displacement signal. This section aims at proving that the image-based extraction methods may also work in target-free conditions, namely, from the natural edges of the structure.

Different from the previous selected ROI of the target, the ROI could be focused on selected areas along the edges of the beam. For simplicity, the selected areas are very close to the LDS or the ROI of the monitored target as shown in Fig. 5.18 We chose the ROI covering the edge of the vibrating beam close to the ROI of the selected target. The displacements measured by vision system based on the method EQI and the spectra based STFT are compared in Fig. 5.19. We have observed that the displacement signals extracted by the 2 selected area are very similar and the spectra based STFT is almost the same. Due to the slight difference of the position of the ROI, the difference is acceptable. This analysis is a powerful support for the practical application of SHM without the

installation of the artificial targets, which eases the application of the SHM using image processing techniques.

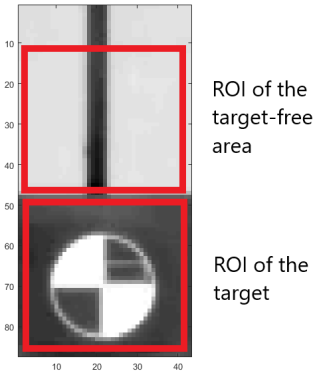
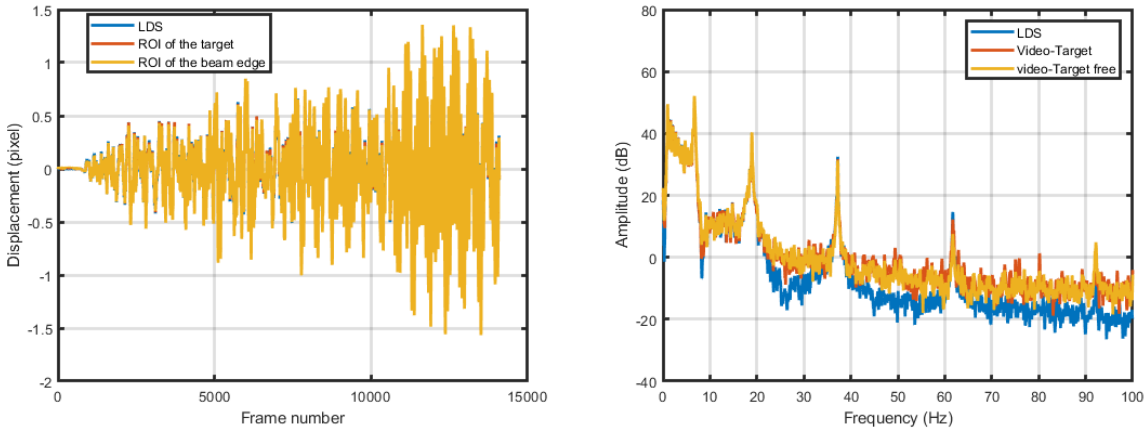


Figure 5.18 – Selected ROI of the target-free area compared with the ROI of the target



(a) Displacement

(b) Spectra

Figure 5.19 – Comparison of the selected ROI of the target-free area compared with the ROI of the target

5.2.3 Outdoor experimental test

All the experiments conducted before are in indoor condition with a stable and luminous light source. For testing the efficiency of the motion extraction algorithms for video-based techniques against the changing light condition, we conducted the same experiment in the outdoor condition without additional light source. The video was regis-

tered with natural light. As shown in Fig. 5.20, we used a natural wall as a background, three targets were installed to the beam structure as in the previous experience. There laser displacement sensors were placed at the same height of the targets on another independent steel structure. A fixed target attached to the background wall was used as the stationary reference to compensate the measured displacements in case the camera was slightly disturbed by the wind or other external environmental conditions. The weather condition affects the estimation of the video-based techniques. We selected a cloudy period to compare the impact of the changing light condition.

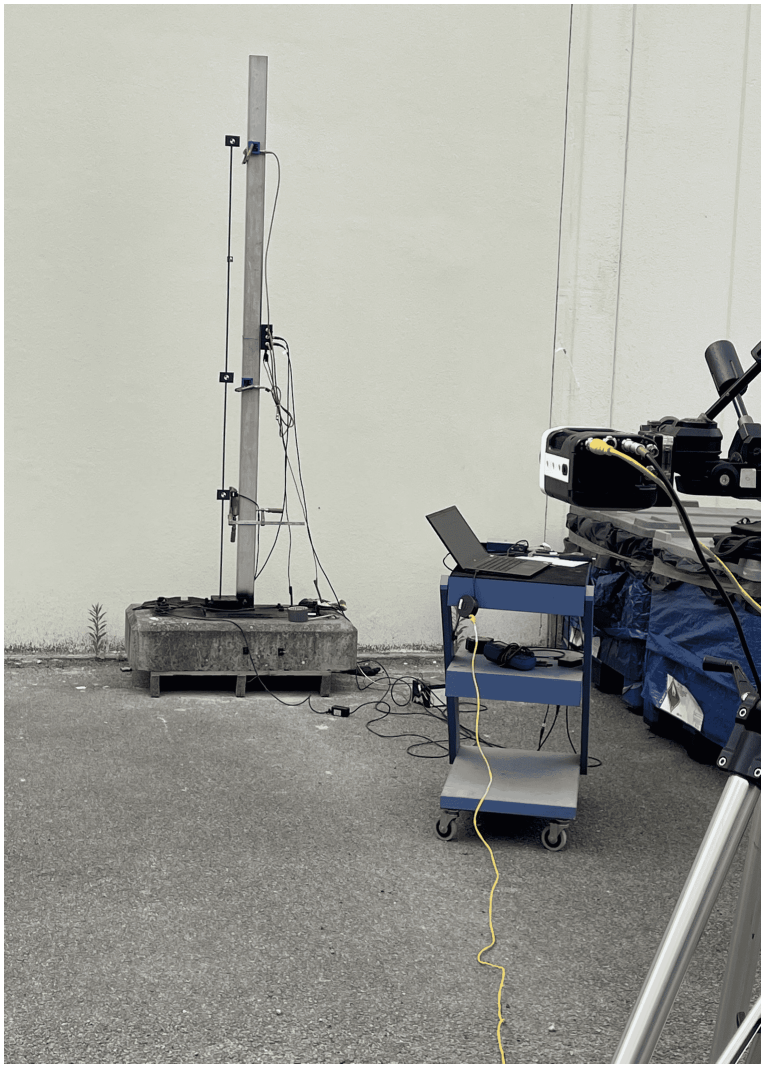


Figure 5.20 – Experimental setup for the outdoor condition

Two conditions have been considered: 1. no cloud comes across and the light condition is relatively unchanged; 2. The cloud comes across for several seconds and the light

condition is changing during this period.

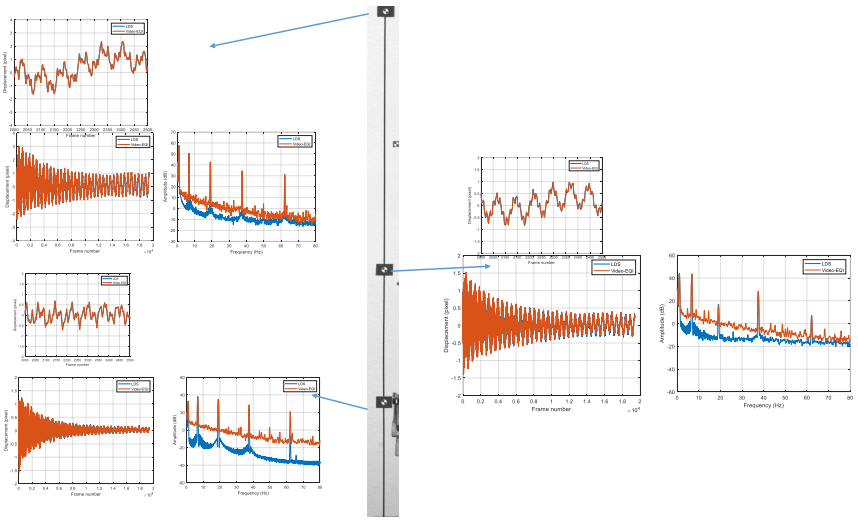


Figure 5.21 – Extracted displacements and spectral analysis for the outdoor condition in the case 1: no cloud comes across and the light condition is relatively unchanged.

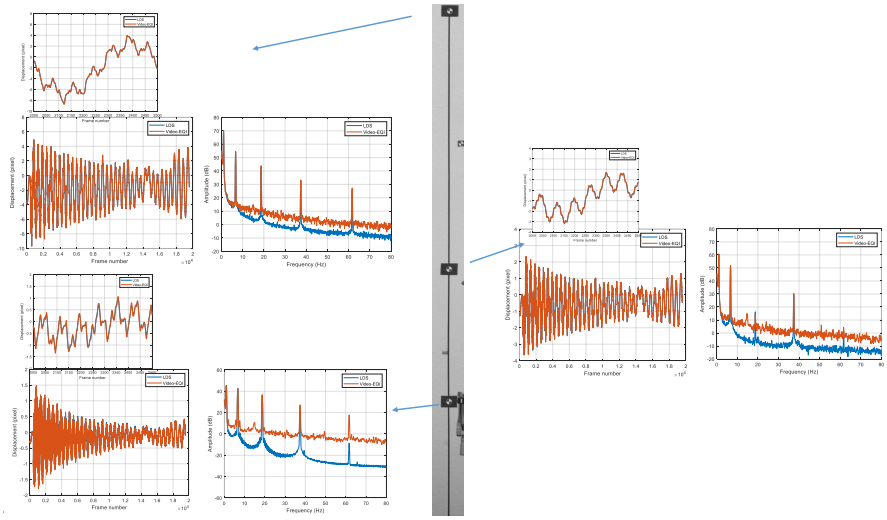


Figure 5.22 – Extracted displacements and spectral analysis for the outdoor condition in the case 2: the cloud comes across for several seconds and the light condition is changing during this period.

In the outdoor environment, the main concern affecting the performance of video-based techniques comes from the natural light and the external influence such as wind. Comparing the first result (no cloud) with indoor experiments as shown in Fig. 5.17, we have noticed that the noise floor obtained by displacement signals of LDS was very

similar in the indoor and outdoor condition. It is logical that the measurements of LDS are almost unaffected by the light condition. By contrast, a redundant frequency at 65.74 Hz is observed during the outdoor experiments for the measurements of LDS. The reason is that the structure to install the LDS and the beam is not ideally independent, and there is a heavy cornerstone to connect the two structures as shown in Fig. 5.20, implying that the installation of the conventional sensors such as LDS or accelerometers is rigorous.

It is noticed that our video-based techniques are slightly affected by the light condition as shown in Fig. 5.21 and 5.22. The obtained spectra from the relatively unchanged light condition or the changing light condition show no big difference. Compared to the spectral of measurements of LDS, the noise floor of the measurements of LDS is always a little better than that of video-based techniques. The difference is obvious for the position 1/4 as the LDS at this position is more accurate than the others.

Tab. 5.10 shows the RMSE for all the motion extraction methods in the outdoor experiments. Compared to the Taylor method, all the other methods have a good improvement on the precision. Our proposed method EQI and EQI-octant have a relatively good performance compared with other reference methods.

Table 5.10 – RMSE Comparison of different motion extraction algorithms using the measurements of laser system as reference in outdoor condition

	Position	Motion extraction algorithms						
		Taylor	EQI	EQI-Octant	QSF	POC	UCC	ECC
No cloud	Position (1/1)	0.0649	0.0507	0.0509	0.0521	0.0487	0.0487	0.0487
	Position (1/2)	0.0517	0.0376	0.0385	0.0371	0.0336	0.0342	0.0360
	Position (1/4)	0.0676	0.0585	0.0591	0.0579	0.0602	0.0579	0.0612
Cloud	Position (1/1)	0.0688	0.0469	0.0481	0.0423	0.0382	0.0389	0.0396
	Position (1/2)	0.0508	0.0311	0.0330	0.0318	0.0352	0.0286	0.0295
	Position (1/4)	0.0662	0.0485	0.0486	0.0487	0.0480	0.0473	0.0497

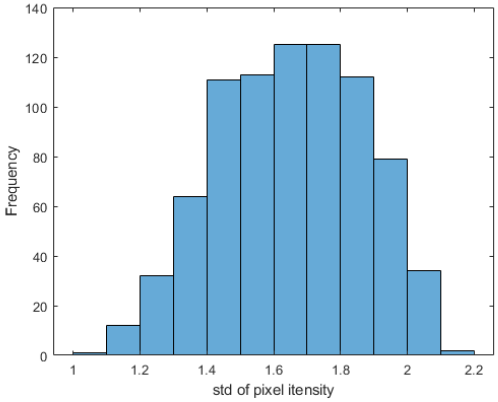
5.3 Performance assessment on the experimental data

In this section, we relate the result of the simulation and the result in the experimental test and analyse the consistency in the estimation of the noise and the performance assessment of different subpixel shift estimation methods.

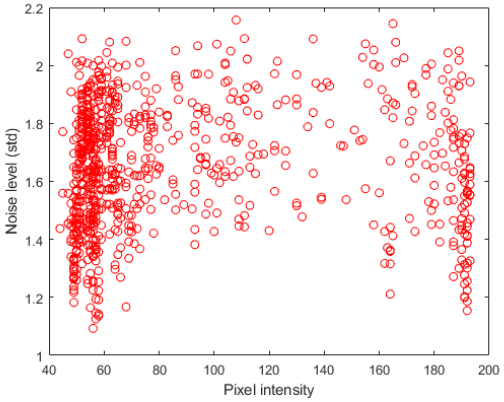
5.3.1 Noise estimation in the experimental test

In the simulation test, we have simulated the synthetic images with two noise models, namely additive Gaussian noise and additive plus signal-dependent multiplicative noise. In the experimental environment, the image formation and the light condition are complicated. We have noticed that if the image is recorded with 8-bit depth (format 'png'), the noise distribution is close to the model of additive Gaussian noise. If the image is recorded with 16-bit depth (format 'tif'), the noise distribution is close to the latter, the model of additive noise and signal-dependent multiplicative noise. In the experimental test, we have observed that the intensity of the pixel of the image is not stable due to the light condition and the CCD camera in the condition where there is no excitation. We have recorded a short video before the excitation with the intention of synchronising the video-based signal and LDS measurement. This can also be applied to analyse the noise in the experimental condition. With the SDK of Phantom camera, we can extract the image in 16 bits from the raw video data 'cine', in the another way, we can convert the ram video data to the video with format 'mp4' with the application of Phantom camera named "PCC" (Phantom camera control application) and then extract the image from video with format 'mp4' to obtain the image in 8 bits. For the duration before the excitation, the standard deviation (std) of gray level in time series for each pixel is calculated, and then we present the relation between noise level (std) and pixel intensity as shown in Fig. 5.23

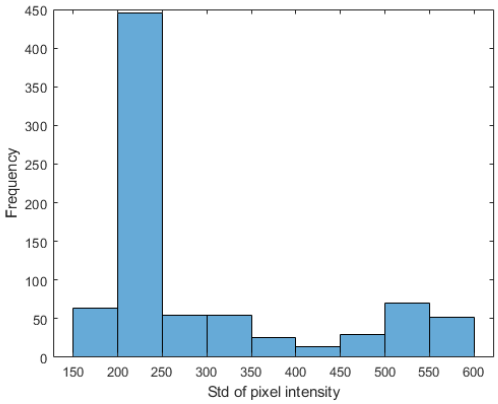
As shown in Fig.5.23, when the image is recorded with 8-bit depth, the histogram of pixel intensity presents a Gaussian distribution with mean value 1.6. The relation between noise level (std) and pixel intensity is random. When the image is recorded with 16-bit depth, the histogram of pixel intensity centers on the standard deviation of pixel intensity in the interval 200-250, and the relation between noise level (std) and pixel intensity presents linear relation. To better compare the 8-bit image and 16-bit image, the result is normalized with division by $2^{\text{n-bit}}$ as shown in Fig. 5.24 and 5.25. This similar linear relation has been mentioned by many authors (Lee et al., 2018; X. Liu et al., 2013) as shown in Fig. 5.26. (X. Liu et al., 2013) explained that the additive white Gaussian noise (AWGN, our first noise model with additive noise) is widely assumed in many image processing algorithms for its simplicity of adding the noise and estimating the noise. However, in the real world, the noise from actual cameras is better modeled as signal-dependent noise (our second noise model with AWGN and multiplicative noise).



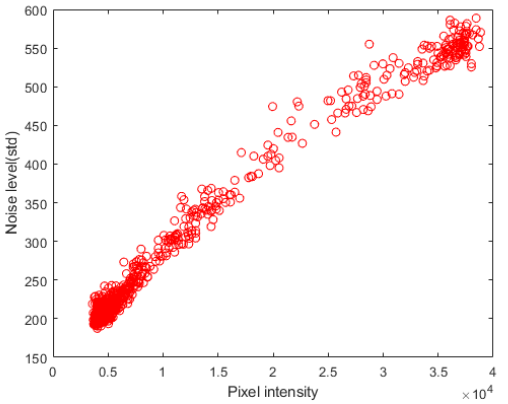
(a) Histogram of pixel intensity std for image 8-bit



(b) Std vs. pixel intensity for image 8-bit

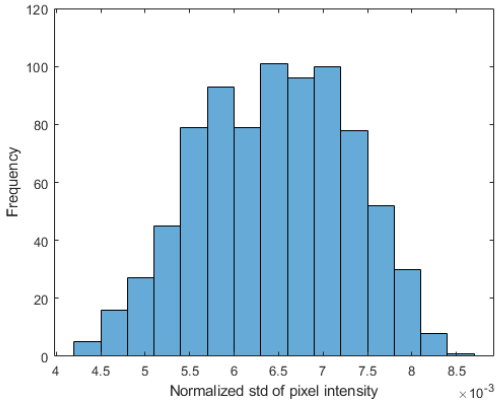


(c) Histogram of pixel intensity std for image 16-bit

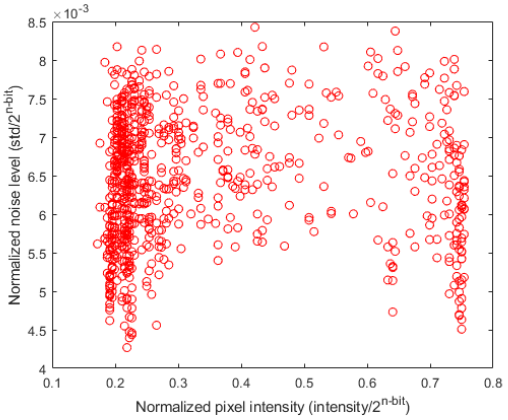


(d) Std vs. pixel intensity for image 16 -bit

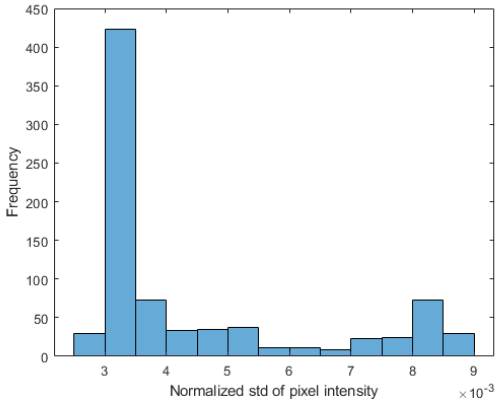
Figure 5.23 – Estimation of noise for the experimental image 8-bit and 16-bit



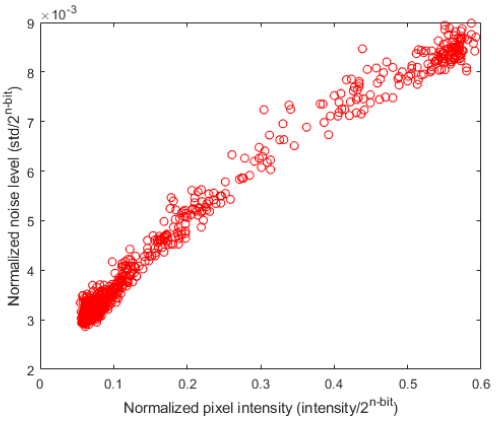
(a) Normalized histogram of pixel intensity std for image 8-bit



(b) Normalized std vs. pixel intensity for image 8-bit

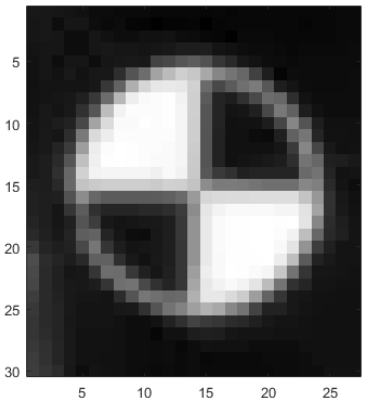


(c) Normalized histogram of pixel intensity std for image 16-bit

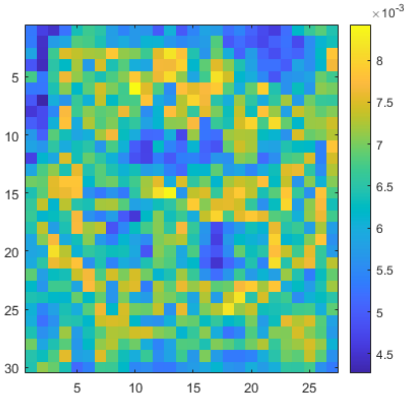


(d) Normalized std vs. pixel intensity for image 16 -bit

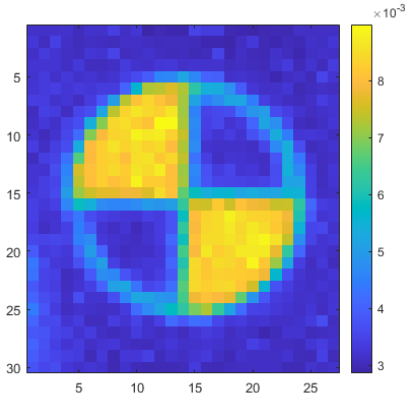
Figure 5.24 – Estimation of noise for the experimental image 8-bit and 16-bit after normalization



(a) Original image



(b) Distribution of std for image 8-bit



(c) Distribution of std for image 16-bit

Figure 5.25 – Comparison of the distribution of std for image 8-bit and 16-bit

In order to find the consistency between the simulation test and experimental test, we analyse the histogram of pixel intensity std and normalized std vs. pixel intensity for simulated image with two noise models: 1. AWGN; 2. AWGN and multiplicative noise. As shown in Fig. 5.27, compared with the result of Fig. 5.24, we found that the experimental images recorded by 8-bit present highly similar histogram and relation of std vs. pixel intensity with the simulated images with AWGN, and the experimental images recorded by 16-bit present highly similar histogram and relation of std vs. pixel intensity with the simulated images with AWGN and multiplicative noise.

In the end, we compared the RMSE of different motion extraction method tested on 8-bit images and 16-bit images to confirm if the bit depth has an effect on the measurement. In Tab. 5.11, we see that the images recorded by 16-bit can improve the precision of

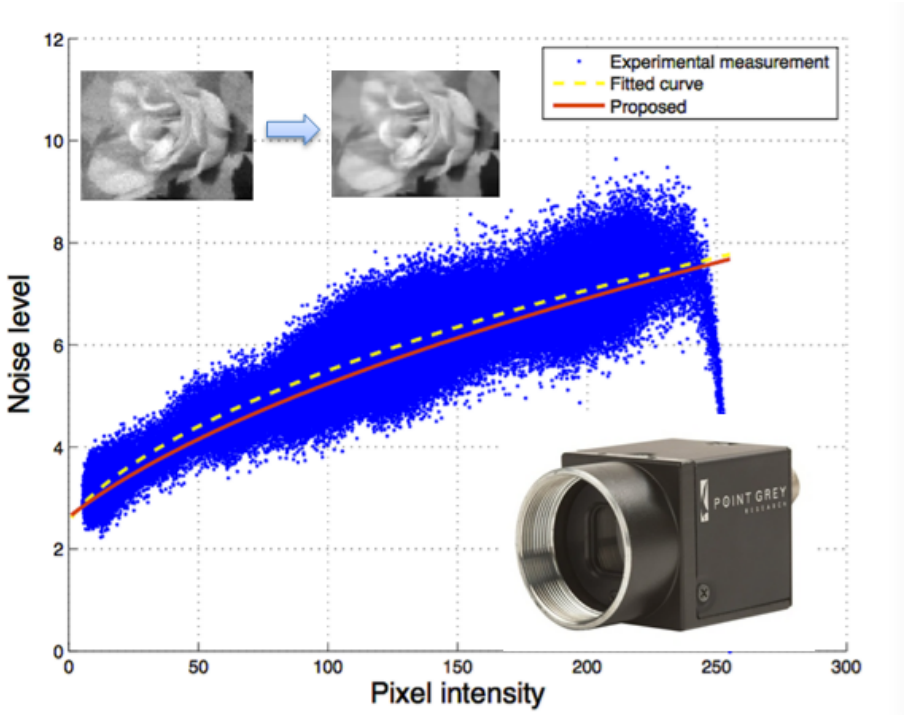


Figure 5.26 – Signal dependent noise model to estimate the noise from the camera in the real world (X. Liu et al., 2013)

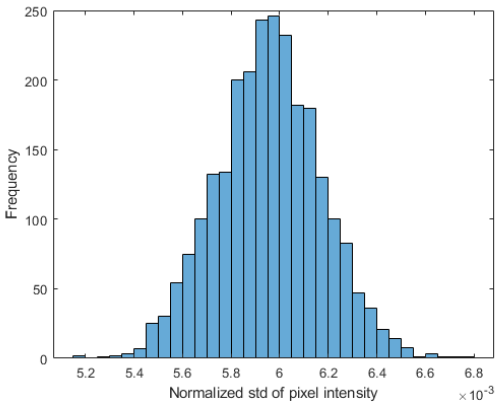
estimation for all motion extraction methods. After the raw video data 'cine' (16-bit) is converted to video format 'mp4' (8-bit), we lose some image details, resulting in a loss of accuracy.

Table 5.11 – RMSE comparison for different methods in the experimental beam test excited with random Gaussian signal for different types of image: 8-bit and 16-bit

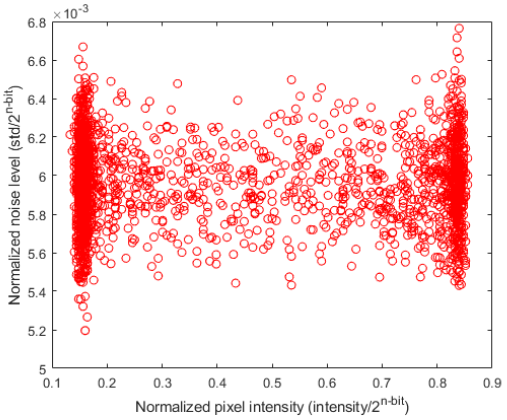
Image format	Motions extraction algorithms						
	Taylor	EQI	EQI-Octant	QSF	POC	UCC	ECC
8-bit	0.0292	0.0284	0.0284	0.0314	0.0284	0.0282	0.0257
16-bit	0.0264	0.0261	0.0261	0.0298	0.0248	0.0257	0.0233

5.3.2 Robustness vs. displacement amplitude

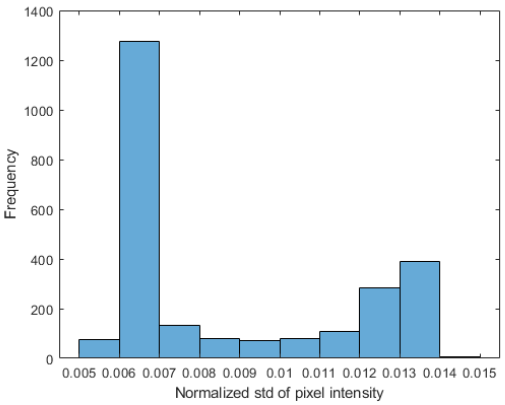
To estimate the robustness of the motion extraction method with respect to displacement amplitude, we present the estimation error vs. displacement amplitude in the experimental condition and simulation condition. Here we mainly introduce the methods



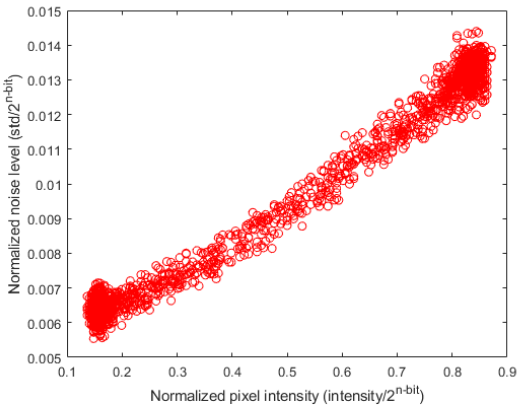
(a) Normalized histogram of pixel intensity std for image with AWGN (additive white Gaussian noise) in the simulation



(b) Normalized std vs. pixel intensity for for image with AWGN (additive white Gaussian noise) in the simulation



(c) Normalized histogram of pixel intensity std for image with AWGN and multiplicative noise in the simulation



(d) Normalized std vs. pixel intensity for for image with AWGN and multiplicative noise in the simulation

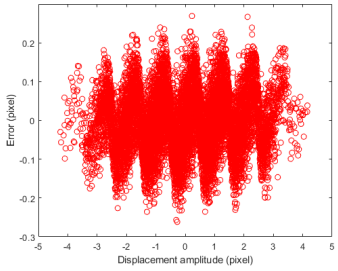
Figure 5.27 – Estimation of noise for simulated image with AWGN and multiplicative noise, $\sigma_\beta = 0.0002$ for the multiplicative noise and $\sigma_N = 1.5$ for the AWGN

Taylor, EQI and QSF, because they all take a strategy of estimation in two steps: pixel shift estimation by ZNCC and subpixel shift estimation. The accuracy vs. displacement is shown in Fig. 5.28. In the simulation test, our proposed method EQI presents the best accuracy and the robustness to the displacement amplitude, as the accuracy is invariant to the subpixel shift. The method Taylor is most sensitive to the displacement amplitude, and it presents the linear relation with the displacement amplitude. When the subpixel shift is close to 0.5 pixel, the estimation error is maximum. This result is reasonable, because the 1st order Taylor approximation is related to the magnitude of subpixel shift and the computation of gradient. When the subpixel is close to 0, we can reach better approximation. The method QSF shows a S-shaped curve which has been confirmed by many authors (Bornert et al., 2012; Su et al., 2018) when they estimate the systematic error. In the experimental result, from the Fig. 5.28(a)(b)(c), it is hard to find the same result as the measuring data in the experimental test is massive (point cloud). After we average the data for each step with 0.1 pixel length in the displacement amplitude as shown in Fig. 5.28(e) we found a similar consistency with the simulation result. The Taylor presents the large estimation error when the subpixel shift amplitude is close to 0.5. The other two methods EQI and QSF represent a similar S-shaped form. When displacement amplitude is large (close to 4 pixels), the measuring error augments obviously.

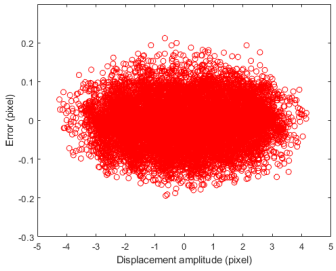
5.4 Application of motion extraction to SHM:

Based on the displacement signals extracted from video image by image processing techniques, we can apply various SHM application such as spectral analysis: natural frequency identification, modal analysis: mode shape identification and detection of deformation. The spectral analysis have been introduced in the previous section, and we have applied a spectra based STFT to identity the natural frequency of the beam. We concluded that the natural frequency identified by motion extraction method has only a slight difference (from 0.0006 Hz to 0.046 Hz) compared with the natural frequency identified by conventional Laser displacement sensor. Furthermore, the noise floor obtained by motion extraction method and LDS is quite similar. The experimental test showed that the displacements extracted by video image were accurate, but a little higher noise floor than the LDS, due to the light condition and the noise in the formation of the image with CCD camera.

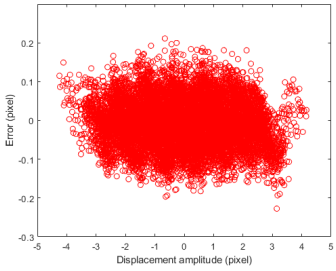
In the following part, we will concentrate on the other SHM applications: modal anal-



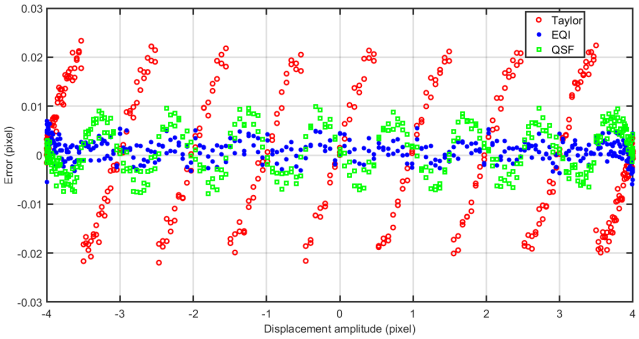
(a) Taylor in experimental test



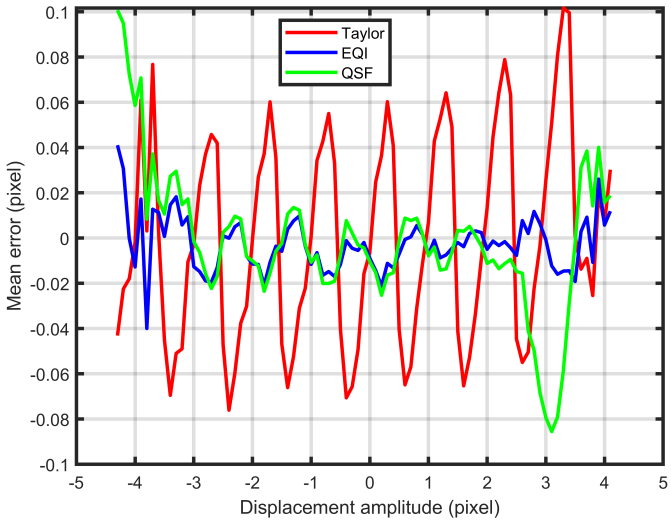
(b) EQI in experimental test



(c) QSF in experimental test



(d) All three methods in the simulation



(e) All three methods in the experimental test with a process averaging the data within every 0.1 pixel units of length in the displacement amplitudes

Figure 5.28 – Accuracy vs. displacement in the experimental test and simulation for methods Taylor, EQI and QSF

ysis and fault detection.

5.4.1 Modal analysis: mode shape identification

In the application, the measurements are made in discrete time. The modal parameters are acquired based on a discrete-time stochastic state-space model presented in the previous section 2.3:

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \omega_k \\ \boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \nu_k \end{cases}$$

where the state vector $\boldsymbol{x}_k$ is composed of the position $z(k)$ and the speed $\dot{z}(k)$ variables, so is the sampled state $\boldsymbol{x}_k \in \mathbb{R}^l$ and l is the system order. $\boldsymbol{y}_k \in \mathbb{R}^{n_s}$ is a vector collecting the outputs of sensors, ω_k and ν_k represent input and output noises respectively. The noise vectors are both unmeasurable vector signals assumed to be zero mean, white noise. The matrix $\boldsymbol{A} \in \mathbb{R}^{l \times l}$ is is the state transition matrix that completely characterises the dynamics of the system by its eigenvalues. The matrix $\boldsymbol{C} \in \mathbb{R}^{n_s \times l}$ is the output matrix which specifies the relation between the internal states and the outside world.

5.4.1.1 Vibration monitoring: modal parameters

As explained in appendix B.1.9, when a state-space model is estimated from sensor data, the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$ can only be estimated up to a similarity transformation by using the subspace identification method (See in appendix B.1.9 for details). The modal parameters are obtained from the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$.

Let λ_j, ϕ_j be respectively the j-th eigenvalue of $\boldsymbol{A}$ and the associated eigenvector, with $j = 1, 2, \ldots, n$. The *mode shape* is defined as

$$\psi_j = \boldsymbol{C}\phi_j \tag{5.1}$$

The discrete time transition matrix $\boldsymbol{A}$ is related to the continuous time transition matrix $\boldsymbol{F}$ by $\boldsymbol{A} = e^{\boldsymbol{F}\tau}$, where τ is the sampling period. Therefore, their corresponding eigenvalues λ_j and μ_j are related by

$$\lambda_j = e^{\mu_j \tau} \tag{5.2}$$

In vibration monitoring, two important parameters, the natural frequency f_j and the

damping ratio ζ_j, are related to μ_j through

$$\mu_j = -2\pi f_j \zeta_j \pm i2\pi f_j \sqrt{1-\zeta_j^2} \tag{5.3}$$

where i is the unit imaginary number, and the sign "$\pm$" is because the complex eigenvalues of a real matrix come always in conjugate pairs. The triples (f_j, ζ_j, ψ_j), for $j = 1, 2, \ldots, n_s$ are called model parameters.

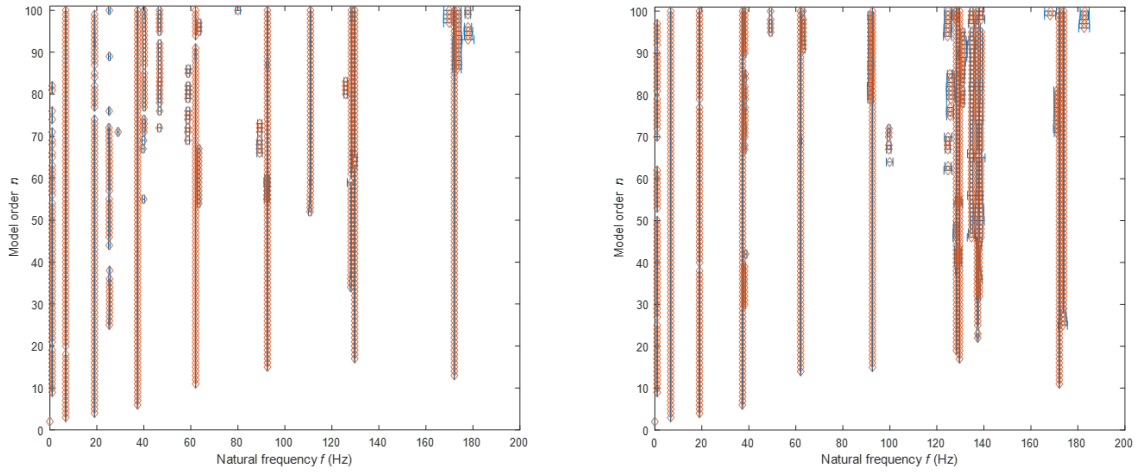
When the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$ is estimated by a subspace identification method, in the SVD step, small singular values are truncated based on some criterion. The resulting $\bar{\boldsymbol{A}}$ matrix leads to a set of complex eigenvalues λ_j, and then to the corresponding natural frequencies and damping ratios f_j, ζ_j.

5.4.1.2 Identification of modal parameters using stochastic subspace identification for the experimental test

In our case, the system (2.50) is observed by our LDS system and video-camera system with measured displacement signals. With the subspace identification algorithm, we can extract the modal parameters from the displacements measured by the video-camera and LDS. Quite often, despite the truncation of small singular values, some of the remaining natural frequencies and damping frequencies and damping ratios f_j, ζ_j are caused by noises. It means that the small singular value truncation in the SVD step of system identification is not sufficiently reliable. The superfluous modes are often removed *manually*, based on the so-called stabilization diagram, which shows the natural frequencies f_j corresponding to different choices of the model order n. Fig. 5.29 shows the results of the stabilization diagram by the displacements measured by LDS and video-camera based EQI. It can be seen that, for the analysis, real physical poles clearly manifest as stable poles from low model order (< 40). However, a high order implies the appearance of spurious modes that may be associated to the noise content of the signals.

From the stabilization diagrams, the modal parameters are chosen. In Table 5.12, the system identification results with their standard deviations σ_f and σ_ζ are summarized. One can observe that between the identification results from the LDS and vision-based measurements, the maximum error is around 0.6% in terms of natural frequency estimation. Figure 5.30 show the mode shapes extracted by LDS and video-camera system. Because of the limitation of the installation of the LDS, we installed 3 laser displacement sensors in the 3 proposed positions (in the top (1/1), in the middle (1/2) and the third

50 cm from the base of the beam (1/4)). Whereas with the application of video-camera based monitoring, we can obtain dense data as each pixel in the image can be considered as a sensor. The image covering the structure of the beam was cropped into nearly 160 ROIs, the displacement signals were extracted from each ROI. As can be seen, we obtain accurate mode shapes compared with the theoretic result. It is noticed that, for large displacement amplitudes (the higher position on the beam), the error on the mode shapes identified is larger than for the small displacement amplitudes (the lower position on the beam)



(a) LDS (b) Video-camera based EQI

Figure 5.29 – The stabilization diagram by the displacements measured by LDS and video-camera based EQI, containing the natural frequencies of the first eight modes with their confidence intervals.

Table 5.12 – Overview of the first eight modes

Mode	Displacement measured by Laser sensor				Vision-based displacement measurements			
	$f(Hz)$	$\sigma_f(Hz)$	$\zeta(\%)$	$\sigma_\zeta(\%)$	$f(Hz)$	$\sigma_f(Hz)$	$\zeta(\%)$	$\sigma_\zeta(\%)$
1	1.0006	0.0071	0.0145	0.8407	0.9982	0.0092	0.0476	0.9035
2	6.7421	0.0099	0.1881	0.1346	6.7419	0.0124	0.2412	0.1644
3	19.0173	0.0127	0.1044	0.0537	18.9960	0.0216	0.2620	0.1547
4	37.3443	0.0182	0.1537	0.0465	37.3406	0.0185	0.1319	0.0443
5	62.0263	0.0272	0.1263	0.0497	62.0245	0.0221	0.1335	0.0391
6	92.5681	0.0168	0.0522	0.0159	92.5742	0.0288	0.0858	0.0456
7	129.6541	0.0288	0.0737	0.0244	129.6795	0.0220	0.0749	0.0217
8	172.1081	0.0255	0.0570	0.0148	172.1133	0.0322	0.0502	0.0164

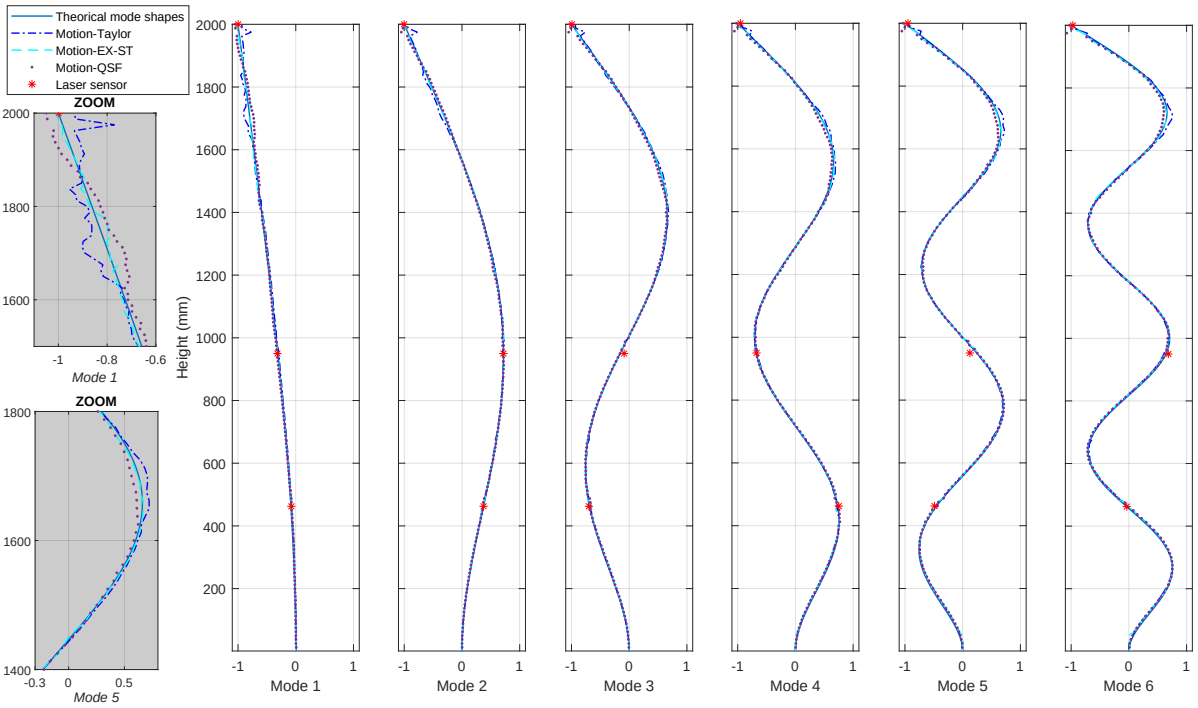


Figure 5.30 – The mode shapes identified by LDS and vision-based measurements

5.4.2 Additive mass detection

For detecting the structural deformation, in our experiments, additive mass, we propose two methods to determine the occurrence of the deformation: 1. identification based on the natural frequencies; 2. identification based on MAC (modal assurance criterion).

The first idea coming in mind is to detect if there is a deformation on the structure

according to the change in the natural frequencies. As shown in Fig. 5.31, based on the modal parameters obtained from SSI on the displacement signals, for each mode, the identified natural frequency is presented as its central point (frequency f) and the interval determined by $[f-2\sigma_f, f+2\sigma_f]$ with 95% confidence interval. With the identified reference interval in the good condition (without deformation), if the identified frequency is still located in this reference interval, then we determine there is no deformation. The whole process of the damage detection based on natural frequencies can be summarized as:

1. Calculate the reference natural frequencies (no damage) f_j^{ref} and their uncertainties σ_{f_j} for each mode j, $j = 1 \dots n_s$ from SSI analysis on the displacement signals.
2. Estimate the actual identified frequency $\hat{f}_j$ and construct the reference interval with 95% confidence interval by $[f_j^{ref} - 2\sigma_{f_j}, f_j^{ref} + 2\sigma_{f_j}]$.
3. Determine the occurrence of damage by comparing if for each mode, the actual estimated frequency $\hat{f}_j$ is well located in the reference interval $[f_j^{ref} - 2\sigma_{f_j}, f_j^{ref} + 2\sigma_{f_j}]$.

In reality, the change of frequency is too subtle, if the damage occur at the node position of the structure. The deformation might not be noticed by analysing the natural frequencies. Thus we introduce the modal assurance criterion (MAC) to realize a supplementary confirmation of the damage detection.

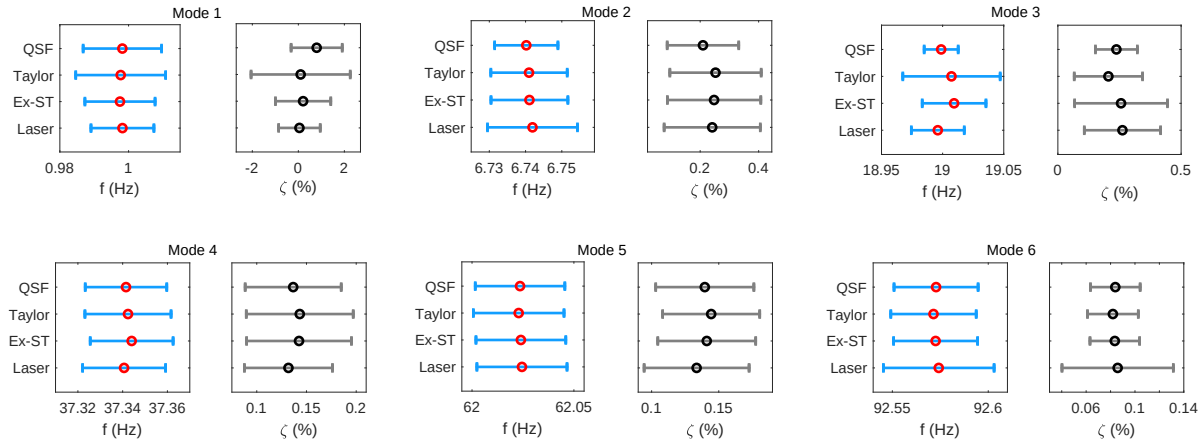


Figure 5.31 – Comparison between the identified natural frequencies (in blue) and damping ratios (in gray) using SSI on the laser sensor displacements and displacements based EX-ST, QSF and Taylor algorithms

We present at first the identified modal parameters from SSI. In Fig. 5.31, we compare the identified natural frequencies and the damping ratios by using the SSI based on the displacements measured by LDS and video-camera. Throughout all the modes, the results obtained by both systems show close correspondence between the LDS displacement signals and displacement signals measured by video-based techniques, regardless of slight variation of the frequency estimation for the third mode, and damping ratios for the first mode. However, each estimation is within the respective confidence interval of the other estimates, proving all the signals are reliable. The system identification results are summarized in Tab. 5.13.

Table 5.13 – Overview of the six modes

Mode	Laser sensor		EX-ST		Taylor		QSF	
	$f(Hz)$	$\zeta(\%)$	$f(Hz)$	$\zeta(\%)$	$f(Hz)$	$\zeta(\%)$	$f(Hz)$	$\zeta(\%)$
1	0.9982± 0.0183	0.0476± 1.8071	0.9975± 0.0204	0.2071± 2.4035	0.9977± 0.0262	0.1003± 4.3103	0.9982± 0.0228	0.7963± 2.2220
2	6.7419± 0.0248	0.2412± 0.3289	6.7411± 0.0212	0.2476± 0.3195	6.7410± 0.0211	0.2525± 0.3125	6.7402± 0.0175	0.2096± 0.2447
3	18.9960± 0.0433	0.2620± 0.3095	19.0094± 0.0521	0.2564± 0.3767	19.0072± 0.0797	0.2050± 0.2771	18.9988± 0.0278	0.2374± 0.1699
4	37.3406± 0.0371	0.1319± 0.0885	37.3440± 0.0370	0.1425± 0.1054	37.3424± 0.0385	0.1432± 0.1075	37.3415± 0.0364	0.1366± 0.0966
5	62.0245± 0.0442	0.1335± 0.0782	62.0240± 0.0441	0.1411± 0.0729	62.0229± 0.0445	0.1444± 0.0723	62.0236± 0.0440	0.1396± 0.0731
6	92.5742± 0.0577	0.0858± 0.0912	92.5725± 0.0438	0.0835± 0.0406	92.5715± 0.0445	0.0819± 0.0416	92.5728± 0.0438	0.0839± 0.0407

Based on the displacement signals, with the mode shapes extracted by SSI algorithm, we can use the modal assurance criterion (MAC) as a modal indicator to decide whether the mode shapes in the monitoring are changed. The MAC between two complex-valued mode shape vectors ψ and ψ_r is calculated as the normalized scalar product of the two vectors and writes (Allemang, 2003):

$$\mathrm{MAC}(\psi, \psi_r) = \frac{|\psi^T \psi_r|^2}{\psi^T \psi \psi_r^T \psi_r} \tag{5.4}$$

where ψ_r is the reference mode shape vector and ψ is another mode shape vector to compare with the reference. The MAC value is real and ranges from 0 to 1. When the MAC estimate is close to 1, it means that the two mode shapes are quite similar, where 1 indicates equal mode shapes.

In the laboratory, a deformation test was conducted. Based on the beam test condition, to generate the deformed model, we added a magnet with 44 g to the specific position of the beam for test with small displacement (less than 1 pixel) and a different magnet with 79 g for the test with large displacement (maximum displacement amplitude $\approx$ 4 pixel) as shown in Fig. 5.32.

To simplify the notation of the experimental condition, we name the experimental condition according to its position of the beam. E.g., P2000 represents the condition where we add the magnet at the position 2000 mm (the top) from the base of the beam. 9 experiences were designed : Mass Free (MF), P2000, P1803, P1507, P1277, P949, P686, P456 and P280. The MAC is calculated between the mode shapes of the mass-free condition (reference mode shapes) and those of deformed condition such as P2000, P1803, etc.. In the experimental tests, the result is shown in Tab. 5.14 from FEM (finite element model) analysis in the simulation.

Table 5.14 – MAC between the reference mode shape obtained from FEM (finite element model) analysis in the simulation tests

Condition	MAC values					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
$P2000$	0.999996	0.993574	0.988291	0.983755	0.980288	0.977836
$P1803$	0.999999	0.998984	0.999773	0.999962	0.999283	0.997411
$P1507$	0.999999	0.999939	0.998921	0.997616	0.999452	0.998785
$P1277$	0.999999	0.999546	0.998771	0.999957	0.997059	0.996809
$P949$	0.999999	0.999677	0.999783	0.998016	0.998776	0.996312
$P686$	0.999999	0.999912	0.998453	0.999715	0.997940	0.995227
$P456$	0.999999	0.999991	0.999498	0.997185	0.997033	0.999730
$P260$	0.999999	0.999999	0.999967	0.999441	0.996738	0.990765

The MAC values in Tab. 5.14 are all close to 1. We can hardly determine if there is a significant change of the mode shapes. In the case that all MAC values are close to 1, it is not easy to distinguish whether the model shape are changed as shown in Fig. 5.31.

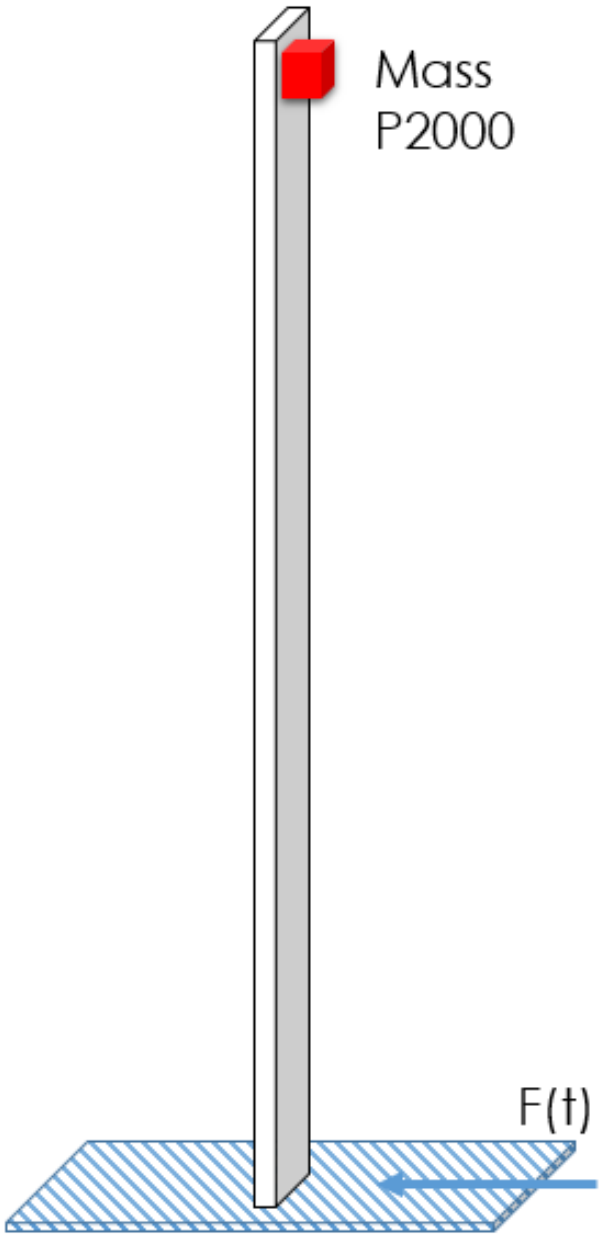


Figure 5.32 – The deformed model with the adding weight of the magnet

In order to realize the monitoring goal, we propose to follow an approach that relies on the MAC with uncertainty quantification with 95% confidence intervals (CI) considering that the scaled and shifted χ^2 approximation of the MAC indicator is applied (Greś et al., 2021). Then we compare the MAC values $\mathrm{MAC}(\psi^j, \psi_r^j)$ with the quantile value q_{MAC}^j for determining the change of j-th mode shapes. The process is described as following:

- If $\mathrm{MAC}(\psi^j, \psi_r^j) \geq q_{\mathrm{MAC}}^j$, the estimated MAC is in the confidence interval, meaning that there is no significant change in the mode shapes.
- If $\mathrm{MAC}(\psi^j, \psi_r^j) < q_{\mathrm{MAC}}^j$, the estimated MAC is beyond the confidence interval, meaning that the damage has led to the significant change in the mode shapes.

All the results are shown in Fig. 5.34, which shows the detection of change of mode shapes based on natural frequency and MAC value, and the result of the normalized frequencies in relation with the mass position of the beam in the simulation is shown in Fig. 5.33. The solid points represent that the mode shape is not changed compared with the q_{MAC}^j. The hollow points represent that the mode shape is changed compared with the q_{MAC}^j. It is noticed that by the application of the 95% confidence interval with the reference identified frequency in the mass-free condition, we can preliminarily determine if the mode shapes are changed or not, but a supplementary process for the confirmation of the change of mode shapes is applying the MAC with uncertainty quantification represented by solid points and hollow points. Compared with the result of simulation for adding additional mass at the different positions of the beam in Fig. 5.33, for example, in the mode 1 in Fig. 5.34, the identified frequencies (the points) follow the blue curve (first mode) presented in Fig. 5.33, so as the other 5 modes. We concluded that the consistency between the simulation and the experimental has been verified.

The following comments on these results shown in Fig. 5.34 can be made:

1. There is no significant change of the mode shapes for mode 1. The main reason are considered as:
   (a) Weak excitation of the first mode due to the limitation of vibrating system. To generate small displacement, the shake table is controlled with a limited band and the low frequency of the band is close to the low natural frequency.
   (b) Limited recording time for identifying the low frequency mode. Our measurement system record the data with the duration about 30 seconds, which may be not enough for analysing the low frequency of modes.

(c) The added mass, used to simulate harmful damage on the structure, had not enough weight to make significant changes in the first mode shape.

2. For 3rd and 5th mode, all the change of mode shapes are correctly labelled.

3. Both detection based on natural frequencies and the MAC with uncertainty quantification is able to detect the deformation. Because if one of all the modes has been found with a distinct change, we can tell the deformation (or damage) has occurred on the structure. We can also combine the two methods in the application of the detection of the deformation. Once the identified natural frequency is changed compared with the reference natural frequency, a supplementary validation by the MAC with uncertainty quantification is made to confirm the judgement.

With this process, we can carry out the monitoring goal based on natural frequencies and the MAC with uncertainty quantification by using SSI analysis on the displacement signals. The damage detection due to the deformation or corrosion can be realized in the SHM application.

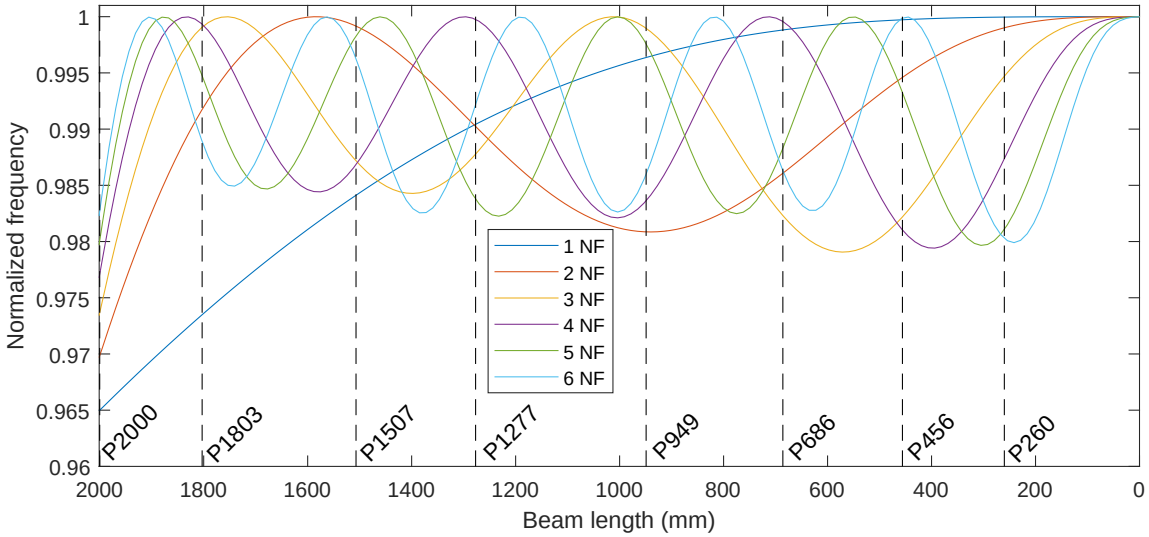
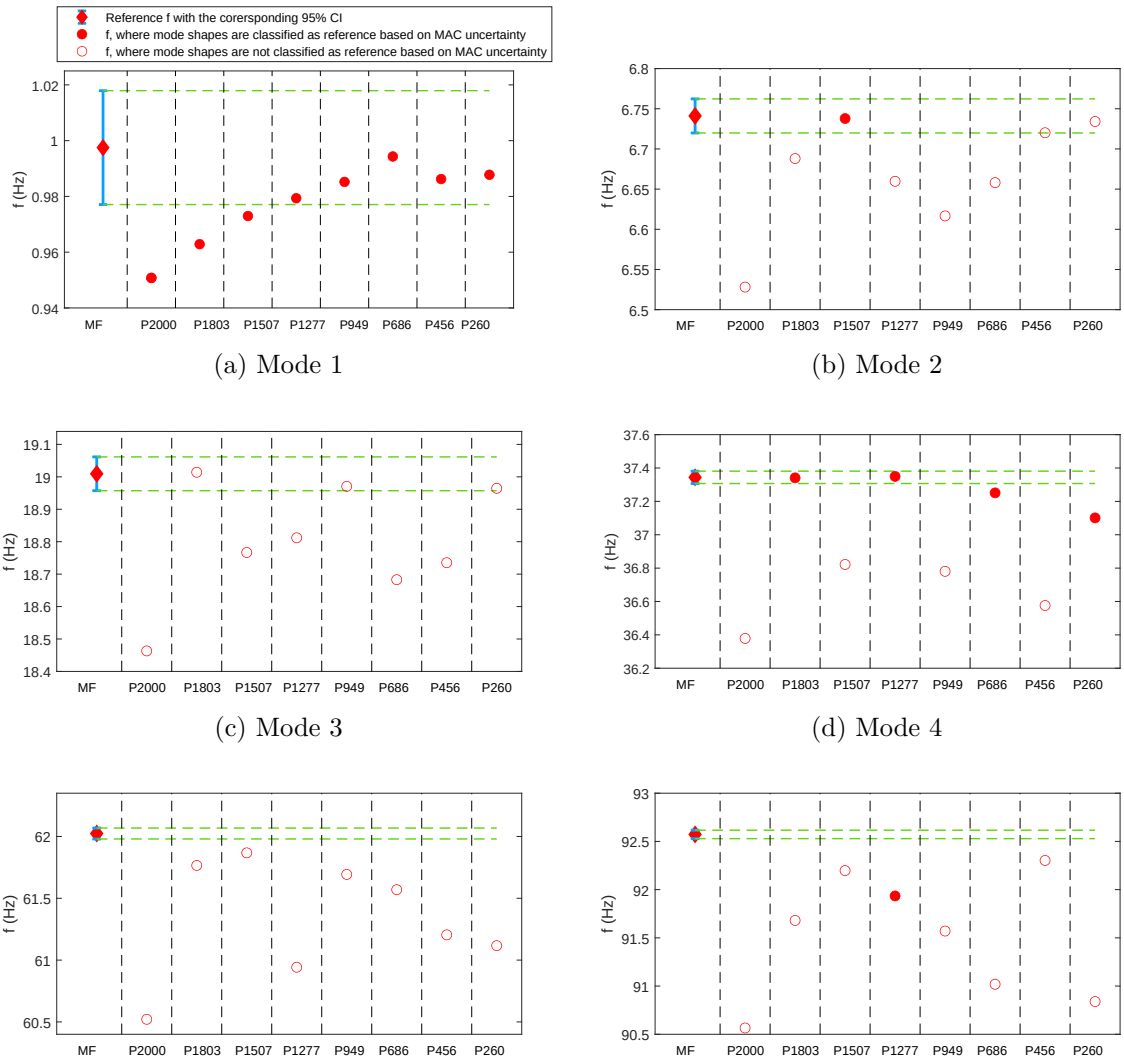


Figure 5.33 – Variations of first six natural frequencies in relation with the mass position in the simulation



(a) Mode 1

(b) Mode 2

(c) Mode 3

(d) Mode 4

(e) Mode 5

(f) Mode 6

Figure 5.34 – The detection of the change of the first 6 mode shapes in all the experimental condition: Mass-free, P2000, P1803, P1507, P1277, P949, P686, P456 and P280 based on the MAC uncertainty (Greś et al., 2021). The solid points represent that the mode shape is not changed compared with the q_{MAC}. The hollow points represent that the mode shape is changed compared with the q_{MAC}

5.5 Conclusion

In this chapter, the field experiments were conducted to validate the motion extraction algorithms and the stochastic subspace identification (SSI) based on the displacement signals and subspace-based fault detection test were pointed out. Two types of experiments

were planed including shake-table test and cantilever beam test.

The displacement signals were correctly measured by the vision-based system compared with the displacements measured by LDS. For the video-based system, several scenarios including the outdoor condition target-free condition were validated to ensure the general purpose of video-based monitoring. A good consistency has been verified between the simulation test and experimental test, concerning the noise model and robustness vs. displacement amplitude. Our proposed method EQI had a great improvement on the measurement accuracy compared with Taylor method in the experimental test and could reach nearly the same level accuracy as other reference subpixel shift estimation methods. With spectral analysis, the natural frequencies could be identified and the noise floor obtained by the video-based system is slightly lower than that by the LDS. The verification measurement showed that the video-based measurements were accurate and could be deployed for the following SHM application.

With the SSI process, the modal parameters were correctly identified with the displacement signals measured by LDS and video-camera. Based on the change in the natural frequencies and MAC with uncertainty quantification, the detection of the deformation was validated.

Chapter 6

Summary, discussion and working perspectives

This thesis involved a multi-disciplinary research which has merged computer vision, signal processing, modal analysis, SHM and experimental design. It aimed at developing video-based techniques to perform non-contact vibration-based SHM for civil engineering infrastructures. This study has relied on image processing techniques allowing the motion-extraction of mechanical structures from video flows.

6.1 Summary

We first introduced in Chapter 1 the existing techniques to perform vibration-based SHM including point-like contact-based methods, i.e., accelerometer, strain gauge and GNSS, and non-contact measurement techniques such as laser (LDS), visible video-based methods and infrared techniques. Video-based methods was selected as a promising recent techniques owing to its handy use, low cost and large-scale information.

Chapter 2 presented the state-of-the-art image processing techniques that enable to estimate the motion between two successive images. We focus on the two-step correlation based techniques. The first step computed the pixel shift between two images with correlation-based techniques. Different subpixel shift computation techniques were then briefly presented, which rely on either the grey level variations or the local phase variations between the reference and the ongoing images. Among others, the optical flow, the gradient-based Taylor method, the quadratic surface fitting (QSF), and the the steerable filter are the most well-known.

In Chapter 3, two new processing methods, entitled "efficient quadrant interpolation" (EQI) and its improved version "efficient octant interpolation" (EQI-Octant), were detailed. They are expected to improve the motion-extraction compared to the selected methods in Chapter 2. Besides, this chapter presented the extension of the steerable fil-

ter method (EX-ST) to multi-pixel motion. Finally, some critical points of the quadratic surface fitting method were detailed.

Chapter 4 presented the performance assessment of the proposed video-based motion-extraction methods on different simulated image data flows. The simplified process to simulate weakly texture images was described at first. The performance of each step of the correlation-based motion extraction technique is separately evaluated. The assessment of the selected motion-extraction methods was performed from simulated synthetic images and the existing DIC challenge dataset, which correspond to weak texture and strong texture images (speckled images), respectively. Subpixel displacement accuracy and large displacement accuracy (larger than one pixel) were successively evaluated. The RMSE computation between the estimated and the true motion over the time record is used to evaluate the performance of algorithms. For convenience, RMSE is converted to pixel fraction. The RMSE behaviour enables to conduct sensitivity analysis in relation with the motion magnitude, the noise level, the slope rate of the contour, the image contrast and the computing time. The proposed EQI method showed the smallest RMSE in noiseless conditions, while most selected algorithms provide similar RMSE at high SNR. QSF and ECC methods are found more robust than others in noisy conditions.

In Chapter 5, a first experimental test has been conducted on a vertical steel beam in both indoor and outdoor conditions. The motion of the vibrating beam was estimated by the selected motion-extraction algorithms and compared with co-located laser displacement sensors. The stochastic subspace identification (SSI) was then performed on the estimated motion data of the beam to compute the corresponding natural resonant frequencies and mode shapes. At last, a SHM experiment was set up to detect changes in the mode shapes from image-based technique. A few magnets were set along the beam to artificially modify the beam's mode shapes. The image-based SHM analysis succeeded in fault detection from both the variations in the beam natural frequencies while the modal assurance criterion (MAC) enables to detect the changes in the mode shapes.

6.2 Main findings

A selection of main findings of this thesis are highlighted hereafter:

— Based on the literature review and analysis of some widespread two-step correlation-based techniques, two new interpolation methods, namely, EQI for Efficient Quadrant Interpolation and octant interpolation (EQI-Octant), have been proposed to

perform motion extraction at the subpixel shift level. The proposed methods take advantage of a preliminary step to determine the subpixel shift direction. This improves the overall accuracy in motion extraction compared to the conventional gradient-based Taylor method.

— The performance of the proposed correlation-based motion-extraction algorithms have been assessed on both simulated image and field image flows. For the sake of clarity, the performance is detailed at both the pixel and the subpixel steps before reporting on the overall performance in terms of pixel RMS. The importance of the subpixel step for the quality of the estimated motion is clearly demonstrated.

  The performance assessment has also allowed illustrating the influence of the noise model, blurring effect, displacement amplitude, experimental set-up, indoor/outdoor conditions, camera parameters and camera types.

— Motion extraction algorithms have enabled to perform the modal analysis of the mechanical structure from video image data flows. Full field image-based SHM processing techniques have been tested to estimate the mode shapes of a mechanical structure and to demonstrate the feasibility of video image-based SHM.

6.3 Discussion on video-based technology for SHM

This section proposes to review the advantages and limitations of the video-based measurement technology for SHM with regards to the conventional point-like sensor techniques such as accelerometer and LDS.

6.3.1 Some advantages

Non-contact measurement:

The thesis has shown that video-based SHM technique represents a handy non-contact measurement technique, which does not require the installation of point-like sensors on the mechanical structure to be monitored.

This may represent a special advantage for either monitoring dangerous or lightweight structures. For the first case, people may be prevented to set up the measuring sensors on the damaged structure. For lightweight structures, the additional weight that represent point-like sensors with supply cables can disturb the modal response of the mechanical system and reduce the SHM sensitivity.

High spatial data density:

By contrast to the point-like sensors, i.e., accelerometers and LDS, the vision-based device can potentially image the full field of view on the structure. A selection of pixel in the image serves as sensors to collect motion signals on the structure, potentially providing a high spatial density of information. In practice, only edge pixels carry motion information, and some segmentation may be required beforehand to select the useful ROIs with useful pixels.

This large quantity of pixel data is particularly well suited to perform data-driven SSI algorithm for SHM application. More data enable to detect the modal changes due to damages on the structure with higher sensitivity.

Cost-effective solution:

The traditional point-like sensors are limited by their expense and the number of installation on a structure, an additional optimisation process is required to install the sensors in the suitable position, whereas the set up of a single camera is usually less expensive and easier.

6.3.2 Some limitations

Line-of-sight areas of the structure :

When the camera is set far away from the structure, video-based technique can only perform motion analysis of the zones which are in straight line of sight with the camera. These zones are not always the most sensitive areas of the structure. For example, the underside of bridge decks may be better suited to SHM purposes.

Larger noise floor :

Video-based techniques depict a higher noise floor compared with the point-like sensor such as LDS or accelerometers. In practice, many noise sources and measurement uncertainties may contribute to the noise floor : image bit-depth, camera vibrations, camera parameters, outdoor conditions, bias on the motion extraction algorithms, etc.

Camera vibrations :

The accuracy of the estimated motion may be limited in indoor/outdoor conditions by the camera's own vibration. Camera vibrations can originate from surrounding sources, and the inner cooling fan of the camera if any. In outdoor conditions, the wind can also slightly vibrate the camera .

Managing outdoor conditions:

In outdoor conditions, the image brightness may be greatly affected by the changeable sky conditions. To avoid image saturation, the diaphragm opening of the camera must be tuned according to the passing of clouds.

It was found that the cross-correlation methods to compute the pixel-level shift motion is quite robust to the changing light conditions. By contrast, the subpixel shift estimation methods is very sensitive to the latter effects.

Another limitation in outdoor conditions may appear when imaging a structure at long range. Depending on the atmosphere conditions, random fluctuations may appear on the refractive index of the air in the line-of-sight. This may result in blurring effect on the video images with random time fluctuations (Feng et al., 2015).

6.4 Working perspectives

The thesis has primarily focused on image-based vibration measurement techniques. Performance have been evaluated on simulated and field image flows. Based on this experience, we have identified hereafter some possible future working issues.

6.4.1 Image-based SHM

The first issue is to improve the measurements accuracy of the video-based motion extraction methods. This may help to perform SHM at longer range and/or from low-cost camera. Within this scope, it is proposed to study the **iterative approach** which is suggested in the literature (Rais et al., 2015; Rytter, 1993; Tuan, 2006). This solution further improves the performance of gradient-based motion-extraction methods and can be then tested on the proposed EQI and octant methods.

The second issue would be to achieve **real time image-based SHM** diagnosis, as it is already reported with conventional accelerometers (Abdullahi et al., 2019). This requires to speed up motion-extraction algorithms, and to define some trade-off between the required motion accuracy and the image size to be processed.

Finally, the image-based SHM technique may be further exploited in the future by **locating the defect along the structure**. This consists in relating the changes in mode shapes due to the defect with the theoretical mechanical response of the structure. The video-based solution may be compared to the conventional point-like sensor solution.

6.4.2 Field experiments :

The laboratory experiment which has been conducted in this thesis with controlled conditions is expected to help at sizing further field measurements on a specific bridge within the next 3 years. The main issue to be worked on consists in managing outdoor conditions by accounting for varying light conditions and camera vibrations.

— **Varying light conditions** have been shown to reduce the performance of image-based motion-extraction algorithms. Some existing technological solutions may alleviate the latter effect, e.g., high dynamic range (HDR) camera, automatic gain control and neutral density filter. The motion-extraction from successive images (i.e., with a moving reference image) may represent an alternative processing to be tested.

— **Camera vibrations:** A first outdoor experiment with the cantilever beam has shown that the wind may induce camera vibrations. To mitigate this discrepancy, different solutions have to be tested including a wind screen and some compensation algorithm. According to the discussion in the previous section, the latter solution requires either to equip the camera with sensors (J. G. Chen et al., 2015) or to find a steady object within the image background.

Imaging devices :

The thesis has also drawn some other promising and further future issues:

— **3D motion extraction:** The thesis has developped a 2D motion signals were extracted. Some image-based motion extraction can be extended to two or more cameras to fulfill the stereo vision algorithm and photogrammetry, e.g., (Lundstrom et al., 2012; Sabato et al., 2020). Nevertheless, this imaging technique is likely to be limited to short working range, less than 10 meter roughly. For image-based SHM applications at longer range, time of flight (ToF) camera (Hansard et al., 2012) may be used but, provided that the millimetric accuracy is matched to the application. Laser-based scanning technology, namely, light detection and ranging (LiDAR), laser Doppler vibrometer (LDV) and terrestrial laser scanner (TLS), may be a powerful alternative imaging technique which may be used as the reference SHM technique to be compared to the low-cost video-based techniques.

— **UAV video-based techniques**. A first experiment with a drone (unmanned aerial vehicle) has been conducted during the PhD. Compared to the fixed camera, the

recorded image flows from UAV were found highly disturbed at low frequency, i.e., below 2 Hz, by the drone flight control process, as already reported in (Kang & Cha, 2018). A motion compensation algorithm must be applied to recover better accuracy.

Appendix A

Additional information on image processing

A.1 Summary of the performances for all the implemented motion extraction algorithms

We summarize the performances of all the motion extraction methods implemented during this thesis project in Tab. A.1.

Table A.1 – Summary of the estimation accuracy (RMSE) of all the implemented motion extraction algorithms based on simulation and experimental tests

	Data	Motion extraction algorithms (RMSE in pixel)						
		Taylor	EQI	EQI-Octant	QSF	POC	UCC	ECC
Simulation	Our simulation	0.0206	**0.0045**	**0.0042**	0.0065	0.0067	0.0072	0.0064
	CCD model	0.0143	0.0049	0.0048	0.0050	**0.0040**	0.0051	0.0056
Experiment (shake-table)	Sinusoidal excitation (S)	0.0270	0.0208	0.0206	**0.0195**	0.0289	0.0206	0.0216
	Sinusoidal excitation (L)	0.0218	0.0217	0.0217	**0.0192**	0.0266	**0.0194**	0.0203
	Random excitation (S)	0.0380	**0.0220**	0.0293	0.0234	0.0545	**0.0219**	0.0273
Experiment (Cantilever beam)	Hammer strike (S)	0.0457	0.0437	0.0475	0.0450	0.0478	0.0479	**0.0399**
	Hammer strike (L)	0.0925	**0.0890**	**0.0890**	0.0902	0.0944	0.0915	**0.0892**
	Random1 (Position 1/1)	0.0506	**0.0396**	**0.0396**	0.0413	0.0480	0.0406	0.0405
	Random1 (Position 1/2)	0.0530	0.0344	0.0344	**0.0302**	0.0536	0.0424	0.0361
	Random1 (Position 1/4)	0.0566	**0.0188**	**0.0188**	0.0204	0.0251	0.0197	0.0275
	Random2 (Position 1/1)	0.0444	**0.0368**	**0.0368**	0.0389	0.0380	0.0370	0.0378
	Random2 (Position 1/2)	0.0639	**0.0469**	**0.0469**	**0.0471**	0.0527	**0.0470**	0.0549
	Random2 (Position 1/4)	0.0230	0.0190	0.0190	0.0199	0.0188	**0.0180**	0.0189
	Random3 (Position 1/1)	0.0444	**0.0368**	**0.0368**	0.0389	0.0380	0.0370	0.0378
	Random3 (Position 1/2)	0.0580	**0.0356**	**0.0356**	**0.0356**	0.0430	0.0351	0.0457
	Random3 (Position 1/4)	0.0373	**0.0280**	**0.0280**	0.0287	0.0283	**0.0279**	0.0306
	Occurrence of the best accuracy	0	10	9	5	1	5	2
	Mean RMSE	0.0432	**0.0314**	0.0321	0.0319	0.0380	0.0318	0.0338
	Avg. Rank	6.5000	**2.1250**	2.9375	3.8125	5.2500	3.1875	4.1875

A.2 An ellipse simulation model by (Chojnacki & Szpak, 2019)

In (Chojnacki & Szpak, 2019), the authors simulated synthetic images of an ellipse target based on an ellipse image formation model. This model mainly consists of blurring (point-spread, through the convolution with a Gaussian kernel), the spatial discretization stage considering that each pixel may cover the ellipse target and the background, photon counting and gray level quantization. The flowchart of the image formation model is shown in Fig. A.1

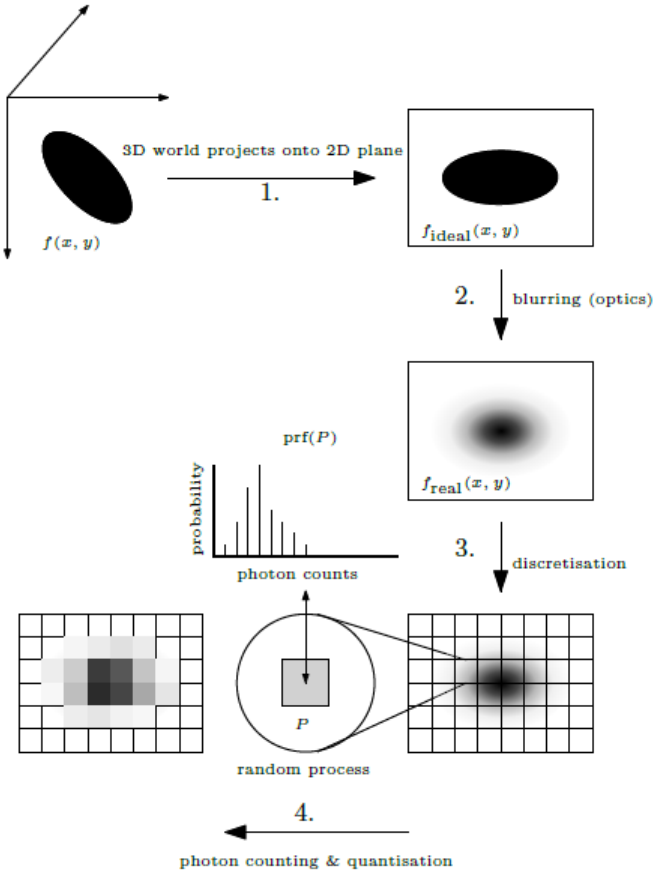


Figure A.1 – An overview of an image formation model. In the first step, the 3D world is projected onto a 2D image plane resulting in an ideal geometric image. The second stage accounts for imprecisions such as geometric distortions and blurring. The third step imposes a grid of pixels, introduces a statistical model for the number of photons that hit each pixel, and accounts for the Poisson noise. The final step models the loss of information due to quantization of the grey-levels. (Chojnacki & Szpak, 2019)

A.3 Expression of X and F matrices for the EQI method in each of the four quadrants or the 8 octants

$$\mathbf{X}_I = \begin{bmatrix} 1 & m & n \\ 1 & m & n+1 \\ 1 & m+1 & n \\ 1 & m+1 & n+1 \end{bmatrix} \qquad \mathbf{F}_I = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n+1,t_1) \\ I_s(m+1,n,t_1) \\ I_s(m+1,n+1,t_1) \end{bmatrix}$$

$$\mathbf{X}_{II} = \begin{bmatrix} 1 & m & n \\ 1 & m & n+1 \\ 1 & m-1 & n \\ 1 & m-1 & n+1 \end{bmatrix} \qquad \mathbf{F}_{II} = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n+1,t_1) \\ I_s(m-1,n,t_1) \\ I_s(m-1,n+1,t_1) \end{bmatrix}$$

$$\mathbf{X}_{III} = \begin{bmatrix} 1 & m & n \\ 1 & m & n-1 \\ 1 & m-1 & n \\ 1 & m-1 & n-1 \end{bmatrix} \qquad \mathbf{F}_{III} = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n-1,t_1) \\ I_s(m-1,n,t_1) \\ I_s(m-1,n-1,t_1) \end{bmatrix}$$

$$\mathbf{X}_{IV} = \begin{bmatrix} 1 & m & n \\ 1 & m & n-1 \\ 1 & m+1 & n \\ 1 & m+1 & n-1 \end{bmatrix} \qquad \mathbf{F}_{IV} = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n-1,t_1) \\ I_s(m+1,n,t_1) \\ I_s(m+1,n-1,t_1) \end{bmatrix}$$

$$\mathbf{X}_{IV} = \begin{bmatrix} 1 & m & n \\ 1 & m & n-1 \\ 1 & m+1 & n \\ 1 & m+1 & n-1 \end{bmatrix} \qquad \mathbf{F}_{IV} = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n-1,t_1) \\ I_s(m+1,n,t_1) \\ I_s(m+1,n-1,t_1) \end{bmatrix}$$

$$\mathbf{X}_V = \begin{bmatrix} 1 & m & n \\ 1 & m+1 & n \\ 1 & m+1/2 & n+1 \\ 1 & m+1/2 & n-11 \end{bmatrix} \qquad \mathbf{F}_V = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m+1,n,t_1) \\ \frac{I_s(m+1,n+1,t_1)+I_s(m,n+1,t_1)}{2} \\ \frac{I_s(m+1,n-1,t_1)+I_s(m,n-1,t_1)}{2} \end{bmatrix}$$

$$\mathbf{X}_{VI} = \begin{bmatrix} 1 & m & n \\ 1 & m & n+1 \\ 1 & m+1 & n+1/2 \\ 1 & m-1 & n+1/2 \end{bmatrix} \qquad \mathbf{F}_{VI} = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n+1,t_1) \\ \frac{I_s(m+1,n,t_1)+I_s(m+1,n+1,t_1)}{2} \\ \frac{I_s(m-1,n,t_1)+I_s(m-1,n+1,t_1)}{2} \end{bmatrix}$$

$$\mathbf{X}_{VII} = \begin{bmatrix} 1 & m & n \\ 1 & m-1 & n \\ 1 & m-1/2 & n+1 \\ 1 & m-1/2 & n-1 \end{bmatrix} \qquad \mathbf{F}_{VII} = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m-1,n,t_1) \\ \frac{I_s(m,n+1,t_1)+I_s(m-1,n+1,t_1)}{2} \\ \frac{I_s(m,n-1,t_1)+I_s(m-1,n-1,t_1)}{2} \end{bmatrix}$$

$$\mathbf{X}_{VIII} = \begin{bmatrix} 1 & m & n \\ 1 & m & n-1 \\ 1 & m+1 & n-1/2 \\ 1 & m-1 & n-1/2 \end{bmatrix} \qquad \mathbf{F}_{VIII} = \begin{bmatrix} I_s(m,n,t_1) \\ I_s(m,n-1,t_1) \\ \frac{I_s(m+1,n,t_1)+I_s(m+1,n-1,t_1)}{2} \\ \frac{I_s(m-1,n,t_1)+I_s(m-1,n-1,t_1)}{2} \end{bmatrix}$$

Appendix B

Subspace method for state-space model estimation

B.1 Mechanical system state-space model

B.1.1 Mechanical system motion equation

We first introduce a linear time-invariant (LTI) model of a mechanical structure with s degrees of freedom (DOF):

$$\boldsymbol{M}\,\ddot{\boldsymbol{z}}(t) + \boldsymbol{D}\,\dot{\boldsymbol{z}}(t) + \boldsymbol{K}\,\boldsymbol{z}(t) = \boldsymbol{u}(t) \tag{B.1}$$

where t denotes the continuous time, the matrices $\boldsymbol{M}, \boldsymbol{D}, \boldsymbol{K} \in \mathbb{R}^{s\times s}$ denote mass, damping and stiffness matrices, respectively. The vectors $\boldsymbol{z}(t) \in \mathbb{R}^s$ and $\boldsymbol{u(t)} \in \mathbb{R}^s$ denote the continuous-time displacements and the unknown external forces, respectively.

B.1.2 The state equation and its general solution

Define

$$\boldsymbol{x} \triangleq \begin{bmatrix} \boldsymbol{z} \\ \dot{\boldsymbol{z}} \end{bmatrix}, \boldsymbol{F} \triangleq \begin{bmatrix} 0 & \boldsymbol{I} \\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{D} \end{bmatrix}, \boldsymbol{G} \triangleq \begin{bmatrix} 0 \\ \boldsymbol{M}^{-1} \end{bmatrix} \tag{B.2}$$

Then Eq. (B.1) is transformed to the state equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}\boldsymbol{x}(t) + \boldsymbol{G}\boldsymbol{u}(t) \tag{B.3}$$

in which $\boldsymbol{x}(t)$ is called the *state vector*.

Given the initial state $\boldsymbol{x}(t_0)$ at the initial instant t_0, then the solution of (B.3) is

$$\boldsymbol{x}(t) = e^{\boldsymbol{F}(t-t_0)}\boldsymbol{x}(t_0) + \int_{t_0}^{t} e^{\boldsymbol{F}(t-s)}\boldsymbol{G}\boldsymbol{u}(s)\,ds, \tag{B.4}$$

B.1.3 The free motion trajectory

The free motion of the mechanical system (B.1), or in the state equation form (B.3), is governed by:

$$\dot{\boldsymbol{x}}_{\mathbf{0}}(t) = \boldsymbol{F}\boldsymbol{x}_{\mathbf{0}}(t) \tag{B.5}$$

Its trajectory is given by the first term of $\boldsymbol{x}(t)$ in (B.4), corresponding to $\boldsymbol{u}(t) = 0$, namely

$$\boldsymbol{x}_{\mathbf{0}}(t) = e^{\boldsymbol{F}(t-t_0)}\boldsymbol{x}(t_0) \tag{B.6}$$

In order to gain some insights into this free motion trajectory, only the case of diagonalizable matrix $\boldsymbol{F}$ is considered here (while most square matrices are diagonalizable, the analysis of the non diagonalizable case remains possible with the Jordan form). Let Λ be the diagonal matrix containing the $n = 2s$ eigenvalues μ_j of the $n \times n$ matrix $\boldsymbol{F}$:

$$\Lambda = \begin{bmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \dots & \mu_n \end{bmatrix} \tag{B.7}$$

Then

$$\boldsymbol{F} = \boldsymbol{T}\Lambda\boldsymbol{T}^{-1}, \tag{B.8}$$

where $\boldsymbol{T}$ is a square matrix whose columns are eigenvectors of $\boldsymbol{F}$. Thanks to important properties of the matrix exponential function,

$$\begin{aligned} \boldsymbol{x}_{\mathbf{0}}(t) &= e^{\boldsymbol{T}\Lambda\boldsymbol{T}^{-1}(t-t_0)}\boldsymbol{x}(t_0) && \text{(B.9)} \\ &= \boldsymbol{T}e^{\Lambda(t-t_0)}\boldsymbol{T}^{-1}\boldsymbol{x}(t_0) && \text{(B.10)} \\ &= \boldsymbol{T}\begin{bmatrix} e^{\mu_1(t-t_0)} & 0 & \dots & 0 \\ 0 & e^{\mu_2(t-t_0)} & \dots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \dots & e^{\mu_n(t-t_0)} \end{bmatrix}\boldsymbol{T}^{-1}\boldsymbol{x}(t_0) && \text{(B.11)} \end{aligned}$$

This result shows that each component of the vector $\boldsymbol{x}_{\mathbf{0}}(t)$ is a linear combination of the (complex) exponential functions $e^{\mu_j(t-t_0)}$. While the imaginary part of μ_j (if any) determines the frequency of an oscillating term, the (negative) real part of μ_j specifies the rate of exponential decay of this term.

In this analysis, the real vector $\boldsymbol{x_0}(t)$ is decomposed into the sum of complex terms (Λ and $\boldsymbol{T}$ are both complex in general). In fact, grouping the terms corresponding to each conjugate pair of eigenvalues would lead to a real term (the complex eigenvalues of the real matrix $\boldsymbol{F}$ always come in conjugate pairs).

B.1.4 Eigenvalues and eigenvectors

Let $\boldsymbol{M}, \boldsymbol{D}, \boldsymbol{K}$ be defined as in E.q (B.1). If a (real or complex) number μ and a vector $\nu \neq 0$ satisfy

$$\mu^2 \boldsymbol{M}\nu + \mu \boldsymbol{D}\nu + \boldsymbol{K}\nu = 0, \tag{B.12}$$

then μ and ν are respectively an eigenvalue and an eigenvector of (B.1). The square matrix $\boldsymbol{F}$ has also eigenvalues μ_j and eigenvectors ζ_j satisfying

$$\boldsymbol{F}\zeta_j = \mu_j \zeta_j. \tag{B.13}$$

To establish a relationship between the above two definitions of eigenvalues and eigenvectors, Eq. (B.12) is rewritten as:

$$-\boldsymbol{M}^{-1}\boldsymbol{K}\nu - \mu \boldsymbol{M}^{-1}\boldsymbol{D}\nu = \mu^2 \nu, \tag{B.14}$$

then

$$\begin{bmatrix} 0 & \boldsymbol{I} \\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{D} \end{bmatrix} \begin{bmatrix} \nu \\ \mu\nu \end{bmatrix} = \mu \begin{bmatrix} \nu \\ \mu\nu \end{bmatrix} \tag{B.15}$$

which is then rewritten as

$$\boldsymbol{F} \begin{bmatrix} \nu \\ \mu\nu \end{bmatrix} = \mu \begin{bmatrix} \nu \\ \mu\nu \end{bmatrix} \tag{B.16}$$

By comparing (B.16) with (B.13), it is then deduced that one of the eigenvalues of $\boldsymbol{F}$, say μ_j, satisfies

$$\mu_j = \mu \tag{B.17}$$

and the associated eigenvector

$$\zeta_j = \begin{bmatrix} \nu \\ \mu\nu \end{bmatrix} \tag{B.18}$$

The eigenvalues μ_j carry important information about the behavior of the mechanical system described by (B.1). The eigenvectors are also important, but in general it is not

possible to fully compute these eigenvectors from available sensor signals.

B.1.5 Discrete time model

In practice, sensor signals are measured at discrete time instants, so the continuous time equation (B.3) will be adapted to discrete time form. Let τ be the sampling period. The simplest way for discretization would be to replace in Eq. (B.3), $\dot{\boldsymbol{x}}(t)$ by $(\boldsymbol{x}((k+1)\tau) - \boldsymbol{x}((k-1)\tau))/2\tau$. A more accurate discretization is described below.

First consider the free motion state equation (B.5). In this case, applying (B.6) with $t_0 = k\tau$ and $t = (k+1)\tau$, then

$$\boldsymbol{x_0}((k+1)\tau) = e^{\boldsymbol{F}\tau}\boldsymbol{x_0}(k\tau) \tag{B.19}$$

This discrete time free motion equation is an *exact* discretization of the continuous time equation (B.5), without *any approximation error*.

To accurately discretize (B.3), by applying (B.4), it is necessary to know the value of $\boldsymbol{u}(t)$ between samping instants:

$$\boldsymbol{x}((k+1)\tau) = e^{\boldsymbol{F}\tau}\boldsymbol{x}(k\tau) + \int_{k\tau}^{(k+1)\tau} e^{\boldsymbol{F}((k+1)\tau - s)}\boldsymbol{G}\boldsymbol{u}(s)\, ds, \tag{B.20}$$

In vibration monitoring problems, $\boldsymbol{u}(t)$ represents exogenous unknown forces (wind, tide wave, etc.). They are typically treated as random disturbances. Let

$$\omega_k \triangleq \int_{k\tau}^{(k+1)\tau} e^{\boldsymbol{F}((k+1)\tau - s)}\boldsymbol{G}\boldsymbol{u}(s)\, ds, \tag{B.21}$$

which is a random sequence (if u(t) is a random process), then

$$\boldsymbol{x}((k+1)\tau) = e^{\boldsymbol{F}\boldsymbol{\tau}}\boldsymbol{x}(k\tau) + \omega_k \tag{B.22}$$

To simplify notations, define

$$\boldsymbol{x}_k \triangleq \boldsymbol{x}(k\tau),\ \boldsymbol{A} \triangleq e^{\boldsymbol{F}\tau}, \tag{B.23}$$

then

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \omega_k. \tag{B.24}$$

B.1.6 Sensor signals

First consider sensors measuring positions and speeds (LDS and radar sensors). In the previous definition, the state vector $\boldsymbol{x}(t)$ is composed of the position $z(t)$ and the speed $\dot{z}(t)$ variables, so is the sampled state $\boldsymbol{x}_k$. Ideally, measured positions and speeds correspond to components of $\boldsymbol{x}_k$. The relationship between the state vector $\boldsymbol{x}_k \in \mathbb{R}^l$ and n_s sensor outputs is usually written as

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \nu_k \tag{B.25}$$

where $\boldsymbol{y}_k \in \mathbb{R}^{n_s}$ is a vector collecting the outputs of sensors, ν_k represents sensor noises, and $\boldsymbol{C}$ is a matrix. For ideal position and speed sensors, each row of $\boldsymbol{C}$ contains a single 1 completed with zero entries.

For sensors measuring accelerations (accelerometers),the situation is slightly more complicated. Rewrite (B.3) as

$$\begin{bmatrix} \ddot{z}(t) \\ \dot{z}(t) \end{bmatrix} = \boldsymbol{F}\boldsymbol{x}(t) + \boldsymbol{G}\boldsymbol{u}(t) \tag{B.26}$$

This equation shows that the acceleration $\ddot{z}(t)$ is linearly related to $\boldsymbol{x}(t)$ with a random disturbance term $\boldsymbol{G}\boldsymbol{u}(t)$. If $\boldsymbol{y}_k$ denotes sensor outputs corresponding to some components of the vector $\ddot{z}(t)$ sampled at discrete time instants, then it can be written as the previous version:

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \nu_k \tag{B.27}$$

with some matrix $\boldsymbol{C}$ different from the previous one introduced in (B.25) for position and speed sensors, and ν_k accounting for both sample random disturbance $\boldsymbol{G}\boldsymbol{u}(t)$ and measurement noises.

As the position and/or speed sensor equation (B.25) and the acceleration sensor equation (B.27) have exactly the same linear form, it is possible to mix the 3 types of sensors in the same equation.

To summarize, the discrete time state-space model is written as

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \omega_k \tag{B.28a}$$

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \nu_k \tag{B.28b}$$

B.1.7 Sensor-based model estimation

The previous sections show that the discrete time state-space model (B.28) is characterized by the matrix $\boldsymbol{A}$ and $\boldsymbol{C}$. In practice, such models are rarely available from physical laws, and when they are available, usually they do not fit well sensor data, due to idealistic assumptions made for deriving the model. Alternatively. the model can be estimated from sensor data.

For control systems, usually data are available for both inputs and outputs. In vibration monitoring problems, typically no input is available. Only sampled "output" data (values of $\boldsymbol{y}_k$) are available.

The difficulty for model estimation is mainly caused by lack of sensors. Usually a small number of component of $\boldsymbol{z}$, $\dot{\boldsymbol{z}}$ and/or $\ddot{\boldsymbol{z}}$ are measured by sensors. If all these components were available through sensor signals, then it would be relatively easy to estimate the matrix $\boldsymbol{M}^{-1}\boldsymbol{D}$ and $\boldsymbol{M}^{-1}\boldsymbol{K}$, by treating $\boldsymbol{u}(t)$ as an unknown random noise. It would be possible to estimate the 3 matrices $\boldsymbol{M}$, $\boldsymbol{D}$ and $\boldsymbol{K}$ if $\boldsymbol{u}(t)$ was also available. Unfortunately, in practice, typically only a small number of sensors, compared to the dimension of $\boldsymbol{z}$, is available.

With a limited number of sensors, it is more convenient to consider model estimation in the state-space form (B.28), where the matrix $\boldsymbol{C}$ has less rows than columns.

B.1.8 State-space indetermination

Now apply a state variable transformation to the state-space model (B.28)

$$\bar{\boldsymbol{x}}_k = \boldsymbol{T_s}^{-1}\boldsymbol{x}_k \tag{B.29}$$

where $\boldsymbol{T_s}$ is any invertible square matrix of appropriate size. Then the state-space model becomes

$$\bar{\boldsymbol{x}}_{k+1} = \boldsymbol{T_s}^{-1}\boldsymbol{A}\boldsymbol{T_s}\bar{\boldsymbol{x}}_k + \boldsymbol{T_s}^{-1}\omega_k \tag{B.30a}$$

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{T_s}\bar{\boldsymbol{x}}_k + \nu_k \tag{B.30b}$$

or, with obvious new notations

$$\bar{\boldsymbol{x}}_{k+1} = \bar{\boldsymbol{A}}\bar{\boldsymbol{x}}_k + \bar{\omega}_k \tag{B.31a}$$

$$\boldsymbol{y}_k = \bar{\boldsymbol{C}}\bar{\boldsymbol{x}}_k + \nu_k \tag{B.31b}$$

This result shows that the same sequence $\boldsymbol{y}_k$ (available from sensors) can be generated either by (B.28) or by (B.31). Because the state variable $\boldsymbol{x}_k$ is not available, there is no way to distinguish between (B.28) and (B.31). For this reason, it is not possible to estimate $\boldsymbol{A}$ and $\boldsymbol{C}$ from sensor signals, but only ${\boldsymbol{T_s}}^{-1}\boldsymbol{A}\boldsymbol{T_s}$ and $\boldsymbol{C}\boldsymbol{T_s}$ for some unknown matrix $\boldsymbol{T_s}$. In other words, $\boldsymbol{A}$ and $\boldsymbol{C}$ can only be estimated up to a similarity transformation.

However, it is easy to show that the matrix $\bar{\boldsymbol{A}} = {\boldsymbol{T_s}}^{-1}\boldsymbol{A}\boldsymbol{T_s}$ has the same eigenvalues as $\boldsymbol{A}$. The importance of eigenvalues has already been explained.

B.1.9 Subspace identification

Assumption 1 *Sensor signals $\boldsymbol{y}_k \in \mathbb{R}^s$ are generated by system* (B.28) *driven by stationary zero mean white noise ω_k and ν_k. The dimension n_s of the state vector is unknown, but is assumed that $n_s < p$ for some given integer p.*

Assumption 2 *The matrix pair $(\boldsymbol{A}, \boldsymbol{C})$ is observable.*

It follows from (B.28) that

$$\begin{aligned}
\boldsymbol{y}_k &= \boldsymbol{C}\boldsymbol{x}_k + \nu_k \\
\boldsymbol{y}_{k+1} &= \boldsymbol{C}\boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{C}\omega_k + \nu_{k+1} \\
&\vdots \\
\boldsymbol{y}_{k+l} &= \boldsymbol{C}\boldsymbol{A}^l\boldsymbol{x}_k + \boldsymbol{C}\boldsymbol{A}^{l-1}\omega_k + \boldsymbol{C}\boldsymbol{A}^{l-2}\omega_{k+1} + \cdots + \boldsymbol{C}\omega_{k+l-1} + \nu_{k+l}
\end{aligned}$$

For some integer $p > n_s$, define double indexed notations

$$\boldsymbol{y}_{k,p} \triangleq \begin{bmatrix} \boldsymbol{y}_k \\ \boldsymbol{y}_{k+1} \\ \vdots \\ \boldsymbol{y}_{k+p-1} \end{bmatrix}, \; \boldsymbol{\varepsilon}_{k,p} \triangleq \begin{bmatrix} \nu_k \\ \boldsymbol{C}\omega_k + \nu_{k+1} \\ \vdots \\ \boldsymbol{C}\sum_{j=1}^{p-1} \boldsymbol{A}^{p-1-j}\omega_{k+k-1} + \nu_{k+p-1} \end{bmatrix} \tag{B.32}$$

and the extended observability matrix

$$\boldsymbol{O}_p = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{p-1} \end{bmatrix} \tag{B.33}$$

then

$$\boldsymbol{y}_{k,p} = \boldsymbol{O}_p \boldsymbol{x}_k + \boldsymbol{\varepsilon}_{k,p} \tag{B.34}$$

Define further

$$\boldsymbol{Y}_{p,N_s} \triangleq \begin{bmatrix} \boldsymbol{y}_{1,p} & \boldsymbol{y}_{2,p} & \dots & \boldsymbol{y}_{N_s,p} \end{bmatrix} \tag{B.35}$$

$$\boldsymbol{X}_{N_s} \triangleq \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_{N_s} \end{bmatrix} \tag{B.36}$$

$$\boldsymbol{E}_{p,N_s} \triangleq \begin{bmatrix} \boldsymbol{\varepsilon}_{1,p} & \boldsymbol{\varepsilon}_{2,p} & \dots & \boldsymbol{\varepsilon}_{N_s,p} \end{bmatrix} \tag{B.37}$$

then

$$\boldsymbol{Y}_{p,N_s} = \boldsymbol{O}_p \boldsymbol{X}_{N_s} + \boldsymbol{E}_{p,N_s} \tag{B.38}$$

The matrix $\boldsymbol{O}_p$ has n_s columns and $\boldsymbol{X}_{N_s}$ has n_s rows (n_s is the dimension of the state vector $\boldsymbol{x}_k$). Then the matrix product $\boldsymbol{O}_p \boldsymbol{X}_{N_s}$ has a rank not larger than n_s. Because $\boldsymbol{Y}_{p,N_s}$ is equal to $\boldsymbol{O}_p \boldsymbol{X}_{N_s}$ plus the noise term $\boldsymbol{E}_{p,N_s}$, its singular values are close to those of $\boldsymbol{O}_p \boldsymbol{X}_{N_s}$. Let

$$\boldsymbol{y}_{k,q} \triangleq \begin{bmatrix} \boldsymbol{y}_k \\ \boldsymbol{y}_{k+1} \\ \vdots \\ \boldsymbol{y}_{k+q-1} \end{bmatrix} \tag{B.39}$$

with some $q > n_s$, and

$$\boldsymbol{\Upsilon}_{q,N_s} \triangleq \begin{bmatrix} \boldsymbol{y}_{1-r,q} & \boldsymbol{y}_{2-r,q} & \dots & \boldsymbol{y}_{N_s-r,q} \end{bmatrix} \tag{B.40}$$

be a matrix built from sensor signals $\boldsymbol{y}_k$ with some integer $r \geq q$ so that each column of the matrix $\boldsymbol{\Upsilon}_{q,N_s}$ is uncorrelated with the corresponding column of $\boldsymbol{E}_{p,N_s}$, i.e.,

$$\mathrm{E}(\boldsymbol{\varepsilon}_{k,p} \boldsymbol{y}_{k-r,q}^T) = 0 \tag{B.41}$$

The fact that $\boldsymbol{\Upsilon}_{q,N_s}$ involves negative instants k is only a matter of convention for choosing the zero instant. The structure of $\boldsymbol{\Upsilon}_{q,N_s}$ is similar to $\boldsymbol{Y}_{p,N_s}$, with the same number of columns N_s, but with a possibly different number of rows q.

Right multiply both sides of Eq. (B.38) by $\boldsymbol{\Upsilon}_{q,N_s}^T/N_s$ and define

$$\boldsymbol{H}_{p,q,N_s} \triangleq \frac{1}{N_s}\boldsymbol{Y}_{p,N_s}\boldsymbol{\Upsilon}_{q,N_s}^T \tag{B.42}$$

$$\boldsymbol{G}_{q,N_s} \triangleq \frac{1}{N_s}\boldsymbol{X}_{N_s}\boldsymbol{\Upsilon}_{q,N_s}^T \tag{B.43}$$

then

$$\boldsymbol{H}_{p,q,N_s} = \boldsymbol{O}_p\boldsymbol{G}_{q,N_s} + \frac{1}{N_s}\boldsymbol{E}_{p,N_s}\boldsymbol{\Upsilon}_{q,N_s}^T \tag{B.44}$$

This operation asymptotically cancels out the noise term, since

$$\frac{1}{N_s}\boldsymbol{E}_{p,N_s}\boldsymbol{\Upsilon}_{q,N_s}^T = \frac{1}{N_s}\boldsymbol{\varepsilon}_{k,p}\boldsymbol{y}_{k-r,q}^T \tag{B.45}$$

and, by the law of large numbers and due to (B.41),

$$\lim_{N_s\to\infty}\frac{1}{N_s}\boldsymbol{E}_{p,N_s}\boldsymbol{\Upsilon}_{q,N_s}^T = \mathrm{E}(\boldsymbol{\varepsilon}_{k,p}\boldsymbol{y}_{k-r,q}^T) = 0 \tag{B.46}$$

Assumption 2 ensures that the extended observability matrix $\boldsymbol{O}_p$ (with $p \geq n_s$) has its rank equal to n_s. For large N_s, the rank of the $n_s \times q$ matrix $\boldsymbol{G}_{q,N_s} = \boldsymbol{X}_{N_s}\boldsymbol{\Upsilon}_{q,N_s}^T/N_s$ is equal to n_s (with $q > n_s$). Then the rank of $\boldsymbol{O}_p\boldsymbol{G}_{q,N_s}$ is n_s, i.e., it has n_s non zero singular values.

Suppose that N_s is large enough so that the noise term in (B.44) is neglected, then

$$\boldsymbol{H}_{p,q,N_s} = \boldsymbol{O}_p\boldsymbol{G}_{q,N_s} \tag{B.47}$$

Let the singular value decomposition (SVD) of $\boldsymbol{H}_{p,q,N_s}$ be

$$\boldsymbol{H}_{p,q,N_s} = \begin{bmatrix} U_1 & U_2 \end{bmatrix}\begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 S_1 V_1^T \tag{B.48}$$

where S_1 is an $n_s \times n_s$ disgonal matrix containing non zero singular values. Then

$$\boldsymbol{O}_p\boldsymbol{G}_{q,N_s} = U_1 S_1 V_1^T \tag{B.49}$$

and

$$U_1 = \boldsymbol{O}_p \boldsymbol{G}_{q,N_s} V_1 S_1^- 1 \tag{B.50}$$

$$= \boldsymbol{O}_p \boldsymbol{T}_s \tag{B.51}$$

with

$$\boldsymbol{T}_s \triangleq \boldsymbol{G}_{q,N_s} V_1 S_1^{-1} \tag{B.52}$$

Each of the matrix factors $\boldsymbol{G}_{q,N_s} V_1 S_1^{-1}$ is of rank n_s, then $\boldsymbol{T_s}$ is an $n_s \times n_s$ invertible matrix.

By recalling the definition of $\boldsymbol{O}_p$ in (B.33), it then follows from (B.51) that

$$U_1 = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{p-1} \end{bmatrix} \boldsymbol{T}_s = \begin{bmatrix} (\boldsymbol{CT_s}) \\ (\boldsymbol{CT_s})(\boldsymbol{T_s}^{-1}\boldsymbol{AT_s}) \\ \vdots \\ (\boldsymbol{CT_s})(\boldsymbol{T_s}^{-1}\boldsymbol{A}^{p-1}\boldsymbol{T_s}) \end{bmatrix} = \begin{bmatrix} (\boldsymbol{CT_s}) \\ (\boldsymbol{CT_s})(\boldsymbol{T_s}^{-1}\boldsymbol{AT_s}) \\ \vdots \\ (\boldsymbol{CT_s})(\boldsymbol{T_s}^{-1}\boldsymbol{AT_s})^{p-1} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{C}} \\ \bar{\boldsymbol{C}}\bar{\boldsymbol{A}} \\ \vdots \\ \bar{\boldsymbol{C}}\bar{\boldsymbol{A}}^{p-1} \end{bmatrix} \tag{B.53}$$

with

$$\bar{\boldsymbol{A}} = \boldsymbol{T_s}^{-1}\boldsymbol{AT_s} \tag{B.54}$$

$$\bar{\boldsymbol{C}} = \bar{\boldsymbol{C}}\boldsymbol{T_s} \tag{B.55}$$

The result of (B.53) means that U_1 is equal to the extended matrix $\boldsymbol{O}_p$ up to a similarity transformation. Then the matrices $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{C}}$ can be extracted from U_1. As the state transformation matrix $\boldsymbol{T_s}$ is unknown, the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$ is only determined up to a similarity transformation.

In practice, for any given data record of finite size N_s, the noise term in (B.44) cannot be fully canceled out, hence the above reasoning is only approximately true. Notably, in addition to the n_s non zero singular values of the matrix $\boldsymbol{H}_{p,q,N_s}$ corresponding to those of $\boldsymbol{O}_p \boldsymbol{G}_{q,N_s}$, the other singular values of $\boldsymbol{H}_{p,q,N_s}$ due to the noise term are close to zero instead of being exactly zero. To summarize, this subspace identification algorithm is as follows.

Subspace identification algorithm

— Collect sensor data to build the matrix $\boldsymbol{\Upsilon}_{q,N_s}$ and $\boldsymbol{Y}_{p,N_s}$.

— Compute $\boldsymbol{H}_{p,q,N_s}$ as Eq. (B.43).

— Perform the SVD of $\boldsymbol{H}_{p,q,N_s}$. Choose a threshold to truncate small singular values. Let U_1 be the matrix containing in its columns the left singular vectors corresponding to the non truncated singular values.

— By using Matlab notations, $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{C}}$ are determined from $U1_1$ as follows:

$$\bar{\boldsymbol{C}} = U_1(1 : n_s, :); \tag{B.56}$$

$$\bar{\boldsymbol{A}} = U_1(1 : (p-1) \times N_s, :) \setminus U_1((N_s + 1) : (p \times n_s), :); \tag{B.57}$$

where n_s and N_s are respectively the dimensions of the state $\boldsymbol{x}_k$ and of the output $\boldsymbol{y}_k$

This is the very basic subspace identification algorithm. Existing variant algorithms essentially differ in the choices of some weighting matrices modifying the results of the SVD, leading to different (unknown) state transformation matrices $\boldsymbol{T_s}$.

For systems involving known inputs (typically control systems), the algorithm includes an extra step to cancel out the input term before the SVD step.

B.1.10 Vibration monitoring: modal parameters

As previously explained, when a state-space model is estimated from sensor data, the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$ can only be estimated up to a similarity transformation.

Let λ_j, ϕ_j be respectively the j-th eigenvalue of $\boldsymbol{A}$ and the associated eigenvector, with $j = 1, 2, \ldots, n_s$. The *mode shape* is defined as

$$\psi_j = \boldsymbol{C}\phi_j \tag{B.58}$$

The n_s pairs (λ_j, ψ_j) are invariant to similarity transformations $\boldsymbol{T_s}$.

For the purpose of vibration monitoring, only complex eigenvalues λ_j are of interest, hence real eigenvalues, if any, are discarded. Hence n_s is an even number in what follows.

As introduced in (B.23), the discrete time transition matrix $\boldsymbol{A}$ is related to the continuous time transition matrix $\boldsymbol{F}$ by $\boldsymbol{A} = e^{\boldsymbol{F}\tau}$, where τ is the sampling period. Therefore, their corresponding eigenvalues λ_j and μ_j are related by

$$\lambda_j = e^{\mu_j \tau} \tag{B.59}$$

In vibration monitoring, two important parameters, the natural frequency f_j and the

damping ration ζ_j, are related to μ_j through

$$\mu_j = -2\pi f_j \zeta_j \pm i2\pi f_j \sqrt{1-\zeta_j^2} \tag{B.60}$$

where i is the unit imaginary number, and the sign "$\pm$" is because the complex eigenvalues of a real matrix come always in conjugate pairs. The triples (f_j, ζ_j, ψ_j), for $j = 1, 2, \ldots, n_s$ are called modal parameters.

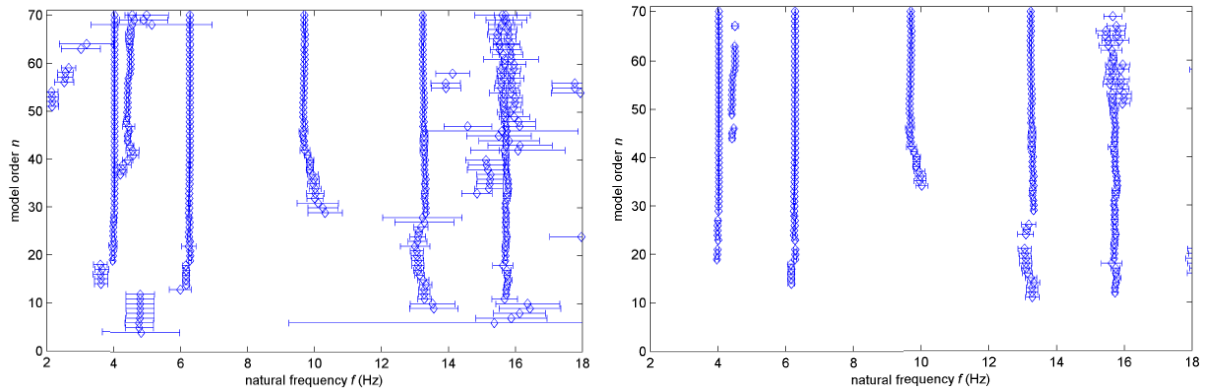
When the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$ is estimated by a subspace identification method, in the SVD step, small singular values are truncated based on some criterion. The resulting $\bar{\boldsymbol{A}}$ matrix leads to a set of complex eigenvalues λ_j, and then to the corresponding natural frequencies and damping ratios f_j, ζ_j.

Quite often, despite the truncation of small singular values, some of the remaining natural frequencies and damping ratios f_j, ζ_j are caused by noises. It means that the small singular value truncation in the SVD step of system identification is not sufficiently reliable. The superfluous modes are often removed *manually*, based on the so-called stabilization diagram, which shows the natural frequencies f_j corresponding to different choices of the model order n_s. A true experimental example is shown in Fig. B.1. In the Fig. B.1a, when the model order n_s varies (along the vertical axis), most natural frequencies (represented by diamonds) remain in vertical lines, whereas some others do not. Only the natural frequencies following vertical lines are kept, as shown in the result of the Fig. B.1b. The other natural frequencies f_j are eliminated, so are the corresponding damping ratios ζ_j and mode shapes ψ_j.

In practice, fault diagnosis by repeated system identification can be applied.

After system identification, as described in the previous section, the monitored system is characterized by the selected set of modal parameters (f_j, ζ_j, ψ_j), which are invariant to state transformation T_s. Some time later, the same procedure of system identification is repeated, leading to another set of modal parameters, which are then compared with the earlier established ones. Any significant evolution of the modal parameters would reveal a fault, characterized by the corresponding modal parameters (f_j, ζ_j, ψ_j), which are well understood by engineers, for the purpose of fault diagnosis. If no fault is detected, the same inspection procedure can be repeated later, and so on. To summarize, both fault detection and fault diagnosis are achieved by monitoring the evolution of modal parameters (f_j, ζ_j, ψ_j), through repeated system identification.

This method has a serious drawback in practice: mode selection is often made manually



(a) Full diagram

(b) Selected natural frequencies

Figure B.1 – Example of stabilization diagram. For each chosen model order n (vertical axis), the corresponding natural frequencies are represented by diamonds. The confidence intervals of the natural frequencies are also indicated. Left: full diagram, right: selected natural frequencies.

by a human operator, based the stabilization diagram. Some automatic algorithms do exist, but they are not reliable enough. There is another drawback: when several natural frequencies change simultaneously because of one or more faults, it is difficult to find the correspondence between the new values and the old values of f_j. This correspondence is essential for the purpose of fault diagnosis. This situation motivated other methods, such as statistical tests on model-data discrepancy.

RÉSUMÉ ÉTENDU

Le plan de la thèse comporte les six chapitres suivants:

— Le **Chapitre 1** introduit le contexte de l'étude, les méthodes vibratoires existantes pour le SHM et les objectifs de la recherche.

— Le **Chapitre 2** décompose le mouvement d'un objet entre deux images aux instants t_1 et t_2 en une composante pixellique et une composante subpixellique. Il propose ensuite un état de l'art des méthodes existantes pour estimer chacune des composante du mouvement. La composante pixellique est évaluée par intercorrélation entre les deux images aux instants t_1 et t_2. Trois familles de méthodes subpixellique permettent d'affiner l'estimation du mouvement: *i)* les méthodes d'interpolation de la luminance dans un ROI, *ii)* l'ajustement d'une surface quadratique (QSF) au voisinage du maxima d'intercorrélation, et *iii)* le calcul de la phase locale de l'image à partir d'un filtre directionnel (*steerable filter*). Enfin, le chapitre présente le principe de l'analyse vibratoire pour le SHM. Il montre comment établir un diagnostic SHM à partir de mesures de mouvement vibratoires enregistrées en différents endroits de la structure.

— Le **Chapitre 3** présente trois contributions de la thèse à l'extraction du mouvement subpixellique. La méthode EQI (et sa variante EQI-Octant) est une méthode d'interpolation linéaire de la luminance dans un voisinage de 4 pixels dans une direction, qui est déterminée par une analyse simple de la fonction d'intercorrélation entre les images. La seconde contribution consiste à étendre l'utilisation du filtre directionnel à l'estimation d'un mouvement pluri-pixellique. Enfin, le chapitre propose une analyse détaillée du fonctionnement de la méthode (QSF) en mettant à jour quelques limites de la méthode.

— Le **Chapitre 4** évalue les performances des méthodes d'extraction de mouvement sélectionnées à partir de flux d'images vidéos simulées. Le chapitre presente tout d'abord deux méthodes de génération d'images simulées, représentant une cible circulaire de faible texture en mouvement. Nous avons également utilisé une base existante d'images représentant une image de speckle décalée de différentes valeurs. L'évaluation des performances est réalisée successivement à l'échelle pixellique et

subpixellique. L'EQM de l'erreur d'estimation entre le mouvement estimé et le mouvement simulé est évaluée à partir de toutes les images du flux vidéo; elle sert à quantifier les performances des algorithmes. Pour des raisons de commodité, l'EQM est exprimée en fraction de pixel. Les résultats de comparaison montrent les meilleures performances de la méthode proposée EQI.

— Le **Chapitre 5** présente les différents essais pour réaliser une analyse SHM expérimentale d'une barre métallique verticale en laboratoire. Il décrit tout d'abord le montage expérimental et le matériel, à savoir la table vibrante, la caméra haute vitesse, l'éclairage, les cibles géométriques pacées sur la barre, le dispositif de mesure de déplacement de référence. La table vibrante permet de générer soit une excitation sinusoïdale pure, soit une excitation aléatoire d'un spectre relativement plat. Des chocs au marteau ont également été réalisés pour générer une source de vibrations transitoire.

  Nous avons tout d'abord complété l'évaluation des algorithmes d'extraction de mouvement à partir de données expérimentales. Pour ce faire, le traitement d'images a porté sur les ROIs centrés sur des cibles géométriques attachées soit à la table vibrante directement, soit à la poutre. Dans une second temps, l'image de la barre est décomposée en 160 ROIs. Le mouvement est estimé à partir des contours naturels de la barre dans chaque ROI, qui fait ainsi office de capteur virtuel. La multiplication des points de mesure permet d'améliorer la précision et la fiabilité de l'estimation des déformées modales de la barre.

  Enfin, une expérience de diagnostic des défauts a été mise en place, en fixant un aimant à différentes hauteurs. L'analyse SHM basée sur l'image permet de détecter les perturbations induites à partir des changements des fréquences naturelles de la poutre, tandis que le critère d'assurance modale (MAC) permet de détecter les changements des formes des modes de vibration.

— Le **Chapitre 6** résume le travail de thèse, et indique les avantages et les limites des méthodes de SHM basées sur l'analyse de flux vidéo. Le chapitre se termine par quelques perspectives de travaux futurs.

Chapitre 1 - Introduction

Contexte

Les techniques de surveillance de santé structurelle (ou techniques SHM pour "Structural Health Monitoring") ont pour objectif d'évaluer le niveau d'intégrité d'infrastructures mécaniques, par l'intermédiaire de l'évolution d'indicateurs établis à partir des mesures de capteurs installés sur la structure. Par extension, le SHM fait également référence aux techniques de traitement de données des signaux issus des capteurs de mesure. En évaluant le niveau de sécurité des infrastructures de génie civil, le SHM joue un rôle essentiel dans les activités économiques.

Les méthodes vibratoires constituent une vaste famille de techniques de SHM, basées sur l'analyse de l'évolution des paramètres modaux de la structure, tels que les fréquences propres, le facteur d'amortissement, et les déformées modales. Elles se sont développées initialement à partir de capteurs ponctuels de type accélérométrique, des méthodes d'identification stochastique permettant d'estimer les paramètres modaux de la structure. L'optimisation du nombre et du positionnement des capteurs est souvent une tâche complexe.

Par contraste, cette thèse propose d'étudier une technique SHM sans contact plus récente, qui fait l'analyse SHM à partir de l'analyse d'images vidéo. En comparaison de la technique conventionnelle à base d'accéléromètres, la technique vidéo offre la possibilité d'extraire le mouvement vibratoire en un plus grand nombre de positions de la structure (inclues dans le champ de vision de la caméra) en évitant l'installation fastidieuse de capteurs ponctuels.

Cette famille de méthodes SHM plus récentes repose sur des techniques de vision par ordinateur qui permettent d'estimer les mouvements vibratoires de la structure avec une précision de l'ordre de la fraction de pixel. En effet, lorsque l'on observe une structure à longue distance avec une caméra vidéo, le mouvement de la structure dans l'image peut correspondre à une dimension inférieure à la taille d'un pixel. Le mouvement vibratoire de la structure (ou d'une partie de celle-ci) entraîne des variations subtiles de l'intensité du pixel. Dans ce contexte, l'objectif des méthodes de traitement d'images consiste à convertir le changement d'intensité du pixel en déplacement sous-pixellique.

Objectifs

L'objectif principal de la thèse est de montrer la faisabilité d'une méthode de suivi de santé structurel (SHM) d'une structure mécanique simple, à partir du traitement d'un flux d'images vidéo de la structure.

Cet objectif nécessite de développer tout d'abord une méthode d'extraction de mouvement d'un objet à l'aide de méthodes de traitement d'images. Une sélection de pixels de l'image (i.e., schématiquement, les pixels appartenant aux contours de la structure mécanique) peut ainsi servir de capteurs virtuels pour déterminer le mouvement vibratoire de la structure dans le plan de l'image, en remplacement des capteurs ponctuels traditionnels du type accéléromètre. L'analyse SHM est effectuée sur le mouvement ainsi estimé, et permet de déterminer les fréquences de résonances et les déformées modales de la structure, ainsi que leur évolution temporelle.

Dans ce contexte, l'état de l'art a pour but de sélectionner quelques techniques de traitement d'image existantes, qui permettent d'obtenir une estimation précise du mouvement. Pour des raisons pratiques, le mouvement comporte une composante pixellique et une composante subpixellique, qui sont estimées successivement. Cette partie du travail de la thèse a également pour objectif de proposer deux nouvelles méthodes d'extraction du mouvement subpixellique, qui présentent de meilleures performances que les méthodes sélectionnées. De fait, les performances des différentes techniques de traitement d'image sont évaluées aux deux échelles (pixel et subpixel) à partir de flux d'images simulées.

Le dernier objectif consiste à mettre en oeuvre en laboratoire la méthodologie de SHM proposée par le biais de 2 séries d'essais expérimentaux. Des séquences vidéos d'une structure métallique simple sont enregistrées à l'aide d'une caméra rapide. Dans un premier temps, la structure est constituée d'une barre métallique verticale soumise à différentes sources vibratoires (table vibrante, choc, vibrations ambiantes). Dans un second temps, la réponse de la barre est perturbée par une masse ponctuelle (aimant), qui est disposée successivement en différents endroits de la barre, pour simuler un désordre mécanique.

Pour l'application SHM, on cherche à démontrer d'une part que les signaux vibratoires estimés par la technique vidéo peuvent être utilisés pour déterminer avec précision les paramètres modaux des structures mécaniques, i.e., fréquences de résonance et déformées modales, à partir des techniques d'identification à sous-espace. D'autre part, l'expérimentation montre que la méthodologie proposée est suffisamment sensible et précise pour réaliser la surveillance de la structure, en détectant les modifications des paramètres modaux de la structure entre les deux séries d'essai. Ces expérimentations

permettent en outre d'illustrer la sensibilité des résultats de la méthode aux différents réglages de la caméra, ainsi que l'importance de la détermination du mouvement subpixellique dans la précision des mesures.

Chapitre 2 - État de l'art

Ce chapitre présente un état de l'art des 2 étapes du traitement de données qui permettent de réaliser l'analyse SHM d'une structure mécanique à partir d'un flux d'images vidéos. Une première partie indique les principales méthodes de traitement d'image qui permettent d'estimer le mouvement vibratoire. Le principe de l'analyse SHM de ces données vibratoires est ensuite succinctement présenté; il s'appuie sur les développements détaillés dans l'annexe B du document principal.

2.1 Extraction du mouvement vibratoire

Nous supposons qu'un objet est observé dans un flux temporel d'images vidéo. L'objet est supposé rigide et avoir un mouvement vibratoire dans le plan de la caméra, avec une faible amplitude de déplacement entre 2 images successives.

Parmi les méthodes existantes de la littérature, ce chapitre se focalise sur des traitements en deux étapes (*coarse-to-fine approach*), qui permettent d'estimer le mouvement avec une grande précision (Pan et al., 2009; D. Zhang et al., 2016; L. Zhang et al., 2015; Z. Zhang et al., 2006).

— Une première étape utilise les techniques de corrélation d'image numérique (*Digital Image Correlation, DIC*) pour estimer le mouvement pixellique, noté $(\Delta m, \Delta n)$.

— La seconde étape estime le décalage subpixellique $(\delta x, \delta y)$ par des méthodes spécifiques. Cette seconde étape fait l'objet de contributions particulières, qui sont détaillées au chapitre 3.

L'amplitude du mouvement total de l'objet (d_x, d_y) s'écrit comme la somme des deux précédents décalages selon:

$$(d_x, d_y) = (\Delta m + \delta x, \Delta n + \delta y)$$

a) Décalage pixellique

La corrélation est calculée dans une région d'intérêt (ROI) de l'image, qui inclut soit l'objet mobile entier (cible), soit un élément saillant de la structure. Les images de N lignes et M colonnes à des instants différents sont notées $I(m, n, t_1)$ et $I(m, n, t_2)$, où les indices (m, n) sont les indices des pixels dans le plan image. Le déplacement pixellique de l'objet entre les deux images est noté $(\Delta m, \Delta n)$ selon les directions horizontale et verticale. Dans le flux d'images vidéos, l'image à l'instant initial t_0 est généralement choisie comme image de référence pour quantifier le mouvement dans les images suivantes.

Dans la littérature, le déplacement pixellique $(\Delta m, \Delta n)$ entre les deux images $I(m, n, t_1)$ et $I(m, n, t_2)$ est donné par la position du maximum de l'intercorrélation 2D, selon l'expression suivante:

$$(\Delta m^*, \Delta n^*) = \arg \max_{\Delta m, \Delta n} \mathrm{CC}(\Delta m, \Delta n)$$

où CC est la corrélation croisée 2D conventionnelle:

$$\mathrm{CC}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} I(m, n, t_1) I(m - \Delta m, n - \Delta n, t_2)$$

avec $\mathbb{T}$ l'ensemble des indices des pixels du ROI.

Lors des tests préliminaires que nous avons réalisés, la fonction de corrélation 2D conventionnelle défini précédemment, montrait un niveau de précision limitée. Le tableau 1 ci-dessous indique les différentes variantes d'intercorrélation que nous avons testées dans cette thèse.

Lors de l'évaluation des méthodes au chapitre 4, nous avons constaté que les fonctions ZNCC et ZCC conduisent à l'estimation la plus précise du décalage subpixellique. Ces deux fonctions sont plus robustes aux variations de luminosité entre images, en comparaison de la méthode conventionnelle CC. (P. Liu & Du, 2019; Pan et al., 2009).

Fonction	Définition
Cross-correlation (CC)	$\mathrm{CC}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} [I(m,n,t_1) I(m-\Delta m, n-\Delta n, t_2)]$
Normalized cross-correlation (NCC)	$\mathrm{NCC}(\Delta m, \Delta n) = \frac{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_1) I(m-\Delta m, n-\Delta n, t_2)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_1)^2} \sqrt{\sum_{(m,n)\in\mathbb{T}} I(m,n,t_2)^2}}$
Zero-mean cross-correlation (ZCC)	$\mathrm{ZCC}(\Delta m, \Delta n) = \sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_1) I_c(m-\Delta m, n-\Delta n, t_2)$
Zero-mean normalized cross-correlation (ZNCC)	$\mathrm{ZNCC}(\Delta m, \Delta n) = \frac{\sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_1) I_c(m-\Delta m, n-\Delta n, t_2)}{\sqrt{\sum_{(m,n)\in\mathbb{T}} I_c(m,n,t_1)^2} \sqrt{\sum_{(m,n)\in\mathbb{T}} I_c(m-\Delta m, n-\Delta n, t_2)^2}}$

Tableau 1: Définitions des fonctions de corrélation croisée utilisées pour l'estimation du décalage pixellique entre 2 ROIs $I(m,n,t_1)$ et $I(m,n,t_2)$

Dans le tableau 1, $\bar{I}(t)$ et $I_c(m,n,t)$ sont la valeur moyenne de l'image et de l'image de centrage zéro, respectivement:

$$\bar{I}(t) = \frac{\sum_{(m,n)\in\mathbb{T}} I(m,n,t)}{MN}$$
$$I_c(m,n,t) = I(m,n,t) - \bar{I}(t)$$

b) Décalage subpixellique

L'objectif de cette seconde étape du traitement consiste à estimer le décalage subpixellique $(\delta x^*, \delta y^*)$ entre l'image de référence décalée $I_s(m,n,t_1)$ et l'image courante $I(m,n,t_2)$.

L'image $I_s(m,n,t_1)$ se déduit de l'image de référence $I(m,n,t_1)$, en appliquant le décalage pixellique $(\Delta m^*, \Delta n^*)$ obtenu à la première étape, selon l'expression :

$$I_s(m,m,t_1) = I(m+\Delta m^*, n+\Delta m^*, t_1)$$

A ce stade du traitement, nous supposons que les images $I_s(m,n,t_1)$ et $I(m,n,t_2)$ diffèrent par un décalage subpixellique, i.e., inférieur à un pixel, noté $(\delta x, \delta y)$. Une relation approximative entre $I_s(m,n,t_1)$ et $I(m,n,t_2)$ s'écrit sous la forme suivante (l'approximation intervient pour tenir compte du bruit):

$$I(m,m,t_2) \approx I_s(m+\delta x, n+\delta y, t_1)$$

Le chapitre 2 détaille ensuite 3 familles d'estimation subpixellique de la littérature:

1. Méthodes basées sur le gradient du niveau de gris de l'image : flux optique (Burton

& Radford, 1978; Chan et al., 2010; Horn & Schunck, 1981; Javh et al., 2017; Lucas & Kanade, 1981)

2. Méthodes basée sur l'interpolation de la corrélation croisée au voisinage du maxima (Hoge, 2003; Pan et al., 2005; Psarakis & Evangelidis, 2005; Tennakoon et al., 2015; Thurman et al., 2008; Ye et al., 2018).
3. Méthodes basées sur la phase locale de l'image, avec en particulier, la méthode par "steerable filtres" (J. G. Chen et al., 2015; Diamond et al., 2017; Freeman & Adelson, 1991).

2.2 Principe de l'analyse vibratoire pour le SHM

Ce paragraphe donne quelques bases pour formuler la dynamique des systèmes mécaniques en termes de modèle d'état et pour en déduire un diagnostic SHM à partir des mesures de mouvement vibratoires enregistrées en différents endroits de la structure. L'annexe B détaille les formulations mathématiques associées.

L'analyse vibratoire pour le SHM repose sur l'analyse modale de la structure, qui permet de déterminer les fréquences de résonance et leur facteur d'amortissement, ainsi que les déformées modales de la structure. L'analyse vibratoire nécessite la détermination des propriétés dynamiques intrinsèques du système mécanique associé. L'équation d'équilibre dynamique d'un système mécanique à plusieurs degrés de liberté peut être décrite comme un système linéaire invariant dans le temps (LTI) à plusieurs degrés de liberté (*Degree Of Freedom*, DOF):

$$\boldsymbol{M}\,\ddot{z}(t) + \boldsymbol{D}\,\dot{z}(t) + \boldsymbol{K}\,z(t) = \boldsymbol{u}(t)$$

où t représente le temps continu, les matrices $\boldsymbol{M}, \boldsymbol{D}, \boldsymbol{K} \in \mathbb{R}^{s\times s}$ désignent respectivement les matrices de masse, d'amortissement et de rigidité, et s est le nombre de degré de liberté du système. Les vecteurs $z(t) \in \mathbb{R}^s$ et $\boldsymbol{u(t)} \in \mathbb{R}^s$ désignent respectivement les déplacements en temps continu et les forces externes inconnues.

En condition expérimentale, seules les signaux des capteurs le long de la structure sont disponibles. Des méthodes d'identification stochastique sont déployées pour estimer les paramètres modaux à partir de l'équation d'état. Dans ce travail, les paramètres modaux de la structure sont déterminés par une méthode d'identification à sous-espace.

Dans ce contexte, le suivi des structures consiste à déterminer les changements qui interviennent dans les valeurs des paramètres modaux. On essaie ainsi de quantifier les variations des fréquences de résonance en particulier, ainsi que les modifications des dé-

formées modales par l'intermédiaire du critère d'assurance modale (MAC) (Greś et al., 2021).

Chapitre 3 - Contributions à l'extraction du mouvement subpixellique

Ce chapitre présente les contributions de la thèse pour chacune des 3 familles d'estimation du mouvement subpixellique indiquées au § 2.1.b.

3.1 Méthode EQI

Cette première méthode cherche à interpoler le niveau de gris de l'image de référence décalée $I_s(m, m, t_1)$ par un plan dans un voisinage de quelques pixels. La méthode d'interpolation présentée dans (D. Zhang et al., 2016) nous sert de référence par sa simplicité. L'équation du plan d'interpolation est déduite d'un développement de Taylor au 1er ordre dans le voisinage conventionnel de Von Neumann, qui comporte 5 pixels dans un voisinage de 3×3 pixels.

En comparaison de la méthode d'interpolation de référence précédente, nous proposons le nouvel algorithme EQI (Efficient Quadrant Interpolation). Le voisinage 3×3 précédent est décomposé en 4 quadrants de 4 pixels chacun incluant le pixel central. EQI ajuste un plan à l'image au sens des moindres carrés, dans le quadrant qui a été sélectionné par une analyse de la corrélation dans le voisinage 3×3 du maximum.

La méthode EQI comprend les trois étapes suivantes:

1. Déterminer la direction approximative de déplacement subpixellique par l'analyse de la matrice de corrélation croisée ZNCC dans le voisinage 3×3 de son extremum;
2. Déterminer les plans locaux qui s'ajustent au sens des moindres carrés aux niveaux de gris du ROI dans le quadrant 2×2 pixels sélectionnée à l'étape précédente;
3. Calculer analytiquement le décalage subpixellique à partir de l'équation du plan ajusté à l'étape précédente.

Les tests réalisées au chapitre 4 sur des images simulées (représentant une cible en mouvement de faible texture) indiquent que la méthode EQI est plus précise que la méthode prise en référence, i.e., dite méthode de Taylor, au détriment d'une augmentation limitée du temps de calcul.

3.2 Extension du filtre directionnel

Dans la littérature, le filtre directionnel (*Steerable filter* ou Steer) permet d'estimer la phase locale dans une image. Cette technique a été utilisée par plusieurs auteurs pour estimer un mouvement entre deux images inférieur au pixel. Pour les déplacements supérieur au pixel, les auteurs réalisent un sous-échantillonnage spatial préalable pour se ramener au cas du déplacement subpixellique précédent.

En comparaison de la littérature, nous proposons la méthode "Extended local phase using steerable filters" (EX-ST) dont le synoptique est représenté Fig. 1. Cette méthode traite les images en pleine résolution sans sous-échantillonnage préalable. Elle nécessite en contrepartie d'utiliser le résultat de déplacement pixellique estimé par les techniques de corrélation croisée.

Les tests réalisés au Chapitre 4 montre que la méthode proposée atteint une meilleure précision pour les faibles amplitudes de déplacements (inférieures au pixel) et reste robuste aux plus grandes amplitude de déplacements (supérieures au pixel).

Tableau 2: Erreur quadratique moyenne (EQM) sur la valeur de déplacement estimé par la méthode Steer originale de (J. G. Chen et al., 2015) et par la méthode EX-ST proposée dans la section § 3.4; résultats obtenus sur des images simulées d'une cible en mouvement sinusoidal d'amplitude 1 pixel et 4 pixels (chap. 4)

Algorithme	**Amplitude de 1 pixel**	**Amplitude de 4 pixels**
Steer	0.0146	NA
EX-ST	0.0122	0.0449

3.3 Analyse critique de la méthode QSF

La méthode d'ajustement de surface quadratique (*Quadratic Surface Fitting*, QSF) est une méthode de la littérature qui consiste à ajuster une surface quadratique au voisinage du pic de corrélation. La position du maximum de l'ellipsoïde indique la valeur du décalage subpixellique. Les résultats du Chapitre 4 sur des images simulées de cibles simples confirment les bonnes performances de cette méthode obtenues par différents auteurs de la littérature, e.g., (Alba et al., 2015; X. Tong et al., 2019; D. Zhang et al., 2016).

Dans ce chapitre, on reprend la base théorique de la méthode QSF, et on propose d'analyser les conditions de son fonctionnement. Deux théorèmes sont proposés pour assurer que la valeur maximale de la surface quadratique est située dans les limites de la grille d'analyse, i.e., le voisinage 3×3 centré sur le maximum de la corrélation ().

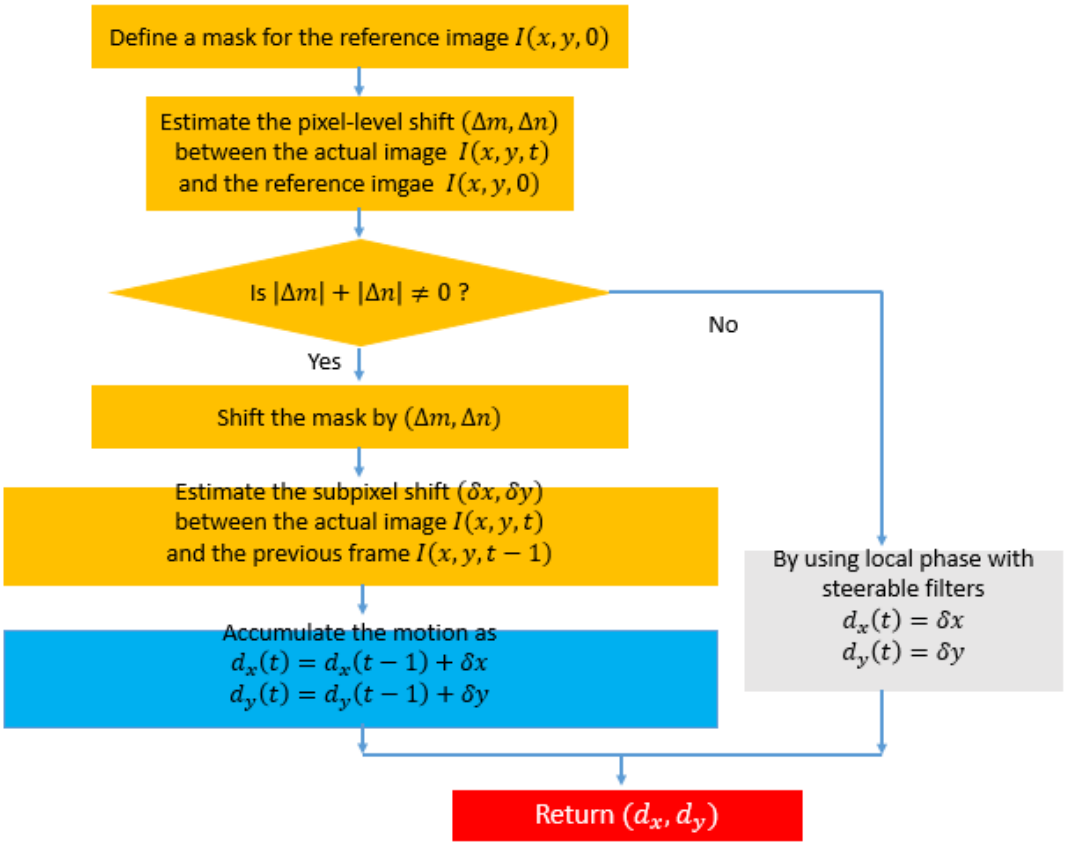


Figure 1: Synoptique de la méthode EX-ST proposée dans la section § 3.4.3, utilisant la phase locale pour estimer les déplacements multi-pixels

Chapitre 4 - Évaluation des performances des méthodes d'extraction de mouvement

Ce chapitre évalue et compare les performances des méthodes d'extraction de mouvement qui ont été présentées aux chapitres 2 et 3 à partir de flux d'images vidéos simulées. Le paragraphe 4.1 résume la génération des images simulées sur des cibles circulaires de faible texture. Le paragraphe 4.2 illustre quelques résultats de comparaison entre méthodes, qui sont détaillés dans le document principal. Le Chapitre 5 complète l'évaluation des méthodes par des tests en laboratoire sur une structure métallique simple.

4.1 Images synthétiques

Les images simulées que nous avons le plus utilisées, correspondent à un flux d'images synthétiques montrant le mouvement sinusoïdale dans une direction d'une cible circulaire claire sur un fond sombre. Les simulations permettent de générer une large variété de flux vidéos et d'illustrer la sensibilité des performances aux paramètres de l'image, notamment, le type et le niveau de bruit, le contraste de l'image, le type de cible, la netteté de l'image, et l'amplitude du mouvement (en pixels).

Les trois types d'images que nous avons traités sont les suivantes:

1. Images *anti-aliasing*: le niveau de gris des pixels du bord de la cible est déterminé selon les travaux de (Chojnacki & Szpak, 2019), qui modélise de façon simplifiée le processus physique de formation de l'image.
2. Images paramétriques : le niveau de gris de la cible est modélisé par une fonction gaussienne généralisée 2D, dont les paramètres contrôlent la transition du bord de la cible, autrement dit la netteté de l'image.
3. Images de forte texture (images de *speckle*) mises à disposition par la Société de Mécanique Expérimentale (SEM) pour le *"DIC challenge"* (Reu, 2011).

L'image synthétique de référence est ensuite utilisée pour générer un flux d'image vidéo. Pour l'évaluation des algorithmes, on suppose un simple mouvement sinusoïdal dans une direction prédéterminée. La durée du flux vidéo et le pas d'échantillonnage temporel sont adaptés à la fréquence du mouvement simulé. L'image synthétique initiale est ensuite décalée de la quantité voulue par l'une des 4 techniques existantes : i.e., transformation analytique, transformée de Fourier, sous-échantillonnage à partir d'une image de très grande résolution, interpolation de l'image par une méthode linéaire, bi-linéaire

ou cubique. Les tests réalisés dans ce chapitre indiquent que la méthode d'interpolation bilinéaire produit les meilleurs résultats.

Enfin, pour rendre les images plus réalistes, nous ajoutons aux images synthétiques un bruit additif et/ou un bruit multiplicatif aux flux d'images vidéos. D'après les tests réalisés avec la caméra rapide CMOS (voir section 5.3.1 du document général), le modèle du bruit peut dépendre du nombre de bits pour quantifier le niveau de gris. Un modèle additif suffit pour représenter des images de 8 bits. Le modèle combinant les 2 types de bruit est plus réaliste pour les images de 12 bits et plus. En outre, il est plus proche de la physique de formation de l'image dans un capteur CMOS d'après Toullier et al., 2020.

4.2 Résultats

Les méthodes d'extraction de mouvement sont évaluées quantitativement en comparant les variations de l'erreur de déplacement estimé et la valeur quadratique moyenne associée entre le déplacement total estimé et le déplacement simulé sur la totalité des images simulées.

Des tests préalables permettent d'évaluer la qualité de l'estimation du déplacement pixellique. On montre ainsi que la méthode ZNCC est la plus robuste au bruit sur des images simulées.

Pour illustrer les performances des différentes méthodes d'estimation de déplacement, nous analysons à la Figure 2 les variations de l'erreur d'estimation de 3 méthodes, i.e., Taylor, EQI et QSF, en fonction de l'amplitude du déplacement. La méthode de Taylor montre une erreur de forte amplitude qui dépend de l'amplitude du mouvement simulé de façon périodique. Par contraste, la méthode EQI montre une erreur de plus faible valeur.

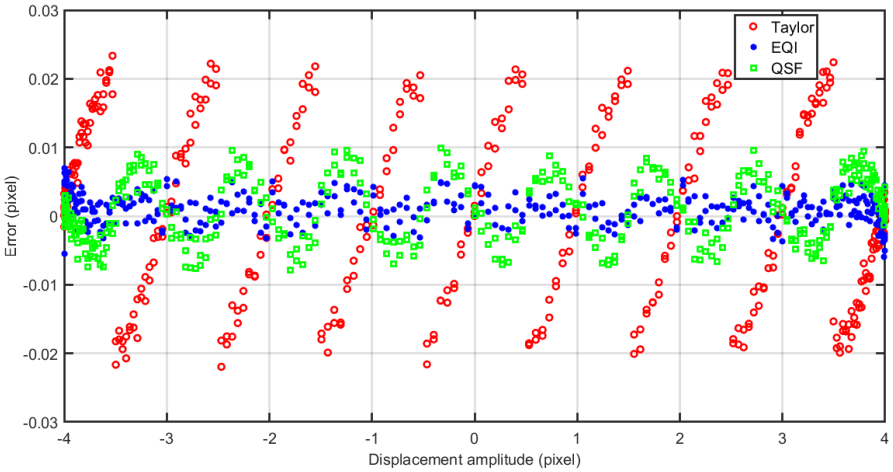


Figure 2: Erreur sur le déplacement estimé en fonction de l'amplitude du mouvement pour les méthodes de Taylor (rouge), EQI (bleu) et QSF (vert); les paramètres de simulation sont décrits au § 4.5 du document général.

Le tableau 3 compare les valeurs RMS de l'erreur de déplacement estimé de toutes les méthodes testées dans le document général. Les méthodes EQI et EQI-Octant proposées au chapitre 3 atteignent la plus faible valeur RMS, au détriment d'un temps de calcul plus long. Par ailleurs, la méthode EQI montre également davantage de robustesse que la méthode de Taylor vis-à-vis d'un flou éventuel de l'image.

Tableau 3: Comparaison des EQM sur le déplacement estimé obtenu par les 7 algorithmes sélectionnés sur un flux d'images vidéos simulées; voir § 4.5 du document général pour le détail des conditions de simulation.

Algorithme	**EQM**$_x$ (10^{-3} pixel)	**EQM**$_{xy}$ (10^{-3} pixel)	**Temps de calcul par trame** (ms)
Taylor	13.4	13.6	4.9
EQI	**3.5**	**4.6**	9.6
EQI-Octant	**4**	**4.6**	9.0
QSF	5.3	5.8	4.7
ECC	5	5.6	18.2
POC	5.1	6.2	**1.1**
UCC	**4**	**4.3**	3.4

Chapitre 5 - Analyse SHM expérimentale d'une barre métallique verticale

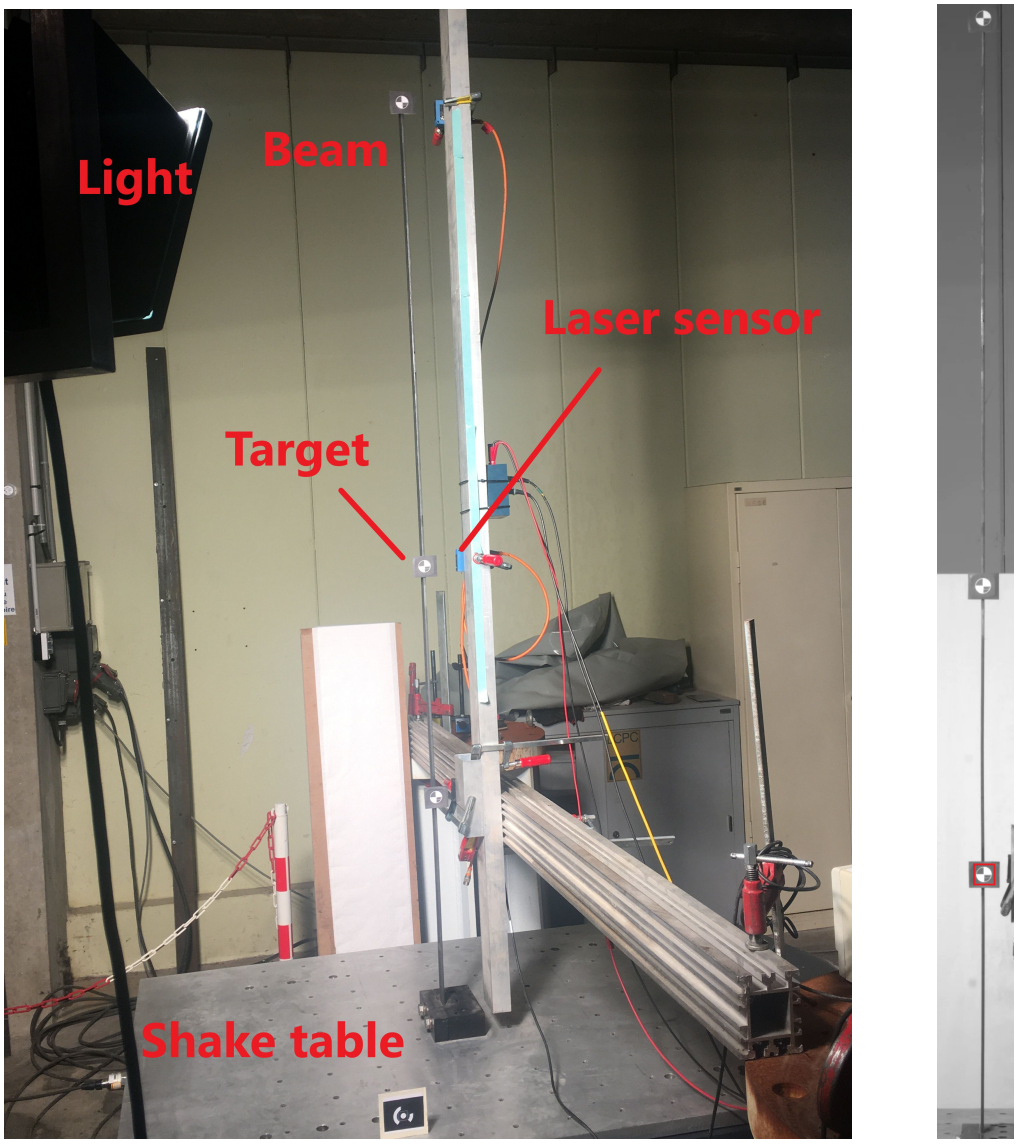
5.1 Montage expérimental et matériels de mesure

La Figure 2 illustre les essais réalisés en laboratoire pour tester les méthodes d'extraction de mouvement sur des images réelles et faire l'analyse SHM d'une structure mécanique simple.

La structure mécanique étudiée est une barre métallique verticale, fixée par sa base à une table vibrante. Pour tester les méthodes d'extraction de mouvement, 3 cibles géométriques en papier sont fixées à la barre à 1/4, 1/2 et 1/1 de la hauteur de la barre. Trois capteurs de déplacement laser (LDS) servent de référence aux mêmes altitudes que les cibles; ils ont été installés sur une poutre sans contact avec la structure étudiée.

Une analyse modale théorique de la barre verticale a été réalisée à l'aide du logiciel SolidWorks. Elle indique l'existence de 8 modes dont le plus élevé est à 170 Hz environ. Cette contrainte implique d'utiliser une caméra rapide de résolution suffisante, ici, une caméra Phantom 4K.

La structure est excitée par différentes sources dans la direction horizontale uniquement, perpendiculairement à la section de la barre: source de vibration ambiante, choc au marteau, lâcher de barre, table vibrante avec un mouvement sinusoïdal ou un bruit blanc. Auparavant, un premier test à vérifier la cohérence des résultats expérimentaux avec les résultats de simulation du chapitre 4. Pour ce faire, une cible a été installée directement sur la table vibrante, qui était excitée par un mouvement sinusoïdal pur.



(a) Experimental setup

(b) Raw frame

Figure 3: Montage experimental en laboratoire : (a- Experimental set up) barre métallique verticale montée sur une table vibrante; trois cibles en papier sont collées sur la barre à 1/4, 1/2 et 1/1 de la hauteur; trois capteurs de déplacement laser (LDS) servent de référence au niveau des cibles; (b- Raw frame) Image de la barre prise avec la caméra rapide Phantom; sur cette image, le fond d'image est volontairement divisé en 2 zones pour varier le contraste avec les contours de la cible.

5.2 Base de données et traitement

Les mesures expérimentales regroupent plusieurs séries d'images vidéos, dont la plupart ont été enregistrées avec la caméra Phantom à une cadence 500 images/seconde. La durée d'enregistrement varie de 10 secondes à quelques minutes, selon le nombre de pixels de l'image vs. la capacité de stockage de la caméra. La technologie CMOS de la caméra Phantom permet de n'acquérir que les pixels inclus dans un cadre restreint de l'image (*image cropping*). La taille du pixel dans le plan objet est de l'ordre de 1 mm.

Les enregistrements sont regroupés selon la source vibratoire utilisée, l'amplitude du mouvement vibratoire, la nature des cibles utilisées, la barre dans son état naturel ou avec l'ajout de masse. On distingue les différentes catégories suivantes :

— Sources vibratoires : signal aléatoire gaussien, choc au marteau, sinusoïde pure, bruit ambiant;

— Amplitude maximal subpixellique ou multi-pixellique;

— ROI constitués des cibles géométriques ou des contours naturels de la barre;

— Barre sans masse ou incluant une masse ponctuelle à différentes hauteurs

— Condition expérimentale intérieure ou extérieure; dans le cas de mesures à l'extérieur, seul le choc marteau et /ou vibration ambiante induite par le vent local étaient possibles.

Le traitement des flux d'images vidéo est réalisé en différé avec Matlab. A titre d'exemple, la Figure 4 illustre le synoptique du traitement du flux vidéo pour la méthode basée sur la phase locale, i.e., méthode *Extended Steerable Filter* (EX-ST) du § 3.4.3). Les performances des algorithmes d'estimation du déplacement des chapitres 2 et 3 sont comparées quantitativement en évaluant la valeur RMS de l'erreur entre le déplacement estimé et la mesure de référence donnée par le laser de déplacement au niveau des trois cibles géométriques. Ce calcul est réalisé au niveau des 3 cibles artificielles fixées à la barre, i.e., en haut, à mi-hauteur et au quart de hauteur de la barre.

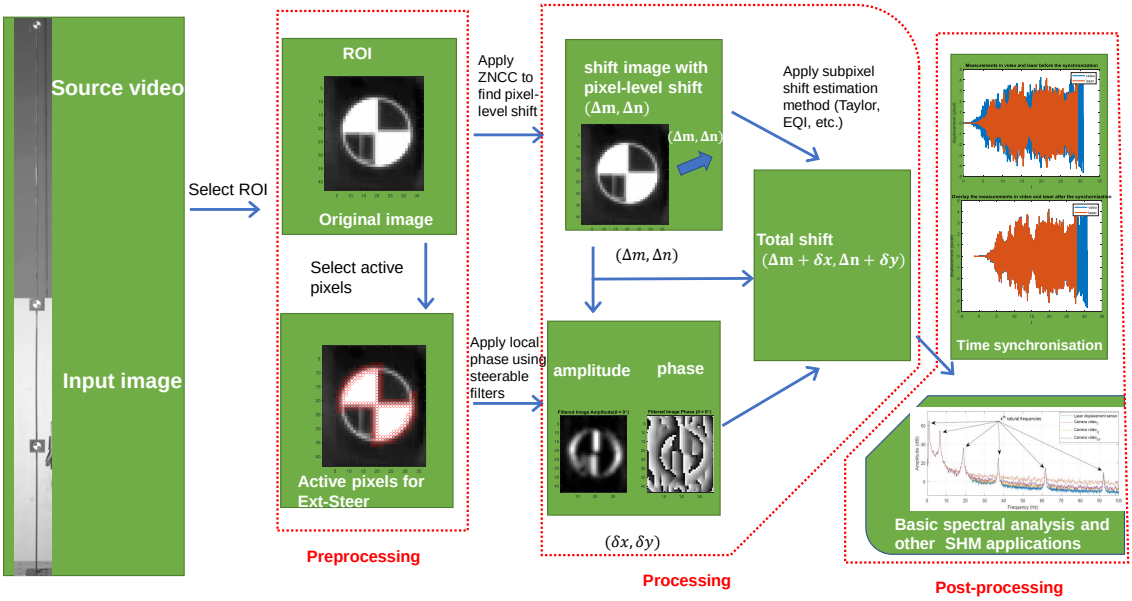


Figure 4: Synoptique du traitement du flux vidéo par la méthode EX-ST (*extended steerable filter* du § 3.4.3) lors de l'expérimentation; a) *Preprocessing*: sélection de zones ROI (cibles géométriques ici), sélection de pixels actifs dans le ROI; b) *Processing*: extraction successive du déplacement pixellique et subpixellique, calcul du déplacement total de la barre dans le ROI; c) *Post-processing*: (haut) Visualisation du mouvement total dans le ROI (en rouge) et comparaison avec la mesure de référence du LDS (en bleu); (bas) Représentation du spectre de vibration à partir du déplacement estimé (en rouge) et de la mesure de référence du LDS (bleu).

5.3 Principaux résultats

Les test réalisés permettent de distinguer les 3 résultats suivants :

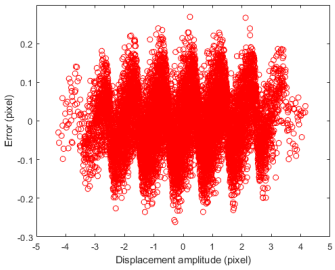
a) Évaluation de la mesure du déplacement de la barre

Les mesures expérimentales ont permis de poursuivre l'évaluation des algorithmes d'extraction de mouvement introduits aux chapitres 2 et 3. Le premier résultat concerne l'expérimentation où la cible géométrique est placée directement sur la table vibrante, et bouge horizontalement dans le plan image selon un mouvement sinusoïdal. Le traitement de ces images permet de vérifier la cohérence entre les résultats expérimentaux et les résultats obtenus en simulation au chapitre 4.

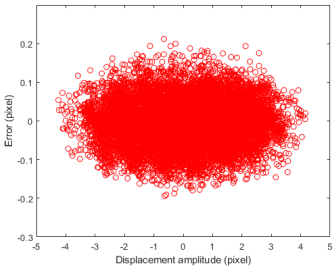
Ensuite, l'évaluation des algorithmes est réalisée sur la barre vibrante à partir des images des 3 cibles géométriques. La valeur RMS de l'erreur entre le déplacement estimé par

traitement d'image et les mesures des capteurs LDS permet d'évaluer quantitativement la précision des algorithmes d'extraction de mouvement. Cette évaluation est réalisée pour les différentes sources de vibrations mécaniques.

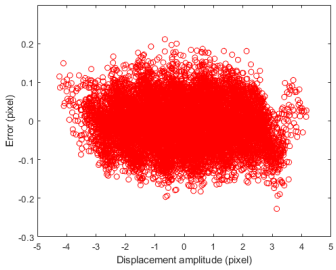
Les tests expérimentaux ont permis de valider tous les algorithmes introduits dans la thèse et de mettre en évidence les meilleures performances de la méthode EQI, qui a été introduite au Chapitre 3.



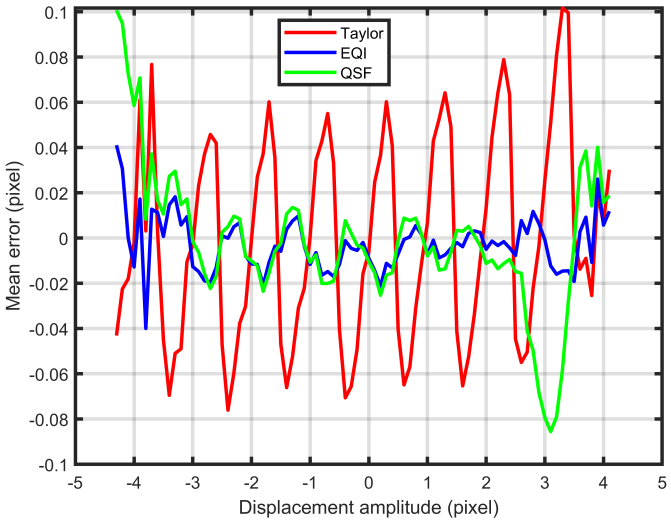
(a) Taylor



(b) EQI



(c) QSF



(d) Toutes les méthodes dans le test expérimental avec un processus de moyennage des données dans toutes les unités de longueur de 0,1 pixel dans les amplitudes de déplacement

Figure 5: L'erreur d'estimation vs. déplacement dans le test expérimental pour des methods Taylor, EQI et QSF

b) Évaluation de l'analyse modale de la barre

Les algorithmes d'identification stochastique à sous-espace sont appliqués sur les estimées de déplacement et permettent d'estimer les paramètres modaux de la structure. Dans un premier temps, cette analyse modale est réalisée sur les 3 cibles géométriques attachées à la barre pour estimer les fréquences de résonance des modes et leur facteur d'amortissement. D'après les résultats du Tableau 4, nous observons qu'entre les résultats d'identification provenant du LDS et ceux issus du traitement des images vidéos, l'erreur relative maximale est d'environ 0.6% sur la fréquence de résonance.

Tableau 4: Fréquences de résonance et amortissements des 8 premiers modes de la barre (et leur écart-type entre parenthèses) obtenus par analyse des données des 3 capteurs de référence LDS (à gauche) et des déplacements extraits des images vidéos sur les cibles géométriques (à droite)

Mode	Mesure de référence (LDS)		Estimée du déplacement par vidéo	
	$f_{mode}(\sigma_f)$, Hz	$\zeta_{damping}(\sigma_\zeta)$, %	$f_{mode}(\sigma_f)$, Hz	$\zeta_{damping}(\sigma_\zeta)$, %
1	1.0006 (0.007)	0.0145 (0.841)	0.9982 (0.009)	0.0476 (0.904)
2	6.7421 (0.010)	0.1881 (0.135)	6.7419 (0.012)	0.2412 (0.164)
3	19.0173 (0.013)	0.1044 (0.054)	18.9960 (0.022)	0.2620 (0.155)
4	37.3443 (0.018)	0.1537 (0.046)	37.3406 (0.019)	0.1319 (0.044)
5	62.0263 (0.027)	0.1263 (0.050)	62.0245 (0.022)	0.1335 (0.039)
6	92.5681 (0.017)	0.0522 (0.016)	92.5742 (0.029)	0.0858 (0.046)
7	129.6541 (0.029)	0.0737 (0.024)	129.6795 (0.022)	0.0749 (0.022)
8	172.1081 (0.026)	0.0570 (0.015)	172.1133 (0.032)	0.0502 (0.016)

Dans une second temps, l'image de la barre est décomposée en 160 ROIs à différentes hauteurs, dans lequel le mouvement est estimé. Chaque ROI agit ainsi comme autant de capteur virtuel. La multiplication des points de mesure le long de la barre permet d'améliorer la précision et la fiabilité de l'estimation des déformées modales de la barre. La Figure 5 illustre le résultat de ce traitement pour les 6 premiers modes de la barre. Les principales différences entre les méthodes d'extraction de mouvement interviennent dans la partie supérieure de la barre. La méthode EX-ST (*Extended Steerable Filter*, décrite au § 3.4.3) semble de ce point de vue la plus proche de la déformée modale théorique de la barre, qui a été calculée sous SolidWorks.

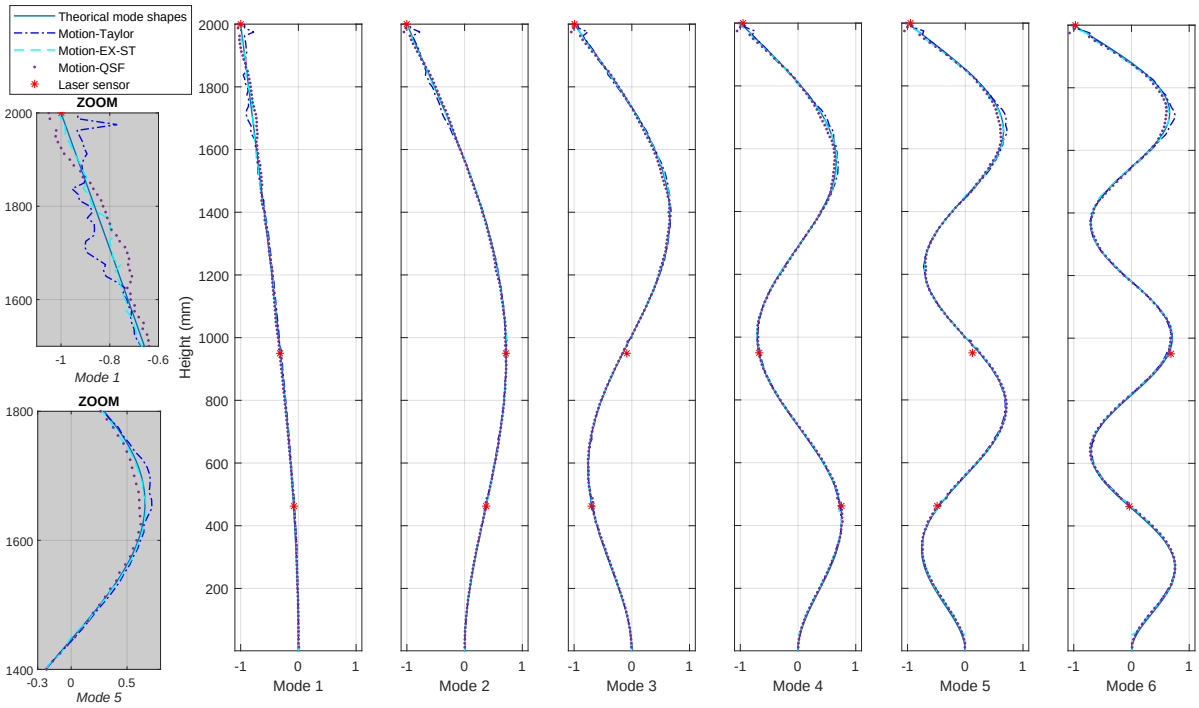


Figure 6: Déformées modales des 6 premiers modes de la barre, identifiées à partir de l'extraction du mouvement dans le flux vidéo sur un ensemble de 160 ROIs le long de la barre; en rouge, figure la déformée modale calculée à partir des 3 points de mesure laser (LDS); les deux figures à gauche en grisé représentent un zoom de la déformée modale des modes 1 et 2 dans la partie supérieure de la barre, avec en comparaison le résultat de la déformée théorique (obtenue par SolidWorks) et de 3 méthodes d'extraction de mouvement (Taylor au §2.2.1.4, EX-ST au §3.4.3, QSF au §3.5)

c) Test SHM

Dans une dernière série d'expérimentations, la réponse modale de la barre est perturbée par l'ajout d'un aimant à 9 hauteurs différentes de la barre, notée P2000 au sommet de la barre, à sa base, notée P280. Un flux d'images vidéo est enregistrée pour chaque position de l'aimant selon le même type de sollicitation vibratoire. Les résultats du paragraphe précédent (sans masse additionnelle) donne l'état initial de la barre.

Pour quantifier le changement des déformées modales entre les différentes positions de l'aimant et l'état initial de la barre, nous proposons deux méthodes : 1. mesure de l'écart entre les fréquences résonance avec et sans masse additionnelle, 2. calcul du critère d'assurance modale (MAC) (Greś et al., 2021).

Les résultats de la Figure 6 montrent que les deux méthodes proposées permettent au minimum de détecter les modifications d'un mode de vibration.

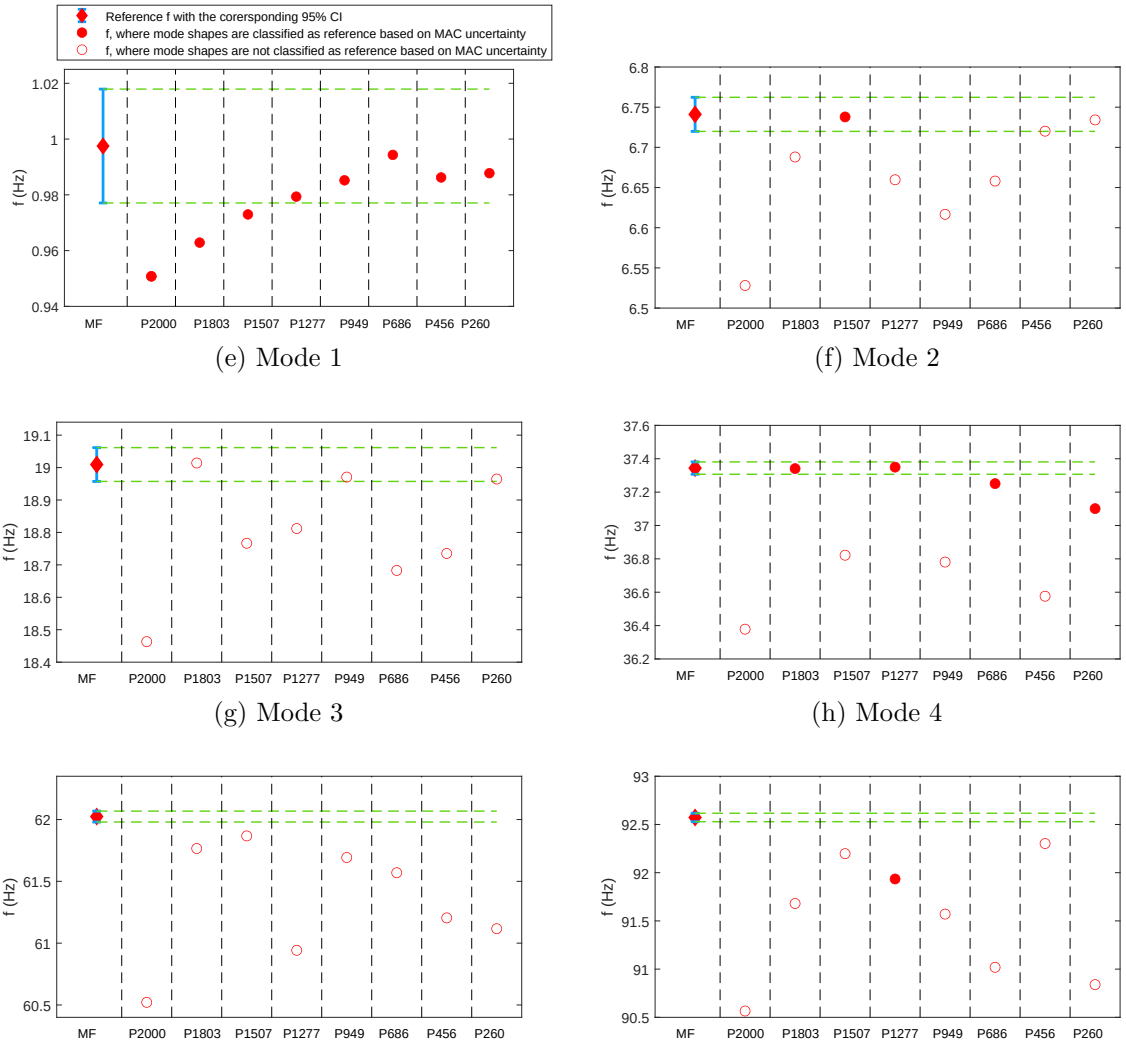
Conclusion

Parmi les diverses techniques vibratoires de SHM (accéléromètre, jauge de contrainte, GNSS, LDS, TLS, GB-SAR), cette thèse a présenté une technologie sans contact basée sur l'analyse d'un flux d'images vidéos. Ces méthodes sont des techniques prometteuses grâce à leur facilité de mise en oeuvre, leur faible coût et leur capacité de mesures plein champ.

Cette technique nécessite au préalable d'extraire le mouvement de la structure à partir de méthodes de traitement d'images. Parmi les méthodes existantes, nous avons proposé une nouvelle méthode intitulée "efficient quadrant interpolation" (EQI) et sa version améliorée EQI-Octant, la méthode à base de "steerable filtres" est étendue aux grands déplacements, et la fiabilité de la méthode QSF est améliorée.

L'évaluation des performances des algorithmes d'extraction de mouvement sélectionnés est réalisée sur des données simulées et expérimentales collectées en laboratoire sur une poutre métallique. Les méthodes proposées ont montré une valeur RMS de l'erreur d'estimation du déplacement plus faible que celle obtenue avec les méthodes existantes (méthode d'approximation de Taylor, QSF, ECC, etc), et une meilleure robustesse aux amplitudes de déplacement.

Finalement, la technique d'identification de sous-espace stochastique a été appliquée aux signaux de mouvement extraits du traitement du flux d'images vidéos. On a pu évalué la précision de l'analyse modale ainsi obtenue à partir des essais expérimentaux. Enfin, le critère d'assurance modale (MAC) a permis de détecter de faibles modifications des déformées modales de la barre à laquelle on a ajouté une masse ponctuelle à différentes hauteurs.



(e) Mode 1

(f) Mode 2

(g) Mode 3

(h) Mode 4

(i) Mode 5

(j) Mode 6

Figure 7: Détection du changement des déformées modales des 6 premières modes de la barre selon la position de la masse additionnelle le long de la barre: état de référence (mass-free), P2000 (sommet), P1803, P1507, P1277, P949, P686, P456 et P280 (base) à partir de la valeur d'incertitude du MAC (*Modal Assurance Criterion*). Les disques pleins indiquent que la forme du mode est inchangée par rapport à la mesure de référence; à l'opposé, les cercles révèlent que la forme du mode est modifiée.

BIBLIOGRAPHY

Abdullahi, S., Mustapha, N., Habaebi, M., & Islam, M., (2019), Accelerometer based structural health monitoring system on the go: developing monitoring systems with ni labview, *International Journal of Online and Biomedical Engineering (iJOE)*, *15*, 32, https://doi.org/10.3991/ijoe.v15i07.10427

Alba, A., Vigueras-Gomez, J. F., Arce-Santana, E. R., & Aguilar-Ponce, R. M., (2015), Phase correlation with sub-pixel accuracy: a comparative study in 1d and 2d, *Computer Vision and Image Understanding*, *137*, 76–87, https://doi.org/https://doi.org/10.1016/j.cviu.2015.03.011

Allemang, R., (2003), The modal assurance criterion - twenty years of use and abuse [cited By 860], *Sound and Vibration*, *378*, 14–21, https://www.scopus.com/inward/record.uri?eid=2-s2.0-0141958862&partnerID=40&md5=63d18f0258b1716061384fdd79949466

Allen, J., (1977), Short term spectral analysis, synthesis, and modification by discrete fourier transform, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *253*, 235–238, https://doi.org/10.1109/TASSP.1977.1162950

Amafabia, D.-a. M., Montalvão, D., David-West, O., & Haritos, G., (2017), A review of structural health monitoring techniques as applied to composite structures, *Structural Durability & Health Monitoring*, *112*, 91–147, https://doi.org/10.3970/sdhm.2017.011.091

Bendat, J. S., & Piersol, A. G., (1993), *Engineering applications of correlation and spectral analysis* (2nd ed.), John Wiley & Sons.

Benveniste, A., & Fuchs, J.-J., (1985), Single sample modal identification of a non-stationary stochastic process, *IEEE Transactions on Automatic Control*, *301*, 66–74, https://doi.org/10.1109/TAC.1985.1103787

Bornert, M., Doumalin, P., Dupré, J.-C., Poilâne, C., ROBERT, L., Toussaint, E., & Wattrisse, B., (2012), Short remarks about synthetic image generation in the context of sub-pixel accuracy of digital image correlation, *ICEM15*, 4 p. https://hal.archives-ouvertes.fr/hal-00836213

Bornert, M., Doumalin, P., Dupré, J.-C., Poilâne, C., Robert, L., Toussaint, E., & Wattrisse, B., (2017), Shortcut in dic error assessment induced by image interpolation

used for subpixel shifting, *Optics and Lasers in Engineering*, *91*, 124–133, https://doi.org/10.1016/j.optlaseng.2016.11.014

Burton, A., & Radford, J., (1978), *Thinking in perspective: critical essays in the study of thought processes*, Methuen.

Carr, J., Baqersad, J., Niezrecki, C., Avitabile, P., & Slattery, M., (2012), Dynamic stress–strain on turbine blade using digital image correlation techniques part 1: static load and calibration, https://doi.org/10.1007/978-1-4614-2422-2_20

Cha, G., Sim, S.-H., Park, S., & Oh, T., (2020), Lidar-based bridge displacement estimation using 3d spatial optimization, *Sensors*, *20 24*, https://doi.org/10.3390/s20247117

Chan, S. H., Võ, D. T., & Nguyen, T. Q., (2010), Subpixel motion estimation without interpolation, *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 722–725, https://doi.org/10.1109/ICASSP.2010.5495054

Chapeleau, X., Blanc, J., Hornych, P., Gautier, J.-L. J., & Carroget, J., (2017), Assessment of cracks detection in pavement by a distributed fiber optic sensing technology, *Journal of Civil Structural Health Monitoring*, *74*, 459–470, https://doi.org/10.1007/s13349-017-0236-5

Chen, J., (2016), *Video camera-based vibration measurement of infrastructure* (Doctoral dissertation).

Chen, J. G., Wadhwa, N., Cha, Y.-J., Durand, F., Freeman, W. T., & Buyukozturk, O., (2015), Modal identification of simple structures with high-speed video using motion magnification, *Journal of Sound and Vibration*, *345*, 58–71, https://doi.org/https://doi.org/10.1016/j.jsv.2015.01.024

Chojnacki, W., & Szpak, Z. L., (2019), Determining ellipses from low-resolution images with a comprehensive image formation model, *Journal of the Optical Society of America A*, *36 2*, 212, https://doi.org/10.1364/josaa.36.000212

Chou, J.-Y., & Chang, C.-M., (2021), Image motion extraction of structures using computer vision techniques: a comparative study, *Sensors*, *21 18*, https://doi.org/10.3390/s21186248

Cost action TU1406, part C, Innovation on Technologies, (2021), https://www.tu1406.eu/

Desdemona project home page, (2021), http://www.desdemonaproject.eu/

Diamond, D., Heyns, S., & Oberholster, A., (2017), Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence,

Measurement, *95*, 166–172, https://doi.org/https://doi.org/10.1016/j.measurement.2016.10.021

Doebling, S. W., Farrar, C. R., & Goodman, R. S., (1996), Effects of measurement statistics on the detection of damage in the alamosa canyon bridge, https://www.osti.gov/biblio/432966

Dong, C., (2019), *Investigation of computer vision concepts and methods for structural health monitoring and identification applications (2019). electronic theses and dissertations* (Doctoral dissertation).

Farrar, C. R., Baker, W. E., Bell, T. M., Cone, K. M., Darling, T. W., Duffey, T. A., Eklund, A., & Migliori, A., (1994), Dynamic characterization and damage detection in the i-40 bridge over the rio grande, https://doi.org/10.2172/10158042

Feng, D., & Feng, M. Q., (2018), Computer vision for shm of civil infrastructure: from dynamic response measurement to damage detection – a review, *Engineering Structures*, *156*, 105–117.

Feng, D., Feng, M. Q., Ozer, E., & Fukuda, Y., (2015), A vision-based sensor for noncontact structural displacement measurement, *Sensors*, *15 7*, 16557–16575, https://doi.org/10.3390/s150716557

Fienup, J., (1997), Invariant error metrics for image reconstruction, *Applied optics*, *36 32*, 8352–8357, https://doi.org/10.1364/AO.36.008352

Fischer, P., Dosovitskiy, A., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., Van Der Smagt, P., Cremers, D., & Brox, T., (2015), Flownet: learning optical flow with convolutional networks.

Freeman, W., & Adelson, E., (1991), The design and use of steerable filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13 9*, 891–906, https://doi.org/10.1109/34.93808

Greś, S., Döhler, M., & Mevel, L., (2021), Uncertainty quantification of the modal assurance criterion in operational modal analysis, *Mechanical Systems and Signal Processing*, *152*, 107457.

Gürkan, K., Dindar, A. A., Akbaba, S., Onur Aykut, N., & Akpmar, B., (2017), Noncontact and real-time displacement measurement system for structural health monitoring, *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)*, 1099–1102.

Hansard, M., Lee, S., Choi, O., & Horaud, R., (2012), *Time of Flight Cameras: Principles, Methods, and Applications*, Springer, https://doi.org/10.1007/978-1-4471-4658-2

Hermans, L., & Van Der Auweraer, H., (1999), Modal testing and analysis of structures under operational conditions: industrial application, *Mechanical Systems and Signal Processing*, *13 2*, 193–216, https://doi.org/https://doi.org/10.1006/mssp.1998.1211

Hoge, W. S., (2003), A subspace identification extension to the phase correlation method, *IEEE Trans Med Imaging*, *22 2*, 277–280.

Horn, B. K., & Schunck, B. G., (1981), Determining optical flow, *Artificial Intelligence*, *17 1*, 185–203, https://doi.org/https://doi.org/10.1016/0004-3702(81)90024-2

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., & Brox, T., (2016), Flownet 2.0: evolution of optical flow estimation with deep networks.

Infrastar training school, (2021), http://trainingschool.infrastar.eu

Irani, M., & Peleg, S., (1991), Improving resolution by image registration, *CVGIP: Graphical Models and Image Processing*, *53 3*, 231–239, https://doi.org/https://doi.org/10.1016/1049-9652(91)90045-L

Jähne, B., (2005), *Digital image processing 6th edition*, Springer.

James, I., George H., Carne, T. G., & Lauffer, J. P., (1993), The natural excitation technique (next) for modal parameter extraction from operating wind turbines, NASA STI/Recon Technical Report N, https://www.osti.gov/biblio/6695620

Javh, J., Slavič, J., & Boltežar, M., (2017), The subpixel resolution of optical-flow-based modal analysis, *Mechanical Systems and Signal Processing*, *88*, 89–99, https://doi.org/https://doi.org/10.1016/j.ymssp.2016.11.009

Kang, D., & Cha, Y.-J., (2018), Autonomous uavs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging, *Computer-Aided Civil and Infrastructure Engineering*, *33 10*, 885–902, https://doi.org/https://doi.org/10.1111/mice.12375

Keys, R., (1981), Cubic convolution interpolation for digital image processing, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *29 6*, 1153–1160, https://doi.org/10.1109/TASSP.1981.1163711

Ko, J. M., Wong, C. W., & Lam, H. F., (1994), Damage Detection in Steel Framed Structures by Vibration Measurement Approach, *Proceedings of the 12th International Modal Analysis*, *2251*, 280.

Konnik, M., & Welsh, J., (2014), High-level numerical simulations of noise in ccd and cmos photosensors: review and tutorial.

Lee, S., Lee, M., & Kang, M., (2018), Poisson-gaussian noise analysis and estimation for low-dose x-ray images in the nsct domain, *Sensors*, *18*, https://doi.org/10.3390/s18041019

Lin, Y., & Lan, Z., (2010), Sub-pixel displacement measurement in digital image correlation using particle swarm optimization, *2010 International Conference on Information, Networking and Automation (ICINA)*, *2*, V2-497-V2–501, https://doi.org/10.1109/ICINA.2010.5636461

Liu, C., Freeman, W., Szeliski, R., & Kang, S. B., (2006), Noise estimation from a single image, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, *1*, 901–908, https://doi.org/10.1109/CVPR.2006.207

Liu, P., & Du, X., (2019), Research and improvement of the formula of correlation coefficient in image processing, *5th International Conference on Advanced Computing, Networking and Security (ADCONS 2019)*, 1–10.

Liu, X., Masayuki, T., & Okutomi, M., (2013), Estimation of signal dependent noise parameters from a single image, *2013 IEEE International Conference on Image Processing*, 79–82, https://doi.org/10.1109/ICIP.2013.6738017

Lucas, B. D., & Kanade, T., (1981), An iterative image registration technique with an application to stereo vision, *Proceedings of the 7th Int. Joint Conference on Artificial intelligence, IJCAI'81*, *2*, 674–679.

Lundstrom, T., Baqersad, J., Niezrecki, C., & Avitabile, P., (2012), Using high-speed stereophotogrammetry techniques to extract shape information from wind turbine/rotor operating data, https://doi.org/10.1007/978-1-4614-2419-2_26

Merainani, B., Döhler, M., Xiong, B., Baltazart, V., Zhang, Q., Mével, L., Dumoulin, J., Rinaldi, C., Ciambella, J., & Gattulli, V., (2021), *Identification techniques by image processing/structural modal analysis using computer vision-based methods*, Desdemona report D5.4.

Merainani, B., Xiong, B., Baltazart, V., Dumoulin, J., Döhler, M., & Zhang, Q., (2021), Experimental investigation of structural modal identification using pixels intensity and motion signals from video-based imaging devices: performance, comparison and analysis, *In* E. Stella (Ed.), *Multimodal sensing and artificial intelligence: technologies and applications ii* (pp. 31–40), SPIE, https://doi.org/10.1117/12.2595019

Meyer, C., (2000), *Matrix analysis and applied linear algebra*, SIAM, https://books.google.fr/books?id=-7JeAwAAQBAJ

Monserrat, O., (2012), *Deformation measurement and monitoring with gb-sar* (Doctoral dissertation).

Nassif, H. H., Gindy, M., & Davis, J., (2005), Comparison of laser doppler vibrometer with contact sensors for monitoring bridge deflection and vibration, *NDT&E Int.*, *38*, 213–218, https://doi.org/10.1016/j.ndteint.2004.06.012

O'Meara, O. T., (1973), *Introduction to quadratic forms*, Springer-Verlag.

Pan, B., Qian, K., Xie, H., & Asundi, A., (2009), Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review, *Measurement Science and Technology*, *20 6*, 062001, https://doi.org/10.1088/0957-0233/20/6/062001

Pan, B., Xie, H., Guo, Z., & Hua, T., (2007), Full-field strain measurement using a two-dimensional Savitzky-Golay digital differentiator in digital image correlation, *Optical Engineering*, *46 3*, 1–10, https://doi.org/10.1117/1.2714926

Pan, B., Xu, B.-Q., Chen, D., & Feng, J., (2005), Sub-pixel registration using quadratic surface fitting in digital image correlation, *Acta Metrologica Sinica*, *26*, 128–134.

Park, H., Lee, H., Adeli, H., & Lee, I., (2007), A new approach for health monitoring of structures: terrestrial laser scanning, *Computer-Aided Civil and Infrastructure Engineering*, *22 1*, 19–30, https://doi.org/10.1111/j.1467-8667.2006.00466.x

Peeters, B., & De Roeck, G., (1999), Reference-based stochastics subspace identification for output-only modal analysis, *Mechanical Systems and Signal Processing*, *13 6*, 855–878, https://doi.org/https://doi.org/10.1006/mssp.1999.1249

Pintelon, R., Guillaume, P., Rolain, Y., Schoukens, J., & Van Hamme, H., (1994), Parametric identification of transfer functions in the frequency domain-a survey, *IEEE Transactions on Automatic Control*, *39 11*, 2245–2260, https://doi.org/10.1109/9.333769

Psarakis, E., & Evangelidis, G., (2005), An enhanced correlation-based method for stereo correspondence with sub-pixel accuracy, *10th IEEE International Conference on Computer Vision (ICCV)*, *1*, 907–912, https://doi.org/10.1109/ICCV.2005.33

Rais, M., Morel, J., & Facciolo, G., (2015), Iterative gradient-based shift estimation: to multiscale or not to multiscale?, *In* A. Pardo & J. Kittler (Eds.), *Progress in pattern recognition, image analysis, computer vision, and applications - 20th iberoamerican congress, CIARP 2015, montevideo, uruguay, november 9-12, 2015, proceedings* (pp. 416–423), Springer, https://doi.org/10.1007/978-3-319-25751-8_50

Ratcliffe, C., (1997), Damage detection using a modified laplacian operator on mode shape data, *Journal of Sound and Vibration*, *204 3*, 505–517, https://doi.org/https://doi.org/10.1006/jsvi.1997.0961

Reu, P. L., (2011), Experimental and numerical methods for exact subpixel shifting, *Experimental Mechanics*, *51*, 443–452.

Rytter, A., (1993), *Vibrational based inspection of civil engineering structures* (Doctoral dissertation No. 44) [Ph.D.-Thesis defended publicly at the University of Aalborg, April 20, 1993 PDF for print: 206 pp.], Denmark, Dept. of Building Technology; Structural Engineering, Aalborg University.

Sabato, A., Valente, N. A., & Niezrecki, C., (2020), Development of a camera localization system for three-dimensional digital image correlation camera triangulation, *IEEE Sensors Journal*, *2019*, 11518–11526, https://doi.org/10.1109/JSEN.2020.2997774

Salawu, O., (1997), Detection of structural damage through changes in frequency: a review, *Engineering Structures - ENG STRUCT*, *19*, 718–723, https://doi.org/10.1016/S0141-0296(96)00149-6

Salawu, O. S., & Williams, C., (1994), Damage Location Using Vibration Mode Shapes, *Proceedings of the 12th International Modal Analysis*, *2251*, 933.

Sharma, V., (2008), *Laser doppler vibrometer for efficient structural health monitoring* (Doctoral dissertation).

Shih, C., Tsuei, Y., Allemang, R., & Brown, D., (1988), Complex mode indication function and its applications to spatial domain parameter estimation, *Mechanical Systems and Signal Processing*, *2 4*, 367–377, https://doi.org/https://doi.org/10.1016/0888-3270(88)90060-X

Shimizu, M., & Okutomi, M., (2002), Precise sub-pixel estimation on area-based matching, *Systems and Computers in Japan*, *33*, 1–10, https://doi.org/10.1002/scj.10098

Singh, S., & Ganotra, D., (2012), Modifications in normalized cross correlation expression for template matching applications.

Smith, P., (1981), Bilinear interpolation of digital images, *Ultramicroscopy*, *6 1*, 201–204, https://doi.org/https://doi.org/10.1016/S0304-3991(81)80199-4

Sohn, H., & Park, B., (2021), Laser-based structural health monitoring, *Encyclopedia of Earthquake Engineering*, https://doi.org/10.1007/978-3-642-36197-5_86-1

Spencer, B. F., Hoskere, V., & Narazaki, Y., (2019), Advances in computer vision-based civil infrastructure inspection and monitoring, *Engineering*, *5 2*, 199–222, https://doi.org/https://doi.org/10.1016/j.eng.2018.11.030

Stone, H., Orchard, M., Chang, E.-C., & Martucci, S., (2001), A fast direct fourier-based algorithm for subpixel registration of images, *Geoscience and Remote Sensing, IEEE Transactions on*, *39*, 2235–2243, https://doi.org/10.1109/36.957286

Su, Y., Zhang, Q., Fang, Z., Wang, Y., Liu, Y., & Wu, S., (2018), Elimination of systematic error in digital image correlation caused by intensity interpolation by introducing position randomness to subset points, *Optics and Lasers in Engineering*, *114*, 60–75, https://doi.org/10.1016/j.optlaseng.2018.10.012

Takita, K., Aoki, T., Sasaki, Y., Higuchi, T., & Kobayashi, K., (2003), High-accuracy subpixel image registration based on phase-only correlation, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *E86-A8*, 1925–1934.

Talai, S. M., Desai, D. A., & Heyns, P. S., (2016), Infrared thermography applied to the prediction of structural vibration behaviour, *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, *13*.

Tennakoon, R., Bab-Hadiashar, A., Cao, Z., Hoseinnezhad, R., & Suter, D., (2015), Robust model fitting using higher than minimal subset sampling, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1.

Thurman, S. T., Fienup, J. R., & Guizar-Sicairos, M., (2008), Efficient subpixel image registration algorithms, *Ol/33/2/ol Pdf*, *33 2*, 156–.

Tong, W., (2005), An evaluation of digital image correlation criteria for strain mapping applications, *Strain*, *41 4*, 167–175, https://doi.org/https://doi.org/10.1111/j.1475-1305.2005.00227.x

Tong, X., Ye, Z., Xu, Y., Gao, S., Xie, H., Du, Q., Liu, S., Xu, X., Liu, S., Luan, K., & Stilla, U., (2019), Image registration with fourier-based image correlation: a comprehensive review of developments and applications, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12 10*, 4062–4081, https://doi.org/10.1109/JSTARS.2019.2937690

Torres, B., Payá-Zaforteza, I., Calderón, P. A., & Adam, J. M., (2011), Analysis of the strain transfer in a new fbg sensor for structural health monitoring, *Engineering Structures*, *33 2*, 539–548, https://doi.org/https://doi.org/10.1016/j.engstruct.2010.11.012

Toullier, T., Dumoulin, J., & Mevel, L., (2020), Study and development of an image synthesis software for the simulation of in-situ 3D scenes monitored by infrared cameras,

QIRT 2020 - 15th Quantitative InfraRed Thermography conference, https://hal.inria.fr/hal-02986528

Tuan, P. Q., (2006), *Spatiotonal adaptivity in super-resolution of under-sampled image sequences* (Doctoral dissertation), http://resolver.tudelft.nl/uuid:66da5bdb-aa60-49e0-b2fe-c35aa0585d3f

Varanasi, M. K., & Aazhang, B., (1989), Parametric generalized gaussian density estimation, *The Journal of the Acoustical Society of America*, *86 4*, 1404–1415, https://doi.org/10.1121/1.398700

Viberg, M., (1995), Subspace-based methods for the identification of linear time-invariant systems [Trends in System Identification], *Automatica*, *31 12*, 1835–1851, https://doi.org/https://doi.org/10.1016/0005-1098(95)00107-5

Wadhwa, N., Rubinstein, M., Durand, F., & Freeman, W. T., (2013), Phase-based video motion processing, *ACM Transactions on Graphics (TOG)*, *32 4*, 1–10.

Wikipedia contributors, (2021), List of bridge failures — Wikipedia, the free encyclopedia [[Online; accessed 21-October-2021]], https://en.wikipedia.org/w/index.php?title=List_of_bridge_failures&oldid=1048684778

Xiong, B., Zhang, Q., & Baltazart, V., (2021), Matlab code of QSF counterexample [Download from http://people.rennes.inria.fr/Qinghua.Zhang/QSFexample.zip].

Xu, M., Li, T., & Ping, X., (2009), A new model of nature images based on generalized gaussian distribution, *2009 WRI International Conference on Communications and Mobile Computing*, *1*, 446–450, https://doi.org/10.1109/CMC.2009.302

Ye, Z., Tong, X., Xu, Y., Gao, S., & Liu, X., (2018), An improved subpixel phase correlation method with application in videogrammetric monitoring of shaking table tests, *Photogrammetric Engineering and Remote Sensing*, *84 9*, 579–592.

Yu, J., Yan, B., Meng, X., Shao, X., & Ye, H., (2016), Measurement of bridge dynamic responses using network-based real-time kinematic gnss technique, *Journal of Surveying Engineering*, *142*, 04015013, https://doi.org/10.1061/(ASCE)SU.1943-5428.0000167

Zhang, D., Guo, J., Lei, X., & Zhu, C., (2016), A high-speed vision-based sensor for dynamic vibration analysis using fast motion extraction algorithms, *Sensors*, *16 4*, https://doi.org/10.3390/s16040572

Zhang, L., Wang, T., Jiang, Z., Qian, K., Liu, Y., Liu, Z., Tang, L., & Dong, S., (2015), High accuracy digital image correlation powered by gpu-based parallel computing, *Optics and Lasers in Engineering*, *69*, 7–12, https://doi.org/https://doi.org/10.1016/j.optlaseng.2015.01.012

Zhang, Z., Kang, Y., Wang, H., Qin, Q., Qiu, Y., & Li, X., (2006), A novel coarse-fine search scheme for digital image correlation method, *Measurement*, *39 8*, 710–718, https://doi.org/https://doi.org/10.1016/j.measurement.2006.03.008

Zhao, T., Zhang, W., & Guo, W., (2019), Digital image correlation analysis of displacement based on corrected three surface fitting algorithm, *Complexity*, *2019*, 1–9, https://doi.org/10.1155/2019/4620858

Zona, A., (2021), Vision-based vibration monitoring of structures and infrastructures: an overview of recent applications, *Infrastructures*, *6 1*, https://doi.org/10.3390/infrastructures6010004

UNIVERSITÉ DE RENNES 1

Titre : Analyse vibratoire par traitement d'images vidéo pour la surveillance de structures de génie civil

Mot clés : Traitement d'images, Extraction du mouvement, Analyse vibratoire, Corrélation d'images, Identification modale, Surveillance de santé structurelle, Génie civil

Résumé : Les techniques vibratoires de surveillance de santé structurelle des ouvrages sont généralement effectuées à partir de capteurs ponctuels tels que des accéléromètres. Par contraste, cette thèse développe une technique vibratoire basée sur l'analyse d'images vidéo de l'ouvrage, évitant ainsi l'installation fastidieuse de capteurs ponctuels. Le traitement du flux d'images vidéo permet d'extraire le mouvement des pixels sur l'ouvrage. Après la revue des méthodes existantes d'extraction de mouvement par traitement d'images, la thèse présente trois contributions : 1. une nouvelle méthode intitulée "EQI", 2. la méthode à base de filtre directionnel est étendue aux grands déplacements, et 3. la fiabilité de la méthode QSF est améliorée.
L'évaluation des performances des algorithmes d'extraction de mouvement sélectionnés est réalisée sur des données simulées et expérimentales collectées en laboratoire sur une poutre métallique. Finalement, la technique d'identification de sous-espace stochastique est appliquée sur les signaux de mouvement estimés sur l'image pour effectuer l'analyse modale, l'identification de la forme de mode et une détection de défaut. Les résultats démontrent la faisabilité des techniques d'extraction de mouvement basées sur le traitement d'images vidéo pour des application de santé structurelle d'ouvrages mécaniques.

Title: Video-based vibration analysis for structural health monitoring in civil engineering

Keywords: Image processing, Motion extraction, Image correlation, Structural health monitoring (SHM), Vibration analysis, Modal parameters, Stochastic subspace identification

Abstract: Vibration-based Structural Health Monitoring techniques are usually performed from point-like sensors such as accelerometers. By contrast, this thesis is focused on developing low cost video-based vibration techniques avoiding tedious point-like sensors installation. Image-based processing then enables to extract the motion signal from the video data flow. After the review of existing image-based motion extraction method, the thesis introduced three contributions: 1. a new method labelled "efficient quadrant interpolation", 2. the steerable filter is extended to large displacement, and 3. the reliability of the quadratic surface fitting method is improved.
The performance assessment of the selected motion-extraction algorithms is performed on both simulated and experimental data collected in laboratory on a cantilever beam. At the end, the stochastic subspace identification technique is applied on the image-based estimated motion signals to perform the modal analysis, the mode shape identification and fault detection. The results demonstrate the feasibility of video-based motion extraction techniques for SHM purposes.