AGH

AGH UNIVERSITY OF KRAKOW

FIELD OF SCIENCE: ENGINEERING AND TECHNOLOGY

SCIENTIFIC DISCIPLINE: AUTOMATION, ELECTRONIC, ELECTRICAL ENGINEERING AND SPACE TECHNOLOGIES

DOCTORAL DISSERTATION

Marker-free optical-flow-based automatic fatigue crack assessment in metallic structures under periodic excitation

Author: Adam Machynia

Supervisor: Ziemowit Dworakowski, PhD, DSc

Completed at: AGH University, Faculty of Mechanical Engineering and Robotics

Kraków, 2025

AGH

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

DZIEDZINA NAUK INŻYNIERYJNO-TECHNICZNYCH

DYSCYPLINA AUTOMATYKA, ELEKTRONIKA, ELEKTROTECHNIKA
I TECHNOLOGIE KOSMICZNE

ROZPRAWA DOKTORSKA

Marker-free optical-flow-based automatic fatigue crack assessment in metallic structures under periodic excitation

Autor: Adam Machynia

Promotor rozprawy: dr hab. inż. Ziemowit Dworakowski, prof. AGH

Praca wykonana: Akademia Górniczo-Hutnicza im. Stanisława Staszica
w Krakowie, Wydział Inżynierii Mechanicznej i Robotyki

Kraków, 2025

I would like to sincerely thank my supervisor, Ziemowit Dworakowski, PhD, DSc, for his guidance, trust, and continuous motivation. His valuable feedback and support greatly shaped this research.

I would also like to thank Krzysztof Holak, PhD and Kajetan Dziedziech, PhD, DSc for the many hours of valuable discussions that guided both my work and my development as a researcher.

Abstract

Fatigue cracking in metallic structures, particularly in aeronautical applications, remains a major cause of durability loss and operational risk. Evaluation of structural condition relies on Structural Health Monitoring (SHM) and complementary Nondestructive Testing (NDT) practices that aim to detect, localize, and assess damage without disrupting operation. Computer Vision (CV) methods are attractive in this context because they are non-contact and some are compatible with low-cost consumer-grade hardware. However, practical deployment is still impeded by several issues, including dependence on manual tuning of method parameters, limited robustness under varying conditions, and a lack of direct interpretability.

This dissertation develops a marker-free, CV-based workflow for automatic fatigue-crack assessment under periodic excitation, exploiting the crack-breathing phenomenon. The pipeline converts dense Optical Flow (OF) time series into frequency-domain amplitude maps, on which a scoring procedure localizes the crack tip automatically. The principal methodological contributions are the formulation of amplitude maps for visualization of breathing-induced motion around cracks, and an automated crack-tip localization procedure operating directly on these maps. In addition, a Local Orientation Inconsistency (LOI) metric is introduced to quantify the directional coherence of estimated motion in video recordings, enabling masking of regions with unreliably estimated motion and guiding Region of Interest (ROI) selection. The impact of phase-based motion magnification is also investigated, demonstrating its usefulness for amplitude maps generation in certain low-excitation cases.

The methodology is validated on multiple experimental specimens, including a compressor, a cantilever beam, and a plate-like aircraft-grade specimen, recorded under diverse setups with different measurement equipment, including a consumer-grade smartphone. The results demonstrate consistent performance across crack-growth stages and camera configurations. Quantitative evaluation against manual measurements shows that the Mean Absolute Error (MAE) of crack-tip localization is in most cases below, or very close to, the resolution of manual assessment, which is one millimeter. Practical aspects are examined through a comparison of key method parameter choices, confirming that the technique operates reliably without requiring precise tuning. The proposed processing pipeline reduces operator input, produces interpretable results, and remains feasible with consumer-grade hardware, supporting practical integration into NDT and SHM workflows.

Streszczenie

Pęknięcia zmęczeniowe w konstrukcjach metalowych, w szczególności w zastosowaniach lotniczych, są jedną z głównych przyczyn poważnych uszkodzeń, zwiększającą ryzyko eksploatacyjne. Ocena stanu konstrukcji inżynieryjnych opiera się na metodach monitoringu strukturalnego oraz badaniach nieniszczących, których celem jest wykrywanie, lokalizacja i ocena uszkodzeń bez wyłączania badanego obiektu z eksploatacji. W tym kontekście techniki oparte o wizję komputerową stanowią wartościowe podejście, ponieważ są bezkontaktowe, a niektóre z nich można stosować z wykorzystaniem relatywnie niedrogiego, ogólnodostępnego sprzętu pomiarowego. Jednakże praktyczne zastosowania są utrudnione ze względu na, między innymi, znaczną zależność technik wizyjnych od ręcznego doboru ich parametrów oraz zmiennych warunków eksperymentalnych, a także brak możliwości bezpośredniej interpretacji wyników.

W niniejszej pracy przedstawiono metodę automatycznej wizualizacji oraz detekcji położenia wierzchołka pęknięcia zmęczeniowego w strukturach metalicznych, poddanym wymuszeniu periodycznemu. Zaproponowana technika bazuje na analizie ruchu związanego z cyklicznym otwieraniem się i zamykaniem pęknięcia zmęczeniowego. Oparta jest ona na wizji komputerowej, ale nie wymaga zastosowania znaczników wizyjnych. Metoda wykorzystuje przekształcenie przepływu optycznego do dziedziny częstotliwości, co pozwala na przygotowanie wizualizacji wzorców ruchu związanych z cyklicznym otwieraniem się i zamykaniem pęknięcia, co daje możliwość automatycznego określenia położenia wierzchołka pęknięcia. Oryginalnymi aspektami rozprawy są przede wszystkim opracowanie wizualizacji ruchu związanego z pęknięciem oraz automatyczna procedura detekcji położenia wierzchołka pęknięcia. Ponadto zaproponowano metrykę LOI (ang. *Local Orientation Inconsistency*) służącą do oceny spójności kierunku ruchu estymowanego z nagrań wideo. Metryka ta pozwala na maskowanie regionów o niepewnej estymacji ruchu oraz wspomaga wybór obszaru odpowiedniego do dalszej analizy. Sprawdzono również wpływ zastosowania techniki wizyjnego wzmacniania ruchu na nagraniach wideo, wykazując jego użyteczność przy wizualizacji pęknięć zmęczeniowych w niektórych przypadkach relatywnie niskiego poziomu amplitudy wymuszenia.

Działanie zaproponowanych metod zweryfikowano na nagraniach przedstawiających trzy obiekty: kompresor, belkę wspornikową oraz próbkę lotniczą będącą strukturą płytową. Dane eksperymentalne zarejestrowano w zróżnicowanych konfiguracjach pomiarowych, z użyciem różnych kamer, w tym kamery smartfona. Uzyskane wyniki dowodzą poprawnego działania zaproponowanego podejścia dla kolejnych etapów propagacji pęknięcia oraz różnych konfiguracji ekspery-

mentalnych. Porównanie względem ręcznych pomiarów, wskazuje, że średni błąd bezwzględny lokalizacji wierzchołka pęknięcia jest w większości przypadków mniejszy lub bardzo zbliżony do rozdzielczości pomiarowej ręcznego pomiaru. Ponadto porównano działanie metody dla różnych wartości jej parametrów, potwierdzając jej poprawne funkcjonowanie bez konieczności ich precyzyjnego dopasowania. Zaproponowana procedura, przy nieznacznym działaniu operatora systemu oraz zastosowaniu sprzętu amatorskiego, zapewnia interpretowalne wyniki, co w przyszłości może pozwolić na jej integrację z procesami monitoringu strukturalnego i badaniami nieniszczącymi.

Contents

Glossary

CM Condition Monitoring

CNN Convolutional Neural Network

CV Computer Vision

DFT Discrete Fourier Transform

DIC Digital Image Correlation

DNN Deep Neural Network

FFT Fast Fourier Transform

LOI Local Orientation Inconsistency

MAE Mean Absolute Error

mROI Measurement Region Of Interest

NDE Nondestructive Evaluation

NDT Nondestructive Testing

OF Optical Flow

ROI Region Of Interest

SHM Structural Health Monitoring

SPC Statistical Process Control

sROI Scoring Region Of Interest

UAV Unmanned Aerial Vehicle

List of Figures

Chapter 1

Introduction

Assessment of the condition and detection of damage in engineering structures is a crucial engineering task aimed at ensuring safety and achieving economic benefits. Over time, structural deterioration, for example, due to material fatigue, corrosion, or environmental exposure, can compromise a structure's safety and functionality, potentially leading to failures. To prevent such outcomes, infrastructure is regularly inspected and monitored to enable early detection of damage. The goal is to minimize downtime and avoid costly unplanned outages by facilitating proactive maintenance before minor issues escalate.

Damage can be understood as a modification of a system's material properties or geometry that degrades its performance, safety, reliability, or remaining service life. It may manifest as an immediate failure or as a slight deterioration that accumulates over time. Its progression can be gradual, as in the case of fatigue or corrosion, or abrupt for example as a consequence of accidental or natural catastrophic events [1].

Among the various forms of structural degradation, fatigue represents a particularly critical failure mechanism, and it has been the subject of extensive study over many years [2]. Fatigue failure in metals proceeds in three stages under cyclic loading. The process begins with crack initiation at stress concentrators such as surface imperfections, sharp notches, or internal flaws that act as stress risers. In the subsequent stage of stable crack propagation, the crack advances incrementally with each cycle, typically perpendicular to the maximum tensile stress. Growth is initially slow at short lengths but accelerates as the crack extends, since the stress at the crack tip increases and the effective load-bearing cross-section is progressively reduced. Once the crack reaches a critical size, the remaining material can no longer sustain the applied loads, resulting in rapid fracture and sudden, final failure [3, 4].

In metallic aircraft structures, fatigue is the dominant mode of failure, more significant than in many other engineering domains, and it has been identified as a major contributing factor in numerous accidents and incidents [4]. Cracks are of particular concern because they reduce the structural stiffness of metallic components, thereby altering modal properties and the dynamic response [5]. Consequently, the analysis of fatigue crack growth is an important task in structural integrity assessment.

Fatigue cracks in aircraft structures are traditionally assessed through scheduled inspections using a range of Nondestructive Testing (NDT) techniques, such as ultrasonic, radiographic, or eddy-current testing [6, 7]. More recently, increasing attention has been directed toward Structural Health Monitoring (SHM) approaches that aim to automate this process and enable more frequent or even continuous monitoring of structural condition. Among these, non-contact techniques, including optical and Computer Vision (CV)-based methods, are of particular interest because they provide full-field measurements while reducing the need for extensive surface preparation or sensor installation. Despite these advances, there remains a need for methodologies that combine non-contact operation with robustness to experimental conditions and the ability to deliver results that are both reliable and directly interpretable for engineering assessment.

The practical challenge, therefore, is to enable early, non-contact assessment of fatigue damage under variable conditions and limited access, while still delivering interpretable evidence with minimal setup or surface preparation. To address this need, this dissertation develops and validates a marker-free, video-based methodology for fatigue-crack assessment under periodic excitation. The approach transforms camera recordings into frequency-domain amplitude maps derived from dense Optical Flow (OF) time series, and subsequently localizes the crack tip through an automated scoring procedure. The methodology is evaluated on an aeronautic plate-like structure and across different cameras, including consumer-grade device.

1.1 Overview of the dissertation

This dissertation is organized into ten chapters, outlined as follows.

Chapter 2: State of the art. Reviews the state of the art. It first surveys the role of CV) in SHM and then examines CV-based crack detection and assessment. The chapter also introduces motion estimation with an emphasis on OF and describes motion magnification techniques.

Chapter 3: Aim and scope of the dissertation. States the aim and scope of the dissertation.

Chapter 4: Experiments. Describes the experimental datasets used in the study.

Chapter 5: Local orientation inconsistency. Introduces the Local Orientation Inconsistency (LOI) metric as an indicator of OF reliability. The chapter details the calculation of the metric and presents experimental results.

Chapter 6: Amplitude maps. Explains the methodology for constructing amplitude maps, presents their variants, and demonstrates their use on the experimental datasets.

Chapter 7: Crack tip localization procedure. Presents an automatic crack-tip localization procedure based on amplitude maps.

Chapter 8: Plate-like structure case study. Validates the full processing pipeline on an aeronautic plate-like structure with a propagating crack. The influence of method parameters is evaluated in this chapter.

Chapter 9: Discussion and future work. Synthesizes findings, discusses practical implications, and outlines directions for future research.

Chapter 10: Concluding remarks. Concludes the dissertation by restating the contributions.

Note on language editing. The author acknowledges the use of ChatGPT tool exclusively for ensuring grammar correctness and language clarity in the preparation of this dissertation. All research design, data processing, analysis, and interpretation were carried out solely by the author.

Chapter 2

State of the art

Over the years, a variety of diagnostic techniques have been developed, ranging from traditional inspection procedures to advanced sensor-based monitoring frameworks. This chapter outlines the main categories of diagnostic techniques, with emphasis on SHM and NDT, and then describes the role of CV within this context, especially for crack detection and assessment. The concepts of OF and motion magnification are discussed in detail together with their applications in structural diagnostics, which are crucial to the methodology developed in this thesis.

A range of complementary approaches has been proposed in the literature, reflecting differences in sensing strategies, deployment scenarios, and application domains. Among these, four main categories of diagnostic techniques are generally distinguished: NDT (also commonly referred to as Nondestructive Evaluation (NDE) [8]), SHM, Condition Monitoring (CM), and Statistical Process Control (SPC). CM has traditionally been applied primarily to rotating machinery, while SPC is associated with chemical and material processes [8]. SHM is defined as "the process of implementing a damage identification strategy for aerospace, civil, and mechanical engineering infrastructure" [9]. In both scope and practice it overlaps with, yet remains distinct from, these related fields.

SHM is understood as a monitoring paradigm based on permanently installed, fixed-position sensors that acquire data online or at regular intervals with minimal human intervention, thereby enabling automated analysis over time. NDT refers to episodic, on-demand inspections using portable instruments that are brought to specific locations to detect and assess defects. Within the damage-identification hierarchy, SHM encompasses detection, localization, classification, assessment, and prediction of structural safety [10], whereas NDT typically focuses on detection, classification, and assessment, with localization often assumed in advance based on inspection planning, prior knowledge, or regulatory requirements. An NDT technique can migrate toward SHM when its sensing hardware becomes sufficiently inexpensive, robust, and durable for permanent installation, and when data acquisition and handling can be automated. However, many NDT techniques are limited in spatial coverage due to the local nature of the underlying physical phenomena, meaning that full incorporation into SHM would require dense sensor networks. In practice, SHM deployments therefore tend to target structurally significant elements or components where damage is more likely to occur, based on prior knowledge [11].

Within SHM, multiple complementary families of methods are in routine use, including vibration-based monitoring [1], guided ultrasonic waves [12], fiber-optic sensing with fiber Bragg gratings [13], and acoustic emission [14]. Another important group relies on CV, which is attractive because these methods are inherently nondestructive and enable non-contact measurements with relatively simple and low-cost hardware, while also providing full-field responses that can scale to large structures, although they remain sensitive to environmental and imaging conditions [15]. Recent reviews emphasize that CV-based SHM is developing rapidly, with growing emphasis on automation and scalability [16].

2.1 Computer Vision in Structural Health Monitoring

A typical CV-based monitoring system consists of two principal components: an image acquisition system and an image processing system [17]. The image acquisition system forms the foundation of a CV-based SHM framework. Its primary function is to capture visual data that reliably represent structural behavior under operational or controlled loading. Core elements include the imaging device (camera), appropriate optics (lens), and the observed object. A wide range of cameras may be employed, from high-resolution or high-speed devices to consumer-grade equipment. To minimize measurement errors, cameras are typically mounted on rigid supports such as tripods, although alternative solutions involve robotic arms or Unmanned Aerial Vehicles (UAVs) [18].

The acquisition setup may also include auxiliary illumination sources to ensure adequate contrast and image quality. Proper calibration of the imaging system, covering camera placement, lens selection, the use of external lighting if necessary, and adjustment of acquisition parameters, is essential for reliable image registration. Calibration depends on the characteristics of the monitored target and on the distance between the camera and the structure.

The image processing system constitutes the second essential component of a CV-based SHM framework [17]. Its role is to transform raw visual data into quantifiable indicators of structural integrity, such as displacement fields, deformation patterns, vibration modes, or measures of crack initiation and propagation. Processing workflows often begin with pre-processing steps, including noise reduction, stabilization, and contrast enhancement, to prepare the captured frames for analysis. Subsequently, feature extraction is often performed, which may involve techniques such as edge or texture analysis, region segmentation, or corner detection [19].

Different classifications of CV applications in SHM have been proposed in the literature. Ferraris et al. [16] distinguish four principal categories: vibration estimation, displacement estimation, general damage detection (covering deformations, deflections, and corrosion), and surface crack analysis. Dong and Catbas [19] instead group applications into local evaluations, such as crack detection, crack propagation monitoring, and corrosion assessment, and global analyses, such as structural response monitoring and load estimation.

2.2 Computer Vision-based crack detection and assessment

CV–based approaches represent an important class of methods for crack detection and assessment. They can be broadly divided into two categories: deep learning methods, particularly those based on Convolutional Neural Networks (CNNs), which have grown rapidly in recent years, and classical image processing techniques that do not involve machine learning. In this thesis, the emphasis is on metallic structures, with particular interest in approaches based on motion analysis.

Most applications of CNNs for crack detection rely on static image segmentation. Examples include fatigue crack detection in metallic components such as steel trestles [20] and large-scale structural elements in buildings [21]. Many studies address steel bridges, including welded joints [22, 23] and steel decks [24]. Crack detection in steel box girders has also been demonstrated using an ensemble of three CNNs [25]. Other applications include turbine inspections using image filtering [26] and wind turbine blade assessment from UAV-acquired images, where cracks and other types of damage were identified [27].

Hybrid strategies have also been investigated. Wang et al. [28] combined scale-invariant feature transform (SIFT) as a preprocessing step for multi-camera images with CNNs for crack detection in steel beams. Yu et al. [29] integrated strain data obtained via Digital Image Correlation (DIC) with CNNs to detect cracks in basalt fiber–reinforced polymer beams. Model-informed deep learning approaches integrate data-driven and model-based strategies. Shu et al. [30] combined vision-based displacement estimation for finite element model calibration with subsequent CNN-based damage identification on a synthetic truss structure. Video-based crack detection has also been explored in safety-critical applications, particularly in metallic components in nuclear power plants, using both deep learning [31] and classical machine-learning approaches [32].

Although deep learning methods are increasingly popular, they also present several limitations. A key challenge is the requirement for large annotated datasets, which are often labor-intensive and expensive to generate. Moreover, such datasets are frequently imbalanced and insufficiently diverse, limiting the generalization capability and robustness of trained models [33]. Another concern is the limited interpretability of deep networks, which makes it difficult to explain their predictions or to ensure reliability in safety-critical contexts. For these reasons, alternative approaches grounded in physical phenomena and based on classical image processing continue to serve as important tools in fatigue crack analysis.

A group of methods not related to machine learning, employs classical image processing techniques, with a particular emphasis on motion analysis. Jiang et al. [34] developed an approach using a high-speed camera system to record beam vibrations and compute a nonlinearity measure for crack detection and localization. Peng et al. [35] proposed a procedure to mitigate the effects of non-uniform illumination in detecting surface damage on wind turbine blades. Damage identification in beam-like structures can also be achieved through the analysis of free vibration recordings, as demonstrated by Garrido et al. [36] on a polypropylene beam.

Another group of methods exploits the crack breathing phenomenon, which occurs in vibrating structures containing a crack. In this process, the regions adjacent to the crack periodically move toward and away from each other, causing the crack to alternately open and close. Kong and Li [37] applied an image overlapping technique to align frames captured at different breathing stages (open and closed), followed by feature and intensity-based registration techniques to detect cracks. Mojidra et al. [38] extended this concept by introducing global motion compensation to account for overall structural movement thereby isolating motion associated with crack breathing. Their approach was implemented for both 2D and 3D videos and validated on steel plate specimens as well as bridge girder to cross frame connection plates subjected to cyclic fatigue loading.

A crucial component of crack assessment is measurement, as crack size directly reflects the severity of the defect and guides decisions regarding further inspection or structural reinforcement [24]. Within CV applications, DIC has emerged as an important technique for monitoring crack propagation. Applications include studies on aluminum alloy specimens [39, 40] and on a model of a single-span steel bridge deck [41]. A notable limitation of DIC, however, is the need to apply speckle patterns to the surface. In these works, surface preparation, specifically painting random speckle patterns, was required to achieve reliable measurements.

2.3 Motion estimation: Optical Flow

Motion estimation refers to techniques for determining movement in image sequences. It is a fundamental problem in CV, supporting tasks ranging from video compression and object tracking to 3D scene understanding. In the context of SHM and NDT, motion estimation enables non-contact measurement of how structures move or deform under loads, which is crucial for vibration analysis. Among the most widely used approaches is OF, which can be categorized as dense, computing a motion vector for every pixel, or sparse, where only selected features such as corners are tracked. Sparse methods are computationally efficient but rely on distinctive features, whereas dense OF provides more detailed information at the cost of higher computational effort [42].

2.3.1 The Optical Flow concept

In computer vision, OF is intuitively understood as a vector field representing the apparent displacement of pixels between two consecutive video frames. It is obtained by comparing two images from a sequence captured of the same scene at different time instants. Motion of objects within the scene, or movement of the camera itself, results in a shift of pixel intensities between frames. The corresponding projected velocity vector field describing this apparent motion is referred to as the OF [43].

In this work, an image is treated as a grayscale intensity function $I(x, y)$, where (x, y) denote pixel coordinates. An image sequence is thus denoted as $I(x, y, t)$, with t indicating time. The schematic in Fig. 2.1 illustrates the OF concept. The flow at location (x, y) and time t is

$$\mathbf{v}(x, y, t) = \begin{bmatrix} u(x, y, t) \\ v(x, y, t) \end{bmatrix}.$$

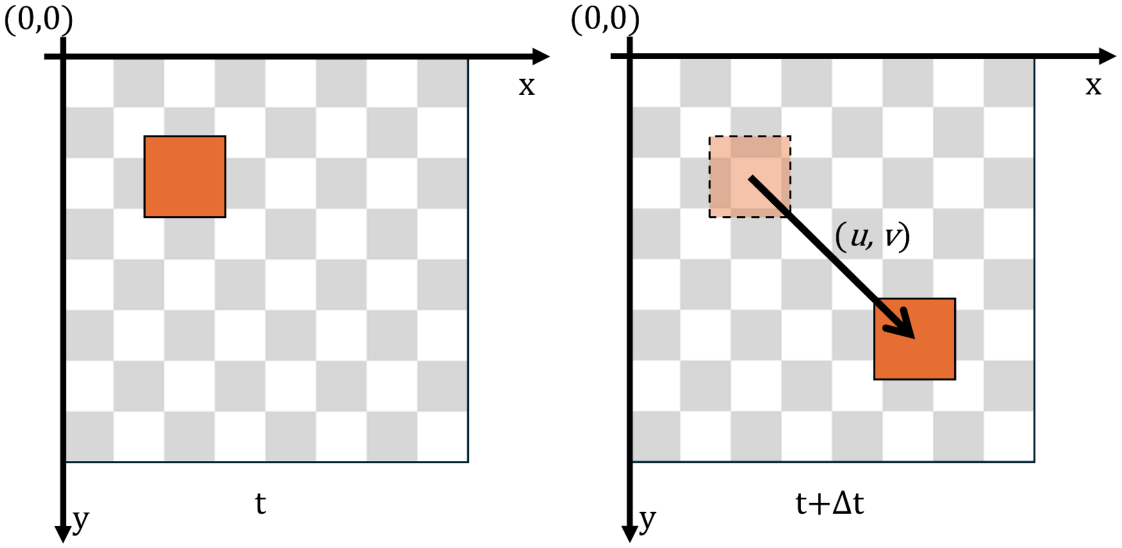


Figure 2.1: Illustration of OF as a translation. A patch at time t moves to a new location at $t + \Delta t$; the vector $\mathbf{v} = (u, v)$ denotes the displacement in pixels. Image coordinates with y-axis downward.

Hereafter, the shorthand $\mathbf{v} = (u, v)$ is used. Vector $\mathbf{v}$ represents the motion of an image patch between two consecutive frames at times t and $t+\Delta t$. While certain OF techniques can be adapted for use with color images, in most cases the data are converted to grayscale prior to processing.

The computation of OF constitutes a fundamental task in computer vision, forming the basis for numerous applications such as motion detection, object tracking, or crowd behavior analysis. Establishing a reliable relationship between the two-dimensional OF field and the actual three-dimensional motion of objects, along with their velocities, is inherently non-trivial. Consequently, the accuracy of OF estimation is of critical importance. Since the introduction of the first methods proposed by Horn and Schunck [44] and Lucas and Kanade [45, 46] in the early 1980s, the field has evolved considerably, with many improved and alternative approaches emerging over the past four decades.

OF estimation relies on several assumptions. One of the most common is the brightness constancy assumption, which states that corresponding pixels in consecutive images should have the same intensity value. This condition, however, is not always satisfied due to factors such as varying illumination, shadows, reflections, or specific properties of the observed object, for example transparency. Another limitation is the so-called aperture problem, which occurs when the motion of objects that are partially visible or lack sufficient texture becomes ambiguous. For instance, the displacement of a straight line can be perceived only in the direction perpendicular to the line. [47]

Consequently, the numerical computation of OF typically requires additional constraints to produce a unique and physically meaningful solution. The exact problem formulation depends on the chosen computation method and the associated assumptions.

OF estimation techniques can be categorized in various ways. A common distinction is between local and global methods. Local approaches compute OF within small image neighborhoods, whereas global methods estimate it over the entire frame, incorporating a smoothness constraint that promotes similar motion estimates for neighboring pixels. Furthermore, algorithms can provide either pixel-level or subpixel-level accuracy, the latter being particularly important when analyzing very small motions. The following sections present the most relevant approaches and selected methods.

2.3.2 Classical Optical Flow methods

Many classical OF computation techniques are based on the brightness constancy assumption:

$$I(x, y, t) = I(x + u,\, y + v,\, t + \Delta t), \tag{2.1}$$

where u and v denote the horizontal and vertical components of the OF between times t and $t+\Delta t$.

Numerous methods are differential techniques, which compute the OF from the spatiotemporal derivatives of image intensity [48]. Linearizing Eq. (2.1) using a first-order Taylor series expansion yields the gradient constraint equation, also referred to as the Optical Flow constraint equation or the motion constraint equation:

$$I_x\, u + I_y\, v + I_t = 0, \tag{2.2}$$

where $I_x = \frac{\partial I}{\partial x}$ and $I_y = \frac{\partial I}{\partial y}$ are the spatial derivatives of image intensity, and $I_t = \frac{\partial I}{\partial t}$ is the temporal derivative.

Equation (2.2) provides a single equation for the two unknown velocity components u and v, making the problem underdetermined. Classical OF methods resolve this ambiguity by introducing additional constraints, most commonly spatial smoothness over the flow field (global methods) [44] or constant motion within a local neighborhood (local methods) [45, 46]. An alternative dense approach, such as Farnebäck's polynomial expansion [49], models the local intensity pattern with a quadratic function to improve robustness. Many classical methods are further enhanced using image pyramids to handle large motions by coarse-to-fine refinement.

Horn–Schunck

One of the first widely used OF computation methods was introduced in 1981 by Horn and Schunck [44], who defined OF as "the distribution of apparent velocities of movement of brightness patterns in an image." Their formulation relies on several assumptions:

- The surface represented in the image is flat.
- The surface is illuminated uniformly.
- The reflectance is smooth and continuous.
- The OF field is smooth across the entire image.

Rather than estimating motion for individual pixels in isolation, the method assumes that the motion field varies smoothly almost everywhere, except at object boundaries where discontinuities may occur. To formalize this, Horn and Schunck introduced a smoothness term in the form of the squared magnitude of the flow gradients that penalizes spatial variations of the flow:

$$\|\nabla u\|^2 + \|\nabla v\|^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2, \tag{2.3}$$

where $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$, and $\frac{\partial v}{\partial y}$ are the spatial derivatives of the flow components. Minimizing this term reflects the assumption that neighboring pixels typically reveal similar motion patterns. It makes the problem well-posed but can oversmooth motion boundaries. This smoothness assumption leads to the formulation of a global optimization problem.

Combining the OF constraint (Eq. (2.2)) with the smoothness term (Eq. (2.3)) yields the following energy functional:

$$E = \iint (I_x u + I_y v + I_t)^2 \, dx\, dy + \alpha^2 \iint \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2\right] dx\, dy, \tag{2.4}$$

where α is a regularization parameter controlling the trade-off between the two terms. The first integral enforces the OF constraint with a quadratic penalty, while the second term regularizes the solution by promoting smoothness.

To minimize Eq. (2.4) with respect to u and v, Horn and Schunck employed the Euler–Lagrange equations and introduced $\overline{u}$ and $\overline{v}$, the neighborhood averages of the flow components. They therefore proposed an iterative update scheme to approximate the solution:

$$u^{k+1} = \overline{u}^k - \frac{I_x\left[I_x\overline{u}^k + I_y\overline{v}^k + I_t\right]}{\alpha^2 + I_x^2 + I_y^2}, \tag{2.5}$$

$$v^{k+1} = \overline{v}^k - \frac{I_y\left[I_x\overline{u}^k + I_y\overline{v}^k + I_t\right]}{\alpha^2 + I_x^2 + I_y^2}, \tag{2.6}$$

where k is the iteration index, and $\overline{u}^k$ and $\overline{v}^k$ denote the neighborhood averages of the flow components from the previous iteration, initially set to zero.

In this scheme, the averaging terms $\overline{u}^k$ and $\overline{v}^k$ enforce the smoothness constraint, while the terms proportional to I_x, I_y, and I_t incorporate the Optical Flow constraint. The iterative process refines the estimates of u and v until convergence, though initializing with zero flow tends to bias the solution toward small displacements, a limitation that later extensions sought to overcome.

The Horn–Schunck method is considered a global approach, as it minimizes a single energy functional over the entire image [48]. It remains one of the most influential OF methods, forming the basis of the variational framework and serving as a common benchmark for evaluating new algorithms. However, it has notable limitations: the basic derivative estimation is sensitive to noise, and the neighborhood averaging in the smoothness term tends to blur motion boundaries [43, 50]. Numerous extensions have therefore been proposed to improve robustness and address these limitations. A major drawback of the quadratic smoothness constraint is its tendency to oversmooth at motion boundaries and its poor performance in the presence of occlusions. To mitigate this, Nagel introduced the concept of oriented smoothness, which enforces smoothness primarily perpendicular to image gradients, thereby preserving flow discontinuities at object edges [51–54].

Subsequent research focused on another limitation of the Horn–Schunck framework, its restriction to small displacements. Brox et al. [55] proposed a variational model that avoided linearization of the data terms and incorporated both brightness and gradient constancy assumptions together with a discontinuity-preserving smoothness constraint. Their formulation provided a theoretical foundation for the coarse-to-fine warping strategy. Later, Brox and Malik [56] extended this framework by integrating descriptor matching into the variational formulation, coupling sparse correspondences from robust descriptors with dense variational refinement. Both refinements aimed to improve the estimation of large displacements.

These developments mark the evolution of the Horn–Schunck framework into a broader class of variational methods.

Lucas–Kanade

The Lucas–Kanade method [45, 46] was originally developed as an image alignment algorithm, aiming to align a template image to an input image [57]. In this approach, the OF vector is assumed to be constant within a local neighborhood. The image is thus divided into small regions in which the flow is considered uniform, making this a local method in contrast to global approaches such as Horn–Schunck. Local methods are generally more robust to noise than global approaches [48].

The method applies a weighted least-squares fit of the OF constraint equation (Eq. (2.2)) over each neighborhood $\mathcal{N}$:

$$\sum_{\mathcal{N}} W^2 \left[I_x \, u + I_y \, v + I_t \right]^2 , \tag{2.7}$$

where W is a window function that increases the influence of pixels near the center of the neighborhood. This yields the following linear system:

$$\begin{bmatrix} \sum W^2 I_x^2 & \sum W^2 I_x I_y \\ \sum W^2 I_y I_x & \sum W^2 I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum W^2 I_x I_t \\ \sum W^2 I_y I_t \end{bmatrix} . \tag{2.8}$$

According to Barron et al. [50], the eigenvalues $\lambda_{1,2}$ of the 2×2 matrix on the left-hand side are compared against a threshold τ:

- If $\lambda_1 \geq \tau$ and $\lambda_2 \geq \tau$, both velocity components u and v are directly estimated.
- If $\lambda_1 \geq \tau$ and $\lambda_2 < \tau$, significant variation exists only along one direction (e.g., an edge), and only the component of motion normal to the edge can be computed.
- If both eigenvalues are below the threshold, the OF is not computed for that location due to insufficient gradient variation.

In practice, the threshold is commonly set to $\tau = 1$. This classification reflects the aperture problem: flow can be reliably estimated only in regions with sufficient intensity variation in more than one direction.

The original Lucas–Kanade method is limited to small displacements due to the reliance on a first-order Taylor approximation. [47]. A common solution is the pyramidal approach [58], which constructs a multi-scale representation of the images by iterative down-sampling. OF is first computed at the coarsest scale (lowest resolution) and then refined progressively at higher resolutions, with each step initialized by the upsampled result from the previous level. This coarse-to-fine strategy significantly extends the method's applicability to larger motions.

Numerous extensions and implementation variants of the Lucas–Kanade method have been proposed. Baker and Matthews [57] compared a range of algorithms that modify specific computation details of the original approach. A notable hybrid method was proposed by Bruhn et al. [59], combining elements of local (Lucas–Kanade) and global (Horn–Schunck) formulations to enhance overall performance.

Farnebäck

The method introduced by Farnebäck [49] builds upon differential OF concepts by locally approximating the image intensity function with quadratic polynomials. Within a neighborhood centered at pixel position $\mathbf{x} = (x, y)^\mathsf{T}$, the image signal is expressed as

$$I(\mathbf{x}) \approx \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} + \mathbf{b}^\mathsf{T}\mathbf{x} + c, \tag{2.9}$$

where $\mathbf{A}$ is a symmetric 2×2 matrix capturing second-order variations, $\mathbf{b}$ is a vector representing the first-order terms, and c is a constant. The coefficients $(\mathbf{A}, \mathbf{b}, c)$ are estimated using weighted least squares with spatial weights that emphasize the central pixels of the neighborhood.

When two consecutive frames are considered, the quadratic coefficients $\mathbf{A}$ remain nearly unchanged under pure translation, while the linear terms shift according to the displacement vector $\mathbf{v}$. This relation provides a local linear system linking the polynomial coefficients of both frames to the motion vector. For each pixel, overlapping neighborhoods yield multiple equations, which are combined in a weighted least-squares minimization to estimate a dense displacement field.

A limitation of the polynomial expansion is its local character, which can lead to poor accuracy for large displacements. To overcome this, Farnebäck proposed two refinement strategies. The first employs an iterative update scheme in which a single set of polynomial coefficients is reused across iterations. The drawback of this approach is that if the initially set displacement (typically zero) is far from the actual value, the iterative procedure does not converge to a correct solution. The second, more reliable strategy is a multi-scale approach in which the images are repeatedly down-sampled, OF is estimated at the coarsest resolution, and the result is propagated to finer scales. At each level the polynomial coefficients are recomputed, which increases the computational cost but enables robust estimation of large displacements. This coarse-to-fine process resembles the pyramidal schemes used with the Lucas–Kanade method.

In addition to intensity- and gradient-based formulations, OF can also be estimated from the local phase of bandpass filter responses. In phase-based OF, video sequences are decomposed with velocity-tuned spatiotemporal Gabor filters. The normal component of motion is obtained from the local phase gradient of each filter's complex response, and the full two-dimensional flow is then recovered by combining these component velocities using local least squares. This local method is relatively robust to smooth contrast changes and small affine deformations [60]. Phase-based OF also motivated later developments in phase-based motion magnification [61], discussed in detail in Section 2.4.

2.3.3 Region-based matching approaches

An alternative to differential methods is region-based matching, where the goal is to establish pixel correspondences between consecutive frames by comparing local image patches. The displacement between matched patches directly yields the OF, which is therefore pixel-accurate. Since the search space is limited to the image extent, the problem has a combinatorial nature and can be efficiently parallelized in hardware, offering potential for real-time performance [47]. Region-based approaches are generally robust due to their discrete formulation, but their accuracy is often limited compared to sub-pixel differential methods.

Regions similarity can be quantified either by maximizing a similarity measure (e.g., correlation) or minimizing a distance measure such as the sum of squared differences (SSD) [50]. Notable early examples include the works of Anandan [62] and Singh [63], while later methods combined coarse correlation-based matching with differential refinement, such as the hybrid approach of Sousa et al. [64]. Hardware-friendly block-matching schemes have also been demonstrated, for example by Liu et al. [65].

An interesting extension is the census transform, originally introduced for visual correspondence [66]. Instead of relying solely on absolute intensity values, this non-parametric transform encodes the relative ordering of a pixel's intensity with respect to its neighbors into a compact bit string. Extensions, such as the introduction of intensity thresholds [67], further improved robustness by distinguishing between similar, higher, or lower neighboring values. More recent studies have established theoretical connections between census-based approaches and gradient constancy assumptions, and have incorporated them into variational models [68].

Overall, region-based matching techniques offer robustness and suitability for hardware implementation, though they are less precise for fine motion estimation compared to differential and variational methods.

2.3.4 Deep learning

Deep learning represents the most recent approach for computing OF. Deep Neural Networks (DNNs) have become widely adopted in computer vision, where large annotated datasets enable the training of models to learn complex mappings from inputs to outputs. A major obstacle in applying DNN to OF estimation was the scarcity of ground-truth flow data, as generating such datasets is labor-intensive and often requires synthetic simulations [69].

Early attempts combined CNNs with variational methods, for instance by using CNN-based feature descriptors for patch matching [70]. The first networks trained for direct dense OF prediction were FlowNetS and FlowNetC [71]. Both models take a pair of consecutive frames as input: FlowNetS stacks the two frames into a six-channel input and follows an encoder–decoder architecture, while FlowNetC processes the frames separately and fuses them through a correlation layer, explicitly modeling feature correspondences. Although these architectures did not outperform state-of-the-art variational methods in accuracy, they demonstrated that DNNs could be successfully applied to OF estimation.

Building on these foundations, FlowNet2 [72] combined multiple FlowNet modules specialized for different motion regimes: FlowNetC and FlowNetS for large displacements, FlowNet-SD for small motions, and a fusion network for integration. The authors also emphasized that the order of presenting training data significantly affects performance. FlowNet2 significantly improved accuracy compared to its base models and reached performance competitive with variational methods, while remaining substantially faster.

Other architectures utilize pyramid-based strategies. Spatial Pyramid Network (SPyNet) [73] consists of a set of separate networks operating at different levels of an image pyramid, refining flow estimates from coarse to fine scales. This structure is more interpretable than FlowNet and resembles spatio-temporal filtering. PWC-Net [74] advanced this concept by incorporating pyramids, warping, and a cost volume layer which provides a more suitable representation for matching.

A more recent development is Recurrent All-Pairs Field Transforms (RAFT) introduced by Teed and Deng [75]. In contrast to coarse-to-fine strategies, RAFT maintains a single fixed-resolution flow field that is iteratively refined. The method constructs an all-pairs correlation volume between image features extracted for every pixel and forms a multi-scale representation by pooling, which allows both small and large displacements to be captured. At each iteration, the current flow estimate is updated using correlation lookups in this volume, enabling accurate refinement without cascading pyramids. This combination of dense all-pairs matching and recurrent refinement achieves state-of-the-art accuracy on benchmark datasets while remaining computationally efficient.

Deep networks have the advantage of not requiring explicit modeling assumptions, and modern architectures often outperform classical methods in both speed and accuracy. However, challenges remain. FlowNet-style encoder–decoders behave as black boxes, offering little interpretability. Furthermore, deep models are vulnerable to input perturbations: even small distortions may cause large errors in the predicted flow [76]. Pyramidal designs such as SPyNet or PWC-Net mitigate some of these issues, showing improved robustness compared to encoder–decoder networks.

2.3.5 Optical Flow applications in Structural Health Monitoring and Nondestructive Testing

In SHM and NDT, OF can be employed as a core element for structural motion estimation, integrated into processing pipelines for tasks such as modal identification or damage detection. Both dense and sparse OF methods have been utilized in this context.

An example of dense OF application is RAFT usage, combined with correlation-based template matching, to estimate structural displacement [77]. Luan et al. [78] demonstrated full-field displacement extraction with dense OF, followed by modal identification on bridges and laboratory specimens. Wu et al. [79] estimated displacements of a laboratory-scale cable-net model using pyramidal Lucas–Kanade tracking at selected points, enabling identification of modal frequencies and mode shapes. Pyramidal Lucas–Kanade OF was also applied to marker points, with subsequent time and frequency-domain analyses enabling modal identification on laboratory frames and a full-scale steel footbridge [80].

Motion compensation in non-fixed camera configurations is another important aspect. Bai et al. [81] used a modified Lucas-Kanade OF for video data acquired by UAV. In order to measure structural vibration of the analyzed object, their work aimed at compensating UAV-induced motion. Camera motion compensation was also applied to handheld smartphone recordings, followed by modal identification and damage detection, where bolt loosening and crack in an acrylic frame were detected from changes in mode shapes [82]. Another example of damage detection is an incorporation of dense Farnebäck OF as the first step in a pipeline for structural anomaly detection based on the analysis of nonlinearity occurrence [83].

2.4 Motion Magnification

The term *motion magnification* was introduced in 2005 by Liu et al. [84], who described the technique as a microscope for motion captured on video. Its purpose is to process a sequence of images in such a way that subtle displacements are visually amplified, enabling the observation of movements that are otherwise barely or not at all perceptible. A naive solution would be to compute pixel-wise displacements, amplify them, and regenerate the video sequence. However, such an approach is inadequate, as it introduces artifacts and causes amplified regions to blend with the original content. A major challenge lies in distinguishing small motions to be magnified from noise that induces pixel intensity variations of a similar magnitude.

2.4.1 Original method

The original motion magnification method [84] was designed to amplify subtle displacements in video sequences while avoiding artifacts caused by camera motion and noise. A first challenge arises from unwanted inter-frame motion introduced by camera vibrations, which makes it essential to distinguish static pixels from those undergoing actual motion. To this end, the method incorporates a fully automatic image registration procedure. The first frame is treated as reference, on which corners are detected using a modified Harris corner detector [85]. For each subsequent frame, OF relative to the reference is computed at these corner locations, initially by minimizing the sum of squared differences and then refined to subpixel accuracy using the Lucas–Kanade method with Shi–Tomasi tracker [86]. The most reliable points are selected probabilistically and used to align all frames to the reference. In addition, histogram equalization in reference to the first frame is applied for all frames to compensate for illumination changes.

Point tracking is then performed, restricted to those detected in the reference frame. The trajectories of tracked points are clustered to group objects with correlated, though not necessarily identical, motion. Normalized correlation is employed as a similarity measure, ensuring that even very small displacements can be associated with larger correlated ones. A dedicated cluster is reserved for stationary objects, representing the background. The desired number of clusters is specified by the user.

The sparse OF field defined by the tracked points in each cluster is interpolated to a dense field, yielding for each pixel a set of possible trajectories corresponding to the defined clusters. Individual pixels are assigned to these clusters by a segmentation. This is formulated as an energy minimization problem solved using graph cuts [87], with probabilities based on motion similarity, color similarity, and spatial connectivity. This segmentation yields a set of separate layers exhibiting similar motion patterns. Pixels with low overall similarity are assigned to an outlier group, which forms a separate layer, as they may include objects showing atypical motion. The resulting layer map may also be manually adjusted if necessary.

Once layers are established, motion magnification is applied by scaling the displacements associated with a selected layer, typically by a factor between 4 and 40. Finally, a new video sequence is rendered. Rendering proceeds layer by layer, either in the order specified by the user or according to the scene depth, starting from the background and progressing to the foreground layers.

This pipeline constituted the first complete framework for motion magnification, providing a foundation for later techniques that aimed to simplify computation and improve robustness.

2.4.2 Eulerian video magnification

While the original motion magnification technique relied on trajectories and layered motion representation, requiring motion estimation and clustering, subsequent work by Wu et al. [88] shifted the perspective toward analyzing temporal variations of pixel intensities. In this formulation, signals are amplified directly, enabling simpler computation and extending applicability to both subtle displacements and temporal changes such as color fluctuations.

The authors observed that not only small displacements carry valuable, visually imperceptible information, but also that pixel-wise intensity variations over time can reveal important signals. Accordingly, they proposed *Eulerian video magnification*, a method that amplifies different kinds of signals in video sequences, simultaneously enhancing temporal intensity fluctuations and highlighting small spatial displacements.

The method operates on temporal sequences of pixel color values at fixed spatial locations, amplifying their variations within a specified frequency band. Although it is based on temporal filtering, the approach can also magnify very small motions. Explicit motion estimation is not required, but motion is revealed by amplifying temporal color changes at fixed positions.

Processing begins with a spatial decomposition of each frame to obtain an image representation at multiple spatial frequency bands. This separates fine-scale details from coarse structures and ensures that amplification remains spatially coherent. The temporal evolution of pixel intensities is then analyzed within each band, allowing selective filtering and magnification of subtle variations while limiting artifacts and noise propagation. Signal magnification is based on a linear approximation with a first-order Taylor expansion of the brightness constancy assumption (see Eq. 2.1), the same foundation underlying classical optical flow formulations. Filtering is applied locally using bandpass filters.

This demonstrates that explicit motion estimation is not always necessary for magnifying displacements. A key achievement of the method is its ability to enhance both motion and purely temporal changes. Applications include physiological monitoring, such as assessing blood circulation or measuring heart rate from subtle skin color variations.

The primary limitation arises at higher frequencies, where only small amplification factors can be applied reliably. Larger multipliers tend to substantially amplify noise, degrading the quality of the magnified signal.

In contrast to the trajectory-based formulation [84], the Eulerian method does not require explicit motion tracking. It is particularly effective for smooth structures and small magnifications, and it does not assume any specific type of motion.

2.4.3 Phase-based motion magnification

The method proposed by Wadhwa et al. [61] is based on the use of steerable pyramids[89, 90], which decompose an image into a multi-scale, multi-orientation, and spatially localized representation. Its basis functions resemble Gabor wavelets, i.e., sinusoids modulated by Gaussian windows.

Within this representation, temporal variations in the phase of complex pyramid coefficients are related to local displacements along the orientation and scale of the filter. By tracking these phase changes over time, motion can be determined without explicitly computing OF. Applying temporal filtering to the phase signals enables selective amplification of displacements within chosen temporal frequency bands.

Compared to Eulerian video magnification [88], the phase-based approach supports significantly larger magnification factors and demonstrates greater robustness to noise. These advantages extend the practical applicability of motion magnification.

2.4.4 Riesz pyramids

An important extension of phase-based motion magnification was introduced by Wadhwa et al. [91], with the introduction of *Riesz pyramids*. This technique relies on the Riesz transform, a two-dimensional generalization of the Hilbert transform previously applied in image and signal analysis[92]. The key contribution is a new image representation that is substantially simpler than even the smallest steerable pyramid, while retaining the ability to support phase-based motion magnification.

The method begins by decomposing the input image into sub-bands corresponding to different spatial scales. Each sub-band is then processed with an approximate Riesz transform, implemented efficiently using two finite difference filters. Unlike steerable pyramids, the Riesz transform is computed entirely in the spatial domain. This makes it possible to avoid artifacts associated with frequency-domain filtering and to implement the full pyramid using compact linear filters. As a result, the computational cost is reduced by a factor of approximately four compared to steerable pyramid–based method.

Despite this simplification, Riesz pyramids provide motion magnification results of comparable quality to those obtained with complex steerable pyramids. At the same time, their efficiency enables real-time performance. These advantages make the Riesz pyramid a practical and scalable representation for phase-based video motion processing. However, they offer limited capabilities for representing complex textures compared to steerable pyramids, which explicitly decompose images into multiple orientations.

2.4.5 Video magnification in the presence of large motions

A crucial limitation of phase-based approaches is their applicability only to very small displacements. In many practical scenarios, however, subtle deformations coexist with larger motions. To address such cases, Elgharib et al. [93] introduced *Dynamic Video Motion Magnification* (DVMAG), a technique designed to amplify small motions even in the presence of large ones.

Large-scale motion is first removed through temporal stabilization of the Region of Interest (ROI), achieved using either KLT tracking [86] or OF combined with regularized low-order parametric motion models. This warping step suppresses dominant motion components while retaining subtle variations.

The second stage applies layer-based magnification. The image is decomposed into foreground, background, and an opacity matte. Magnification is then performed selectively on the foreground and opacity matte using Eulerian techniques [61, 88], while the background remains unchanged. To address texture discontinuities introduced by amplified foreground displacements, missing background regions are filled via texture synthesis.

This framework enables amplification of small displacements in the presence of large motions, reduces artifacts at object boundaries, and supports larger magnification factors than previous techniques. Moreover, magnification can be restricted to user-specified ROI, ensuring visual integrity of the surrounding scene.

2.4.6 Learning-based video motion magnification

Deep learning methods represent a newer class of approaches to video motion magnification. Oh et al. [94] proposed the first *learning-based* approach, in which convolutional filters are trained directly from data rather than being hand-crafted. The network follows a modified encoder–decoder architecture comprising three components. The encoder performs a spatial decomposition of the input frames to obtain a representation suitable for magnification. A dedicated *manipulator* block then modifies this representation to apply the desired magnification. Finally, the decoder reconstructs the output frames with the amplified motion.

To enable training, the authors generated a synthetic dataset designed to accurately model subpixel displacements. The network was trained in a simplified two-frame setting, where the task is to magnify the difference between consecutive frames. The method achieves lower levels of edge artifacts and greater robustness to noise compared to earlier phase-based techniques.

Since then, learning-based video motion magnification has advanced considerably. Lado-Roigé and Pérez [95] proposed a transformer-based model built on the *Swin Transformer* architecture, achieving improved performance. Singh et al. [96] introduced a multi-domain network that combines frequency and spatial domain processing to reduce distortions in magnified videos. In subsequent work, Singh et al. [97] developed *hierarchical motion magnification*, which employs a multi-scale manipulator and a multi-resolution decoder to progressively generate magnified outputs from lower to higher resolutions, together with a contrastive loss that improves robustness to noise artifacts. Byung-Ki et al. [98] presented *axial motion magnification*, which enables selective amplification along user-specified directions through a dedicated motion separation module.

2.4.7 Motion magnification applications in Structural Health Monitoring and Non-destructive Testing

The application of motion magnification techniques in SHM has received growing attention, primarily through the use of phase-based methods to visualize and analyze subtle structural responses.

Chen et al. [99] incorporated phase-based motion magnification for modal identification from high-speed video. Their work demonstrated the visualization of operational deflection shapes of a cantilever beam and a pipe in selected frequency bands. Harmanci et al. [100] extended operational deflection shapes estimation to three dimensions by combining phase-based motion magnification with marker-based particle tracking in a single-camera multi-view setup. An example of more general application is a pipeline for structural vibration measurement that integrates phase-based OF with phase-based motion magnification, demonstrated on a pedestrian bridge [101].

The approach has also been applied to damage detection. Sarrafi et al. [102] used phase-based motion magnification to extract resonant frequencies and operational deflection shapes. By observing frequency down-shifts and changes in operational deflection shapes, they demonstrated the detection of simulated damage in a wind-turbine blade. Choi and Han [103] formulated the crack detection task as an optimization problem by comparing natural frequencies derived from vision-based measurements using phase-based motion magnification with those obtained from an analytical model, subsequently employing a genetic algorithm to estimate crack location and depth. To further improve operational deflection shapes visualization in this context, Canny edge detection and morphological post-processing were used. Similarly, Civera et al. [104] detected cracks in a cantilever beam by analyzing changes in operational deflection shapes obtained with phase-based motion magnification. Their approach required a baseline of an intact structure and also involved time–frequency analysis of displacement histories to identify mass changes.

Beyond crack detection, motion magnification has been employed for broader characterization of material and structural properties. Davis et al. [105] extracted small motions using a phase-based method to compute full-field motion spectra from video, enabling inference of resonant frequencies, damping, and material parameters such as Young's modulus and stiffness.

Liu et al. [106] repurposed phase-based motion magnification to enhance nonlinear ultrasonic modulation for fatigue prognosis. Their framework converted a one-dimensional ultrasonic signal into a synthetic two-dimensional video, applied phase-based motion magnification to amplify sidebands, and then reconstructed a magnified signal. This process enabled the estimation of the remaining fatigue life of a steel padeye.

These examples illustrate the range of motion magnification applications in SHM, including modal identification, damage detection, material characterization, and fatigue life prognosis.

2.5 Summary and research positioning

This chapter has reviewed CV-based SHM approaches for crack detection and assessment, covering both deep learning methods and classical image processing techniques, as well as subsequent developments in OF and motion magnification. Taken together, these studies demonstrate significant progress while also highlighting remaining challenges.

Deep learning, particularly with CNNs, has significantly advanced CV applications in SHM by enabling automated detection of surface damages from visual data, making this category of methods an attractive alternative to classical approaches [107]. Despite these advances, classical CV techniques remain highly relevant in SHM. Deep learning approaches typically demand large annotated datasets, substantial computational resources, and careful optimization to ensure robustness under varying conditions. Such requirements are often difficult to satisfy in field deployments. Moreover, their black-box nature and limited explainability are still perceived as major obstacles to adoption in safety-critical applications. By contrast, classical methods generally require little or no training data and produce interpretable features directly linked to structural behavior. Consequently, they continue to play a critical role, both as practical solutions in challenging environments and as benchmarks that ensure methodological rigor in SHM research [108].

Despite their value, both deep learning–based and classical CV approaches for crack detection and crack tip localization face important limitations. These include dependence on surface visual quality, the need for surface preparation, or, in the case of deep learning, the requirement for large annotated datasets. The classical image processing methods discussed in the literature [34–38] are designed primarily for detecting the presence of cracks, without extending to the assessment of their length. Consequently, they have not been validated for monitoring crack evolution over time. DIC-based techniques [39–41] enable measurement of crack growth, but require surface preparation, typically through the application of speckle patterns, and therefore cannot be regarded as fully non-contact.

Another significant limitation is the reliance on manual tuning of method parameters, often requiring thresholds or detection criteria tailored to individual experimental setups. This case-specific configuration restricts both scalability and generalizability. Moreover, only a minority of studies address evaluation under varying experimental conditions or across different hardware configurations. Most existing techniques have been demonstrated on a single type of structure, commonly small laboratory specimens, under controlled conditions, with limited evaluation across diverse geometries, materials, or sensing environments.

Taken together, these limitations emphasize the need for approaches that are non-contact, applicable to diverse structures, and capable of assessing crack growth, while operating with minimal configuration. This gap motivates the research presented in this thesis.

Chapter 3

Aim and scope of the dissertation

The aim of this dissertation is to *develop and validate a marker-free, computer vision-based methodology for automatic fatigue-crack assessment in metallic structures under periodic excitation, designed for integration into NDT or SHM practice.*

To achieve the aforementioned aim, the following tasks were undertaken.

Literature study of the state of the art. The literature is reviewed to situate the proposed methodology within ongoing developments in NDT and SHM. The review covers applications of CV to crack detection and assessment, including both classical image-processing techniques and modern deep-learning approaches. OF algorithms are examined in detail, and motion-magnification techniques are also surveyed. The state of the art highlights both progress and remaining limitations. These insights motivate the formulation of the research gap addressed in this dissertation.

Identification of datasets for evaluation of methods The dissertation employs three laboratory datasets: (i) a compressor recorded over its operating cycle, used primarily to study motion-estimation quality; (ii) a cantilever beam with a crack subjected to controlled sinusoidal excitation, which provided a basis for designing amplitude maps for crack visualization; and (iii) an aircraft-grade, plate-like panel subjected to cyclic loading to elicit crack breathing, which enabled validation of the methodology across multiple sensing setups. In this experiment, the acquisition covered both professional high-speed cameras and consumer-grade smartphones.

Formulation of a motion-estimation quality measure. A motion-estimation quality measure, the LOI, is formulated to quantify the directional coherence of dense OF orientation within local neighborhoods and over time. The resulting LOI field serves two roles: (i) masking unreliable areas, such as textureless or highly reflective zones, before further analysis, and (ii) guiding ROI selection so that subsequent processing is restricted to areas with sufficiently consistent orientations, thereby ensuring high motion estimation quality.

Formulation of amplitude maps for visualization of crack-breathing motion. To visualize crack-breathing induced motion under periodic excitation, amplitude-at-peak maps are formulated from dense OF time series. After single-point stabilization that renders one side of the scene quasi-static, pixelwise OF signals (magnitude and selected directional components) are transformed via Discrete Fourier Transform (DFT), and a narrow frequency neighborhood around the known excitation is isolated. The amplitude at the fundamental peak is then assembled into a spatial map that highlights regions moving relative to the chosen reference. Both magnitude-based and directional-component variants are produced, yielding interpretable visualizations that emphasize crack breathing while suppressing unrelated motion. The impact of motion magnification on the informativeness of amplitude maps is explicitly analyzed, demonstrating its potential to enhance visualization under certain conditions while identifying its limitations.

Development of an automatic crack tip localization procedure. An automatic procedure is developed to determine the crack tip position directly from amplitude maps. The method relies on scanning successive regions of the map and computing a score profile that reflects differences in motion patterns between adjacent segments. This profile exhibits a characteristic shape: a low baseline in crack-free regions followed by a distinct rise within the crack zone. In this way, the procedure identifies the crack tip location without manual inspection of the amplitude map.

Validation of the methodology across experimental setups and devices. The complete crack tip localization methodology is validated across heterogeneous recordings, including two camera views of the plate-like structure and smartphone videos. Cross-device comparability is ensured by consistent Measurement Region of Interest (mROI) definitions. Quantitative accuracy is evaluated against manual references using Mean Absolute Error (MAE), complemented by qualitative checks on the appearance of the maps and the shape of the score profiles. The results demonstrate that the method generalizes across acquisition hardware and settings, including consumer-grade smartphone cameras, while preserving interpretability.

Analysis of practical configuration and repeatability. The dissertation includes an analysis of practical configuration choices and repeatability. Sensitivity studies quantify how reference-point placement, template size, and reference score length influence stabilization quality, contrast in amplitude maps, and localization accuracy. From these studies, recommended defaults and guidelines for parameter selection are extracted to support consistent use across experiments and measurement setups. Together with mROI definition supported by LOI, these outcomes prepare the methodology for application to new recordings and facilitate its adoption in inspection workflows within SHM or NDT.

Note on software. Data processing in this dissertation was conducted using MATLAB [109] together with selected toolboxes: Computer Vision Toolbox [110], Image Processing Toolbox [111], Robotics System Toolbox [112].

Chapter 4

Experiments

This chapter presents the experimental datasets used to develop, validate, and demonstrate the proposed techniques. All datasets were acquired through controlled laboratory experiments conducted on different structural components.

Three separate experiments are described. The compressor experiment (Section 4.1) was primarily used for the development of a LOI index. The cantilever beam experiment (Section 4.2) features an aluminum beam with a crack and provides high-resolution visual data for the initial development and preliminary configuration of the crack analysis method. Finally, the plate-like structure experiment (Section 4.3) involves an aircraft panel and captures successive stages of fatigue crack propagation under cyclic loading. This setup was used to further refine and validate the crack assessment method.

The experiments were designed and conducted by the following researchers.

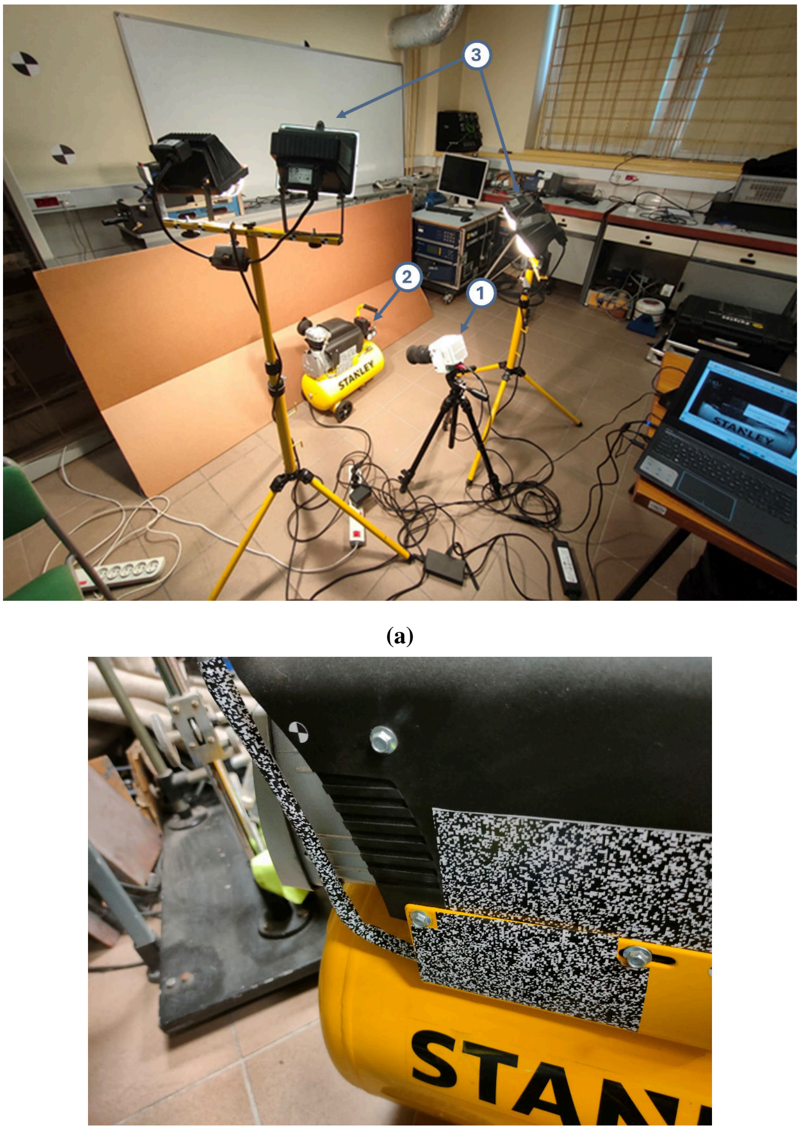
- The compressor experiment: Jakub Spytek, PhD; Kajetan Dziedziech, PhD, DSc; and Krzysztof Holak, PhD. This experiment was conducted within the scope of project LIDER/26/0103/L-9/17/NCBR/2018 funded by the Polish National Center for Research and Development.

- The cantilever beam experiment: Kajetan Dziedziech, PhD, DSc and Krzysztof Holak, PhD. This experiment was conducted within the scope of project LIDER/26/0103/L-9/17/NCBR/2018 by the Polish National Center for Research and Development.

- The plate-like structure experiment: Krzysztof Holak, PhD; Ziemowit Dworakowski, PhD, DSc; Michał Dziendzikowski, PhD; Kamil Majchrowicz, PhD; and the author. This experiment was conducted within the scope of project LIDER13/0132/2022 by the Polish National Center for Research and Development.

4.1 Compressor experiment

The compressor experiment was originally designed to support the development of a damage detection approach that lies beyond the scope of this dissertation. Accordingly, the specific damage scenarios investigated are not discussed here. For detailed results of that study, interested readers may refer to the work of Spytek et al. [113]. The complete experimental setup is shown in Fig. 4.1a.



(a)

(b)

Figure 4.1: Compressor experiment: (a) overview of the setup: 1. high-speed camera, 2. compressor, 3. lamps; and (b) close-up view of compressor components with applied random speckle patterns and a circular marker used for reliable motion estimation.

Air compressors present a particular challenge for analysis due to their inherently variable dynamic behavior. As internal air pressure increases during operation, the functioning of individual components may change. Moreover, the presence of multiple movable and interacting elements adds further complexity to the system's behavior.

The operating compressor was recorded using a Phantom VEO-E 340L high-speed camera at a sampling frequency of 600 Hz, with a recording duration of 1 s per sequence. The image resolution was 2560 × 1600 pixels. To facilitate accurate motion estimation, printed sheets with optical random speckle patterns were glued to the visible surfaces of the compressor and air tube. In addition, several circular markers were placed at selected locations to enable localized motion tracking (see Fig. 4.1b). A representative video frame captured during the compressor experiment is shown in Fig. 4.2.

Figure 4.2: Representative video frame captured during the compressor experiment.

The experimental cycle began with the tank at atmospheric pressure (0 bar) and continued as the compressor operated until the maximum pressure of 8 bar was reached. Twelve recordings were captured during this process, enabling data collection across a range of pressure conditions reflecting typical operating scenarios.

This experimental dataset provided the basis for developing the OF direction-based technique for motion estimation quality assessment described in Chapter 5.

4.2 Cantilever beam experiment

This experiment was conducted to analyze the motion of a cracked cantilever beam subjected to controlled harmonic excitation. The aluminum specimen, measuring 300 mm in length with a rectangular cross-section of 20 mm × 10 mm, was rigidly fixed at one end using a pneumatic clamp. A fatigue crack was initiated from a manually introduced notch and extended approximately halfway through the beam's width. The analysis focused on a zoomed-in field of view centered around the cracked area.

The beam was excited at its free end using an electrodynamic shaker (The Modal Shop SmartShaker K2004E01), which includes an integrated amplifier. A 10 Hz sinusoidal signal was supplied to the shaker by an Agilent 33522A function generator. The experimental setup is illustrated in Fig. 4.3.

High-speed video recordings were acquired using a Phantom VEO-E 340L camera, equipped with a Nikon AF Nikkor 50 mm f/1.8D lens and extension rings (Meike N-AF 20 mm and Meike N-AF 12 mm) to achieve the required magnification. Each sequence was recorded at 600 frames

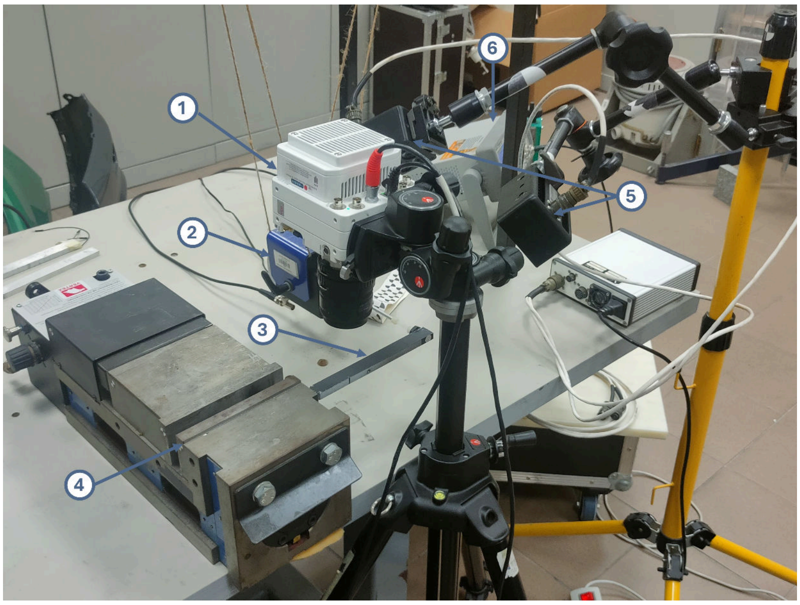


Figure 4.3: Configuration of the cantilever beam experimental setup: 1. high-speed camera, 2. electrodynamic shaker, 3. specimen, 4. pneumatic clamp, 5. lamps, 6. waveform generator.

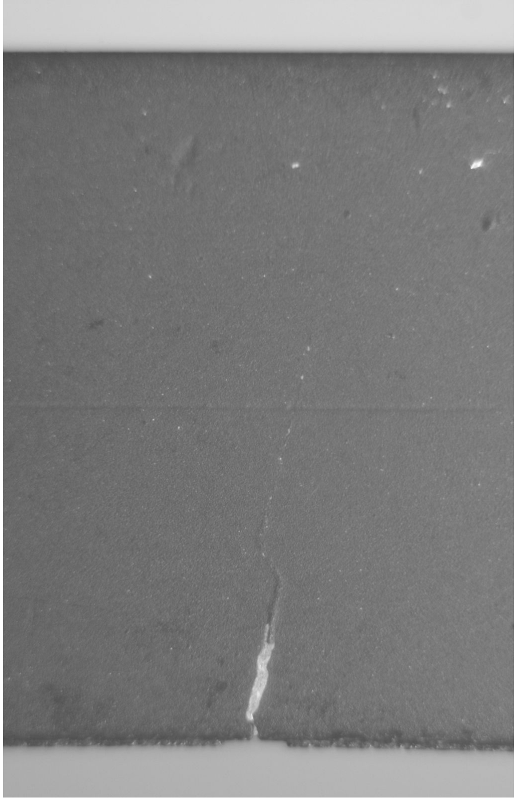


Figure 4.4: Representative video frame captured during the beam experiment. The open crack is visible in the lower part of the beam.

per second, resulting in 7429 frames per video, with a resolution of 1024×1600 pixels. This corresponds to approximately 12 seconds of acquisition time. An exemplary frame from the raw video data is presented in Fig. 4.4.

Six video sequences were collected in total, each corresponding to a different excitation amplitude controlled via the input voltage to the shaker. The applied voltage levels, listed in Table 4.1, range from 40 mV to 240 mV. As expected, higher input voltages resulted in increased vibration amplitudes of the beam.

Table 4.1: Excitation amplitude level for each video.

Video No.	Excitation amplitude level [mV]
1	40
2	80
3	120
4	160
5	200
6	240

4.3 Plate-like structure experiment

The third experiment involved fatigue testing of a structural specimen extracted from an aircraft structure, specifically a section of the tail boom skin near the fuselage attachment area, made of an aluminum alloy equivalent to 2024-T4 and containing reinforcements and rivet joints. A network of sensors was installed on the structure for monitoring purposes. These included a crack propagation gauge on the outer surface (see Fig. 4.5a) and eddy current as well as piezoelectric (PZT) sensors mounted internally. Sensor-based measurements are beyond the scope of this thesis and are not analyzed further. However, during the experiment, manual evaluation of crack propagation was carried out using a magnifying lens and ruler measurements.

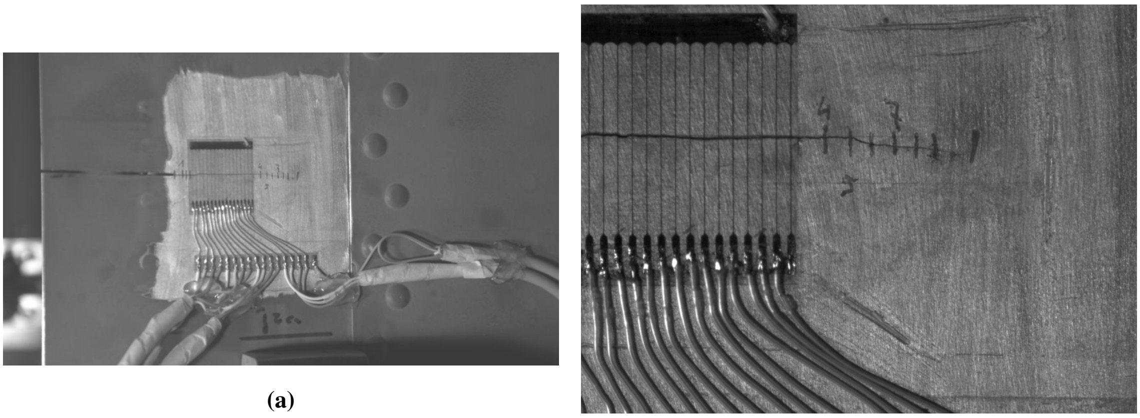
The specimen underwent repeated cycles of sinusoidal loading with a cycle-constant amplitude and a stress ratio of $R = \sigma_{\text{min}}/\sigma_{\text{max}} = 0.1$, where σ_{max} denotes the maximum tensile stress in the cycle and σ_{min} the minimum stress. The loading frequency was set to 5 Hz. As the fatigue crack advanced, the load amplitude was gradually adjusted across runs to maintain stable crack propagation.

High-speed video recordings were acquired from the external side of the specimen using a pair of cameras (Fig. 4.5b). The first device, a Phantom VEO-E 340L camera (white, labeled 4 in Fig. 4.5b), was equipped with a Sigma f/2.8 EXDG 24–70 mm lens and positioned 62 cm from the specimen. The second camera, a Phantom v9.1 (black, labeled 3 in Fig. 4.5b), used a SIGMA f/2.8 DG MACRO 105 mm lens and was placed 50 cm away. This dual-camera configuration enabled basic verification of the influence of the acquisition setup on the results.

(a)

(b)

Figure 4.5: Configuration of the fatigue experiment setup: (a) test specimen and (b) experimental setup layout: 1. fatigue testing machine, 2. specimen, 3. second high-speed camera, 4. first high-speed camera.



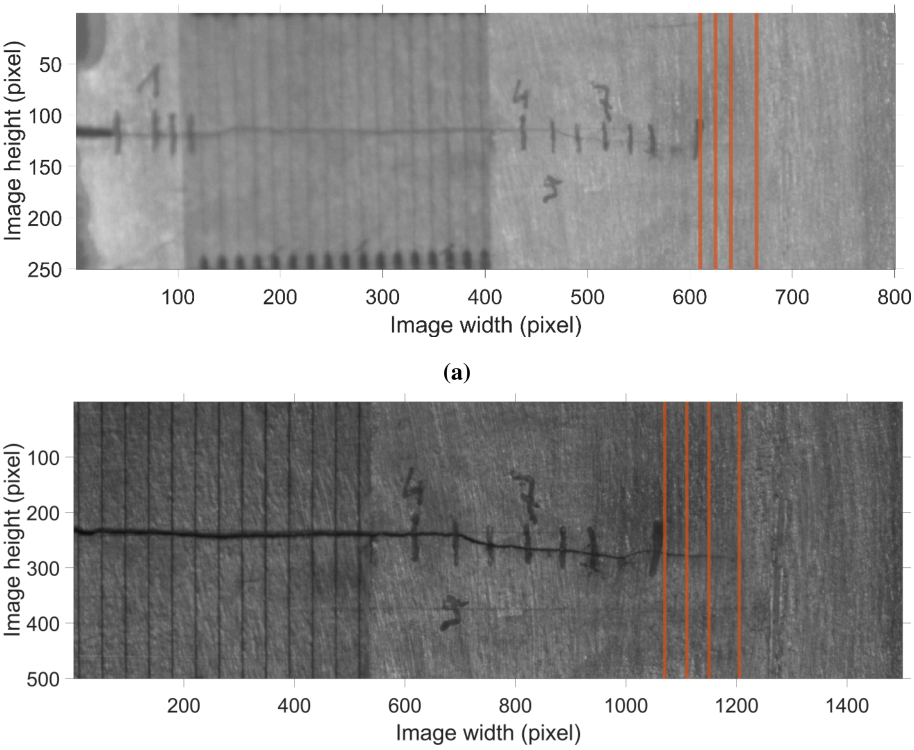
(a)

(b)

Figure 4.6: Representative video frames captured during the plate-like structure experiment: (a) the first camera and (b) the second camera.

Both cameras operated at a frame rate of 210 frames per second. The VEO-E 340L camera recorded at a resolution of 2560×1440 pixels, while the v9.1 captured video at 1632×1200 pixels. Surface coating was selectively removed near the crack to improve visibility. To ensure the video data captured relevant motion features, recordings were initiated only after the crack tip had propagated beyond the location of the crack gauge (see Fig. 4.5a). The exemplary video frames for this experiment are shown in Fig. 4.6a for the first camera and in Fig. 4.6b for the second camera.

The resulting dataset consisted of a series of video pairs documenting successive stages of fatigue crack growth. Each recording pair captured a distinct phase of crack propagation. The crack tip positions between runs was identified manually using a magnifying glass and are shown in Fig. 4.7a for the first camera and in Fig. 4.7b for the second camera.



(a)

(b)

Figure 4.7: Approximate crack tip localization in the plate-like structure experiment: (a) view from the first camera, and (b) view from the second camera.

In addition to the high-speed camera recordings, selected stages of crack propagation were also documented using a smartphone. These recordings were acquired handheld, without a fixed mounting or tripod. The device used was a Motorola Moto G100 (XT2125-4) equipped with a camera with an f/1.7 aperture. The smartphone captured video at a resolution of 3840×2160 pixels and a frame rate of 30 frames per second. These recordings were used to verify the applicability of the developed techniques when applied to consumer-grade equipment and non-professional acquisition setups.

Chapter 5

Local Orientation Inconsistency

Computer vision-based motion estimation techniques rely on the presence of visual features that can be effectively tracked. As a result, different surface types, observed under varying environmental and operational conditions, offer different levels of suitability for motion estimation. Consequently, reliable motion estimates cannot be guaranteed for all pixel locations (x, y) within a video frame. To assess the quality of motion estimation across the video frames, a processing approach based on evaluating the consistency of OF vector directions within local neighborhoods is proposed, referred to as the Local Orientation Inconsistency (LOI). Since coherent motion is typically expected within contiguous regions of a moving object, low directional variability among neighboring pixels serves as an indicator of higher estimation reliability.

5.1 Motion estimation quality assessment procedure

Building upon the assumption that consistent motion within local neighborhoods indicates reliable motion estimation, the proposed procedure quantifies spatial directional coherence and aggregates the results over time. The methodology is detailed in the following subsections.

As the analysis focuses on the consistency of the estimated OF direction in the vicinity of each pixel, the OF orientation $\theta(x, y, t) = \operatorname{atan2}(v, u)$ is the quantity under analysis. It represents the direction of the OF vector $\mathbf{v}$ at pixel location (x, y) in frame t, expressed in radians. The core idea is to evaluate the similarity between the OF direction at a given pixel and those of its neighboring pixels.

5.1.1 Definition of the evaluation subset

The OF computed for each frame is analyzed sequentially using a sliding window approach. For every pixel location $(x, y) \in \mathbb{Z}^2$ a square window $\Omega(x, y)$ of size $w_s \times w_s$ is defined

$$\Omega(x,y) = \big\{(i,j) \in \mathbb{Z}^2 \,\big|\, |i-x| \le r,\ |j-y| \le r\big\}, \qquad r = \tfrac{w_s-1}{2}, \tag{5.1}$$

where (i, j) denote the integer-valued coordinates of pixels within the window centered at (x, y).

To prevent unreliable values from influencing the subsequent analysis, all pixels with zero OF magnitude $\|\mathbf{v}(x, y, t)\|$ are excluded. This ensures that only regions exhibiting motion contribute to the reliability assessment. To avoid self-comparison and to exclude static regions, the evaluation subset is defined as:

$$\Omega'(x, y, t) = \{(i, j) \in \Omega(x, y) \setminus \{(x, y)\} \mid \|\mathbf{v}(i, j, t)\| > 0\}, \tag{5.2}$$

Only neighbors with non-zero OF magnitude are retained for further analysis.

5.1.2 Angular deviation

Within the selected window $\Omega'(x, y, t)$, directional consistency is assessed by computing the mean absolute angular difference between the central pixel and each of its valid neighborhood pixels. The local angular deviation metric $\delta(x, y, t)$ is defined as:

$$\delta(x, y, t) = \frac{1}{|\Omega'(x, y, t)|} \sum_{(i,j) \in \Omega'(x,y,t)} \left| \operatorname{angdiff}\left(\theta(i, j, t),\, \theta(x, y, t)\right) \right|, \tag{5.3}$$

where $\operatorname{angdiff}(\cdot, \cdot)$ computes the signed angular difference between two orientations, wrapped to the interval $[-\pi, \pi]$ [112].

This measure yields low values in regions where the OF directions are spatially coherent, such as on rigid, uniformly moving surfaces, and higher values in areas where the estimated directions vary significantly, typically due to occlusions, textureless regions, or discontinuities in the motion field. Therefore, $\delta(x, y, t)$ serves as a localized, per-frame indicator of motion estimation reliability.

5.1.3 Temporal averaging

To enhance robustness and suppress frame-to-frame noise, the per-frame angular deviation metric is temporally averaged over a sequence of N_f frames. The aggregated metric is defined as:

$$\overline{\delta}(x, y) = \frac{1}{N_{\text{val}}(x, y)} \sum_{t=1}^{N_f} \delta(x, y, t), \tag{5.4}$$

where $N_{\text{val}}(x, y)$ represents the number of valid observations available at that location across the sequence.

Although the total number of frames is fixed at N_f, the number of valid temporal samples at a given pixel may be smaller due to the exclusion of frames in which the deviation metric could not be computed, because the spatial window lacked a sufficient number of valid (i.e., non-static) neighbors. Consequently, $N_{\text{val}}(x, y) \leq N_f$ in general.

The resulting scalar field $\overline{\delta}(x, y)$ captures the mean local angular deviation observed over the entire sequence. This field is referred to as the LOI. It enables stable patterns of motion estimation quality to emerge, while suppressing frame-specific noise and artifacts. The LOI is subsequently used to identify and exclude regions where the motion estimation is deemed unreliable.

Since angular differences are wrapped to the interval $[-\pi, \pi]$ before taking the absolute value, the values of $\overline{\delta}(x, y)$ lie in the range:

$$0 \leq \overline{\delta}(x, y) \leq \pi. \tag{5.5}$$

5.2 Configuration

All videos are limited to a duration of one second, so the number of frames is given by $N_f = f_s{\cdot}1\,\mathrm{s}$, where f_s denotes the frame rate. Dense OF is estimated using the Farnebäck algorithm [49], The Farnebäck method is among the most commonly implemented dense OF algorithms, owing to its availability in popular libraries such as OpenCV and MATLAB. Its accessibility, together with its robustness and computational efficiency, has made it a practical choice, and it has been adopted in the development of the methods presented in this thesis. Farnebäck algorithm is applied with its default configuration. The setup employs a three-level image pyramid with a scale factor of 0.5. The neighborhood window size for OF computation is set to 5 pixels, and the averaging filter size is set to 15 pixels. For the LOI calculation, a square sliding window of size $w_s = 7$ is used. To handle the sliding window near the boundaries of the OF orientation array, zero-padding is applied at the array edges.

5.3 Results

This section presents the LOI analysis results for the compressor, beam, and plate-like structure experiments. In each case, the LOI is visualized as an image, where color encodes the score value at each pixel location. Lighter shades (approaching white) indicate low LOI values and thus higher OF reliability, whereas darker regions (approaching black) correspond to increased directional variability and reduced reliability of the motion estimation. A reversed `bone` colormap is used for visualization [109]. Although this mapping, from light to darker colors, may appear unconventional, it was chosen for visual clarity and for improved readability in print.

5.3.1 Local Orientation Inconsistency evaluation on the compressor experiment

Fig. 5.1 presents the results of the motion estimation quality assessment for the compressor experiment, shown across twelve consecutive pressure levels. Since the LOI images do not exhibit significant variation across the different pressure conditions, they are discussed collectively.

The most apparent continuous bright region, indicating low LOI values and thus high OF reliability, corresponds to the surface covered with a random speckle pattern, which facilitates accurate motion estimation. Additional areas with low scores are primarily associated with the edges of compressor components and surfaces exhibiting natural texture. A distinctly bright region is also visible at the location of a glued marker in the top-left part of the images.

Interestingly, the manometer appears clearly on the right-hand side of the images, as it contains numerous visual edges that support effective OF estimation. In contrast, regions corresponding to

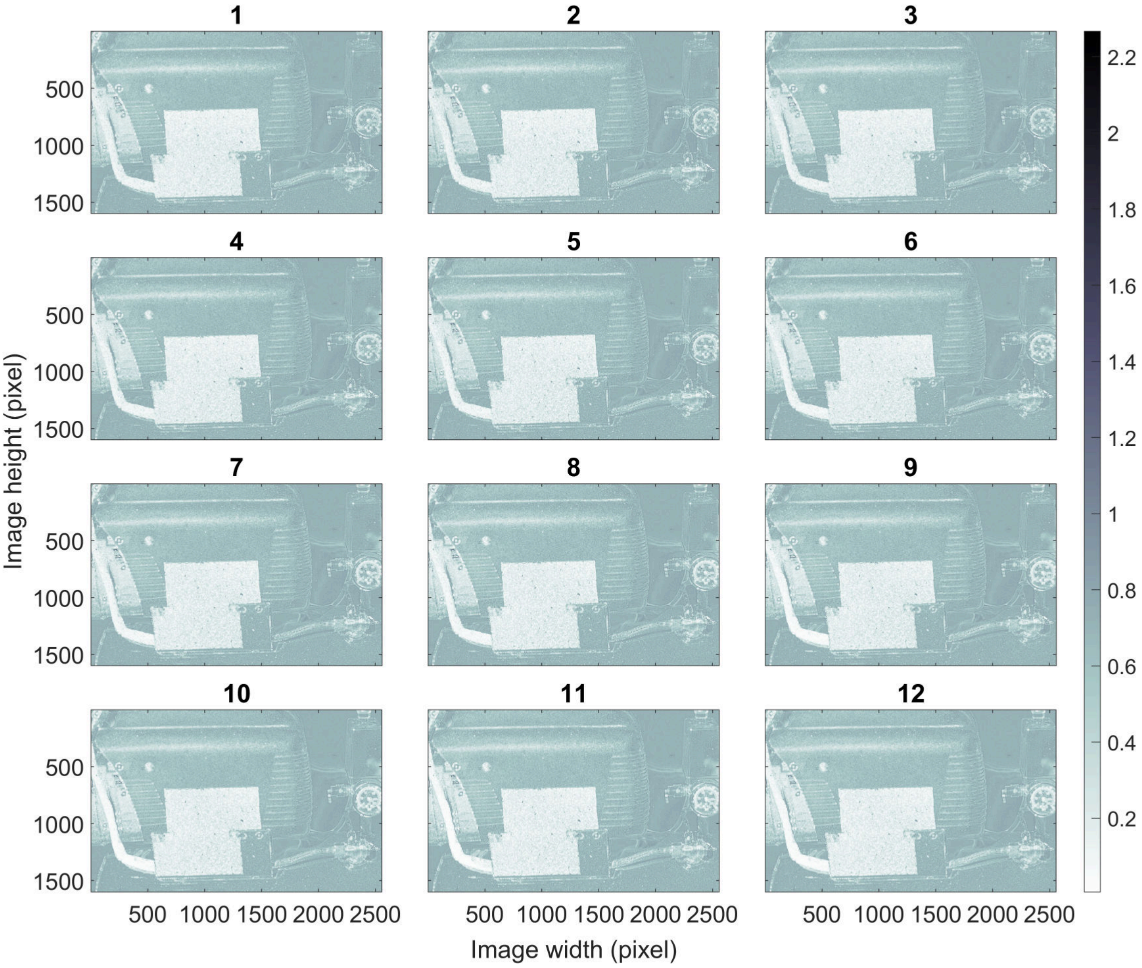


Figure 5.1: Results of the motion estimation quality assessment for the compressor experiment. The LOI is shown for twelve consecutive pressure levels.

the compressor's smooth cover and tank, where texture is absent and reflections may occur, exhibit significantly higher LOI values, resulting in darker shading in the visualization.

These results demonstrate that the proposed methodology effectively identifies regions of reliable motion estimation. Consequently, further processing is introduced to make the LOI field practically applicable.

Averaging over instances and masking

Since the videos were recorded from the same camera position and orientation, the resulting field of view remained consistent across all recordings, under varying operational conditions (i.e., different pressure levels in the air tank). This allows the corresponding LOI images to be aggregated. A straightforward approach is to compute the pixel-wise average of the LOI images obtained at different pressure levels. The result of this averaging procedure is shown in Fig. 5.2.

As the LOI maps for each pressure level are highly similar, the averaged image does not exhibit noticeable differences compared to the individual instances. While this aggregation step is not critical, it may contribute to increased applicability under varying operational conditions. In particular, averaging enables a single, universal region that remains valid under slight condition changes, allowing the same area to be analyzed consistently across cases rather than redefined per case.

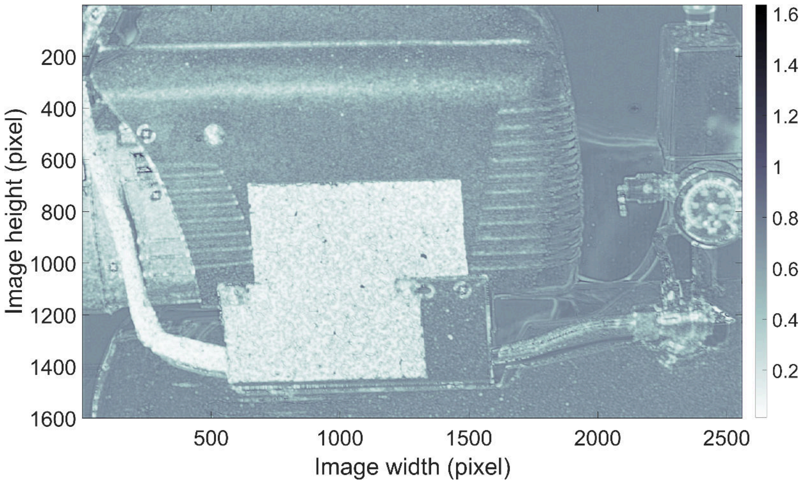


Figure 5.2: Results of the motion estimation quality assessment for the compressor experiment. The LOI is averaged over twelve different pressure levels.

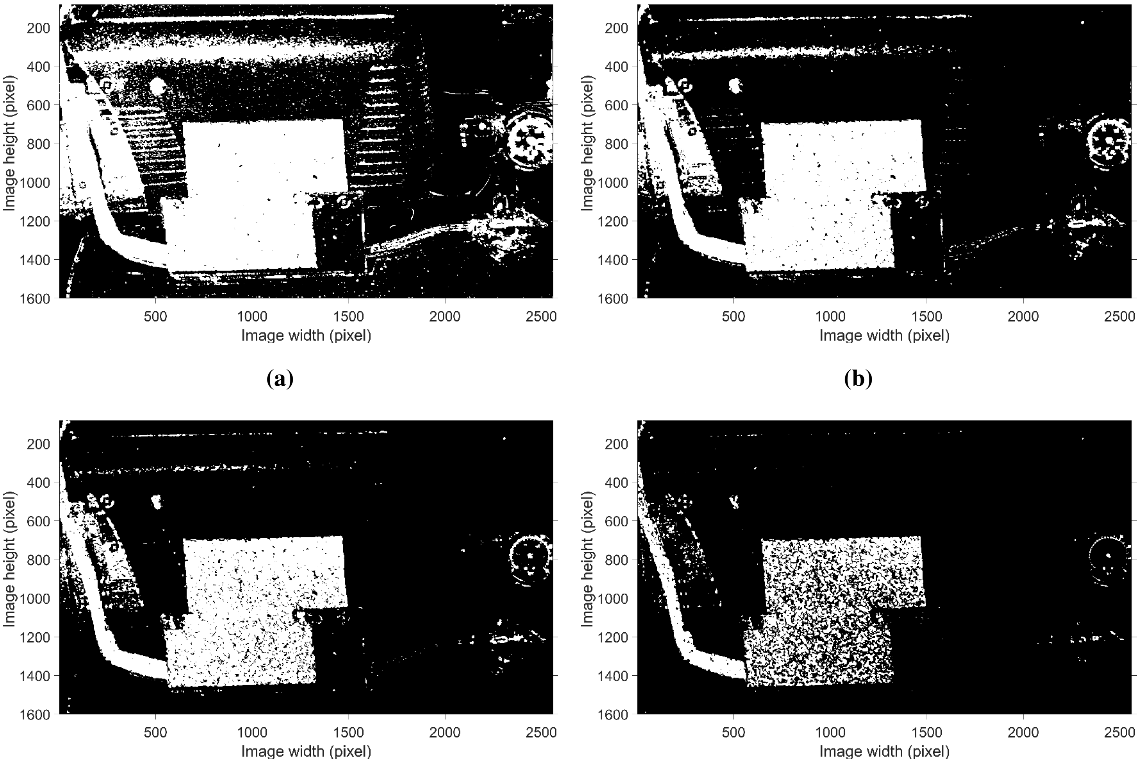
Binary mask

To directly utilize the information obtained about motion estimation quality, a binary mask is introduced based on the LOI score. It is generated through simple thresholding: all pixels with a score above the threshold t_m are rejected (set to 0), while all others are accepted (set to 1). Binary masks obtained from the LOI field averaged over different operational conditions are shown in Fig. 5.3. Each mask corresponds to a different threshold t_m, reflecting an acceptable orientation deviation of $30°$, $20°$, $15°$, and $10°$, respectively.

The choice of threshold t_m has a considerable impact on the resulting binary mask. The most lenient threshold, $t_m = 30°$ (Fig. 5.3a), results in a relatively inclusive mask, allowing many areas outside the speckled regions to pass through. As the threshold is lowered, the presence of such regions decreases. However, even at the strictest threshold of $t_m = 10°$ (Fig. 5.3d), a few scattered elements remain.

The two rectangular areas with random speckle patterns, which should ideally be preserved entirely, are not perfectly uniform. Small gaps or holes are visible due to local LOI values exceeding the threshold. While this effect is minor for higher thresholds (Figs. 5.3a and 5.3b), it becomes more pronounced at lower thresholds (Figs. 5.3c and 5.3d). This behavior reflects the fact that even within the speckled regions, motion estimation may exhibit localized inconsistencies.

The choice of threshold t_m may depend on the specific requirements of a given application. If discontinuities or isolated false-positive regions in the binary mask are unacceptable, the mask can be refined using morphological operations, however, such post-processing is beyond the scope of this work.



(a) (b) (c) (d)

Figure 5.3: Binary masks for the compressor experiment based on the LOI averaged over different operational conditions. Threshold t_m is set to: (a) 30°, (b) 20°, (c) 15°, (d) 10°. Small gaps or holes appear where local LOI values exceed the threshold: minor at 30° and 20°, but more pronounced at 15° and 10°, reflecting localized inconsistencies even within the speckled regions.

Ultimately, the binary mask serves to exclude regions where motion estimation quality is insufficient for further analysis. It can be used either for manual selection of a reliable ROI for inspection or for automatic masking, as demonstrated in work by Spytek et al. [113]. In addition, feedback provided by the LOI field and the resulting mask can be employed to evaluate the quality of an applied random speckle pattern and to double-check the suitability of the experimental configuration, ensuring that the recorded data are suitable for subsequent analysis.

5.3.2 Local Orientation Inconsistency evaluation on the beam experiment

The results for the beam experiment are shown in Fig. 5.4, where the LOI is presented for six consecutive excitation amplitude levels, resulting in different motion amplitudes. In each case, the beam surface is clearly distinguishable from the background as a brighter region, indicating lower LOI values.

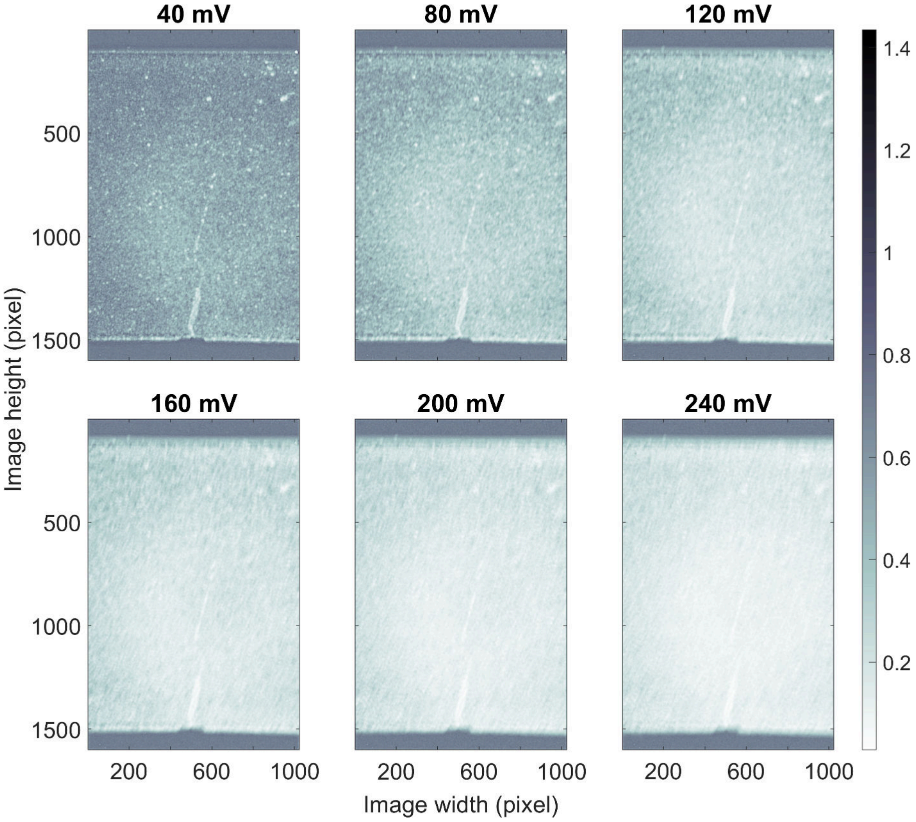


Figure 5.4: Results of the motion estimation quality assessment for the beam experiment. The LOI is shown for six consecutive excitation amplitude levels.

As the excitation amplitude increases, this contrast becomes more pronounced. The LOI values on the beam surface decrease, while the background values remain relatively high and unchanged, enhancing the visibility of the beam.

An additional observation is the presence of a brighter region near the center of the beam surface. This area corresponds to the crack, which introduces visual edges that support more accurate motion estimation. This suggests that the LOI field may, in favorable conditions, also be useful for coarse defect detection.

These results highlight the significant impact of motion amplitude on the LOI values, with higher excitation levels leading to improved reliability and clearer separation between the beam and the background.

5.3.3 Local Orientation Inconsistency evaluation on the plate-like structure experiment

Fig. 5.5 presents the results of the motion estimation quality assessment for the plate-like structure experiment. The LOI maps obtained from the first-camera view (Fig. 5.5a) exhibit very similar patterns across the four crack-propagation stages. Two distinct regions can be identified on the specimen surface: a darker area, corresponding to the painted part, where the LOI values are higher, and a brighter area, where the coating has been removed, characterized by lower LOI values. The latter region provides a more favorable texture for motion estimation. The left edge of the plate-like structure is also clearly delineated by a narrow bright band.

Additional bright spots appear at several positions and correspond to components of the measurement setup. These elements contain strong visual features that the OF algorithm can reliably track. In contrast, isolated red regions (visible in stages 2–4) indicate pixels where no motion was estimated throughout the entire sequence. In this experiment, those regions correspond to a highly reflective background area.

Crucially, the difference between the LOI values on the painted surface and those in the background is small, implying that the reliability of motion estimation over the coated area is limited. Hence, for the current acquisition and processing configuration, robust analysis can be expected only for the uncoated section of the plate-like structure, while the painted region remains unsuitable for further evaluation.

The LOI maps obtained from the second-camera view (Fig. 5.5b) remain largely unchanged across the four crack-propagation stages. Since this field of view includes only the uncoated part of the specimen, the pronounced contrast observed between coated and uncoated regions in the first-camera results is absent.

Several additional elements arising from the measurement setup are nevertheless visible. In the upper-left corner, a darker patch indicates reduced motion estimation quality; it corresponds to a dark, nearly uniform component of the crack-propagation gauge that provides few trackable features. In the lower-left corner, wires appear as elongated, darker elements. Although such slender structures typically offer strong gradients suitable for tracking, their local motion is likely decoupled from the global motion of the plate-like structure, and possibly out of phase with it. Consequently, the resulting discrepancies in local OF orientation lead to elevated LOI values and thus a lower reliability score.

This observation illustrates a characteristic limitation if the LOI metric is considered as a general-purpose indicator of motion estimation quality: it is most effective when the scene consists of rigid bodies undergoing coherent motion. When independently moving objects or flexible parts are present within the analysis window, the metric assigns them higher values, indicating lower reliability, even though such objects may still be individually trackable. However, for the practical applications, this behavior is intentional and advantageous, as it allows the analysis to concentrate on the main structure while omitting regions affected by occlusions or external elements.

5.4 Discussion

This chapter introduced the LOI metric for assessing OF reliability and demonstrated its behavior across compressor, beam, and plate-like structure experiments. The following discussion highlights its practical implications, observed dependencies, and limitations.

LOI as an indicator of OF reliability The results across the three experimental settings indicate that the LOI is a useful and interpretable indicator of OF reliability. Low LOI aligns with regions that provide consistent local texture supporting reliable motion estimation, such as areas with applied speckle patterns, whereas elevated LOI highlights areas where orientation estimates are inconsistent or unavailable. In general, the LOI metric can serve as a preprocessing step to identify and exclude unreliable regions in motion-based analyses, thereby potentially improving the reliability of measurements that rely on random speckle patterns or on OF.

Cross-experimental observations Across experiments, the empirical behavior of LOI is consistent with physical and visual characteristics of the scenes. In the compressor recordings, the lowest LOI appears over the speckled surfaces and along strong geometric edges, whereas smooth painted parts and reflective components yield higher LOI. In the beam experiment, increasing excitation amplitude leads to lower LOI on the beam and relatively unchanged values in the background, which enhances separability. In the plate-like structure recordings, the uncoated surface exhibits lower LOI than the painted area. Measurement equipment exhibiting motion patterns different from those of the main structure can introduce localized anomalies, which are correctly identified as regions of lower reliability.

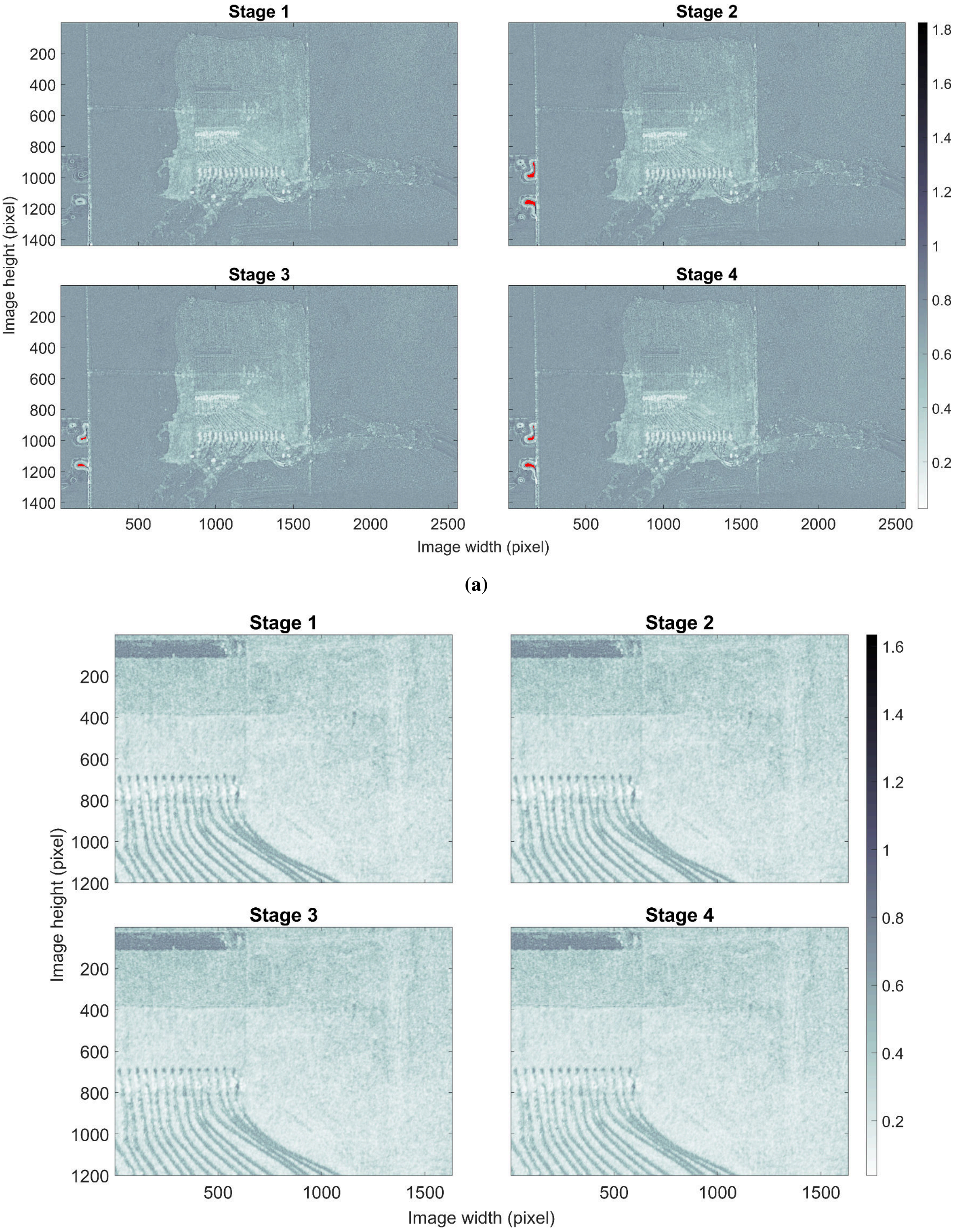
Motion amplitude impact The beam experiment also highlights the influence of motion amplitude on estimation quality. Motion resulting from higher excitation amplitudes appears to be estimated more reliably, as indicated by lower LOI values on the beam surface and clearer contrast with the background. It should be noted, however, that even at the highest excitation levels the displacements remain relatively small, yet the results suggest that the applied excitation levels are sufficient for reliable motion estimation. The LOI field can be used to assess whether a given excitation level provides adequate motion for reliable estimation.

Surface texture, illumination, and background artifacts Surface appearance and illumination have a direct effect on LOI. Natural texture and applied speckle promote consistent local orientation estimates and hence lower LOI, while smooth coatings, low-texture zones, and reflections tend to increase LOI. In addition, the background, although static, may still exhibit apparent motion caused by imperfections in image acquisition, for example due to lighting variations or other recording artifacts. This effect contributes to the relatively high LOI values typically observed in background regions, despite the absence of actual motion. A possible solution to this issue is to weight the LOI metric with the magnitude of estimated motion in order to suppress such effects.

Independently moving or flexible elements Further, independently moving or flexible elements illustrate an important property of the metric. Slender objects such as wires or parts of gauges may have strong image gradients that are trackable in isolation, yet their motion can be decoupled from the global motion of the examined structure. The resulting mismatch in local OF orientation elevates LOI in these areas. This behavior is desirable for masking because it prevents such regions from influencing analyses that target the motion of the primary structure.

Mask thresholding and aggregation Threshold selection for the binary mask should reflect the noise level, available texture, and desired coverage. Midrange values of t_m can balance the retention of useful regions with the suppression of spurious areas. When holes or isolated artifacts are problematic, light morphological post-processing is expected to improve mask usability. Where recordings of the same field of view are available across operating conditions, averaging LOI fields can potentially increase robustness, provided that the viewpoint is fixed and the scene is well registered.

The results presented in this chapter demonstrate that the LOI field can be used to identify regions of reliable motion estimation. On this basis, a suitable region for defining the mROI is selected, which is then applied in the analyses presented in Chapter 8.



(b)

Figure 5.5: Results of the motion estimation quality assessment for the plate-like structure experiment. The LOI is shown for four consecutive crack propagation stages: (a) first camera view, (b) second camera view. The red region indicates an area where no motion was estimated, and therefore no score was assigned.

Chapter 6

Amplitude maps

This chapter introduces a video-based methodology for crack assessment that combines dense OF analysis with frequency-domain processing to create amplitude maps. The proposed technique utilizes the crack-breathing phenomenon. Section 6.1 details each processing stage, starting from video stabilization within a predefined mROI, through Fast Fourier Transform (FFT)-based extraction of motion amplitudes, to the construction of maps that accentuate discontinuities. Section 6.2 then specifies all implementation choices for the cantilever-beam experiment, while Section 6.3 systematically compares four processing variants: raw OF magnitude, motion-magnified OF magnitude, the horizontal (crack-perpendicular) OF component, and its motion-magnified counterpart. By examining the influence of excitation level, motion magnification, and directional OF component, the chapter demonstrates the conditions under which amplitude maps most clearly reveal cracks and lays the groundwork for the quantitative scoring procedure developed in the next chapter.

6.1 Methodology of amplitude maps preparation

The crack assessment technique proposed in this work leverages the crack breathing phenomenon. In contrast to an undamaged structure, where the direction of motion remains uniform, a crack introduces a local discrepancy in motion direction between adjacent regions. By analyzing this relative motion, the presence and location of a crack can be identified.

Structural vibrations are captured using high-speed video recordings. The first stage involves preparing amplitude maps that visualize the underlying motion patterns. The overall processing pipeline is illustrated schematically in Fig. 6.1. The initial step consists of video stabilization, aimed at suppressing the dominant motion of the entire object and isolating local relative displacements induced by the crack. This is accomplished by defining a mROI, which restricts the spatial domain of the analysis and enhances computational efficiency. A single reference point, located on one side of the expected crack, is then selected to stabilize the video. Since the procedure accounts only for translational motion, excluding rotation, one reference point suffices.

Around the reference point, a square template of size $w \times w$ is established for matching. This ensures consistent frame alignment during processing. In practice, the method is robust to the specific choice of template, provided that it is sufficiently large to capture image texture and ensure reliable correlation. The influence of template size and location is discussed further in section 8.4. Template tracking across frames is performed using normalized cross-correlation [114], at pixel level accuracy. The mROI is updated for each frame based on the tracked reference position to maintain consistent spatial alignment throughout the video sequence. Although sub-pixel motions can be tracked by DIC, this update is performed with pixel accuracy only, which may introduce a pixel-level error.

Following stabilization, dense OF is computed across the entire frame for each pixel, enabling motion analysis at full image resolution. From this point, several processing strategies are possible.

The most general approach involves analyzing the magnitude of the OF. The results obtained using this method are presented in Section 6.3.1.

An alternative strategy incorporates motion magnification applied to the stabilized video prior to OF computation. The magnitude of the resulting OF field is then analyzed. This technique is particularly advantageous when dealing with barely visible movements and is demonstrated in Section 6.3.2.

Another approach involves analyzing the component of the OF that is approximately perpendicular to the expected crack orientation. This method enhances sensitivity to the crack breathing phenomenon. However, it requires prior knowledge of the crack orientation, information that is often available in practice, though it remains case-dependent. Alternatively, a rough estimation of the crack orientation may be derived from preliminary OF magnitude analysis. Results of this analysis are presented in Section 6.3.3.

In all of the approaches described, the processing differs in the selection of the specific OF component and in whether motion magnification is applied. The temporal signal of the selected OF component at each pixel is denoted as $\rho(x, y, t)$, where (x, y) represents the spatial coordinates and t denotes time. A FFT is applied along the temporal axis t:

$$P(x, y, f) = \mathcal{F}_t\{\rho(x, y, t)\}. \tag{6.1}$$

Rather than examining a single frequency bin, the amplitude corresponding to the excitation frequency is extracted by identifying the peak of the DFT spectrum in the vicinity of that frequency. Due to possible slight shifts in the peak frequency relative to the nominal excitation, the amplitude map $Amp_f(x, y)$ is defined as the maximum value within a narrow frequency window ϵ_f centered at f:

$$Amp_f(x, y) = \max\{P(x, y, \epsilon_f)\}. \tag{6.2}$$

The window ϵ_f is defined to span approximately ± 1 Hz around the examined frequency f. This amplitude map provides a novel image representation that significantly enhances the visibility

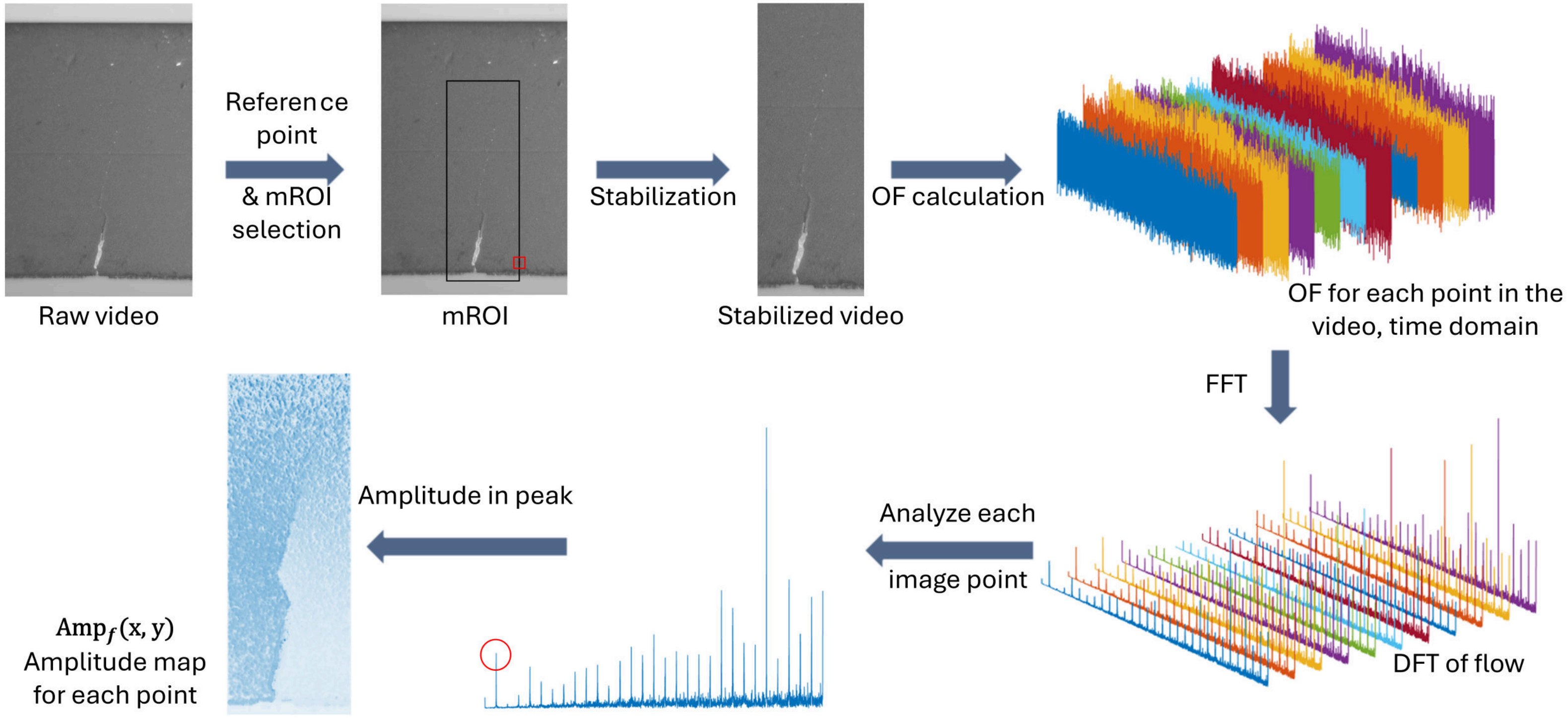


Figure 6.1: Schematic representation of the processing workflow for amplitude map generation. Initially, a reference point and a mROI are selected in the raw video to enable video stabilization. Next, OF is computed for every pixel in the stabilized sequence. The resulting OF signal at every image coordinate is treated as a time series, and a FFT is applied. The amplitude at the peak of the DFT corresponding to the excitation frequency is extracted to construct a new image representation.

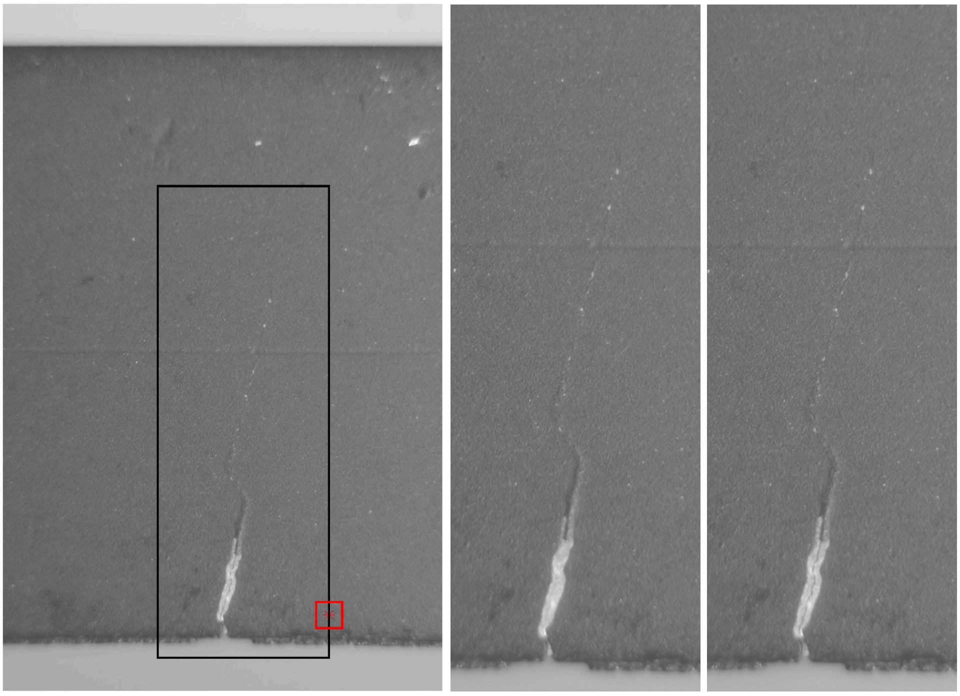
of structural discontinuities. Depending on the application, this representation may serve as the basis for further processing, as discussed in Chapters 7 and 8.

6.2 Processing and method configuration

The processing and method configuration are presented for the data acquired during the experiment with the cantilever beam (Section 4.2).

6.2.1 Video stabilization

The initial step involves cropping the image to a predefined mROI, which spans 400 pixels in width and 1100 pixels in height. The mROI is selected to encompass one edge of the beam. For video stabilization, a reference point is placed in the lower right corner of the mROI. A square template centered at this reference point is used for frame-to-frame matching. The template size w is configured relative to the image dimensions; in this case, it is set to 61 pixels to ensure robust and accurate correlation. The configuration is illustrated in Fig. 6.2a, where the mROI is outlined in black, the reference point is indicated by a red asterisk, and the template area is marked with a red square.



(a) **(b)** **(c)**

Figure 6.2: Example frames from the cantilever beam video sequence. (a) Frame with the defined mROI (black rectangle), reference point (red asterisk), and template for matching (red square). (b) Stabilized frame showing the crack in a closed state. (c) Stabilized frame showing the crack in an open state.

Selecting a reference point on one side of the anticipated crack effectively suppresses relative motion in its immediate vicinity, rendering that part of the structure quasi-static in the stabilized frames. In this setup, the procedure compensates for the dominant vertical displacement of the cantilever beam, enabling the detection of subtle relative displacements near the crack tip. The choice of the reference location thus determines which part of the object remains visually static in the resulting sequence.

6.2.2 Amplitude maps

In the subsequent step, dense OF is computed using the Farnebäck method [49]. The algorithm is applied with its default configuration, using the following parameter settings: the image pyramid comprises three levels with a scale factor of 0.5; the neighborhood size is set to 5 pixels, and the averaging filter size is defined as 15 pixels. Depending on the chosen strategy, either the magnitude of the OF or its component perpendicular to the crack is selected for further analysis.

A FFT is then applied independently at each pixel to the time series of the selected OF component. The analysis focuses on the spectral peak in the DFT corresponding to the known excitation frequency of 10 Hz. An example of the resulting amplitude map for the OF component in the horizontal direction, obtained at an excitation level of 240 mV, is shown in Fig. 6.3. To facilitate comparison, this and all subsequent maps are scaled to the $[0, 1]$ range. Importantly, interpretation relies on relative amplitude differences within each individual map, rather than on absolute scaling.

For visualization, the `sky` colormap is used [109]. It spans from light shades close to white at low values to more saturated blue shades at higher values. Although such a progression, from nearly white to more colorful, may appear unusual compared to conventional dark-to-light scales, it was chosen for clarity and to ensure readability in print.

Color intensity in the amplitude maps reflects motion magnitude relative to the reference point: brighter areas correspond to regions with minimal relative motion (effectively static), while darker regions represent areas undergoing oscillatory displacement. The presence of the crack is indicated by a sharp transition between dark and light regions. The upper portion of the image appears uniform, as it does not contain the crack. The bottom edge is also uniformly bright, reflecting the absence of significant motion in that part of the frame, which lies outside the region of structural interest.

6.3 Results

This section presents a comparative analysis of amplitude maps generated using different variants of the proposed processing pipeline. Each variant is evaluated at multiple excitation amplitude levels, with the goal of assessing its effectiveness in visualizing the crack. The influence of motion magnification and the choice between OF magnitude and directional components are examined.

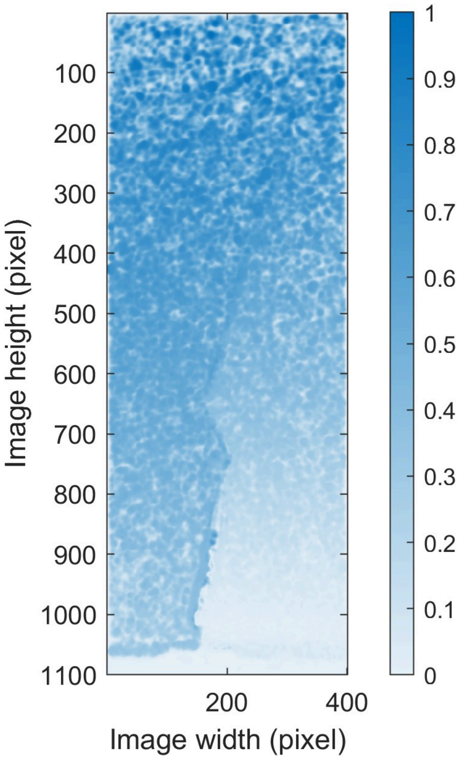


Figure 6.3: Amplitude map corresponding to 240 mV excitation, derived from the OF in the horizontal direction.

6.3.1 Optical Flow magnitude

Amplitude maps obtained directly from the OF magnitude at various excitation levels are shown in Fig. 6.4.

At lower excitation amplitude levels, a darker region appears near the center of the image, most visible at the two lowest excitation amplitude levels. This region corresponds to the crack but likely arises from artifacts produced by the motion estimation process and the crack's opening behavior. While such features could potentially be exploited for crack localization, their reliability is questionable due to sensitivity to lighting conditions and the choice of method hyperparameters. For higher excitation amplitude levels, slight differentiation between the two sides of the crack can be observed, although the effect is subtle and not sufficiently robust for reliable detection.

6.3.2 Amplified Optical Flow magnitude

Given the limitations of direct OF magnitude analysis, an additional processing step is proposed to enhance detectability.

A motion magnification technique is applied to the stabilized video prior to OF computation. Specifically, phase-based motion magnification as introduced by Wadhwa et al. [61] is employed. An amplification factor of 3 is used for all videos, as higher values introduce visual artifacts. To magnify motion around the excitation frequency of 10 Hz, the amplification band is set to a range from 5 to 15 Hz. After magnification, the standard processing pipeline is followed, beginning with

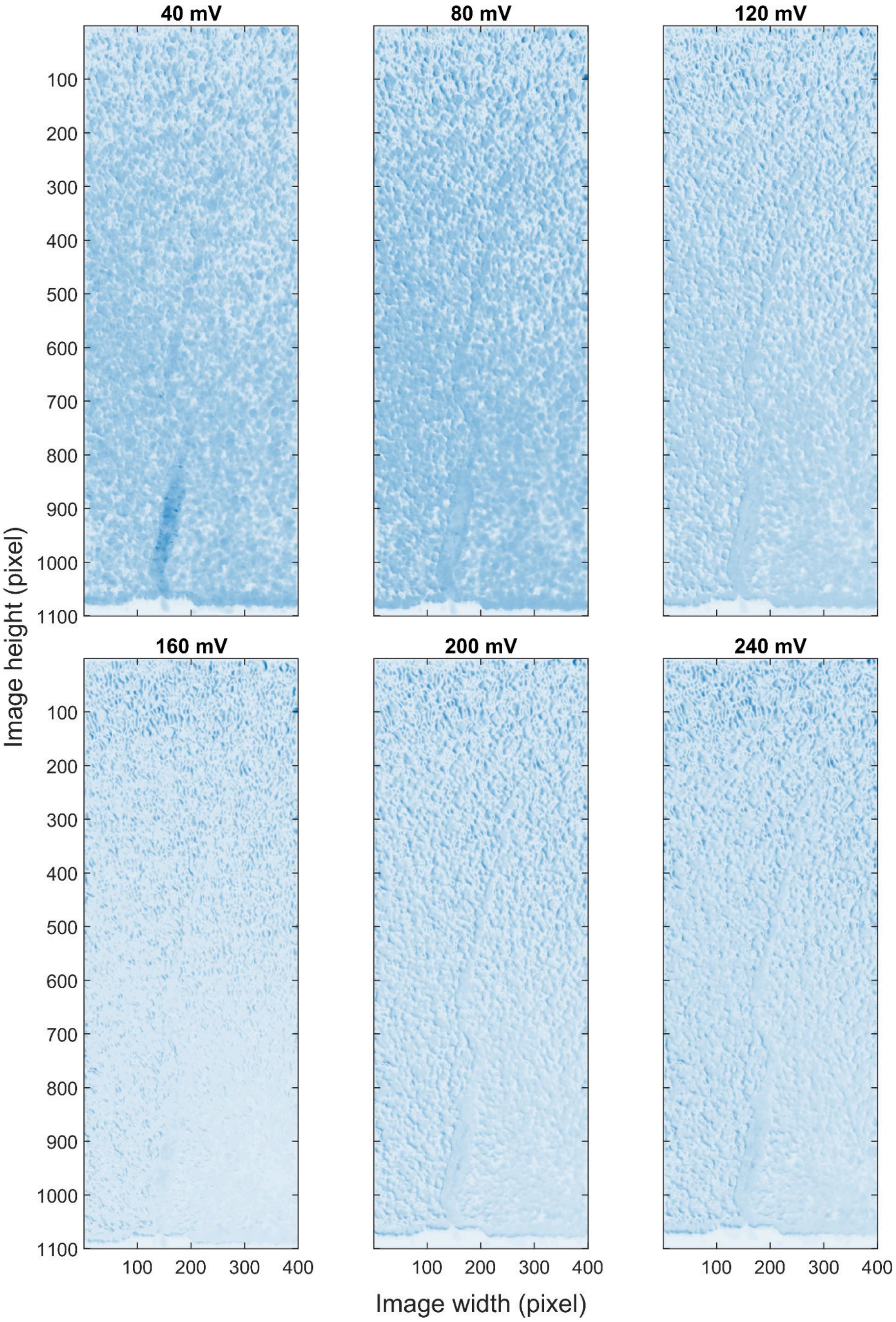


Figure 6.4: Amplitude maps computed from the OF magnitude at consecutive excitation amplitude levels.

OF calculation. The resulting amplitude maps derived from the OF magnitude are presented in Fig. 6.5.

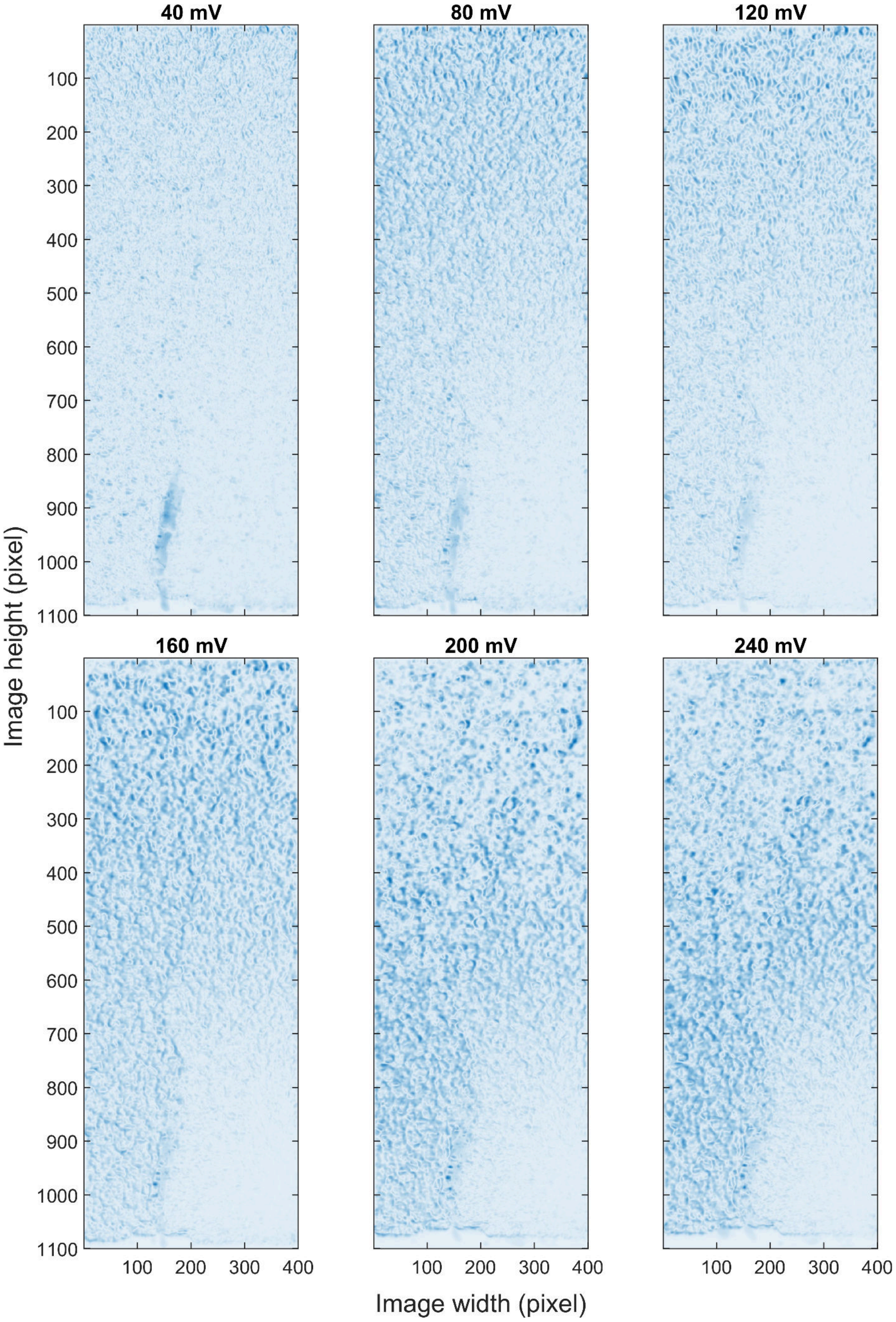


Figure 6.5: Amplitude maps computed from the OF magnitude of the magnified video at consecutive excitation amplitude levels.

Incorporating motion magnification improves the quality of the amplitude maps. Two distinct regions typically emerge: a brighter area corresponding to the visually stationary region of the structure, and a darker area indicating the part moving with respect to the reference point. The contrast becomes more distinct as the excitation amplitude increases. However, only the lower part of the crack, starting approximately between pixel rows 600 and 700, is clearly noticeable.

6.3.3 Directional Optical Flow component

An alternative approach involves analyzing the component of the OF that is perpendicular to the expected crack orientation. In this case, the horizontal direction in the image is used. Figure 6.6 presents the resulting amplitude maps across various excitation amplitude levels.

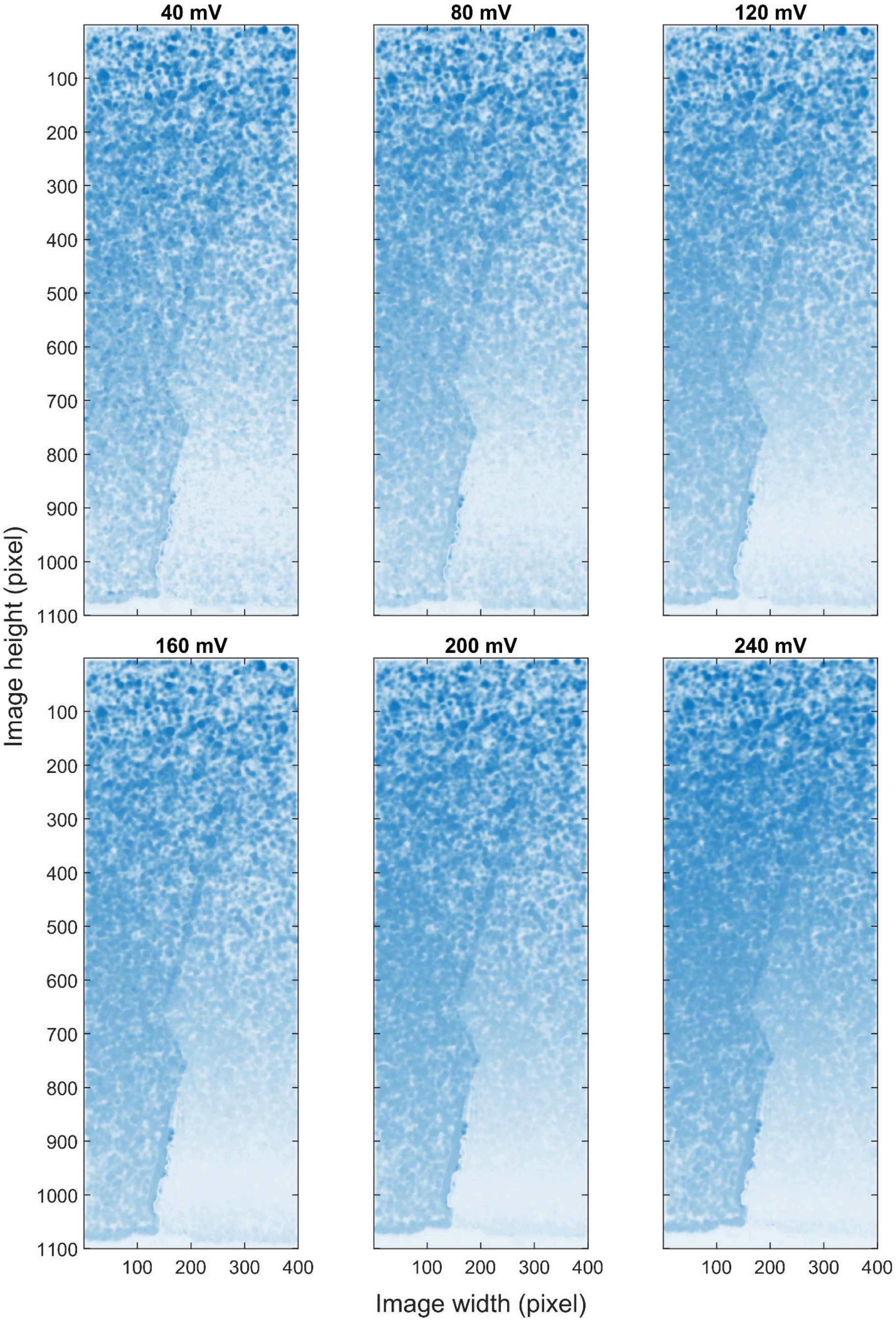


Figure 6.6: Amplitude maps computed from the horizontal OF component at consecutive excitation amplitude levels.

As the excitation amplitude increases, the contrast between the left and right sides of the map becomes more distinct, enhancing the visibility of the crack. Notably, the discontinuity remains visible even at the lowest excitation level of 40 mV. Selecting the OF component perpendicular to the crack significantly improves the quality of the generated maps.

The distinct color separation between regions enables crack localization without the need for explicit edge detection algorithms, which are often sensitive to noise and prone to artifacts. A dedicated scoring procedure for quantitative analysis is presented in the following chapter 7.

6.3.4 Amplified directional Optical Flow component

For the sake of completeness of the comparative evaluation, amplitude maps are additionally generated from the horizontal OF component after applying motion magnification. The results are shown in Fig. 6.7.

In this case, motion magnification does not lead to an improvement in map quality. At lower excitation amplitudes (40–120 mV), the results are similar to those obtained without magnification. For higher excitation levels, the map quality deteriorates due to distortions that appear in the upper regions of the image. Therefore, the use of motion magnification is not recommended when analyzing the OF component perpendicular to the crack. For small displacements, it provides negligible benefit, and for larger displacements, it may degrade the quality of the results.

6.4 Discussion

This chapter demonstrates that amplitude maps constructed from dense OF by applying frequency-domain processing provide an effective visual representation for crack assessment under periodic excitation.

Assumptions. The approach assumes a dominant periodic component associated with structural excitation and the crack-breathing phenomenon. Targeting a narrow frequency neighborhood ϵ_f around the known excitation frequency f (Section 6.1) leverages this periodicity to suppress unrelated motion. While this is advantageous under controlled excitation, performance may degrade if the excitation is broadband, nonstationary, or strongly time-varying.

Role of video stabilization and reference selection. Stabilization using a single reference point (Section 6.2) effectively removes global translation and renders one side of the scene quasi-static, which clarifies relative motion near the discontinuity. However, it also anchors interpretation: bright regions indicate quasi-static areas with respect to the chosen reference, and dark regions denote motion relative to that anchor. The placement of the reference point within the mROI therefore determines which side appears visually stationary and should be selected consistently across datasets. Reliable template tracking requires sufficient local texture and a template size w large enough to ensure proper matching. The influence of w is discussed further in Section 8.4.

Directional versus OF magnitude amplitude maps. Using the OF magnitude alone (Section 6.3.1) yielded limited and sometimes ambiguous contrast between cracked and uncracked

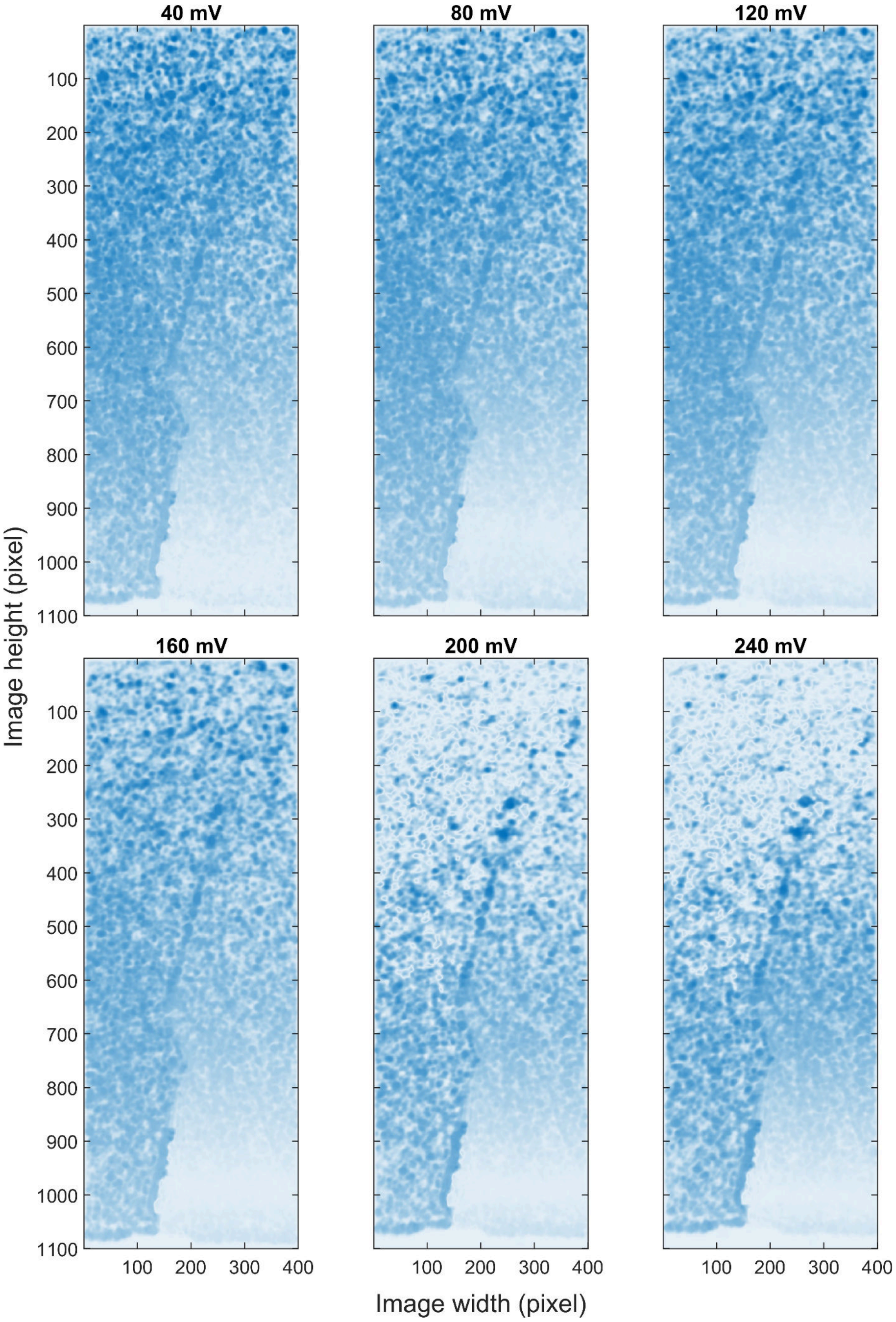


Figure 6.7: Amplitude maps computed from the horizontal OF component of the magnified videos at consecutive excitation amplitude levels.

regions in the amplitude maps, particularly at low excitation. Conversely, maps built from a directional OF component perpendicular to the expected crack orientation (Section 6.3.3) provided consistently higher contrast across map regions and clearer delineation of the discontinuity across excitation levels (see Fig. 6.6). This indicates that aligning the analysis with the expected relative motion induced by crack breathing is more informative than relying on undirected motion magnitude. The pixel-level accuracy of stabilization may introduce motion artifacts, which can be a potential cause of the limited performance of the OF magnitude amplitude maps.

Effect of motion magnification. Introducing phase-based motion magnification (Sections 6.3.2 and 6.3.4) improved separation of quasi-static and moving regions when amplitude maps were computed from the OF magnitude at low excitation. However, for the directional component variant, magnification did not increase the contrast of the resulting maps and, at higher excitation levels, introduced distortions in the upper parts of the maps (Fig. 6.7). In practice, magnification is most beneficial when the underlying motion is barely visible and the analysis uses the OF magnitude; for the directional-component maps, it is generally unnecessary and may be counterproductive at larger amplitudes.

Frequency-domain design. Evaluating $|P(x, y, f)|$ within a narrow neighborhood ϵ_f around f (Section 6.1) provides robustness to small spectral shifts and discretization effects. The choice of ϵ_f should reflect the frequency resolution determined by frame rate and record length. A wider band improves tolerance to frequency drift and leakage but risks admitting nearby frequency components.

Limitations. While the results demonstrate the potential of amplitude maps for crack visualization, several limitations should be acknowledged. The method relies on high-quality video recordings with sufficient texture and contrast for reliable OF estimation; smooth or reflective surfaces may reduce accuracy. Stabilization is limited to translation compensation using a single reference point, which may be insufficient in cases with significant rotational or non-rigid motion. The approach also assumes a known dominant excitation frequency, enabling the focus on a narrow spectral band, and its applicability under broadband or nonstationary excitation remains unexplored. Furthermore, artifacts, especially near discontinuities, may arise from OF estimation itself, from insufficient texture, or from illumination changes. These effects were most apparent in magnitude-based maps at low excitation. Motion magnification can further amplify such artifacts and may introduce additional distortions.

Overall, the comparative evaluation in this chapter shows that amplitude maps can effectively highlight crack-related motion, particularly when constructed from a directional OF component. Nevertheless, qualitative inspection alone has inherent limitations. The next chapter therefore introduces a quantitative scoring procedure designed to exploit the contrast patterns in amplitude maps for crack-tip localization.

Chapter 7

Crack tip localization method

While amplitude maps offer visually informative representations of structural responses, they require either manual interpretation or further processing to extract specific defect-related information. To address this limitation, an automated crack tip localization procedure is proposed. In order to evaluate such a procedure, experiments in which the actual crack position is known are required. This ground truth enables a direct assessment of localization accuracy and allows for a reliable evaluation of the method. For this reason, the plate-like structure experiment was used to verify the crack tip localization procedure.

The overall concept is illustrated in Fig. 7.1. The method exploits the contrast in amplitude map values, which typically form two distinct zones on either side of the crack. These zones are assessed using pairs of Scoring Region of Interests (sROIs), positioned on opposite sides of the map, as depicted in Fig. 7.2. The first pair of sROIs is shown using black squares. A sliding window approach is employed, where a 100×100 pixel sROI is shifted incrementally along a line situated 10 pixels inward from the image edge. This margin helps mitigate boundary noise effects, with the scanning area defined by the red rectangle.

For each horizontal pair of regions $sroi_n$ and $sroi'_n$, where n denotes the index of the pair, the mean amplitude values are computed, and the score is defined as the absolute difference between these means:

$$score(n) = |\overline{Amp_f(sroi_n)} - \overline{Amp_f(sroi'_n)}|. \tag{7.1}$$

Depending on the specific case, the index n may correspond to horizontal or vertical movement across the image. The score is assigned to the central coordinate of each sROI pair.

Low score values indicate similar intensity levels between the regions, suggesting a crack-free area. In contrast, high score values reflect substantial differences in amplitude, typically due to the presence of a crack. Hence, this absolute difference metric offers a robust and interpretable measure for identifying crack-affected regions.

Figure 7.3 presents the computed scores for the cantilever beam under varying excitation amplitudes. The X-axis reflects the vertical coordinate in the image. In the initial segment, where the crack is absent, the score remains consistently low. A notable increase is observed between ap-

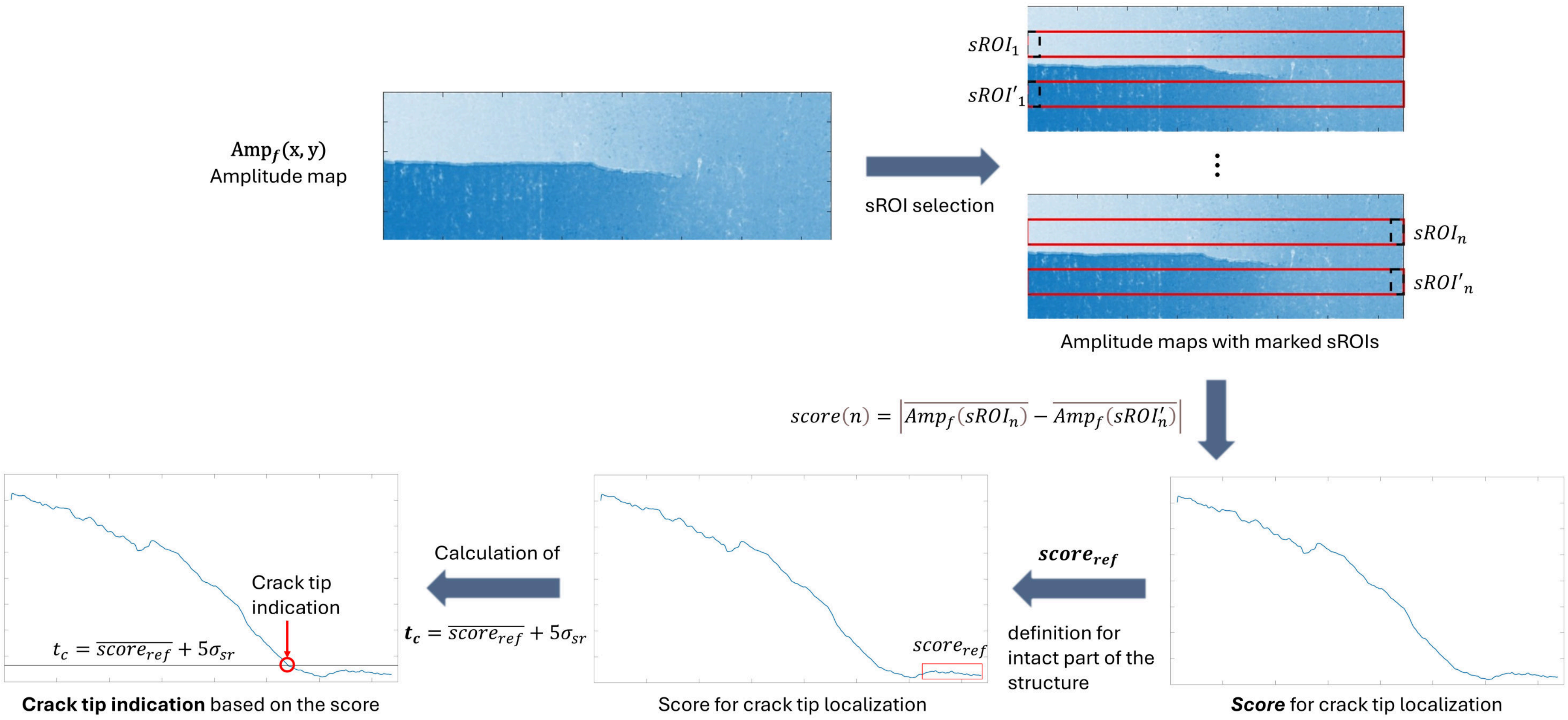


Figure 7.1: Schematic representation of the crack tip localization approach using amplitude maps. Pairs of sROIs are moved along the map edges, allowing for comparison of amplitude intensities. A score is computed as the absolute difference between the mean values of each sROI pair. A threshold derived from a reference crack-free region is then used to localize the crack tip automatically.

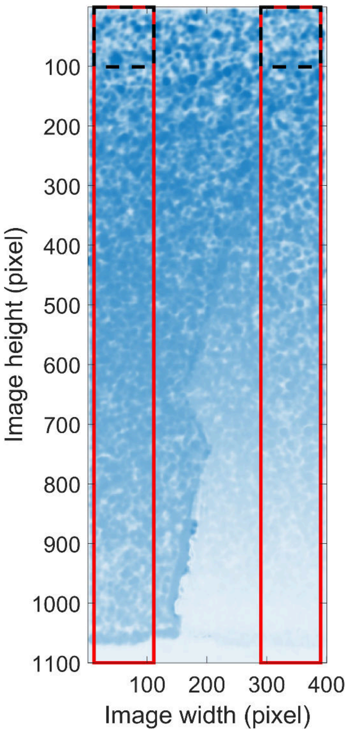


Figure 7.2: Example pair of sROIs (black squares) moved along the edges of the amplitude map, confined within the red scanning rectangle.

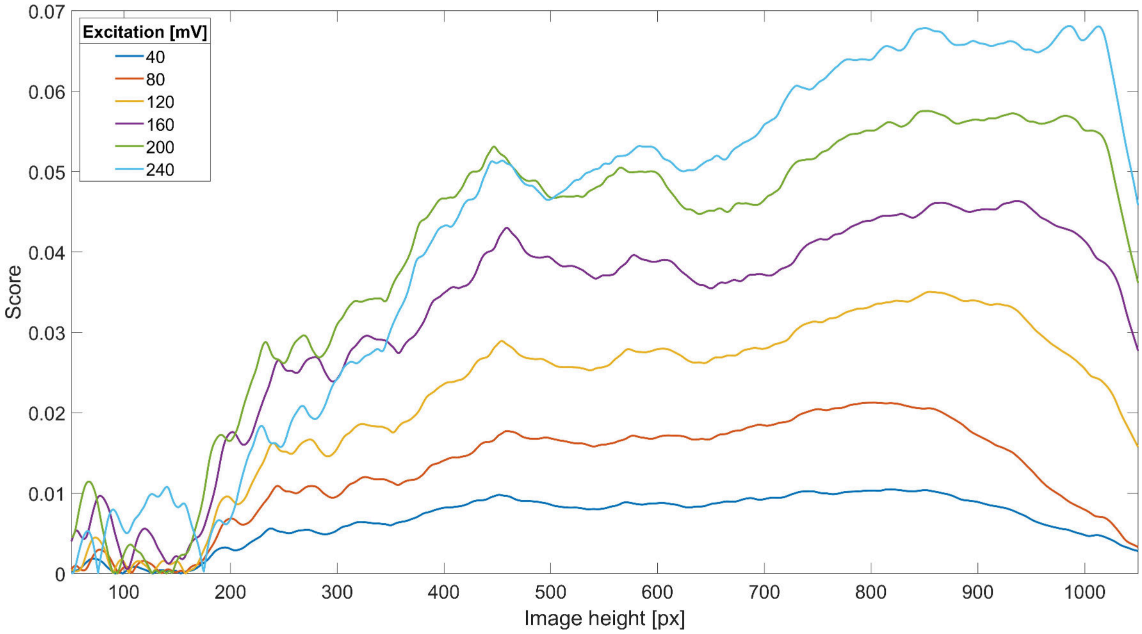


Figure 7.3: Score values obtained from the cantilever beam experiment. The x-axis corresponds to the vertical position in the image. Each point represents the center of an sROI. Multiple excitation levels are shown for comparison.

proximately pixels 200 and 450, indicating the crack region. The score's magnitude is positively correlated with the excitation amplitude; stronger excitation yields higher scores, thereby improving crack tip visibility. After the rise associated with the crack, the score remains at an elevated level with some fluctuations. It then drops sharply once the map reaches the region beyond the beam.

These results demonstrate that the magnitude of motion, which depends on the excitation level, significantly influences the score. Nonetheless, the method consistently yields sufficient contrast for crack detection across all tested conditions. The automatic crack tip localization procedure will be further examined in the following experiment.

Chapter 8

Plate-like structure case study

This chapter presents the validation of the proposed methodology on a plate-like structure specimen undergoing fatigue testing. Two independent video datasets, recorded with different cameras and capturing four distinct stages of crack propagation, are analyzed with the aim of localizing the crack tip from amplitude maps derived from the vertical component of the OF, which is approximately perpendicular to the crack. The construction of amplitude maps follows the procedure introduced in earlier chapters; maps of the OF magnitude are also included for completeness, although the primary analysis relies on the directional (vertical) component. In this chapter, results after motion magnification are not presented, since for crack tip localization the most informative representation is given by the directional OF component without motion magnification. Manual crack-tip annotations serve as the reference, while automatic localization is obtained by thresholding score profiles introduced in Chapter 7. To examine robustness to acquisition conditions, complementary smartphone recordings are assessed for selected crack propagation stages with closer and more distant views and mild out-of-plane rotation. Finally, a parameter study evaluates three factors: the placement of the reference point for stabilization, the template size w in the stabilization matching step, and the length of the reference segment of the score profile used for crack tip indication. For the last of these, results are compared using a single (global) threshold and stage-wise thresholds.

8.1 Configuration

In the case of the first camera, the region of interest was restricted to a manually defined mROI measuring 800 px by 250 px. This selection was guided by the LOI score obtained for this experiment and presented in Chapter 5. The selected mROI is illustrated in Fig. 8.1a. For the second camera, the mROI was defined as a 1500 px by 500 px window (see Fig. 8.1b). In both recordings, the mROIs were chosen to include comparable regions of the test object, enabling meaningful comparison between the two datasets. The approximate scaling factor between the first and second camera images is 2.5.

In each case, a reference point for stabilization was selected in the top-left corner of the mROI. Template matching for stabilization was performed using a square window of size $w = 61$ pixels. Both videos had a duration of 10 seconds.

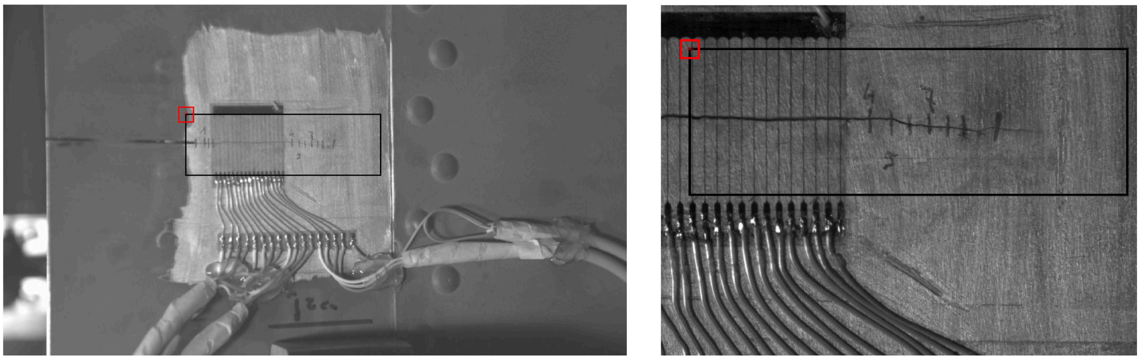

(a) **(b)**

Figure 8.1: Illustrative video frames from the plate-like structure experiment with mROI indicated by the black rectangle. The red square marks the stabilization template, and the red asterisk denotes the reference point. (a) The first camera, (b) the second camera.

The size of the sROI was set to 50 by 50 pixels for the first camera. For the second camera, due to higher resolution, the height of the sROI was increased to 100 pixels, while the width remained 50 pixels. Margins of 50 and 100 pixels, respectively, were applied at the edges of the amplitude maps to ensure consistent evaluation regions.

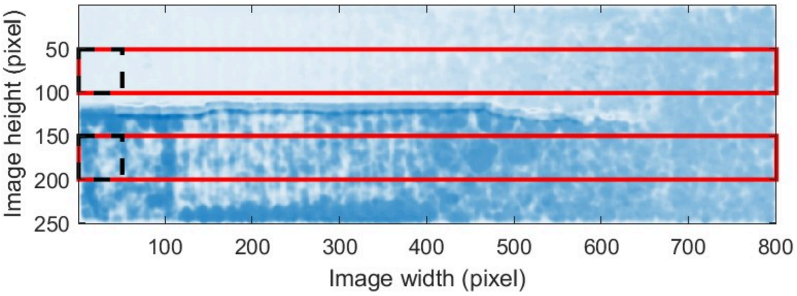
Fig. 8.2a and Fig. 8.2b illustrate the placement of sROIs in both recordings. The black squares represent individual sROI pairs, which were systematically shifted within the red rectangle pixel by pixel. At each position, the absolute difference between the mean values of the two sROIs was computed according to Eq. 7.1.

8.2 Amplitude maps

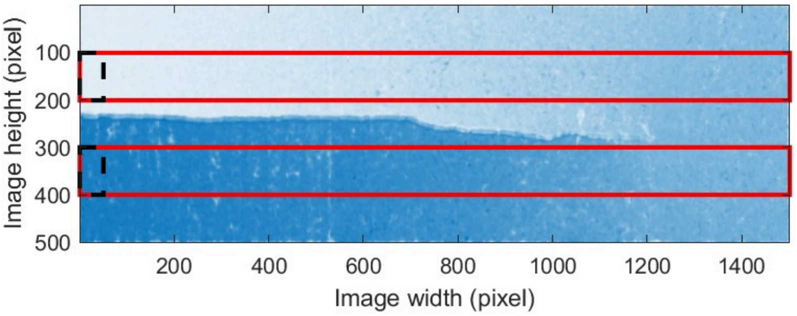
This section presents two variants of the amplitude maps. First, the vertical component maps highlight motion perpendicular to the crack, which is most affected by the crack breathing phenomenon. Second, for completeness, maps of the OF magnitude are included. Both sets are shown for four successive stages of crack propagation and for each camera, allowing direct comparison of the results.

8.2.1 Directional Optical Flow component

Fig. 8.3 presents the amplitude maps of the vertical component of the OF, which is approximately perpendicular to the crack. Four distinct stages of crack propagation are captured. To support visual interpretation, representative video frames showing the mROI are provided in Fig. 8.3a for the first camera and Fig. 8.3b for the second. The corresponding amplitude maps are shown in Fig. 8.3c and Fig. 8.3d.



(a)



(b)

Figure 8.2: sROIs used in the plate-like structure experiment: (a) first camera, (b) second camera. The black square denotes the sROI, which is shifted pixel by pixel within the red rectangle during the scoring process.

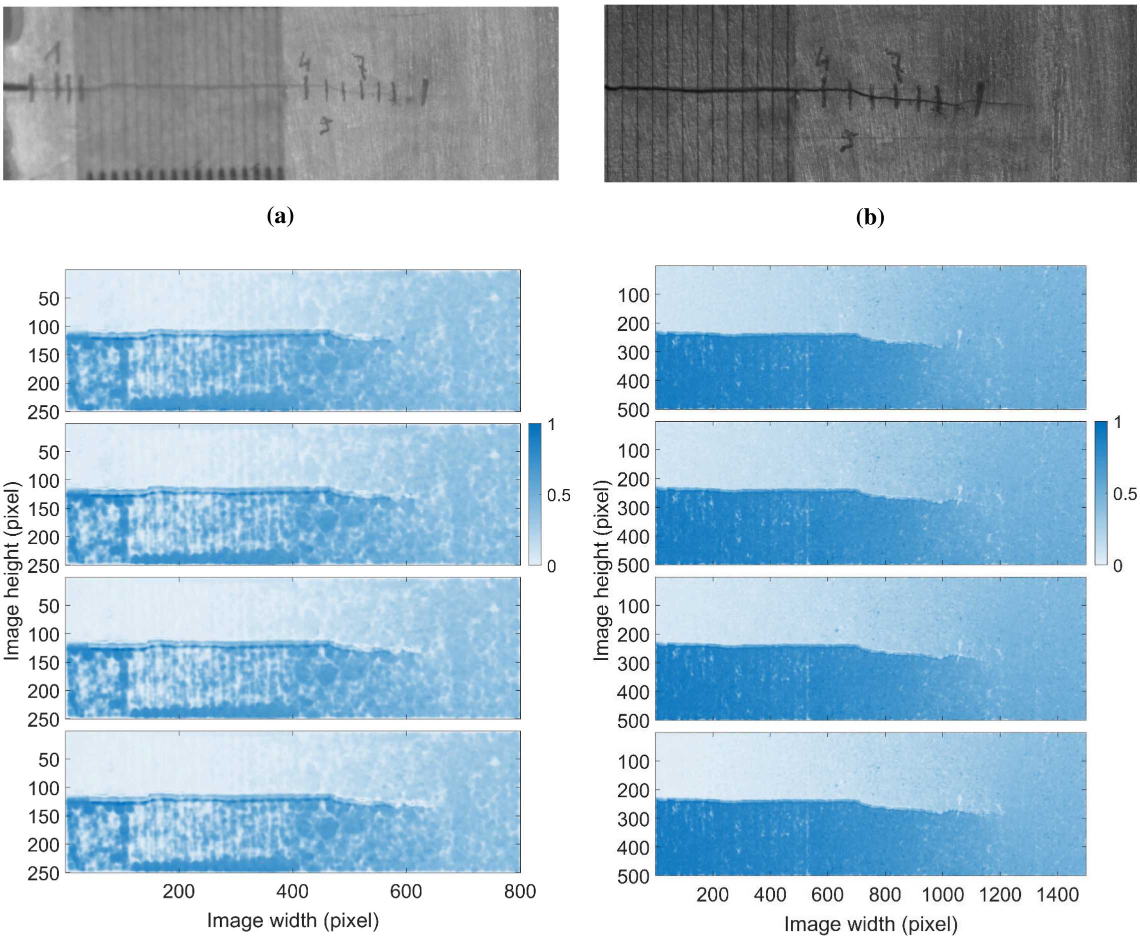
All amplitude values were scaled to the range $[0, 1]$. The maps consistently reveal a structural division: a brighter upper region indicating very limited relative motion, and a darker lower section that exhibits movement with respect to the stabilization point. Additionally, a region of intermediate intensity appears toward the right edge of each map. This area represents an uncracked segment of the specimen.

In the maps from the first camera (Fig. 8.3c), certain inconsistencies are visible, notably bright patches within the lower region. These artifacts are attributed to localized inaccuracies in motion estimation, likely influenced by surface irregularities and the presence of embedded sensors. In contrast, the amplitude maps obtained from the second camera (Fig. 8.3d) demonstrate improved spatial uniformity, with sharper delineation between moving and relatively static regions and fewer distortions. The differences between the two cameras are attributed primarily to image resolution, with higher resolution providing finer texture detail and supporting more reliable motion estimation.

The resulting amplitude maps enable manual identification of the crack tip and clearly reflect the evolution of the damage across successive stages.

8.2.2 Optical Flow magnitude

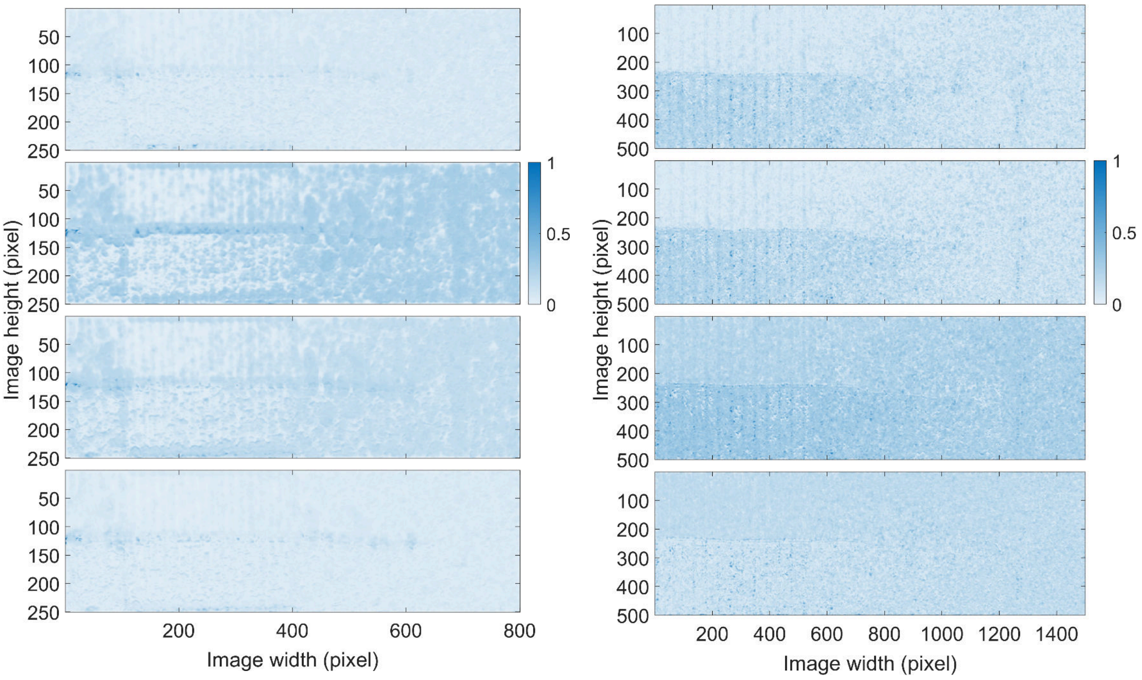
In addition to the vertical component amplitude maps of the OF, the amplitude maps representing the OF magnitude are presented in Fig. 8.4. For the first camera (Fig. 8.4a), the maps are dominated by artifacts caused by imperfections in motion estimation, primarily due to the presence of additional equipment on the surface and the motion estimation performed along the crack border



(a) (b) (c) (d)

Figure 8.3: Results from the plate-like structure experiment. Video frames illustrating the analyzed mROI: (a) first camera, (b) second camera. Directional amplitude maps corresponding to four stages of crack propagation, with values scaled to the $[0, 1]$ range: (c) from the first camera and (d) from the second camera.

itself. In contrast, the maps from the second camera (Fig. 8.4b) show two distinct regions on either side of the crack. However, the contrast between regions is limited and fades prematurely. The substantial differences between the results obtained from the two cameras are likely attributable to the much higher image resolution provided by the second camera, for a similar mROI selection, due to differences in the field of view. Only the results from the second camera are considered usable under this configuration, though their overall quality remains low. Consequently, the magnitude amplitude maps are not analyzed further.



(a) **(b)**

Figure 8.4: Amplitude maps representing the magnitude of the OF for the plate-like structure experiment: (a) first camera, (b) second camera.

8.3 Crack tip localization

Directional amplitude maps of the vertical OF component provide sufficient contrast to enable visual estimation of the crack tip's location. As the crack advances, the observable transition between regions in the maps gradually extends to the right, clearly indicating the progression of the crack tip.

Fig. 8.5 presents the computed scores derived from the amplitude maps. The x-axis indicates the horizontal pixel coordinate corresponding to the center of the evaluated sROI, while the y-axis shows the score. Different stages of crack propagation are represented by distinct colors, and corresponding vertical dashed lines mark approximate crack tip positions based on manual inspection.

For the first camera (Fig. 8.5a), the score is relatively high in the initial segment across all tested cases. In this region, noticeable fluctuations are present, reflecting reduced motion-estimation accuracy influenced by surface artifacts. After approximately the 480th pixel, the score begins to decline. Between pixels 600 and 700, it stabilizes at a low level, corresponding to an undamaged area of the structure. This region forms a baseline for detecting the presence of the crack.

To automate the detection of the crack tip, a threshold-based method is proposed. First, the mean score $\overline{score_{ref}}$ is calculated over the reference region, defined in general as the last 10% of the score profile (corresponding in this case to 75 pixels), from the first amplitude map. The standard deviation σ_{sr} of this region is also computed. The threshold t_c is defined as:

$$t_c = \overline{score_{ref}} + 5\sigma_{sr}. \tag{8.1}$$

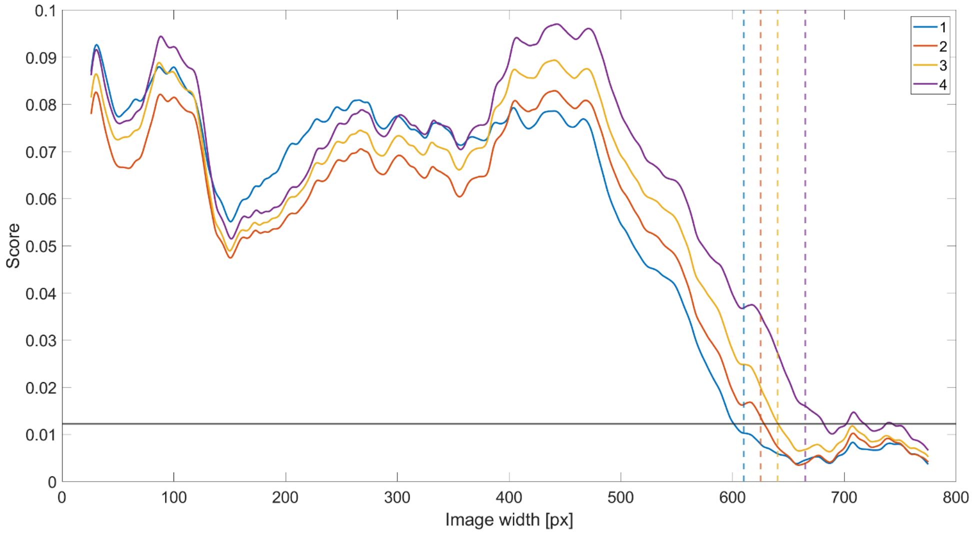
The intersection of the score with the threshold t_c (black horizontal threshold line shown in Fig. 8.5a) indicates the crack tip. The systematic rightward shift of these intersection points across successive recordings aligns with the propagation of the crack. The automatically detected crack tip positions for the respective test cases correspond to manually identified locations, which lie between the 610^{th} and 665^{th} pixels.

The scores corresponding to the second camera are shown in Fig. 8.5b. Irregular variations are visible before the 600^{th} pixel, primarily due to surface disturbances and the presence of measurement equipment, which degrade motion-estimation accuracy; these appear as irregular spots in the amplitude maps (Fig. 8.3d). Despite these local anomalies, the method's overall reliability is not compromised. A gradual decline in the score is observed throughout the map, except for the terminal segment after approximately the 1200th pixel, which corresponds to the undamaged reference zone. Unlike the sharp drop observed with the first camera, the transition here is more gradual. This difference is attributed to reduced relative motion magnitude near the crack tip, resulting from the crack breathing phenomenon. As a result, the subtler displacement variations lead to a more gradual decline in the score.

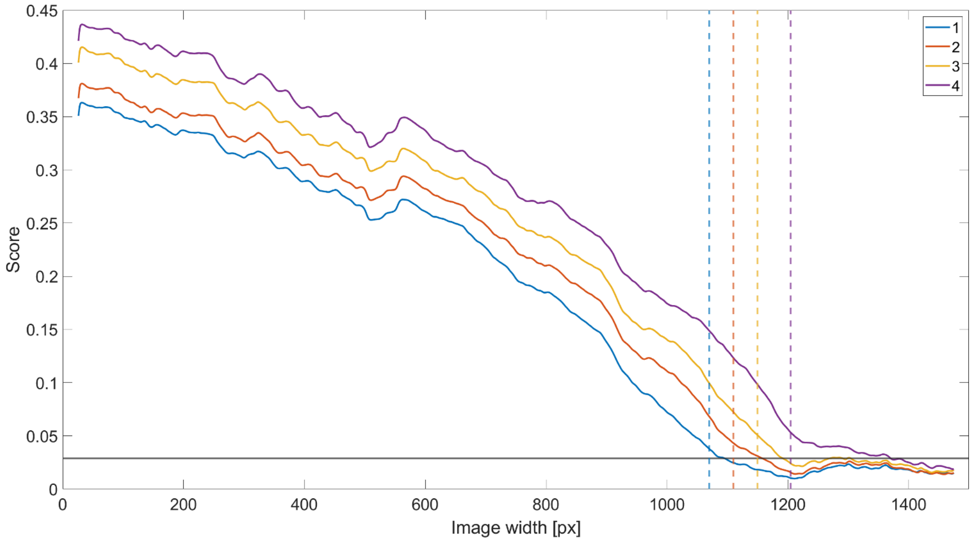
For threshold determination, the reference region is again defined as the last 10% of the score profile (corresponding in this case to the last 145 pixels) in the first amplitude map. The indicated crack tip positions are slightly shifted rightward, while in the final case the shift is more pronounced.

To quantify the accuracy of the automatic crack tip localization, the errors are calculated as the difference between the detected and manually assessed crack tip positions for both cameras. The results are expressed in pixels as well as in millimeters, with the latter obtained from approximate image scaling factors. Since the camera optical axis was not perfectly perpendicular to the specimen surface (although nearly so), this conversion should be treated as a coarse estimate rather than a precise physical measurement. Table 8.1 presents the errors for four consecutive crack propagation stages together with the MAE values.

For the first camera, the MAE is 0.68 mm, which is lower than the resolution of the manual measurement. For the second camera, the MAE is 2.52 mm; however, this value is notably influenced by the fourth crack tip position. When considering only positions one to three, the MAE



(a)



(b)

Figure 8.5: Scores computed for the plate-like structure experiment: (a) results from the first camera, (b) results from the second camera. The horizontal line indicates the threshold t_c calculated using Eq. 8.1. Vertical dashed lines denote approximate crack tip positions for different propagation stages.

Table 8.1: Error across four consecutive crack propagation stages for the first and second cameras.

	Stage 1	Stage 2	Stage 3	Stage 4	**MAE**
Camera 1					
Pixels (px)	−9	3	1	17	**7.5**
Millimeters (mm)	−0.82	0.27	0.09	1.55	**0.68**
Camera 2					
Pixels (px)	26	48	42	166	**70.5**
Millimeters (mm)	0.93	1.71	1.50	5.93	**2.52**

decreases to 1.38 mm, which is only slightly above the resolution of the manual measurement. The increased error for the more advanced crack tip is attributed to the use of the score from the video corresponding to the first crack tip position as the uncracked reference, rather than recalculating it separately for each propagation stage. This approach was chosen because of its greater potential for application in crack growth monitoring, where updating the reference during propagation may not be feasible, and it will be further discussed in Section 8.4.3.

For the second camera, the score profiles decrease gradually, whereas for the first camera the drop is more rapid, which appears to be more suitable for thresholding. Although the higher error observed with the higher image resolution and the much clearer amplitude maps may seem counterintuitive, in this case the lower image resolution may have acted as an averaging factor that facilitated the analysis.

8.4 Parameters

The results presented in the previous sections were obtained using manually selected parameter values. In this section, the influence of several key parameters is examined. A formal optimization was not performed; instead, the parameter space was sampled to demonstrate the robustness of the method with respect to various choices. In particular, the following aspects are discussed: the localization of the reference point and the template size used for stabilization, as well as the proportion of the score profile taken as the reference region. Each of these parameters is addressed in the subsequent subsections.

8.4.1 Reference point localization

To examine the sensitivity of the method to reference point placement, ten alternative points were randomly selected within a range of ± 25 pixels around the originally chosen position. For each reference point, the entire processing pipeline was repeated, and the final results were evaluated in terms of the absolute error between the automatically predicted and manually measured crack tip locations. The errors obtained for both cameras are summarized in Table 8.2.

For the first camera, the variation in error with respect to reference point localization remains relatively small. In most cases the MAE is below 1 mm, and in the remaining cases it only slightly exceeds this value. For the second camera, the MAE is approximately 2.5 mm for the majority of reference point localizations, with only two cases yielding noticeably lower errors (1.35 and 1.54 mm). Overall, within the sampled ± 25 px neighborhood, the method exhibits robust behavior with respect to reference-point localization, although the second camera consistently shows higher error levels.

8.4.2 Template size

Another factor that may influence the performance of the method is the template size w used for stabilization. The template must be large enough to contain sufficient texture for reliable matching. To investigate this aspect, several values of w were tested while keeping the reference point fixed, and the resulting crack tip localization errors were compared. The errors for both cameras are summarized in Table 8.3. In each case, the smallest template size that still yielded valid results (i.e., the template could be matched such that the mROI remained within the video frame) is included, and from $w = 21$ px upward the parameter space was sampled in increments of 10 px.

For the first camera, although some variation in MAE is observed across template sizes, a practical choice is to use larger templates to ensure stable matching. In particular, $w \geq 61$ px yields consistently low errors (MAE 0.68–0.70 mm). While some smaller windows (e.g., $w = 19$ px) also produced low errors in this dataset, these outcomes may reflect particularly favorable local texture and are unlikely to generalize. Accordingly, selecting $w \in [61, 71]$ px is a reasonable default for this configuration.

For the second camera, the MAE is essentially flat across the tested range (2.52–2.60 mm), with identical values for $w = 41$–71 px. Errors are dominated by the fourth stage, and variations with w are negligible compared to stage-to-stage differences. This limited sensitivity to template size is likely a consequence of the smaller field of view of the second camera, which increases the pixel-to-millimeter ratio and thus allows more effective texture capture even for moderate template sizes.

The choice of $w = 61$ px, used in the preceding analysis, was therefore appropriate. These results confirm that once the template size is sufficiently large, further increases do not materially affect the outcome.

8.4.3 Reference score length

The final parameter investigated concerns the proportion of the score profile used to define the reference region for threshold calculation. Two approaches are compared. In the first, a single threshold t_c is derived from the reference region of the first-stage score profile and then applied consistently to all subsequent stages; this approach is hereafter referred to as *global thresholding*.

Table 8.2: Error obtained across four consecutive crack propagation stages for different reference point localizations, reported in pixels and millimeters. For each reference point, the MAE is also shown. S1–S4 denote consecutive crack propagation stages.

Camera 1

Ref. point	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
Original	−9	3	1	17	**7.50**	−0.82	0.27	0.09	1.55	**0.68**
1	7	14	14	23	**14.50**	0.64	1.27	1.27	2.09	**1.32**
2	−16	−3	−10	−13	**10.50**	−1.45	−0.27	−0.91	−1.18	**0.95**
3	−14	−2	−7	−15	**9.50**	−1.27	−0.18	−0.64	−1.36	**0.86**
4	−8	3	2	17	**7.50**	−0.73	0.27	0.18	1.55	**0.68**
5	−14	−4	−6	−3	**6.75**	−1.27	−0.36	−0.55	−0.27	**0.61**
6	−1	10	11	23	**11.25**	−0.09	0.91	1.00	2.09	**1.02**
7	−16	−4	−8	−10	**9.50**	−1.45	−0.36	−0.73	−0.91	**0.86**
8	1	14	15	18	**12.00**	0.09	1.27	1.36	1.64	**1.09**
9	−15	−18	−8	−9	**12.50**	−1.36	−1.64	−0.73	−0.82	**1.14**
10	−14	−1	−7	−13	**8.75**	−1.27	−0.09	−0.64	−1.18	**0.80**

Camera 2

Ref. point	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
Original	26	48	42	166	**70.5**	0.93	1.71	1.50	5.93	**2.52**
1	28	51	50	168	**74.25**	1.00	1.82	1.79	6.00	**2.65**
2	25	30	39	165	**64.75**	0.89	1.07	1.39	5.89	**2.31**
3	1	14	27	130	**43.00**	0.04	0.50	0.96	4.64	**1.54**
4	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**
5	25	44	39	165	**68.25**	0.89	1.57	1.39	5.89	**2.44**
6	26	48	46	166	**71.50**	0.93	1.71	1.64	5.93	**2.55**
7	25	34	39	165	**65.75**	0.89	1.21	1.39	5.89	**2.35**
8	29	53	52	190	**81.00**	1.04	1.89	1.86	6.79	**2.89**
9	24	37	37	164	**65.50**	0.86	1.32	1.32	5.86	**2.34**
10	0	13	26	112	**37.75**	0.00	0.46	0.93	4.00	**1.35**

Table 8.3: Error across four consecutive crack propagation stages for different template sizes w, reported in pixels and millimeters. For each template size, the MAE is also shown. The original template size $w = 61$ is highlighted. S1–S4 denote consecutive crack propagation stages.

Camera 1										
Template size w (px)	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
19	−1	5	5	20	**7.75**	−0.09	0.45	0.45	1.82	**0.70**
21	−8	4	4	19	**8.75**	−0.73	0.36	0.36	1.73	**0.80**
31	−13	−20	4	18	**13.75**	−1.18	−1.82	0.36	1.64	**1.25**
41	−13	−19	3	20	**13.75**	−1.18	−1.73	0.27	1.82	**1.25**
51	−13	−19	3	14	**12.25**	−1.18	−1.73	0.27	1.27	**1.11**
61	−9	3	1	17	**7.50**	−0.82	0.27	0.09	1.55	**0.68**
71	−9	4	1	17	**7.75**	−0.82	0.36	0.09	1.55	**0.70**

Camera 2										
Template size w (px)	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
19	26	49	43	166	**71.00**	0.93	1.75	1.54	5.93	**2.54**
21	28	52	44	167	**72.75**	1.00	1.86	1.57	5.96	**2.60**
31	27	49	43	166	**71.25**	0.96	1.75	1.54	5.93	**2.54**
41	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**
51	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**
61	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**
71	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**

In the second, a separate threshold t_c is determined for each stage individually, using its own score profile; this approach is hereafter referred to as *stage-wise thresholding*. For the first camera, the reference region was varied from 50 to 100 px, in increments of 5 px, while for the second camera it was varied from 50 to 240 px, in increments of 10 px. This selection ensured that the reference region was always located beyond the crack tip, so that it represented an intact part of the specimen. The resulting errors are summarized in Tables 8.4–8.7.

Table 8.4: Error for the first camera across four consecutive crack propagation stages for different reference score lengths (global thresholding), reported in pixels and millimeters. For each score length, the MAE is also shown. The originally used length of 75 px is highlighted. S1–S4 denote consecutive crack propagation stages.

Score (px)	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
50	−10	2	0	16	**7.00**	−0.91	0.18	0.00	1.45	**0.64**
55	−9	3	0	16	**7.00**	−0.82	0.27	0.00	1.45	**0.64**
60	−9	3	1	17	**7.50**	−0.82	0.27	0.09	1.55	**0.68**
65	−8	4	1	17	**7.50**	−0.73	0.36	0.09	1.55	**0.68**
70	−9	3	1	17	**7.50**	−0.82	0.27	0.09	1.55	**0.68**
75	−9	3	1	17	**7.50**	−0.82	0.27	0.09	1.55	**0.68**
80	−8	4	1	17	**7.50**	−0.73	0.36	0.09	1.55	**0.68**
85	−9	3	1	17	**7.50**	−0.82	0.27	0.09	1.55	**0.68**
90	−10	2	−1	15	**7.00**	−0.91	0.18	−0.09	1.36	**0.64**
95	−11	2	−1	15	**7.25**	−1.00	0.18	−0.09	1.36	**0.66**
100	−10	2	−1	15	**7.00**	−0.91	0.18	−0.09	1.36	**0.64**
MAE	9.27	2.82	0.82	16.27		0.84	0.26	0.07	1.48	

For the first camera, the results are highly stable across the tested reference lengths. Under global thresholding (Table 8.4), the MAE across stages remains in a narrow range of 0.64–0.68 mm, with most of the error attributable to stage 4. Stage-wise thresholding (Table 8.5) yields nearly identical MAE (0.64–0.68 mm). In this case, the reduction of MAE in stage 4 (from 1.48 to 1.07 mm) is accompanied by an increase in stage 3 errors. Nevertheless, the latter remain well below the resolution of manual measurement, whereas the improvement in stage 4 is closer to that resolution and therefore of greater practical relevance. Overall, the performance in terms of MAE is essentially unchanged, but the stage-wise strategy may offer a slight advantage in practice, at the cost of permanently reserving a portion of the field of view for reference purposes.

For the second camera, the influence of reference score length is much more pronounced. With global thresholding (Table 8.6), short reference regions yield very large errors, including undefined results for stage 4 at 50 px, where the entire score profile remained above the threshold. Extending the reference reduces the error substantially, with the MAE across stages decreasing from more

Table 8.5: Error for the first camera across four consecutive crack propagation stages for different reference score lengths (stage-wise thresholding), reported in pixels and millimeters. For each score length, the MAE is also shown. S1–S4 denote consecutive crack propagation stages.

Score (px)	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
50	−10	−1	−3	−10	**6.00**	−0.91	−0.09	−0.27	−0.91	**0.55**
55	−9	−1	−3	−10	**5.75**	−0.82	−0.09	−0.27	−0.91	**0.52**
60	−9	−1	−4	−10	**6.00**	−0.82	−0.09	−0.36	−0.91	**0.55**
65	−8	−2	−6	−12	**7.00**	−0.73	−0.18	−0.55	−1.09	**0.64**
70	−9	−4	−8	−13	**8.50**	−0.82	−0.36	−0.73	−1.18	**0.77**
75	−9	−3	−8	−13	**8.25**	−0.82	−0.27	−0.73	−1.18	**0.75**
80	−8	−3	−7	−13	**7.75**	−0.73	−0.27	−0.64	−1.18	**0.70**
85	−9	−3	−7	−12	**7.75**	−0.82	−0.27	−0.64	−1.09	**0.70**
90	−10	−4	−7	−12	**8.25**	−0.91	−0.36	−0.64	−1.09	**0.75**
95	−11	−4	−7	−12	**8.50**	−1.00	−0.36	−0.64	−1.09	**0.77**
100	−10	−4	−7	−12	**8.25**	−0.91	−0.36	−0.64	−1.09	**0.75**
MAE	9.27	2.73	6.09	11.73		0.84	0.25	0.55	1.07	

than 6 mm for short regions to approximately 2 mm at lengths around 190–210 px. It is worth noting, however, that in terms of physical dimensions a 50 px score length for the first camera corresponds to roughly 130 px for the second camera. In fact, beginning from 120 px the MAE drops considerably and stabilizes at lower values. Stage-wise thresholding (Table 8.7) provides a marked improvement: MAE across stages decreases from over 5 mm at 50 px to consistently sub-millimeter values once the reference length exceeds 120 px. In this range, errors across all stages are balanced and remain below the resolution of manual measurement. These results indicate that, unlike the first camera, the second camera setup benefits from careful selection of the reference length, with stage-wise thresholding offering the most accurate and stable localization.

In summary, the influence of reference score length differs markedly between the two camera setups. For the first camera, errors remain stable across the tested range of 50–100 px, and both global and stage-wise thresholding yield comparable MAE values, with only minor differences between stages. For the second camera, the tested range is wider (50–240 px), and because of the higher pixel-to-millimeter ratio, short reference lengths correspond to physically much smaller regions than those tested with the first camera. At these scales, global thresholding produces large errors, but accuracy improves steadily as the reference is extended. Stage-wise thresholding further enhances performance, reducing errors to sub-millimeter levels once the reference length exceeds approximately 120 px. These findings indicate that the method is robust to the specific choice of reference length, provided the region is sufficiently long to represent an intact part of the specimen, though the minimal required length is case-specific.

Table 8.6: Error for the second camera across four consecutive crack propagation stages for different reference score lengths (global thresholding), reported in pixels and millimeters. For each score length, the MAE is also shown. The originally used length of 145 px is added here for clarity and highlighted. S1–S4 denote consecutive crack propagation stages. — marks undefined values.

Score (px)	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
50	89	86	280	—	**—**	3.18	3.07	10.00	—	**—**
60	72	76	263	265	**169.00**	2.57	2.71	9.39	9.46	**6.04**
70	70	75	262	233	**160.00**	2.50	2.68	9.36	8.32	**5.71**
80	65	69	72	226	**108.00**	2.32	2.46	2.57	8.07	**3.86**
90	47	66	57	202	**93.00**	1.68	2.36	2.04	7.21	**3.32**
100	43	64	55	199	**90.25**	1.54	2.29	1.96	7.11	**3.22**
110	35	58	51	190	**83.50**	1.25	2.07	1.82	6.79	**2.98**
120	27	49	43	166	**71.25**	0.96	1.75	1.54	5.93	**2.54**
130	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**
140	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**
145	26	48	42	166	**70.50**	0.93	1.71	1.50	5.93	**2.52**
150	25	47	41	165	**69.50**	0.89	1.68	1.46	5.89	**2.48**
160	24	46	41	165	**69.00**	0.86	1.64	1.46	5.89	**2.46**
170	22	45	40	164	**67.75**	0.79	1.61	1.43	5.86	**2.42**
180	14	41	36	162	**63.25**	0.50	1.46	1.29	5.79	**2.26**
190	12	39	35	139	**56.25**	0.43	1.39	1.25	4.96	**2.01**
200	12	38	35	132	**54.25**	0.43	1.36	1.25	4.71	**1.94**
210	12	38	35	138	**55.75**	0.43	1.36	1.25	4.93	**1.99**
220	13	40	35	161	**62.25**	0.46	1.43	1.25	5.75	**2.22**
230	13	40	36	161	**62.50**	0.46	1.43	1.29	5.75	**2.23**
240	13	41	36	162	**63.00**	0.46	1.46	1.29	5.79	**2.25**
MAE	32.67	52.48	75.19	176.40		1.17	1.87	2.69	6.30	

Table 8.7: Error for the second camera across four consecutive crack propagation stages for different reference score lengths (stage-wise thresholding), reported in pixels and millimeters. For each score length, the MAE is also shown. S1–S4 denote consecutive crack propagation stages.

Score (px)	Pixels (px)					Millimeters (mm)				
	S1	S2	S3	S4	**MAE**	S1	S2	S3	S4	**MAE**
50	89	85	259	187	**155.00**	3.18	3.04	9.25	6.68	**5.54**
60	72	77	58	164	**92.75**	2.57	2.75	2.07	5.86	**3.31**
70	70	74	53	163	**90.00**	2.50	2.64	1.89	5.82	**3.21**
80	65	67	43	130	**76.25**	2.32	2.39	1.54	4.64	**2.72**
90	47	62	36	106	**62.75**	1.68	2.21	1.29	3.79	**2.24**
100	43	56	33	100	**58.00**	1.54	2.00	1.18	3.57	**2.07**
110	35	45	29	45	**38.50**	1.25	1.61	1.04	1.61	**1.38**
120	27	32	25	19	**25.75**	0.96	1.14	0.89	0.68	**0.92**
130	26	28	24	15	**23.25**	0.93	1.00	0.86	0.54	**0.83**
140	26	26	23	11	**21.50**	0.93	0.93	0.82	0.39	**0.77**
150	25	23	21	7	**19.00**	0.89	0.82	0.75	0.25	**0.68**
160	24	22	20	4	**17.50**	0.86	0.79	0.71	0.14	**0.63**
170	22	19	18	0	**14.75**	0.79	0.68	0.64	0.00	**0.53**
180	14	11	15	−5	**11.25**	0.50	0.39	0.54	−0.18	**0.40**
190	12	9	13	−10	**11.00**	0.43	0.32	0.46	−0.36	**0.39**
200	12	8	10	−13	**10.75**	0.43	0.29	0.36	−0.46	**0.38**
210	12	8	9	−15	**11.00**	0.43	0.29	0.32	−0.54	**0.39**
220	13	9	9	−17	**12.00**	0.46	0.32	0.32	−0.61	**0.43**
230	13	11	10	−19	**13.25**	0.46	0.39	0.36	−0.68	**0.47**
240	13	12	10	−21	**14.00**	0.46	0.43	0.36	−0.75	**0.50**
MAE	33.00	34.20	35.90	52.55		1.18	1.22	1.28	1.88	

8.5 Results for smartphone recordings

This section presents results obtained from the smartphone recordings. Two propagation stages are considered: the second and the fourth. Figs. 8.6 and 8.7 correspond to the second stage but differ in camera placement: the former employs a more distant view, whereas the latter provides a closer view. Fig. 8.8 presents the fourth propagation stage.

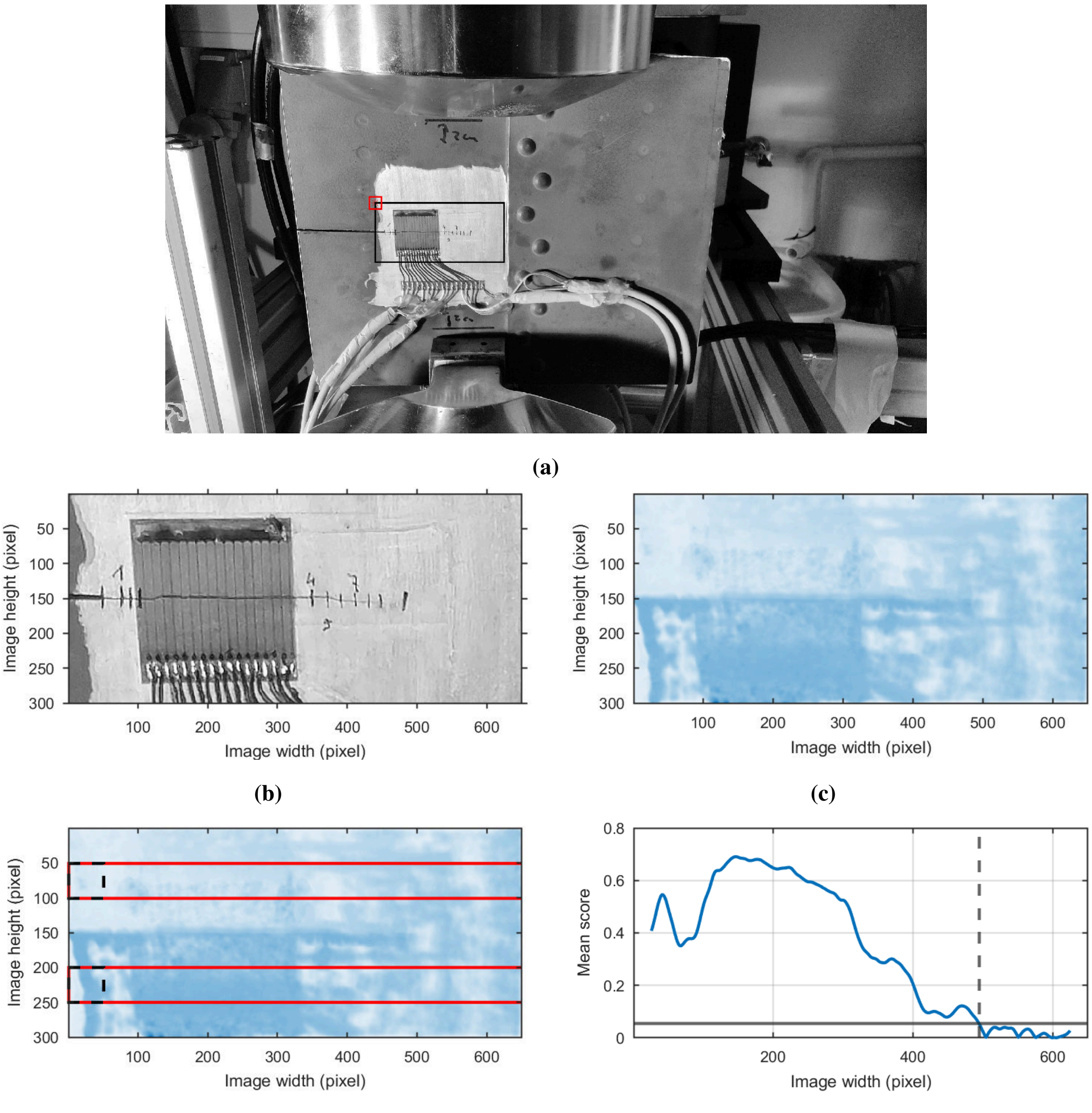
In all three cases, the mROIs were chosen to cover approximately the same physical region of the specimen, as shown in Figs. 8.6a, 8.7a, and 8.8a. The mROI is indicated by a black rectangle, the reference point by a red asterisk, and the stabilization template by a red square. Only the amplitude maps of the vertical OF component are analyzed, as these are the most informative.

Due to differences in field of view and image resolution, the sROI size was set to 50×50 pixels for the distant view configuration (Figs. 8.6d and 8.8d) and 100×50 pixels for the close-up configuration (Fig. 8.7d). The stabilized frames, which correspond to the selected mROIs, are shown in Figs. 8.6b, 8.7b, and 8.8b. For all smartphone videos, the last 10% of the score is used as the uncracked reference area for calculating the threshold t_c.

First, results from two smartphone placements at crack stage number two are discussed. Fig. 8.6c presents the amplitude map for the distant view. The map appears slightly blurred, however, two distinct regions in the upper and lower parts are still distinguishable, particularly in the left half of the image. This corresponds to the location of the crack propagation gauge. The higher contrast in the left part of the map, where the gauge is present, and the more blurred appearance in the right part, are also reflected in the score (Fig. 8.6e) as a rapid drop after the 300th pixel. The crack tip is indicated at pixel 495, which matches perfectly with the manually assessed crack tip location.

The amplitude map for the close-up view of stage two is shown in Fig. 8.7c, and the corresponding score is presented in Fig. 8.7e. The map is very clear, with two distinct regions that are sharply delineated. In the score profile, the values are initially high and then decrease gradually, becoming flat toward the end. The crack tip is located at approximately pixel 1230, while the score-based detection indicates it at pixel 1252. It is worth noting that, in this case, the smartphone was held at a slight angle during recording, so the crack is not perfectly horizontal in the frames. Nevertheless, the method still provides correct localization. Compared to the distant-view configuration, the close-up view yields a noticeably clearer amplitude map and a smoother, more interpretable score profile. Despite this, the crack tip localization was correct in both cases, and in the distant view it even appears slightly more accurate, further supporting the claim that the method is robust.

The amplitude map for stage four is shown in Fig. 8.8c, and the corresponding score is presented in Fig. 8.8e. Two distinct regions are again clearly visible in the map. The score transitions smoothly from high to low values, without abrupt changes. The crack tip is located at pixel 430, while the score-based detection indicates it at pixel 396. Although this recording was made with a field of view similar to that of the distant-view configuration for stage two, the setups are not identical, so a direct comparison between the two stages is not possible.


(a)

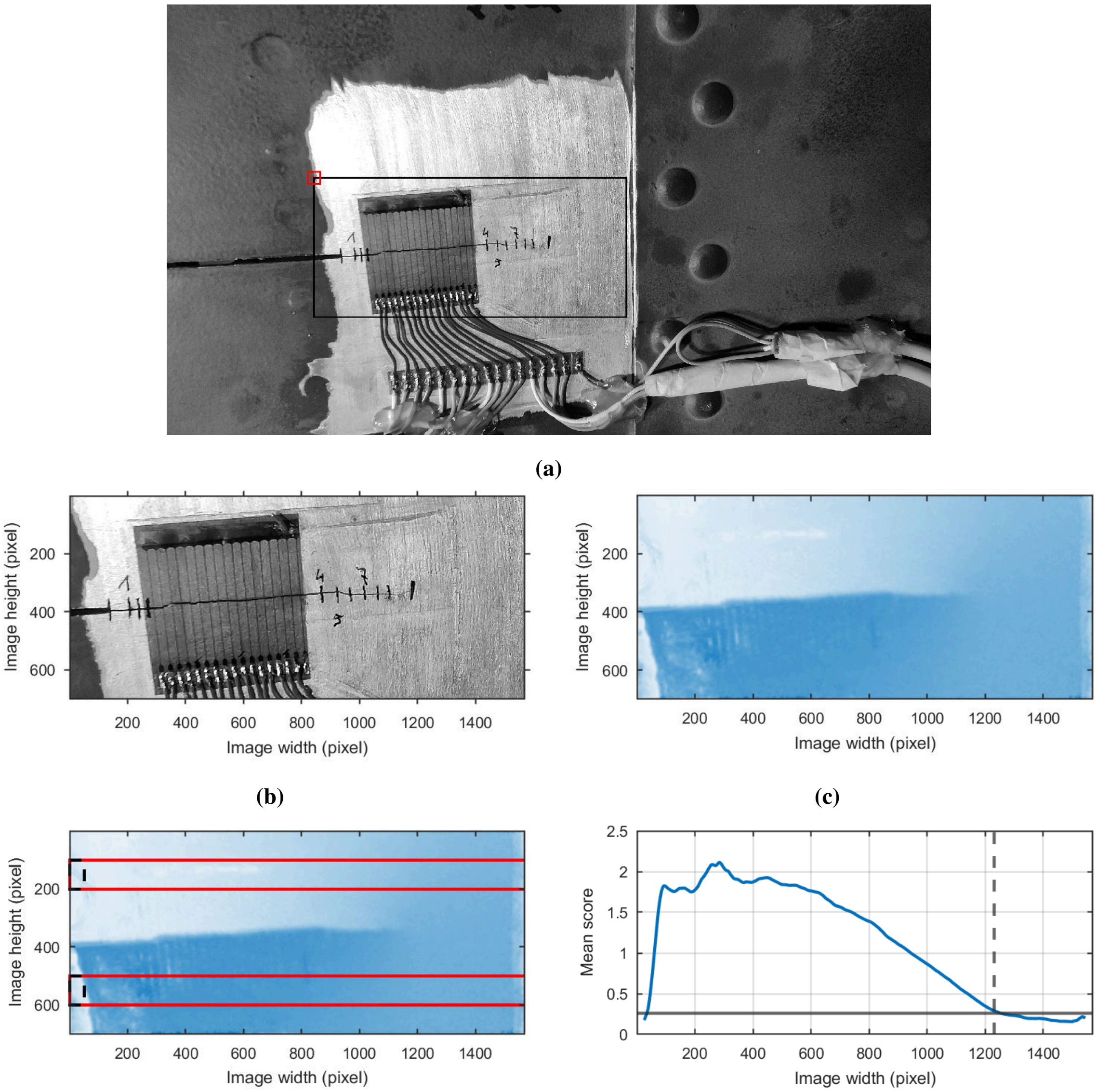
(b)

(c)

(d)

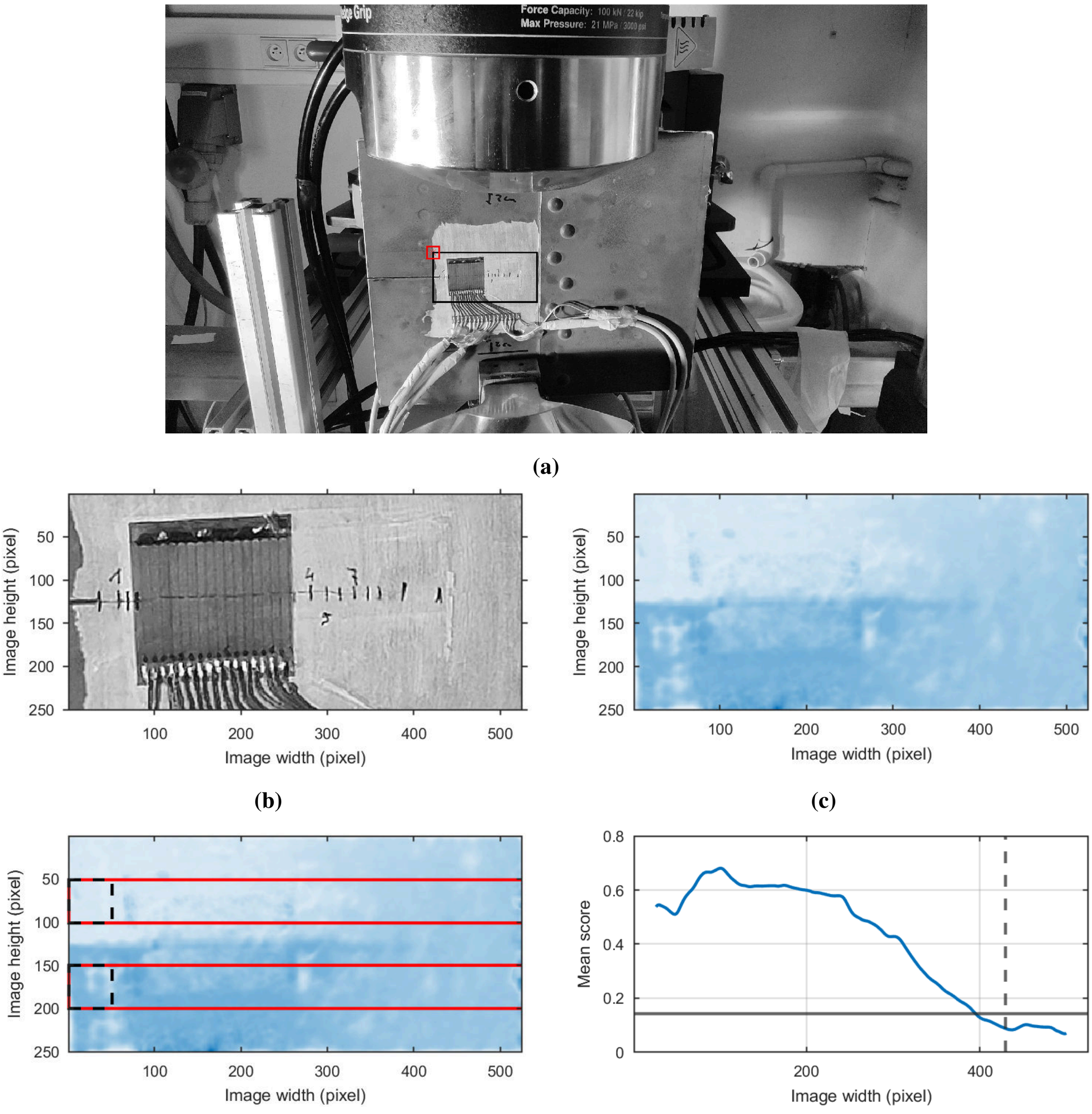
(e)

Figure 8.6: Results for the plate-like structure experiment using the smartphone video, shown for the second propagation stage (distant view configuration): (a) raw frame with the mROI (black rectangle), reference point (red asterisk), and stabilization template (red square) indicated; (b) corresponding frame after stabilization; (c) amplitude map of the vertical OF component; (d) amplitude map with the sROIs marked (the black square shifted pixel by pixel within the red rectangle); (e) score as a function of pixel position. The horizontal line indicates the threshold t_c calculated using Eq. 8.1, and the vertical dashed line denotes the approximate crack tip position.



(a) (b) (c) (d) (e)

Figure 8.7: Results for the plate-like structure experiment using the smartphone video, shown for the second propagation stage (close-up view configuration): (a) raw frame with the mROI (black rectangle), reference point (red asterisk), and stabilization template (red square) indicated; (b) corresponding frame after stabilization; (c) amplitude map of the vertical OF component; (d) amplitude map with the sROIs marked (the black square shifted pixel by pixel within the red rectangle); (e) score as a function of pixel position. The horizontal line indicates the threshold t_c calculated using Eq. 8.1, and the vertical dashed line denotes the approximate crack tip position.



(a)

(b)

(c)

(d)

(e)

Figure 8.8: Results for the plate-like structure experiment using the smartphone video, shown for the fourth propagation stage: (a) raw frame with the mROI (black rectangle), reference point (red asterisk), and stabilization template (red square) indicated; (b) corresponding frame after stabilization; (c) amplitude map of the vertical OF component; (d) amplitude map with the sROIs marked (the black square shifted pixel by pixel within the red rectangle); (e) score as a function of pixel position. The horizontal line indicates the threshold t_c calculated using Eq. 8.1, and the vertical dashed line denotes the approximate crack tip position.

The presented results confirm that the proposed method is applicable to smartphone recordings, even under non-ideal acquisition conditions such as varying camera distances, changes in field of view, and slight rotation of the device during recording. In all tested configurations, the amplitude maps provided sufficient contrast for effective analysis, and the score-based approach yielded crack tip indications that, while not perfectly aligned in every case, were consistently within a very small margin of the manually assessed positions. These findings demonstrate that the method remains reliable and accurate when applied to lower-grade, handheld recording equipment, making it suitable for practical use outside laboratory conditions.

8.6 Discussion

The results obtained for the plate-like specimen demonstrate that amplitude maps of the vertical OF component enable reliable crack-tip localization across multiple propagation stages and heterogeneous acquisition setups, namely two fixed cameras and a smartphone. The directional maps consistently reveal two regions with distinct motion patterns, separated by a visually clear transition, and the associated score profiles provide a compact one-dimensional indicator from which a simple threshold yields accurate crack-tip positions (see Figs. 8.3 and 8.5).

Directional vs. OF magnitude amplitude maps. Consistent with Chapter 6, the directional maps, specifically the vertical component in this case, proved more informative for crack assessment than OF magnitude maps. For the first camera, magnitude maps were dominated by artifacts associated with surface-mounted instrumentation and the crack border, while for the second camera their contrast was limited and diminished prematurely. In both cases, they were inferior to the vertical-component maps and were therefore not analyzed further.

Score profiles. The score profiles derived from directional amplitude maps provide a one-dimensional indicator of motion-pattern dissimilarity extracted from the map representation. For the first camera, a rapid drop toward low values near the crack tip is followed by a flat segment corresponding to the intact region, which makes thresholding straightforward. For the second camera, the transition is more gradual, consistent with smaller relative motion near the crack tip due to crack breathing, nonetheless, the intact region is still captured as a low, stable segment (see Fig. 8.5).

Global vs. stage-wise threshold. The fixed threshold t_c proved effective across stages and cameras. For the first camera, the choice between a single (global) threshold and stage-wise thresholds had little impact on overall MAE (Tables 8.4–8.5), though stage-wise thresholds reduced late-stage errors. For the second camera, stage-wise thresholds are advantageous once a sufficiently long reference region was used (Tables 8.6–8.7). In monitoring scenarios where frequent re-referencing may be impractical, a global threshold remains attractive as a simpler approach.

Conversely, stage-wise thresholding is potentially more suitable under variable experimental conditions, such as changes in lighting or camera placement between measurements. In practice, two operational modes can be distinguished: global thresholding if the crack is not expected to extend into the reference segment, and stage-wise thresholding when reliable redefinition of the reference is feasible.

Length of the reference segment. Sensitivity to the reference segment length is not a critical factor. For the first camera, MAE is stable over 50–100 px and low for both global and stage-wise thresholding (Tables 8.4–8.5). For the second camera, the tested range is wider (50–240 px), and the shorter pixel lengths correspond to physically smaller regions than those considered for the first camera due to the higher px/mm ratio. At these short physical lengths, global thresholding yields large errors, whereas extending the reference steadily improves accuracy. Once the reference exceeds about 120 px (roughly comparable in physical scale to $\sim$50 px for the first camera), errors converge to similarly low levels and stage-wise thresholds reach sub-millimetric errors (Tables 8.6–8.7). Thus, when compared at similar physical scales, both cameras exhibit comparable behavior. The apparent differences arise from the disparate px/mm ratios and the broader range explored with the second camera. Overall, the method is robust once the reference segment is sufficiently long to represent an intact region, while the minimal required length is case-specific.

Robustness to reference-point placement and template size. The stabilization reference-point location has limited influence within a ± 25 px neighborhood (Table 8.2), supporting robustness to modest variations in placement provided the reference patch contains stable texture. Similarly, the template size w used for matching exhibits weak sensitivity once it exceeds the scale needed to capture sufficient texture. For the first camera, $w \geq 61$ px yields consistently low errors, whereas occasional success at smaller w likely reflects favorable local texture and should not be overinterpreted. For the second camera, MAE remains identical across all tested values of $w \geq 41$ px (Table 8.3).

Smartphone recordings and viewpoint variability. Smartphone recordings, covering distant and close-up views with mild in-plane rotation, demonstrate the method's robustness to non-ideal conditions. Despite differences in field of view, resolution, and slight rotation, the directional maps remained interpretable and the score-based crack tip localization aligned well with manual annotations (Figs. 8.6–8.8). These findings suggest that the method can extend beyond laboratory-grade cameras and operate effectively even under less controlled acquisition settings, which is promising for practical deployment.

Error assessment. Pixel-to-millimeter conversions were reported to provide physical context but should be regarded as coarse approximations rather than precise measurements. In particular, sub-millimetric differences fall below the resolution of manual crack-tip annotations (Table 8.1) and should be interpreted with caution. When the MAE approaches the annotation resolution, practical differences become negligible.

Limitations. Several limitations of the proposed approach should be acknowledged. First, the analysis was conducted under laboratory conditions, and performance under more complex operational and environmental conditions remains to be established. Second, although the method avoids manual parameter tuning, it still requires stabilization with a reference point and is sensitive to the presence of sufficient surface texture. Third, errors were evaluated against manual crack-tip annotations, which themselves have limited resolution and introduce uncertainty. Finally, the current implementation relies on dense OF computation, which is computationally demanding and may restrict real-time applicability without further optimization. These limitations highlight directions for future research, including validation on larger specimen sets, testing under in-service conditions, and exploration of more efficient motion-estimation algorithms.

Practical guidance. The results provide several practical insights for applying the method. First, the directional component of the OF should be prioritized where possible, as it consistently yields clearer amplitude maps than the magnitude. Stabilization was found to be robust to modest variations in reference-point placement, provided the selected patch contained sufficient texture and the template size was sufficiently large. The length of the reference segment used for thresholding is also important: it should be long enough to represent an intact region, and in this study lengths of at least 50 px in the first camera or 120 px in the second proved sufficient. Finally, with regard to thresholding strategies, both global and stage-wise approaches are viable, with the global approach offering simplicity and the stage-wise approach providing slightly improved accuracy when reference updates are feasible. These guidelines can facilitate adaptation of the method to different setups and support its application beyond laboratory conditions.

Chapter 9

Discussion and future work

This chapter presents a general discussion that synthesizes the findings from the compressor, beam, and plate-like structure experiments, unifying the outcomes reported in the discussion sections of Chapters 5, 6, and 8. The discussion concentrates on core aspects: the role of LOI as a reliability indicator and the pipeline for crack visualization and localization. It reviews empirical behavior, practical settings, and scope of applicability, and it comments on robustness across measurement setups without reliance on fine-tuned parameters. The chapter is structured as follows: Section 9.1 addresses the role and behavior of LOI; Section 9.2 discusses the crack assessment methodology; and Section 9.3 outlines directions for future work.

9.1 Local Orientation Inconsistency

The first part of the discussion focuses on the LOI metric introduced in Chapter 5. This measure was designed to assess the reliability of motion estimation by quantifying the directional consistency of local OF orientations. The following perspectives synthesize the cross-experimental findings, covering its role within the processing pipeline, empirical behavior across datasets, procedures for mask formation, and the assumptions and limitations that define its scope of applicability.

Role in the pipeline. LOI can serve either as a standalone motion-estimation reliability assessment or as a preprocessing stage preceding crack assessment. As defined in Chapter 5, LOI quantifies local directional inconsistency of the OF orientation field $\theta(x, y, t)$ by comparing each pixel's orientation with that of its neighbors in a spatial window and aggregating this measure over time. In practice, it can be used to suppress pixels where motion direction is inconsistent or undefined, thereby reducing the influence of spurious elements on subsequent analyses. Moreover, LOI guided the selection of the mROI, which further supported crack visualization and crack tip localization.

Empirical behavior across datasets. Across the compressor, beam, and plate-like structure experiments, low LOI values consistently aligned with uniformly moving, textured, trackable areas such as surfaces covered with random speckle patterns, where local OF orientations varied little across neighboring pixels and frames. Elevated LOI highlighted regions where motion estimation was unreliable: smoothly painted parts with no texture, visually homogeneous backgrounds, reflections, or areas affected by image artifacts. In the beam experiment, increasing the excitation amplitude improved conditions for motion estimation and reduced LOI on the specimen, whereas background elements retained elevated values. In the plate-like structure experiment, independently moving or flexible components not belonging to the target structure were observed to form coherent motion patterns distinct from the specimen. LOI helped to identify such regions as less reliable for analyses focused on the main structure.

Mask formation and practical settings. The LOI field can be converted into a binary mask, with the threshold chosen according to the scene's noise level, available texture, and the desired coverage of reliable regions. If the threshold is set too strictly, holes or isolated artifacts may appear. When multiple recordings under different operating conditions are available, averaging their LOI fields can provide a more general indication that remains valid across slightly varying conditions. The spatial neighborhood should be kept small enough to avoid mixing orientations across discontinuities, yet large enough to stabilize the local measure.

Assumptions and scope of applicability. The LOI depends on the presence of sufficient local texture to support consistent orientation estimates. In texture-poor or smoothly painted regions, LOI values become elevated, which correctly signals unreliable motion. An important consideration is the case of thin or flexible elements whose motion is locally coherent but oriented differently from their immediate surroundings. In such cases, LOI is elevated even though the underlying motion estimation is correct, yet this outcome is consistent with the metric's assumption that reliable motion corresponds to directional consistency within a neighborhood, and it helps to filter out such elements so that the analysis can concentrate on the main body of the structure. Similarly, in regions affected by reflections or other image artifacts, elevated LOI values indicate poor motion estimation quality and allow such unreliable areas to be excluded from further analysis. Overall, LOI should therefore be regarded as a supportive reliability indicator rather than a strict motion estimation quality metric.

9.2 Crack visualization and assessment

The second part of the discussion addresses the methodology of crack assessment developed in Chapters 6–8. This approach combines stabilization, frequency-domain processing of dense OF, and automated analysis of amplitude maps to visualize and localize fatigue cracks under periodic excitation. The following aspects synthesize the cross-experimental findings, emphasizing the role of directional amplitude maps, stabilization, frequency selection, thresholding procedures, and validation across measurement setups. Together, these elements suggest the robustness and indicate the practical applicability of the proposed crack-assessment pipeline within NDT or SHM systems.

Directional versus magnitude amplitude maps. Amplitude maps constructed from directional OF components oriented approximately perpendicular to the crack provided clearer visualization of the discontinuity than those based on overall OF magnitude. This effect was observed consistently in the beam and plate-like structure experiments, where directional maps emphasized the motion associated with the crack breathing. These findings indicate that directional components are preferable when available. Nevertheless, magnitude-based maps remain a useful tool for preliminary assessment and can, for example, support identification of the likely crack orientation to guide subsequent directional processing.

Motion magnification. Phase-based motion magnification was evaluated as a potential enhancement to amplitude maps. It proved beneficial for magnitude-based maps under low excitation amplitudes, where small displacements were otherwise difficult to observe, and magnification improved the visibility of crack-related motion. For directional amplitude maps, however, magnification did not enhance visualization and in some cases introduced artifacts, especially at higher excitation levels. The analysis therefore supports the use of motion magnification as an optional aid for magnitude maps in low-amplitude scenarios, but not as a default stage in the directional processing pipeline.

Stabilization and reference selection. Stabilization was essential for suppressing global motion and rendering one side of the specimen quasi-static. In this work, stabilization was implemented in a single-point configuration, which compensates for translation but does not address more complex motions such as rotation. While sufficient for the examined datasets, such limitations may become relevant in more complex scenarios and are therefore noted as an area for future work. The placement of the reference point influenced the resulting maps, as the chosen anchor determines which regions appear static and which display relative motion. In the plate-like structure experiment, stable and repeatable results were obtained when the reference point was placed on a rigid part of the specimen. Template size also affected performance: very small templates were sensitive to noise and texture scarcity, whereas once the template was sufficiently large, the stabilization became robust and did not alter the outcome of crack-tip localization.

Frequency-domain design. The methodology consistently relied on isolating a narrow frequency neighborhood around the known excitation frequency. This approach allowed the analysis to focus on the periodic nature of crack breathing. The beam experiment demonstrated that the amplitude at the fundamental peak grew with excitation amplitude, enhancing crack visibility. A sufficiently long temporal window ensured reliable peak extraction, while the frequency neighborhood ϵ_f needed to be narrow enough to exclude nearby components but tolerant to small deviations in excitation frequency. These choices were shown to generalize across datasets without the need for case-specific tuning.

Thresholding and reference segment length. Automatic crack-tip localization relied on applying a threshold t_c to score profiles derived from amplitude maps. Experiments demonstrated that a global threshold could be used consistently across growth stages, with only minor differences in most cases compared to stage-wise adjustments. This simplifies practical application by reducing the need for repeated tuning. The length of the reference segment used to compute the baseline score was also important: segments that were too short led to inadequately adjusted thresholds, while segments that were sufficiently long produced stable and interpretable results. At the same time, the segment should not extend too far toward the crack, as this may distort the baseline. The plate experiment confirmed that keeping the reference segment adequately long, but still clearly separated from the crack region, represents a reliable approach.

Method parameter sensitivity. It should be emphasized that many aspects of the pipeline, including motion-estimation settings and parameters of subsequent processing stages, could likely be improved by careful fine-tuning of the method's configurable parameters. Such optimization was not the primary aim of this work, as the proposed crack assessment methodology is intended to remain general and not dependent on precise configuration of the method's parameters. The experimental results demonstrated that consistent performance could be achieved with various non-optimized settings, which is advantageous for potential deployment. Further exploration of method parameter optimization is therefore identified as a direction for future work.

Error assessment. Quantitative evaluation against manually annotated crack tip positions showed that the procedure was capable of reliably localizing crack tips across cameras and growth stages. Given that the resolution of manual measurement was one millimeter, most localization errors were below or close to this value. The method maintained consistent accuracy across different measurement setups and parameter configurations, indicating that it is sufficiently robust for practical application. The remaining errors in crack tip localization are likely attributable to imperfections in OF estimation and the single-point stabilization approach. It should also be noted that manual annotations may not be perfectly aligned with the actual crack tip due to human error.

Hardware generalization and deployability. The methodology generalized well across different acquisition setups, including high-speed cameras and consumer-grade smartphone recordings. Smartphone footage was more affected by noise and by handheld acquisition without a tripod, yet the overall score profiles and crack-tip localization results remained consistent with those obtained from high-speed cameras. This confirms that the approach does not require specialized equipment and can be deployed with accessible hardware, broadening its applicability.

Applicability and integration. The results demonstrate that the proposed method performs effectively even under non-optimal acquisition conditions. In the plate-like structure experiment, two different camera positions and image resolutions were tested, with the cameras oriented roughly perpendicular to the specimen surface; exact alignment was not required for the method to operate effectively. Despite these variations, the method remained effective and successfully localized the crack tip. Similarly, in the smartphone recordings, the camera optical axis was not perfectly perpendicular to the crack, yet the method still yielded reliable results. These findings indicate that strict alignment of the camera is not required, which supports the method's applicability in practical monitoring scenarios, positioning it as a complementary tool within NDT or SHM, particularly for scenarios where periodic excitation can be introduced or is naturally present. Its strengths lie in providing interpretable visualizations of crack breathing together with automatic crack tip localization, thereby supporting maintenance decisions through reliable tracking of crack growth over time.

9.3 Future work

Future work should focus on extending the applicability of the proposed pipeline, reducing operator dependence, and strengthening validation and uncertainty quantification. The directions below follow from the scope boundaries and limitations observed in Chapters 5–8 and the synthesis in Sections 9.1–9.2. While many stages could likely benefit from fine-tuning, the methodology is intended to remain general and not dependent on precise parameter configuration; the goal is therefore to establish robust defaults and straightforward procedures rather than case-specific tuning.

Stabilization and reference selection refinements. The present implementation relies on single-point stabilization, which compensates for translation but does not address more complex global motions. Extending stabilization to multi-point variants should be investigated for scenarios where the specimen exhibits more complex movement. A possible solution could be incorporating a homography transformation, which has been used, for example, for CV-based deflection measurement [115, 116]. In addition, employing a subpixel-accurate stabilization approach could further improve performance. In parallel, automatic identification of a suitable reference point on a rigid, undamaged region together with automatic selection of template size based on local texture would reduce operator input and improve repeatability across recordings.

Robustness across acquisition setups and specimens. Experiments indicate that the method generalizes from high-speed cameras to smartphone recordings and can tolerate non-ideal conditions such as handheld acquisition or imperfect alignment of the camera optical axis. Further research is needed to explore the limitations of the current approach when applied to more complex geometries, such as curved or irregular surfaces, as well as to different materials.

Automation and implementation. Future work should focus on reducing dependence on manual parameter choices and improving computational efficiency. This includes automatic tracking of the excitation frequency and adaptive selection of key parameters, such as threshold and reference-segment length, to ensure stable performance under varying conditions. Reducing reliance on predefined orientation is also important; automated selection of directional components, for example by inferring crack orientation from preliminary magnitude maps, would remove manual steps. Regarding implementation, optimization of the processing pipeline, including downsampling strategies and more efficient motion estimation algorithms, is needed to make the method suitable for near real-time applications. In parallel, end-to-end automation covering the definition of mROI and sROI should be developed so that the methodology can operate with minimal user input. These developments are essential for integration of the pipeline into NDT or SHM, where automated crack tip localization can directly inform maintenance decisions.

Taken together, these directions target greater robustness, reduced operator dependence, and practical deployability, while at the same time broadening the range of scenarios and configurations in which the methodology can be applied.

Chapter 10

Concluding remarks

The aim of this dissertation was to develop and validate a marker-free, computer-vision-based methodology for automatic fatigue-crack assessment in metallic structures under periodic excitation, with the intention of integration into NDT or SHM practice. The work addressed this aim by designing and validating a complete pipeline that transforms raw video recordings into interpretable amplitude maps, followed by automated crack-tip localization.

The proposed processing pipeline enables fatigue-crack analysis based on video recordings by combining image processing techniques with frequency-domain analysis. Amplitude maps are constructed from dense OF time series, providing a clear visual representation of breathing crack motion and delineating regions on either side of the crack. To enable automated crack tip localization, a scoring procedure is introduced to evaluate the difference in average amplitude values between selected sROIs on both sides of the crack. This enables assessment of crack presence and progression. In addition, the LOI metric was formulated as an indicator of motion-estimation reliability, supporting masking and selection of suitable ROIs for analysis.

Validation across different experiments demonstrated consistent performance of the proposed method. In the cantilever beam experiment, the excitation amplitude was found to affect the score magnitude and the overall appearance of the maps. The plate-like structure experiment with a propagating crack further confirmed that the procedure operated consistently in varied scenarios, with the scoring reliably indicating the crack tip and enabling automatic tracking of crack growth. The procedure proved resilient to surface imperfections, measurement noise, and recording conditions, indicating potential for generalization.

On the basis of the presented pipeline and experimental findings, the main contributions of this dissertation can be summarized as follows.

Main contributions.

- Formulation of amplitude maps from dense OF time series for visualization of crack-breathing motion.
- Development of an automated crack-tip localization procedure based on amplitude maps.

- Formulation of the LOI metric as an interpretable indicator of motion estimation reliability, enabling masking and ROI selection.
- Assessment of the impact of motion magnification on amplitude maps and crack visualization.
- Case study of amplitude maps and crack-tip localization on an aircraft-grade plate-like structure using consumer-grade smartphone recordings.

List of publications related to the dissertation.

- Adam Machynia, Ziemowit Dworakowski, Kajetan Dziedziech, Michał Dziendzikowski, Krzysztof Holak, **Vision-based automatic fatigue crack visualization and assessment in metallic structures**, *Measurement*, Volume 257, Part A, 2026,
- Jakub Spytek, Adam Machynia, Kajetan Dziedziech, Ziemowit Dworakowski, and Krzysztof Holak. **Novelty detection approach for the monitoring of structural vibrations using vision-based mean frequency maps.** *Mechanical Systems and Signal Processing*, 185, 2023.

The author would like to acknowledge the financial support from the Polish National Center for Research and Development within the scope of the projects LIDER/26/0103/L-9/17/NCBR/2018 and LIDER13/0132/2022.

Bibliography

[1] Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, Moncef Gabbouj, and Daniel J. Inman. A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mechanical Systems and Signal Processing*, 147:107077, 2021.

[2] Walter Schütz. A history of fatigue. *Engineering Fracture Mechanics*, 54(2):263–300, 1996.

[3] René Alderliesten. *Introduction to Aerospace Structures and Materials*. TU Delft OPEN Publishing, 2018.

[4] Russell Wanhill, Simon Barter, and Loris Molent. *Fatigue Crack Growth Failure and Lifing Analyses for Metallic Aircraft Structures and Components*. SpringerBriefs in Applied Sciences and Technology. Springer Dordrecht, Dordrecht, 1 edition, 2019.

[5] Christoph Kralovec and Martin Schagerl. Review of structural health monitoring methods regarding a multi-sensor approach for damage assessment of metal and composite structures. *Sensors*, 20(3):1–25, 2020.

[6] Patricia A. Howell. Nondestructive Evaluation (NDE) Methods and Capabilities Handbook. Technical report, NASA Langley Research Center, 2020.

[7] Fredrik Bjørheim, Sudath C. Siriwardane, and Dimitrios Pavlou. A review of fatigue damage detection and measurement techniques. *International Journal of Fatigue*, 154:106556, 2022.

[8] C. R. Farrar, N. Dervilis, and K. Worden. The Past, Present and Future of Structural Health Monitoring: An Overview of Three Ages. *Strain*, 61(1), 2025.

[9] Charles R. Farrar and Keith Worden. An Introduction to Structural Health Monitoring. In Arnaud Deraemaeker and Keith Worden, editors, *New Trends in Vibration Based Structural Health Monitoring*, pages 1–17. Springer Vienna, Vienna, 2010.

[10] Charles R. Farrar and Keith Worden. Introduction. In *Structural Health Monitoring: A Machine Learning Perspective*, pages 1–16. Wiley, 2012.

[11] P. Gardner, R. Fuentes, N. Dervilis, C. Mineo, S.G. Pierce, E.J. Cross, and K. Worden. Machine learning at the interface of structural health monitoring and non-destructive evaluation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2182):20190581, 2020.

[12] Mohad Tanveer, Muhammad Umar Elahi, Jaehyun Jung, Muhammad Muzammil Azad, Salman Khalid, and Heung Soo Kim. Recent Advancements in Guided Ultrasonic Waves for Structural Health Monitoring of Composite Structures. *Applied Sciences*, 14(23):11091, 2024.

[13] Mohammad Hany Yassin, Mohamad Hussein Farhat, Reza Soleimanpour, and Michel Nahas. Fiber Bragg grating (FBG)-based sensors: a review of technology and recent applications in structural health monitoring (SHM) of civil engineering structures. *Discover Civil Engineering*, 1(1):151, 2024.

[14] Noor Ghadarah and David Ayre. A Review on Acoustic Emission Testing for Structural Health Monitoring of Polymer-Based Composites. *Sensors*, 23(15):6945, 2023.

[15] Dongming Feng and Maria Q Feng. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection – A review. *Engineering Structures*, 156:105–117, 2018.

[16] Claudia Ferraris, Gianluca Amprimo, and Giuseppe Pettiti. Computer Vision and Image Processing in Structural Health Monitoring: Overview of Recent Applications. *Signals*, 4(3):539–574, 2023.

[17] Yizhou Zhuang, Weimin Chen, Tao Jin, Bin Chen, He Zhang, and Wen Zhang. A Review of Computer Vision-Based Structural Deformation Monitoring in Field Environments. *Sensors*, 22(10):3789, 2022.

[18] John Mark Go Payawal and Dong-Keon Kim. Image-Based Structural Health Monitoring: A Systematic Review. *Applied Sciences*, 13(2):968, 2023.

[19] Chuan-Zhi Dong and F. Necati Catbas. A review of computer vision–based structural health monitoring at local and global levels. *Structural Health Monitoring*, 20(2):692–743, 2021.

[20] Chunbao Xiong, Sida Lian, and Wen Chen. Detection and Location of Steel Structure Trestle Surface Cracks Based on Consumer-grade Camera System. *KSCE Journal of Civil Engineering*, 27(3):1150–1165, 2023.

[21] Qinghua Han, Xuan Liu, and Jie Xu. Detection and Location of Steel Structure Surface Cracks Based on Unmanned Aerial Vehicle Images. *Journal of Building Engineering*, 50:104098, 2022.

[22] Cao Vu Dung, Hidehiko Sekiya, Suichi Hirano, Takayuki Okatani, and Chitoshi Miki. A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Automation in Construction*, 102:217–229, 2019.

[23] Dalei Wang, Yiqing Dong, Yue Pan, and Rujin Ma. Machine vision-based monitoring methodology for the fatigue cracks in U-Rib-to-deck weld seams. *IEEE Access*, 8:94204–94219, 2020.

[24] Lexin Zhang, Zhiyu Jie, Zhong Xian Li, Wei Lu, Hao Zheng, and Wanzhen Wang. Intelligent recognition and measurement for fatigue cracks in orthotropic steel decks: A comparative study of algorithms. *Measurement*, 247:116867, 2025.

[25] Chun Zhang, Le Wan, Ruo-Qing Wan, Jian Yu, and Rui Li. Automated fatigue crack detection in steel box girder of bridges based on ensemble deep neural network. *Measurement*, 202:111805, 2022.

[26] Mahtab Mohtasham Khani, Sahand Vahidnia, Leila Ghasemzadeh, Y. Eren Ozturk, Mustafa Yuvalaklioglu, Selim Akin, and Nazim Kemal Ure. Deep-learning-based crack detection with applications for the structural health monitoring of gas turbines. *Structural Health Monitoring*, 19(5):1440–1452, 2020.

[27] Abhishek Reddy, V. Indragandhi, Logesh Ravi, and V. Subramaniyaswamy. Detection of Cracks and damage in wind turbine blades using artificial intelligence-based image analytics. *Measurement*, 147:106823, 2019.

[28] Sen Wang, Chang Liu, and Yinhui Zhang. Fully convolution network architecture for steel-beam crack detection in fast-stitching images. *Mechanical Systems and Signal Processing*, 165:108377, 2022.

[29] Shanshan Yu, Jian Zhang, Chengpeng Zhu, Zeyang Sun, and Shuai Dong. Full-field deformation measurement and cracks detection in speckle scene using the deep learning-aided digital image correlation method. *Mechanical Systems and Signal Processing*, 209:111131, 2024.

[30] Jiangpeng Shu, Congguang Zhang, Xiyuan Chen, and Yanbo Niu. Model-informed deep learning strategy with vision measurement for damage identification of truss structures. *Mechanical Systems and Signal Processing*, 196:110327, 2023.

[31] Fu-Chen Chen and Mohammad R. Jahanshahi. NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion. *IEEE Transactions on Industrial Electronics*, 65(5):4392–4400, 2018.

[32] Fu-Chen Chen, Mohammad R. Jahanshahi, Rih-Teng Wu, and Chris Joffe. A texture-Based Video Processing Methodology Using Bayesian Data Fusion for Autonomous Crack

Detection on Metallic Surfaces. *Computer-Aided Civil and Infrastructure Engineering*, 32(4):271–287, 2017.

[33] Haiyan Zhuang, Yikai Cheng, Man Zhou, and Zhenjun Yang. Deep learning for surface crack detection in civil engineering: A comprehensive review. *Measurement*, 248:116908, 2025.

[34] Mian Jiang, Wenan Zhang, and Qinghua Lu. A nonlinearity measure-based damage location method for beam-like structures. *Measurement*, 146:571–581, 2019.

[35] Yeping Peng, Weijiang Wang, Zhen Tang, Guangzhong Cao, and Shengxi Zhou. Non-uniform illumination image enhancement for surface damage detection of wind turbine blades. *Mechanical Systems and Signal Processing*, 170:108797, 2022.

[36] Hernán Garrido, Ramón Codina, Fernanda de Borbón, and Daniel Ambrosini. Damage identification in beams using burst video-records during free-vibration. *Mechanical Systems and Signal Processing*, 200:110539, 2023.

[37] Xiangxiong Kong and Jian Li. Non-contact fatigue crack detection in civil infrastructure through image overlapping and crack breathing sensing. *Automation in Construction*, 99:125–139, 2019.

[38] Rushil Mojidra, Jian Li, Ali Mohammadkhorasani, Fernando Moreu, Caroline Bennett, and William Collins. Vision-based fatigue crack detection using global motion compensation and video feature tracking. *Earthquake Engineering and Engineering Vibration*, 22:19–39, 2023.

[39] Vipin Chandra and Pritam Chakraborty. Automated crack extension measurement method for fracture and fatigue analysis using digital image correlation. *Engineering Fracture Mechanics*, 305:110182, 2024.

[40] Yan Zhao, Dianyin Hu, Qicheng Liu, Rongqiao Wang, and Jianguang Bao. High resolution and real-time measurement of 2D fatigue crack propagation using an advanced digital image correlation. *Engineering Fracture Mechanics*, 268:108457, 2022.

[41] Derui Li, Bin Cheng, Linze Shi, Enqi Zhang, and Qibin Zhao. Propagation measurement for visually thin fatigue crack using homography mapping error and digital image correlation. *Mechanical Systems and Signal Processing*, 220:111628, 2024.

[42] Jonas Wulff and Michael J. Black. Efficient sparse-to-dense optical flow estimation using a learned basis and layers. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 120–130. IEEE, 2015.

[43] Amar Mitiche and J. K. Aggarwal. *Computer Vision Analysis of Image Motion by Variational Methods*. Springer Cham, 1 edition, 2014.

[44] Berthold K. Horn and Brian G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[45] Bruce D. Lucas and Takeo Kanade. Iterative Image Registration Technique With an Application To Stereo Vision. In *Proceedings of Imaging Understanding Workshop*, pages 121–130, 1981.

[46] Bruce D. Lucas. *Generalized Image Matching by the Method of Differences*. PhD thesis, Carnegie-Mellon University, Pittsburgh, PA, 1984.

[47] Andreas Wedel and Daniel Cremers. *Stereo Scene Flow for 3D Motion Analysis*. Springer London, 1 edition, 2011.

[48] Joachim Weickert, Andrés Bruhn, and Christoph Schnörr. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods. *International Journal of Computer Vision*, 61:211–231, 2005.

[49] Gunnar Farnebäck. Two-Frame Motion Estimation Based on Polynomial Expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis*, pages 363–370, Halmstad, Sweden, 2003.

[50] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 12:43–77, 1994.

[51] Hans-Hellmut Nagel. Constraints for the estimation of displacement vector fields from image sequences. In *IJCAI'83: Proceedings of the Eighth international joint conference on Artificial intelligence*, pages 945–951, Karlsruhe, F.R.G., 1983.

[52] Hans-Hellmut Nagel and Wilfried Enkelmann. An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(5):565–593, 1986.

[53] Hans-Hellmut Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33(3):299–324, 1987.

[54] Hans-Hellmut Nagel. Optical flow estimation and the interaction between measurement errors at adjacent pixel positions. *International Journal of Computer Vision*, 15:271–288, 1995.

[55] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In Tomás Pajdla and Jiří Matas, editors, *Computer Vision - ECCV 2004 Lecture Notes in Computer Science*, volume 3024, pages 25–36. Springer Berlin Heidelberg, 2004.

[56] Thomas Brox and Jitendra Malik. Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513, 2011.

[57] Simon Baker and Iain Matthews. Lucas-Kanade 20 Years On: An Unifying Framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[58] Jean-Yves Bouguet. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm. Technical report, 2001.

[59] Andrés Bruhn, Joachim Weickert, and Christoph Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):1–21, 2005.

[60] David J. Fleet and Allan D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, 1990.

[61] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. Phase-based video motion processing. *ACM Transactions on Graphics*, 32(4):1–10, 2013.

[62] P Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989.

[63] Ajit Singh. An estimation-theoretic framework for image-flow computation. In *[1990] Proceedings Third International Conference on Computer Vision*, pages 168–177, Osaka, Japan, 1990.

[64] A M R Sousa, José Xavier, Mário Vaz, J J L Morais, and Vítor Filipe. Measurement of displacement fields with sub-pixel accuracy by combining cross-correlation and optical flow. *Mecânica Experimental*, 21:101–109, 2012.

[65] Min Liu and Tobi Delbruck. Block-matching optical flow for dynamic vision sensors: Algorithm and FPGA implementation. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4. IEEE, 2017.

[66] Ramin Zabih and John Woodfill. Non-parametric Local Transforms for Computing Visual Correspondence. In Jan-Olof Eklundh, editor, *Computer Vision — ECCV ’94. ECCV 1994. Lecture Notes in Computer Science*, pages 151–158, Berlin, Heidelberg, 1994. Springer, Berlin, Heidelberg.

[67] Fridtjof Stein. Efficient Computation of Optical Flow Using the Census Transform. In Carl Edward Rasmussen, Heinrich H Bülthoff, Bernhard Schölkopf, and Martin A Giese, editors, *Pattern Recognition. DAGM 2004. Lecture Notes in Computer Science*, pages 79–86, Berlin, Heidelberg, 2004. Springer, Berlin, Heidelberg.

[68] David Hafner, Oliver Demetz, and Joachim Weickert. Why is the Census Transform Good for Robust Optic Flow Computation? In Arjan Kuijper, Kristian Bredies, Thomas Pock, and Horst Bischof, editors, *Scale Space and Variational Methods in Computer Vision. Lecture Notes in Computer Science*, volume 7893, pages 210–221, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[69] David Benjamin De Jong, Federico Paredes-Valles, and Guido Cornelis Henricus Eugene C.H.E. De Croon. How Do Neural Networks Estimate Optical Flow? A Neuropsychology-Inspired Study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8290–8305, 2022.

[70] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid. DeepFlow: Large Displacement Optical Flow with Deep Matching. In *2013 IEEE International Conference on Computer Vision*, pages 1385–1392. IEEE, 2013.

[71] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning Optical Flow with Convolutional Networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, volume 2015 Inter, pages 2758–2766. IEEE, 2015.

[72] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1647–1655. IEEE, 2017.

[73] Anurag Ranjan and Michael J Black. Optical Flow Estimation Using a Spatial Pyramid Network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2720–2729. IEEE, 2017.

[74] Deqing Sun, Xiaodong Yang, Ming-yu Liu, and Jan Kautz. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8934–8943. IEEE, 2018.

[75] Zachary Teed and Jia Deng. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science*, volume 12347, pages 402–419. Springer International Publishing, 2020.

[76] Anurag Ranjan, Joel Janai, Andreas Geiger, and Michael Black. Attacking Optical Flow. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2404–2413. IEEE, 2019.

[77] Xiang Gao, Xiaodong Ji, Yi Zhang, Yuncheng Zhuang, and Enjian Cai. Structural displacement estimation by a hybrid computer vision approach. *Mechanical Systems and Signal Processing*, 204:110754, 2023.

[78] Lele Luan, Yang Liu, and Hao Sun. Extracting high-precision full-field displacement from videos via pixel matching and optical flow. *Journal of Sound and Vibration*, 565:117904, 2023.

[79] Jinzhi Wu, Zenan Ma, Yu Xue, Jie Qin, Deqing You, and Guojun Sun. Displacement monitoring and modal parameter identification of cable net structure based on feature optical flow and binocular stereo vision. *Structures*, 76:108914, 2025.

[80] Kemal Hacıefendioğlu, Volkan Kahya, Maria Giuseppina Limongelli, Fatih Yesevi Okur, Ahmet Can Altunışık, Tunahan Aslan, Selenay Pembeoğlu, Cemile Duman, Ayşecan Bostan, and Houssein Aleit. Applications of optical flow methods and computer vision in structural health monitoring for enhanced modal identification. *Structures*, 69:107414, 2024.

[81] Xin Bai, Rongliang Xie, Ning Liu, and Zi Zhang. Structural Vibration Detection Using the Optimized Optical Flow Technique and UAV After Removing UAV's Motions. *Applied Sciences*, 15(11):5821, 2025.

[82] Wei-Han Cheng, Cheng-En Tsai, and Hsin-Haou Huang. Vision-based algorithm for structural response measurement using movable camera and damage localization. *Measurement*, 232:114678, 2024.

[83] Sifan Wang and Mayuko Nishio. Anomaly detection in structural dynamic systems via nonlinearity occurrence analysis using video data. *Mechanical Systems and Signal Processing*, 216:111506, 2024.

[84] Ce Liu, Antonio Torralba, William T Freeman, Frédo Durand, and Edward H Adelson. Motion magnification. In *ACM SIGGRAPH 2005 Papers*, volume 24, pages 519–526, New York, NY, USA, 2005. ACM.

[85] C Harris and M Stephens. A Combined Corner and Edge Detector. In C. J. Taylor, editor, *Procedings of the Alvey Vision Conference 1988*, pages 23.1–23.6. Alvey Vision Club, 1988.

[86] Jianbo Shi and Tomasi. Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, volume 169, pages 593–600. IEEE Comput. Soc. Press, 1994.

[87] Y Boykov, O Veksler, and R Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[88] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, 31(4):1–8, 2012.

[89] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D.J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, 1992.

[90] E.P. Simoncelli and W.T. Freeman. The steerable pyramid: a flexible architecture for multiscale derivative computation. In *Proceedings., International Conference on Image Processing*, volume 3, pages 444–447. IEEE Comput. Soc. Press, 1995.

[91] Neal Wadhwa, Michael Rubinstein, Fredo Durand, and William T Freeman. Riesz pyramids for fast phase-based video magnification. In *2014 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2014.

[92] Michael Felsberg and Gerald Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, 2001.

[93] Mohamed A Elgharib, Mohamed Hefeeda, Frédo Durand, and William T Freeman. Video Magnification in Presence of Large Motions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4119–4127, Boston, MA, USA, 2015. IEEE Computer Society.

[94] Tae-Hyun Oh, Ronnachai Jaroensri, Changil Kim, Mohamed Elgharib, Frédo Durand, William T. Freeman, and Wojciech Matusik. Learning-Based Video Motion Magnification. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018 Lecture Notes in Computer Science()*, volume 11208, pages 663–679. Springer International Publishing, 2018.

[95] Ricard Lado-Roigé and Marco A. Pérez. STB-VMM: Swin Transformer based Video Motion Magnification. *Knowledge-Based Systems*, page 110493.

[96] Jasdeep Singh, Subrahmanyam Murala, and G. Sankara Raju Kosuru. Multi Domain Learning for Motion Magnification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13914–13923. IEEE, 2023.

[97] Jasdeep Singh, Santosh Kumar Vipparthi, Subrahmanyam Murala, G. Sankara Raju Kosuru, and Hasan Al-Marzouqi. Hierarchical motion magnification. *Neurocomputing*, 650:130869, 2025.

[98] Kwon Byung-Ki, Oh Hyun-Bin, Kim Jun-Seong, Hyunwoo Ha, and Tae-Hyun Oh. Learning-based Axial Video Motion Magnification. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül" Varol, editors, *Computer Vision – ECCV 2024. Lecture Notes in Computer Science*, volume 15112, pages 179–195. Springer Nature Switzerland, 2025.

[99] Justin G Chen, Neal Wadhwa, Young Jin Cha, Frédo Durand, William T Freeman, and Oral Buyukozturk. Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration*, 345:58–71, 2015.

[100] Yunus Emre Harmanci, Utku Gülan, Markus Holzner, and Eleni Chatzi. A novel approach for 3D-structural identification through video recording: Magnified tracking. *Sensors*, 19(5):1–16, 2019.

[101] Qiankun Zhu, Depeng Cui, Qiong Zhang, and Yongfeng Du. A robust structural vibration recognition system based on computer vision. *Journal of Sound and Vibration*, 541:117321, 2022.

[102] Aral Sarrafi, Zhu Mao, Christopher Niezrecki, and Peyman Poozesh. Vibration-based damage detection in wind turbine blades using Phase-based Motion Estimation and motion magnification. *Journal of Sound and Vibration*, 421:300–318, 2018.

[103] Andrew Jaeyong Choi and Jae-Hung Han. Frequency-based damage detection in cantilever beam using vision-based monitoring system with motion magnification technique. *Journal of Intelligent Material Systems and Structures*, 29(20):3923–3936, 2018.

[104] M. Civera, L. Zanotti Fragonara, P. Antonaci, G. Anglani, and C. Surace. An Experimental Validation of Phase-Based Motion Magnification for Structures with Developing Cracks and Time-Varying Configurations. *Shock and Vibration*, 2021, 2021.

[105] Abe Davis, Katherine L Bouman, Justin G Chen, Michael Rubinstein, Oral Büyüköztürk, Frédo Durand, and William T Freeman. Visual Vibrometry: Estimating Material Properties from Small Motions in Video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):732–745, 2017.

[106] Peipei Liu, Zhanxiong Ma, Jinho Jang, and Hoon Sohn. Motion magnification-based nonlinear ultrasonic signal enhancement and its application to remaining fatigue life estimation of a steel padeye. *Mechanical Systems and Signal Processing*, 200:110525, 2023.

[107] Billie F. Spencer, Sung-Han Sim, Robin E. Kim, and Hyungchul Yoon. Advances in artificial intelligence for structural health monitoring: A comprehensive review. *KSCE Journal of Civil Engineering*, 29(3):100203, 2025.

[108] Ronghua Fu, Zhimeng Huang, Maosen Cao, Drahomír Novák, Chunhui Xie, and Jinwen Huang. Optimizing deep learning-driven computer vision for civil infrastructure defect Identification: Challenges and strategies. *Engineering Applications of Artificial Intelligence*, 158:111521, 2025.

[109] The MathWorks Inc. Matlab version: 24.2 (r2024b), 2024.

[110] The MathWorks Inc. Computer vision toolbox, 2024.

[111] The MathWorks Inc. Image processing toolbox, 2024.

[112] The MathWorks Inc. Robotics system toolbox, 2024.

[113] Jakub Spytek, Adam Machynia, Kajetan Dziedziech, Ziemowit Dworakowski, and Krzysztof Holak. Novelty detection approach for the monitoring of structural vibrations using vision-based mean frequency maps. *Mechanical Systems and Signal Processing*, 185:109823, 2023.

[114] Wilhelm Burger and Mark J. Burge. *Digital Image Processing: An Algorithmic Introduction*. Springer Cham, 3 edition, 2022.

[115] Ziemowit Dworakowski, Piotr Kohut, Alberto Gallina, Krzysztof Holak, and Tadeusz Uhl. Vision-based algorithms for damage detection and localization in structural health monitoring. *Structural Control and Health Monitoring*, 23(1):35–50, 2016.

[116] Piotr Kohut, Krzysztof Holak, Adam Martowicz, and Tadeusz Uhl. Experimental assessment of rectification algorithm in vision-based deflection measurement system. *Nondestructive Testing and Evaluation*, 32(2):200–226, 2017.