

**FEDERAL UNIVERSITY OF TECHNOLOGY - PARANÁ**

**THIAGO ESTRELA KALID**

**DEVELOPMENT OF VISUAL ODOMETRY METHOD FOR  
IMMERSION INSPECTIONS**

**CURITIBA**

**2025**

**THIAGO ESTRELA KALID**

**DEVELOPMENT OF VISUAL ODOMETRY METHOD FOR  
IMMERSION INSPECTIONS**

**DESENVOLVIMENTO DE MÉTODO DE ODOMETRIA VISUAL PARA  
INSPEÇÕES POR IMERSÃO**

Bachelor's Thesis submitted in fulfillment of the requirements for the degree of Bachelor in Electrical Engineering from the Electrical Engineering program at the Federal University of Technology – Paraná (UTFPR).

Mentor: Prof. Dr. Thiago Alberto Rigo  
Passarin

**CURITIBA**

**2025**



[4.0 Internacional](#)

This license allows sharing, remixing, adaptation, and creation based on the work, even for commercial purposes, as long as credit is given to the author(s). Content created by third parties, cited and referenced in this work, is not covered by the license.

**THIAGO ESTRELA KALID**

**DEVELOPMENT OF VISUAL ODOMETRY METHOD FOR  
IMMERSION INSPECTIONS**

Bachelor's Thesis submitted in fulfillment of the requirements for the degree of Bachelor in Electrical Engineering from the Electrical Engineering program at the Federal University of Technology – Paraná (UTFPR).

Approval date: 12/February/2025

---

Thiago Alberto Rigo Passarin  
Doctor of Electrical Engineering  
Federal University of Technology - Paraná

---

Daniel Rodrigues Pipa  
Doctor of Electrical Engineering  
Federal University of Technology - Paraná

---

Jean Carlos Cardozo Da Silva  
Doctor of Electrical Engineering  
Federal University of Technology - Paraná

**CURITIBA  
2025**

## AGRADECIMENTOS

Esse trabalho é o fechamento de uma jornada de alguns anos de graduação. Vim do interior da Bahia sem conhecer ninguém no Paraná, e terminei essa jornada com grande honra em ter conhecido pessoas maravilhosas! Apesar de haver apenas um autor, várias pessoas foram cruciais para que esse trabalho fosse possível.

Aos meus pais, Ricardo e Christyanne, pelo constante apoio e confiança absoluta. Ao meu pai agradeço por ter sido sempre um professor e ter me ensinado muito não só o que tange à ciência e engenharia, mas sim à vida. E à minha mãe, agradeço pelo seu amor e carinho irrestrito; tenho certeza de que você estaria orgulhosa de mim.

À minha companheira, Bianca, por sua amizade e amor. Seu apoio me deu forças quando precisei, e em seu abraço sempre achei uma casa. Obrigado por ter me ajudado a manter o equilíbrio em minha vida. Sem você, garanto que nada disso seria possível.

À minha família, cada um de vocês foi importantíssimo no apoio emocional para essa conquista. Obrigado por terem me apoiado, apesar de meus péssimos hábitos de demora em respondê-los.

Aos meus colegas do laboratório LASSIP, pelas conversas enriquecedoras e pela camaradagem. À Tatiana Prado, pela amizade e orientação; você é uma excelente profissional e um dia espero me tornar semelhante. A Lucas Ocampos, Mateus Müller, Everton Trento e Felipe Gutierrez, pela amizade que começou no laboratório e estendeu-se para a vida.

A todos os meus amigos da Bahia e do Paraná, esse texto iria ser gigantesco se fosse citar cada um de vocês. Saibam que cada um de vocês foi essencial pelo apoio nessa saga. Sou grato por todas as conversas e gargalhadas que dei ao lado de vocês, vocês tornaram toda a experiência mais leve.

Ao Dr. Thiago A. R. Passarin, por ter sido um excelente orientador durante a iniciação científica e TCC. Seus conselhos e sugestões foram primordiais para melhorar o trabalho e guiar-me para me tornar um excelente pesquisador como o senhor.

À UTFPR à Petrobras pelo apoio estrutural e financeiro, por meio do laboratório LAS-SIP vinculado ao projeto AUSPEX, e à ANP pelo financiamento, por meio do PRH 21.1. Também, à FAPESP, Brasil, Processo nº 2024/10493-9, pelo apoio financeiro.

## ABSTRACT

Non-Destructive Testing (NDT) Ultrasound (US) inspections are commonly used for structural health assessment of vital parts of the infrastructure of the oil & gas industry. Subsea pipelines are especially prone to inner and outer corrosion, as they usually operate in harsh and corrosive environments. During an inspection, moving the US transducer is common practice as it increases the probability of detecting a discontinuity or flaw. This routine is called a sweep and requires information about the performed trajectory to properly trigger the acquisition and reconstruct spatial (3D) models of the specimen. This role is performed by an odometer, a device capable of estimating displacement based on some sort of measurement. The most common type is mechanical odometers, but commercially available alternatives for underwater 2D displacements are too bulky and complex. In this work it was proposed a visual odometer—a new category of odometer for NDT that, based on image processing, estimates 2D and 3D displacements during an inspection. The proposed solution employs a monocular camera pointed toward the specimen encased in a waterproof vessel. A Raspberry Pi 4 was used for image processing and communication to the US acquisition system. Experiments were made to test the proposed solution under controlled and freehand trajectories. Through controlled well-known trajectories, the visual odometer in a standalone operation estimated 2D and 3D trajectories with cumulative error ranging from 0.05 % to 2.80 % of the traveled distance. The uncontrolled routine simulated the conditions found in a real-world pipeline inspection along with the US acquisition system. It demonstrated the system's ability to be part of a broader inspection apparatus. Overall, the experiments helped to sustain the idea that the proposed solution could be part of a reliable odometry system. If noted improvements are made, the visual odometer offers good potential for use in subsea pipeline inspections.

**Keywords:** visual odometry; displacement estimation; monocular camera; non-destructive testing; real-time.

## RESUMO

Ensaios não destrutivos (END) por ultrassom (US) são comumente empregados para monitorar as condições estruturais de partes vitais da infraestrutura da indústria de óleo e gás. Tubulações submarinas são especialmente suscetíveis à corrosão interna e externa, uma vez que estão submetidas a ambientes de difícil operação e corrosivos. Durante uma inspeção, é comum mover o transdutor para aumentar a probabilidade de detectar uma falha ou descontinuidade na tubulação. Esse tipo de rotina é chamado de varredura e requer informação do deslocamento feito para permitir a coleta correta de dados e a reconstrução de um modelo espacial do corpo de prova. Essa função geralmente é desempenhada pelo odômetro, um dispositivo capaz de estimar deslocamento baseado em algum tipo de medição. O tipo mais comum são os odômetros mecânicos, porém soluções comercialmente disponíveis para ensaios de submersão com deslocamentos 2D são muito grandes e complexas. Nesse trabalho, é proposto um odômetro visual—uma nova categoria de odômetro para END que, baseado em processamento de imagem, estima deslocamento 2D e 3D durante uma inspeção. A solução proposta faz uso de uma câmera monocular apontada para o corpo de prova sob inspeção, encapsulada em um compartimento à prova d'água. Um Raspberry Pi 4 foi usado para processar as imagens e comunicar os dados com o sistema de aquisição por US. Experimentos foram feitos para testar a solução proposta em trajetórias definidas e trajetórias arbitrárias aleatórias. Através de trajetórias controladas, o odômetro, operando de maneira independente, obteve erros acumulados entre 0,05 % e 2,80 % da distância percorrida. A trajetória arbitrária simulou condições semelhantes às encontradas durante uma inspeção de campo, onde o emprego de aquisição por US é usado em paralelo. Os experimentos mostraram que a solução proposta é capaz de fazer parte de um aparato mais complexo de inspeção. Em vias gerais, os experimentos reforçaram a ideia de que a solução proposta pode ser parte de um sistema de odometria confiável. Se as sugestões propostas forem implementadas, o sistema de odometria visual tem grande potencial de uso em inspeções de tubulações submarinas.

**Palavras-chave:** odometria visual; estimativa de deslocamento; câmera monocular; ensaios não destrutivos; tempo real.

## LIST OF FIGURES

<b>Figure 1 – Steps involved in performing an ultrasound sweep.</b>	<b>12</b>
<b>Figure 2 – Mechanical odometers commercially available working alongside an ultrasound transducer.</b>	<b>13</b>
<b>Figure 3 – Elements of an ultrasonic NDT measurement system.</b>	<b>17</b>
<b>Figure 4 – Diagram showing a quadrature odometer operation.</b>	<b>18</b>
<b>Figure 5 – Components of an optical quadrature encoder.</b>	<b>19</b>
<b>Figure 6 – The basic model of a pinhole camera.</b>	<b>19</b>
<b>Figure 7 – The same lens with different apertures.</b>	<b>20</b>
<b>Figure 8 – Illustration of the camera lens focal length.</b>	<b>20</b>
<b>Figure 9 – Classic Euler angles around X,Y, and Z axes.</b>	<b>21</b>
<b>Figure 10 – A scene represented with three different exposures. Maximum detail is achieved when the exposure is properly set. Highlighted in blue are features lost due to saturation during overexposure and in red during underexposure.</b>	<b>25</b>
<b>Figure 11 – A scene represented with two sharpness values. (a) is sharper than (b).</b>	<b>25</b>
<b>Figure 12 – Illustration of the Depth-Of-Field. Optimal focus is achieved in the focal plane, but objects within the <i>Depth-Of-Field</i> are still in focus. (a) represents a deep DOF, while (b) a shallow DOF.</b>	<b>26</b>
<b>Figure 13 – Operational difference between global and rolling shutter architecture.</b>	<b>27</b>
<b>Figure 14 – Artifacts occurred by movement of the observer.</b>	<b>27</b>
<b>Figure 15 – The same scene is represented by different numbers of pixels. Image (a) has 256 times more pixels than (b), thus capturing more details.</b>	<b>28</b>
<b>Figure 16 – Boxplots of the absolute errors for different subpixel methods.</b>	<b>30</b>
<b>Figure 17 – Phase of the cross power spectrum matrix of two shifted 2D signals. (a) is reconstructed from an SVD factorization, and (b) is reconstructed from a rank-one approximation.</b>	<b>32</b>
<b>Figure 18 – Steps of phase unwrapping algorithm by Itoh (1982). Wrapping is expected when first-order finite difference is close to <math>\pm 2\pi</math>.</b>	<b>33</b>
<b>Figure 19 – Spectral leakage caused by rectangular windowing.</b>	<b>35</b>
<b>Figure 20 – Example of spatial windowing and its result on the spectra.</b>	<b>35</b>

<b>Figure 21 – System architecture</b>	37
<b>Figure 22 – Visual odometer hardware.</b>	38
<b>Figure 23 – Ultrasound transducer working alongside the visual odometer.</b>	39
<b>Figure 24 – Steps performed to compute 2D displacement (<math>\hat{a}_n, \hat{b}_n</math>) from two consecutive frames <math>f_{n-1}</math> and <math>f_n</math>.</b>	40
<b>Figure 25 – Planar specimen underwater and the trajectories performed: closed-loop (red), single-x (blue) and single-y (green).</b>	44
<b>Figure 26 – Experimental setup and result for a planar specimen in dry environment.</b>	44
<b>Figure 27 – Experimental setup and result for a cylindrical specimen underwater.</b>	45
<b>Figure 28 – A similar frame captured in a dry environment and underwater.</b>	45
<b>Figure 29 – Steps of phase unwrapping algorithm by Itoh (1982) yielding estimation error. The algorithm was unable to unwrap data in some places (highlighted with blue circles). All points showed in 29b are supposed to be within <math>[-\pi, \pi]</math> if the operation was successful.</b>	46
<b>Figure 30 – An instant of the real-time operation. A video of the entire operation is available on YouTube.</b>	47

## LIST OF TABLES

<b>Table 1 – Specifications of the linear phased array from Imasonic (IMASONIC SAS, France).</b>	<b>39</b>
<b>Table 2 – Pixel size estimation results.</b>	<b>43</b>
<b>Table 3 – Cumulative displacements, in millimeters, measured for the 2D trajectories reconstructed from the data provided by the visual odometer.</b>	<b>44</b>

## **LIST OF ABBREVIATIONS AND ACRONYMS**

### **Acronyms**

BIPM	International Bureau of Weights and Measures
CPU	Central Processing Unit
DOF	Depth-Of-Field
GPIO	General Purpose Input/Output
IMU	Inertial Measurement Unit
LED	Light-Emitting Diode
NDT	Non-Destructive Testing
RANSAC	RANDom SAmple Consensus
ROV	Remotely Operated Vehicle
RPi	Raspberry Pi
SI	International System of Units
SLAM	Simultaneous Localization And Mapping
SVD	Singular Value Decomposition
TCP	Transmission Control Protocol
TTL	Transistor-Transistor Logic
US	Ultrasound

## CONTENTS

<b>1 INTRODUCTION . . . . .</b>	<b>12</b>
<b>1.1 Objectives . . . . .</b>	<b>13</b>
<b>1.2 General Objectives . . . . .</b>	<b>13</b>
<b>1.3 Specific Objectives . . . . .</b>	<b>13</b>
<b>1.4 Motivation . . . . .</b>	<b>14</b>
<b>1.5 Methodology . . . . .</b>	<b>14</b>
<b>1.6 Work Outline . . . . .</b>	<b>15</b>
<b>2 THEORETICAL BACKGROUND . . . . .</b>	<b>16</b>
<b>2.1 Non-Destructive Testing . . . . .</b>	<b>16</b>
2.1.1 Ultrasound Inspection . . . . .	16
<b>2.2 Odometry . . . . .</b>	<b>17</b>
2.2.1 Mechanical Odometry . . . . .	18
2.2.2 Visual Odometry . . . . .	18
2.2.2.1 Pinhole camera . . . . .	19
2.2.2.2 Monocular camera . . . . .	20
2.2.3 Inertial Measurement Unit . . . . .	20
2.2.3.1 Euler Angles . . . . .	21
2.2.3.2 Quaternions . . . . .	22
<b>2.3 Image Capturing . . . . .</b>	<b>24</b>
2.3.1 Exposure . . . . .	24
2.3.2 Sharpness . . . . .	25
2.3.2.1 Focus . . . . .	25
2.3.2.2 Motion-related Distortions . . . . .	26
2.3.3 Amount of Details . . . . .	28
2.3.3.1 Number of Pixels . . . . .	28
<b>2.4 Displacement Estimation . . . . .</b>	<b>29</b>
2.4.1 Phase correlation . . . . .	30
2.4.1.1 SVD . . . . .	31
2.4.1.2 Phase unwrapping . . . . .	32
2.4.1.3 Three-dimensional trajectory . . . . .	33

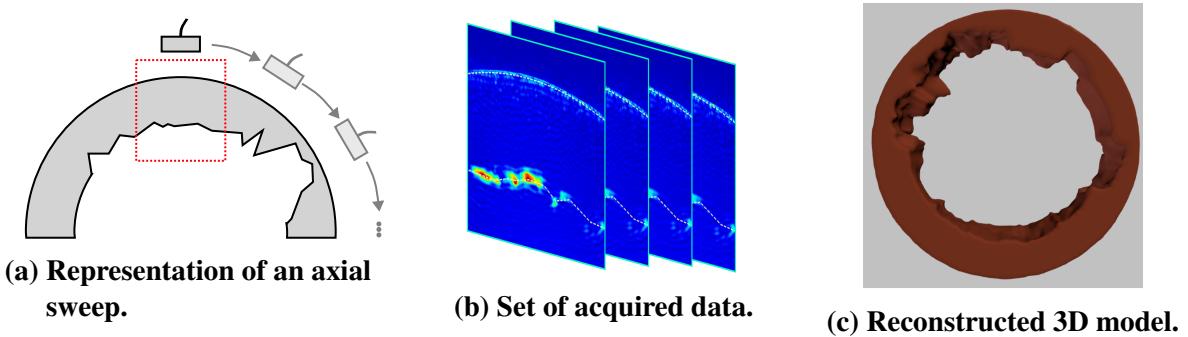
2.4.1.4	Windowing and practical effects . . . . .	34
<b>2.5</b>	<b>Commentaries . . . . .</b>	<b>35</b>
<b>3</b>	<b>MATERIALS AND METHODS . . . . .</b>	<b>37</b>
<b>3.1</b>	<b>System Architecture . . . . .</b>	<b>37</b>
<b>3.2</b>	<b>Hardware . . . . .</b>	<b>38</b>
3.2.1	Inertial Module . . . . .	38
3.2.2	Pulser Module . . . . .	38
3.2.3	Ultrasound acquisition . . . . .	39
<b>3.3</b>	<b>Software . . . . .</b>	<b>40</b>
<b>3.4</b>	<b>Commentaries . . . . .</b>	<b>41</b>
<b>4</b>	<b>RESULTS AND DISCUSSION . . . . .</b>	<b>42</b>
<b>4.1</b>	<b>Experimental Setup . . . . .</b>	<b>42</b>
<b>4.2</b>	<b>Adjustments . . . . .</b>	<b>42</b>
4.2.1	Focus . . . . .	42
4.2.2	Displacement . . . . .	43
<b>4.3</b>	<b>Experiments . . . . .</b>	<b>43</b>
4.3.1	Controlled trajectories . . . . .	43
4.3.2	Operation along US acquisition system . . . . .	46
<b>4.4</b>	<b>Commentaries . . . . .</b>	<b>47</b>
<b>5</b>	<b>CONCLUSIONS . . . . .</b>	<b>48</b>
<b>5.1</b>	<b>Future Works . . . . .</b>	<b>48</b>
	<b>REFERENCES . . . . .</b>	<b>50</b>

## 1 INTRODUCTION

The infrastructure of the oil & gas industry is composed of many critical components commonly exposed to harsh environments. For instance, submarine pipelines that transport oil & gas are prone to inner and outer corrosion. Regular inspections of structural health are crucial to avoid critical failure of the components and stoppages in production. One way of inspecting submarine pipelines is the Ultrasound (US) Non-Destructive Testing (NDT) (HO *et al.*, 2020).

During the inspection of a pipeline, the exact location of a flaw is generally unknown. Moving the transducer during the inspection is a common practice to analyze a wide area and increase the probability of detecting a flaw. In NDT, this type of routine is called a sweep (Figure 1a). Intuitively, knowing the trajectory done in a sweep is essential to not only allow the acquisition system to properly shoot the transducer at constant intervals (Figure 1b), but also to reconstruct a trustworthy model during post-processing (Figure 1c).

**Figure 1 – Steps involved in performing an ultrasound sweep.**



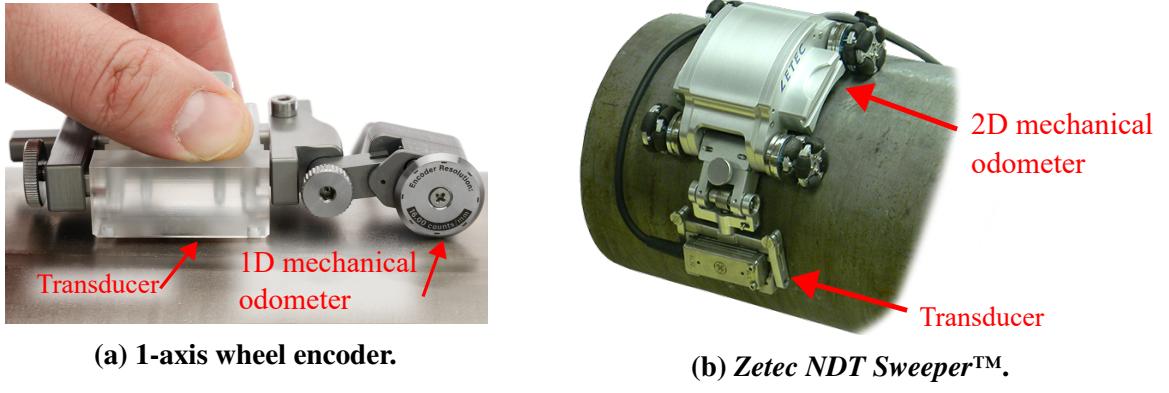
Source: Personal Collection.

One way of estimating the position of the transducer is through processing data from multiple sensors. The position can be estimated in 1, 2, or 3-dimensional space by a device called an odometer. The most common type of odometer in NDT is a 1D mechanical wheel encoder (Figure 2a), as it stands out by its resolution, reliability, and easy usability (Dezhi Zheng *et al.*, 2015). This device is composed of a wheel in contact with the surface and a stationary base mechanically coupled to the transducer. As the transducer moves, the wheel rotates, and the embedded sensors identify and convert it to the observer's (transducer) relative position.

Despite its advantages, when mechanical odometers are associated to compose 2D tracking systems, they become too bulky and complex; see, for instance, *Zetec NDT Sweeper™* (Figure 2b). In other areas outside NDT, cameras have been successfully used to solve these issues as part of a 2D odometry system (KAZIK; GOKTOGAN, 2011; NOURANI-VATANI; BORGES, 2011). Usually these solutions compares pairs of successive and continuously acquired photos to estimates the relative shift between them, and thus, infer the observer's position.

The use of cameras as part of visual odometry system is well explored in areas such as virtual reality (XIA *et al.*, 2017), robotics (MAIMONE; CHENG; MATTHIES, 2007), Simul-

**Figure 2 – Mechanical odometers commercially available working alongside an ultrasound transducer.**



Source: Adapted from (OLYMPUS, 2014) and (ZETEC, 2019), respectively.

taneous Localization And Mapping (SLAM) (GEIGER; LENZ; URTASUN, 2012). Usually, monocular cameras are employed and pointed toward the direction of the movement (the horizon) in an environment where sunlight is abundant (NISTÉR; NARODITSKY; BERGEN, 2004; HE *et al.*, 2020; NISTÉR; NARODITSKY; BERGEN, 2006). In environments where high light dispersion and low sunlight is expected (thus low exposure), such as the ocean bed (ZHOU *et al.*, 2022), the horizon-pointed approach leads to poor results even with artificial illumination. If the observer (monocular camera) moves over an object (such as an oil pipe), pointing the camera directly towards the floor during the movement is an alternative. This perspective change halves the number of degrees of freedom of the displacement during the estimation step, while making artificial scene illumination easier. These are opportunities that could be potentially exploited to compose a fast and robust visual odometry system (GILLES; IBRAHIMPASIC, 2021).

## 1.1 Objectives

## 1.2 General Objectives

Propose a 2D odometry system based on image processing applied to NDT ultrasound underwater inspections.

## 1.3 Specific Objectives

1. Research the state of the art in visual odometry;
2. Select the most suitable method for 2D visual odometry in marine NDT environment;
3. Develop an odometer prototype to implement and compare multiple algorithms asynchronously and disjoined from the US transducer;

4. Integrate IMU data to convert 2D displacement into 3D;
5. Perform experiments and data collection in controlled routines;
6. Enhance the odometer mechanical prototype and optimize the code to allow real-time 2D position estimation;
7. Analyze the results and evaluate the viability of visual odometry as a 2D real-time position estimator;

## 1.4 Motivation

In submarine NDT inspections, Remotely Operated Vehicle (ROV) are used to spare humans from unnecessary harmful conditions. This type of environment requires the employed technologies to be sealed, resistant to high pressure, compact, and reliable.

Systems based on mechanical odometry have numerous weaknesses related to the nature of their operating principle: the presence of moving parts implies greater difficulty in sealing; its size is far bigger than the transducer (Figure 2b) and errors from kinetic origin, such as slippage of the odometer wheel (REINA *et al.*, 2006), will exist and cannot be easily identified by the operator.

The proposed solution, based on visual odometry, has several advantages over mechanical odometry precisely due to the lack of moving parts in its operating principle. This way, it is possible to encapsulate the entire CPU, IMU, lights, and camera in a sealed pressure-proof vessel—as long as there is a transparent window for the camera and lighting to properly work, the system will operate without major problems. Furthermore, because displacement estimation is done via software, the system is expected to be much more flexible and easier to improve over time.

## 1.5 Methodology

The research will be done in three main phases:

- Familiarization with visual odometry methods;
- Proposition and testing of an asynchronous solution;
- Proposition and testing of a synchronous solution working alongside the ultrasound acquisition system.

The first phase will encompass specific objectives 1 to 2. The goal is to find, through the bibliography review, the most suitable position estimation method for the challenges asso-

ciated with the application. The criteria for the comparison will be computational complexity, robustness to noise, and implementation simplicity.

The second phase of the project will take place through specific objectives 3 to 4. Tests will be carried out performing controlled movement, obtaining data in the form of a set of photos acquired during the displacement and orientation data from the IMU. With prior information about the displaced path, it will be possible to evaluate the algorithm estimation error to fine-tune the implementation and physical setup.

The third phase of the project will put objectives 5 to 6 into practice. The implementation of the displacement estimation algorithm is expected to work well in asynchronous mode, so during this phase, efforts will be focused on upgrading the software to handle real-time position estimation. The ultimate goal is to handle the position information to the acquisition system so it can properly shoot the transducer during the sweep.

## 1.6 Work Outline

The rest of this work is organized into the following chapters:

- Chapter 2 presents the theoretical background for NDT, odometry, image capturing and displacement estimation algorithm;
- Chapter 3 exposes the materials and methods involved in the proposed visual odometry solution. Among those, the system architecture, mechanical parts, electronic hardware, and algorithm parametrization are exposed;
- Chapter 4 shows the results obtained by two sets of experiments; one set to evaluate accuracy, and the other to emulate real-world operation;
- Chapter 5 contains conclusions about the presented work, as well as suggestions for future works.

## 2 THEORETICAL BACKGROUND

This chapter provides the theoretical background necessary to understand visual odometry properly and its importance in an ultrasound inspection. Section 2.1 shows basic concepts regarding non-destructive testing using ultrasound. Section 2.2 presents the concept of odometry, the traditional methods used in this area, and how the method herein proposed is situated in this area. Section 2.3 exposes important concepts in photography, which will later be useful to enhance the proposed method and identify improvement gaps. Finally, in Section 2.4, a brief review of displacement methods is shown, and the proposed method is deeply explained.

### 2.1 Non-Destructive Testing

Non-destructive testing is the set of techniques used to inspect an object to evaluate its material characteristics or the presence of flaws without destroying its usability (HELLIER *et al.*, 2003). Many industrial sectors use NDT during different stages of a product's life cycle: during production to ensure standards are met (RIEDER *et al.*, 2014; EVERTON *et al.*, 2016) and after deployment to ensure safe and reliable operation (MAIERHOFER, 2003; CARVALHO *et al.*, 2008).

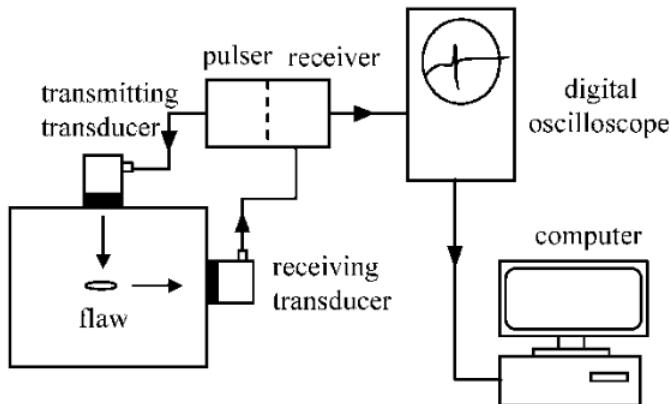
Ultrasound inspection is the most used of the numerous NDT techniques. It presents a quick, compact, safe, and cost-effective method for detecting flaws in manufactured parts. Numerous commercially available US systems can operate underwater and under high pressure, which makes them very suitable for submarine inspections (LEI *et al.*, 2009).

#### 2.1.1 Ultrasound Inspection

Figure 3 illustrates the main components of a US measurement system used on NDT. It comprises a US emitter, a US receiver, a pulser-receiver module, a digital oscilloscope, and a computer. A US acquisition follows the following steps:

1. The pulser module excites the US emitter with an electrical signal similar to a Delta of Dirac;
2. The emitter transforms the electrical signal into acoustic pressure, generating an acoustic wave;
3. The acoustic wave travels through the medium and interacts with the discontinuity (flaw), which reflects part of the energy toward the receiver;
4. The receiver transforms the perceived acoustic pressure into an electrical signal;
5. The electrical signal is converted from analog to digital format;

**Figure 3 – Elements of an ultrasonic NDT measurement system.**



**Source:** (SCHMERR, 2016)

6. Finally, the acquisition system running in the computer reads and stores the acquired digital signal.

Although the emitter and receiver are represented as two distinct single-element transducers in Figure 3, the same element can perform both roles.

During a US inspection, the technician often needs to inspect not only a point in space in a static inspection (e.g. Figure 3) but also an area where a flaw is most likely to be. In this circumstance, multiple acquisitions are made while the transducer is shifted. The rate at which the data is acquired during a sweep is determined by a regular spatial interval a priori defined (e.g., 10 mm per acquisition), which must be somehow measured during the inspection.

## 2.2 Odometry

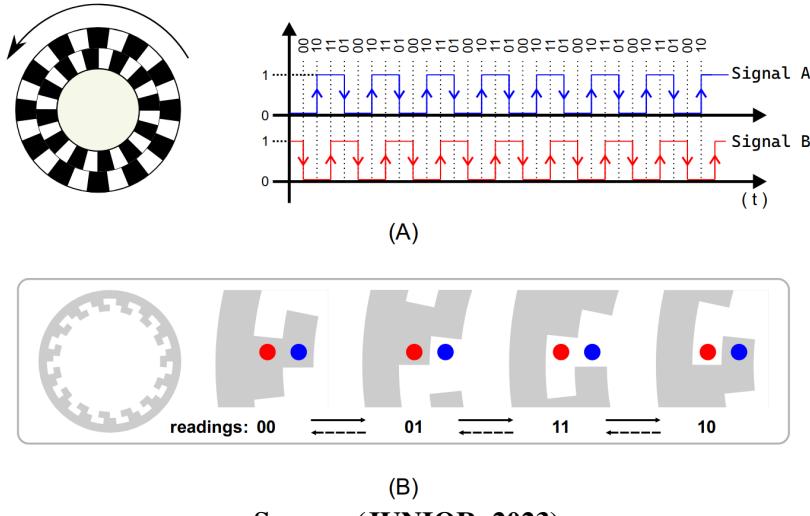
Odometry is a research area focused on estimating displacement based on data acquired by different sensors (AQEL *et al.*, 2016). It is especially useful in environments where global positioning systems (e.g., satellite localization) cannot be used, such as indoors or underwater (HE *et al.*, 2020).

An odometer is a device that measures relative displacement based on data acquired by different sensors. It will estimate displacement so the US acquisition system can fire the transducer according to the user configuration. The odometer output is an electrical signal composed of pulses. Based on the analysis of the pulsing output, instantaneous and absolute displacement can be estimated. The pulse output can be expressed in numerous ways: linear, incremental, absolute, and quadrature counting, the former being the most commonly used.

Quadrature odometers output two square-wave electrical signals (Figure 4). They have equal amplitude, variable frequency, and a 90-degree phase difference between each other. The displacement is proportional to the number of times the wave transitions from amplitude; the

more transition borders, the bigger the displacement. The direction in which the displacement has occurred is estimated based on which wave is leading or lagging.

**Figure 4 – Diagram showing a quadrature odometer operation.**



Source: (JUNIOR, 2023)

### 2.2.1 Mechanical Odometry

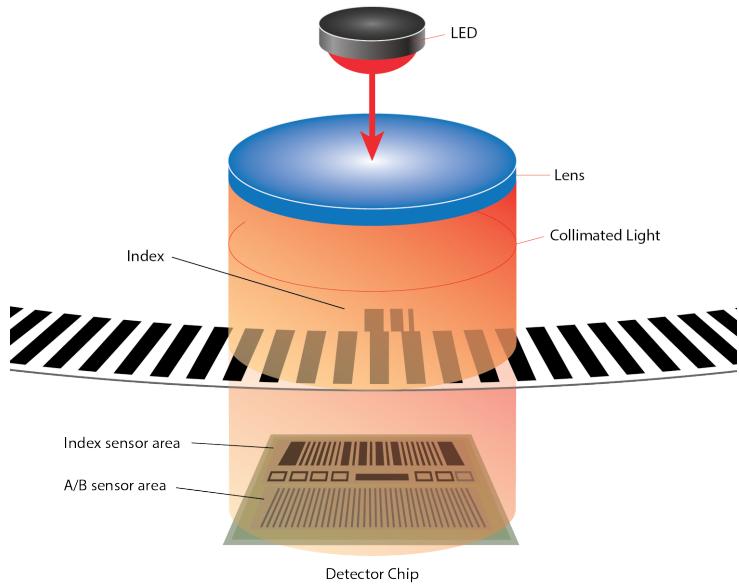
Mechanical odometers are devices that estimate displacement based on monitoring a moving mechanical part, commonly a wheel. This wheel rotates proportionally to the displacement performed. The quadrature wheel odometer has an engraved pattern, one Light-Emitting Diode (LED), and two sensors (Figure 5). The components are arranged so that the sensor's output is the two square waves, thus allowing displacement to be estimated.

### 2.2.2 Visual Odometry

A visual odometer is a system capable of estimating the displacement of an observer by comparing pairs of consecutive photos (NISTÉR; NARODITSKY; BERGEN, 2004). This solution is handy in environments where mechanical odometers are hard to use, such as submarine inspections.

To capture photos, which will later be processed to estimate displacement, a monocular camera is the most compact, cheap, and used device (HE *et al.*, 2020). During this work, only monocular cameras were used. Thus, "cameras" and "monocular cameras" are used interchangeably when the first is used alone. To understand the operating principle of monocular cameras, it is helpful to break down the most straightforward type of camera: the pinhole camera.

**Figure 5 – Components of an optical quadrature encoder.**

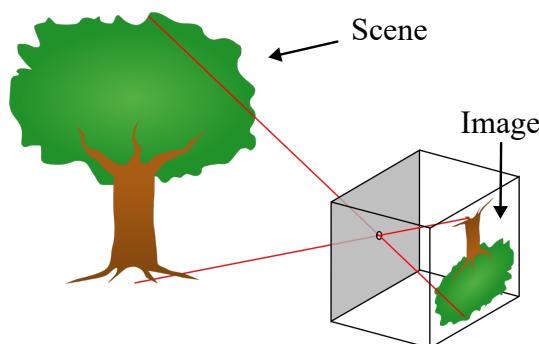


**Source: (DIGITAL, 2020)**

#### 2.2.2.1 Pinhole camera

The pinhole camera is a box with a single small hole. The small hole, a *pinhole*, allows light to enter the box and projects part of the scene into the opposing wall (Figure 6). If one desires to capture the image, a sensor should be placed in the region where the image is projected. A scene is the complex arrangement of objects and light that could be captured by a camera and converted into a 2D representation, an image (highlighted in Figure 6).

**Figure 6 – The basic model of a pinhole camera.**



**Source: Adapted from (KIRKPATRICK; FRANCIS, 2006)**

The pinhole camera is rudimentary and inflexible, making it difficult to capture good images under different circumstances. If the subject is too far away or the external light is low, nothing can be done.

**Figure 7 – The same lens with different apertures.**



Source: Adapted from (KOEPPIK, 2019).

### 2.2.2.2 Monocular camera

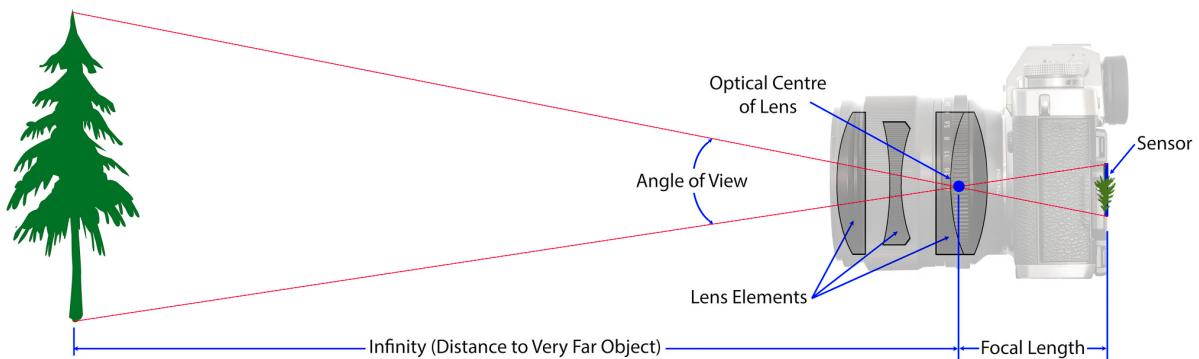
The monocular camera is an enhanced version of the pinhole camera. It comprises a photosensitive element (in our case, a digital sensor), a body, and a lens. Unlike the pinhole, the camera's mechanical apparatus allows the user to control several physical characteristics that affect the final image: aperture, exposure time and focal length.

Aperture is the size of the lens "pupil" (Figure 7). A lower aperture translates to a bigger "hole," allowing more light to reach the camera sensor.

Exposure time or speed is the amount of time the camera sensor is exposed to light during the capture of a photo. The more time the sensor is exposed to light, the brighter the image will be.

The camera lens is an apparatus composed of a series of optical lenses that allow the user to zoom into the scene. The focal length is the distance between the optical center of the lens and the camera sensor (Figure 8). Some lenses have their focal length fixed, but those that do not can be adjusted within certain limits to achieve the desired zoom and focus.

**Figure 8 – Illustration of the camera lens focal length.**



Source: (GRAY, 2023)

### 2.2.3 Inertial Measurement Unit

An Inertial Measurement Unit (IMU) is an electronic device often used to estimate the orientation of an observer. They are instrumental in odometry systems as they can leverage

displacement estimation for a more accurate value or even transform 2D displacements to 3D space. It is mainly composed by:

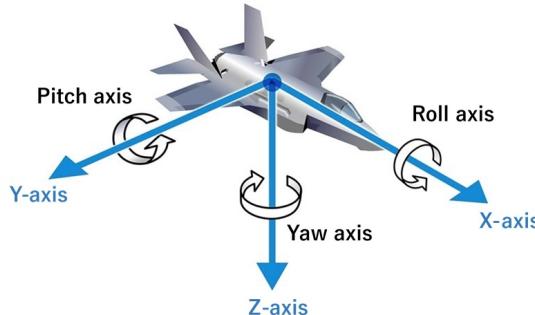
- An accelerometer which measures inertial velocity rate, usually in  $m/s^2$ ;
- A magnetometer which measures rotation or angular rate, usually in  $deg/s$ ;
- A gyroscope which measures earth's magnetic field in  $\mu T$ .

The obtained orientation or *attitude* can be expressed in two different representations: Euler Angle or Quaternion.

#### 2.2.3.1 Euler Angles

Euler angles are a representation of rigid body rotation in three-dimensional space. There are three possible rotation angles: Roll, Pitch, and Yaw, which are rotations around the X, Y, and Z axes (Figure 9).

**Figure 9 – Classic Euler angles around X,Y, and Z axes.**



**Source: (GLORIA, 2016)**

When it comes to rotations, the order in which the operations are performed matters for the final result. The most common order is yaw rotation, pitch rotation, and roll rotation (or z-y-x rotation). Suppose we have a vector  $\mathbf{v} = (v_x, v_y, v_z)^T$  which we desire to rotate to obtain  $\mathbf{r} = (r_x, r_y, r_z)^T$ . They are both separated by an roll ( $\alpha$ ), pitch ( $\beta$ ) and yaw ( $\gamma$ ) angles.

One way to achieve the roll, pitch and yaw rotation is through rotation matrices (STRANG, 2022):

$$\mathbf{R}_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix}, \quad (1)$$

$$\mathbf{R}_y(\beta) = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix}, \quad (2)$$

and

$$\mathbf{R}_z(\gamma) = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (3)$$

which can be combined into

$$\mathbf{R}(\alpha, \beta, \gamma) = \mathbf{R}_x(\alpha) \cdot \mathbf{R}_y(\beta) \cdot \mathbf{R}_z(\gamma), \quad (4)$$

ultimately reducing the entire rotation to a single inner product:

$$\mathbf{r} = R(\alpha, \beta, \gamma) \cdot \mathbf{v}. \quad (5)$$

Although the Euler notation is very intuitive, its conception has some hidden pitfalls. They are discontinuous; thus, composing successive 3D rotations is harder using Euler angles (BERNARDES; VIOLET, 2022). Additionally, they are prone to a phenomenon called *gimbal lock*, which is an inability to represent the current orientation dependent on the observer's previous orientation (MANSUR *et al.*, 2020). One solution to the previous problems is to change the orientation representation from Euler angles to Quaternions.

### 2.2.3.2 Quaternions

Quaternions are four tuple vectors capable of representing any rigid body rotation in 3D space (MANSUR *et al.*, 2020). They are composed of one real component  $q_0$  and three complex components  $q_1$ ,  $q_2$ , and  $q_3$ . A quaternion  $q$  is defined as (NIVEN, 1942):

$$\mathbf{q} = \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} \quad (6)$$

We can rotate a vector  $\mathbf{v} = [v_x, v_y, v_z]^T$  to obtain  $\mathbf{r} = [r_x, r_y, r_z]^T$  based on a orientation described by a quaternion  $\mathbf{q} = [q_1, q_2, q_3, q_4]^T$  through

$$\begin{bmatrix} 0 \\ r_x \\ r_y \\ r_z \end{bmatrix} = \mathbf{q} \otimes \begin{bmatrix} 0 \\ v_x \\ v_y \\ v_z \end{bmatrix} \otimes \mathbf{q}^* \quad (7)$$

where  $\otimes$  is the Hamilton product.

Similar to Euler notation, we can define three quaternions that describe rotations around the x,y, and z axes:

$$\mathbf{R}_x(q) = \mathbf{q} \otimes \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \otimes \mathbf{q}^* = \begin{bmatrix} 0 \\ q_0^2 + q_1^2 + q_2^2 + q_3^2 + q_4^2 \\ 2(q_1 q_2 + q_0 q_3) \\ 2(q_1 q - 3 - q_0 q_2) \end{bmatrix}, \quad (8)$$

$$\mathbf{R}_y(q) = \mathbf{q} \otimes \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \otimes \mathbf{q}^* = \begin{bmatrix} 0 \\ 2(q_1 q_2 - q_0 q_3) \\ q_0^2 - q_1^2 + q_2^2 - q_3^2 \\ 2(q_2 q_3 + q_0 q_1) \end{bmatrix} \quad (9)$$

and

$$\mathbf{R}_z(q) = \mathbf{q} \otimes \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \otimes \mathbf{q}^* = \begin{bmatrix} 0 \\ 2(q_0 q_2 + q_1 q_3) \\ 2(q_2 q_3 + q_0 q_1) \\ q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}, \quad (10)$$

which could be combined into a single rotation matrix:

$$\mathbf{R}(q) = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_0 q_2 + q_1 q_3) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \quad (11)$$

Therefore, finding the rotated vector  $\mathbf{r}$  through Eq. 7 is equivalent to

$$\mathbf{r} = \mathbf{R}(q) \cdot \mathbf{v} \quad (12)$$

or

$$\mathbf{r} = \mathbf{q} \otimes \mathbf{v} \otimes \mathbf{q}^*. \quad (13)$$

## 2.3 Image Capturing

Visual odometry relies on comparing successive images. If they are easily distinguishable, the algorithm is more likely to compute an accurate estimate. However, if the images are too dark and blurry, the algorithm will estimate poorly. Three characteristics drastically affect image quality: exposure, sharpness, and the amount of detail.

### 2.3.1 Exposure

Exposure is the amount of light per unit of area the photo-sensor captures. Due to its limitations, the sensor can only convert a limited range of the light spectrum into digital pixel brightness. The ratio between the lowest and highest pixel brightness that the camera can acquire is called the sensors *dynamic range*. An image's regions outside the dynamic range (on the lower or upper end) will be saturated. If the pixels are overall saturated (or near saturation), the image is considered to be over or underexposed (Figure 10a and 10c, respectively). It is desired to avoid at all costs this phenomenon, as the non-linearity creates unrecoverable "blind spots"(highlighted in Figure 10a and 10c).

To ensure proper scene exposure, one must control, whenever possible, the environment illumination, camera aperture, exposure time, and gain.

The illumination, aperture, and exposure time directly affect the amount of light that reaches the sensor. Higher exposure is achieved through more external illumination, lower aperture, or longer exposure time.

The gain is split into two: analog and digital. The analog gain measures the sensor's sensibility to the light source. The bigger the analog gain, the more "sensible"the sensor will be to the same light source, leading to a higher brightness value. On the other hand, digital gain is a value that multiplies the acquired signal in post-processing. In real life, it is desired to keep the gain as low as possible, as it amplifies not only desired signal features but also undesired, such as noise.

When the proper balance between aperture, focal length, and exposure time is done, an ideal exposure is achieved. It results in an image with the scene's features optimally highlighted and increased detail (Figure 10b).

**Figure 10 – A scene represented with three different exposures. Maximum detail is achieved when the exposure is properly set. Highlighted in blue are features lost due to saturation during overexposure and in red during underexposure.**



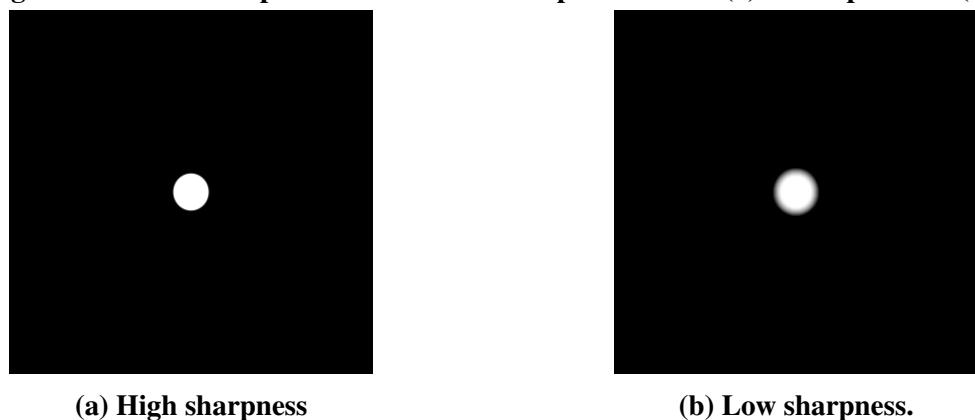
**Source: Personal Collection.**

### 2.3.2 Sharpness

Sharpness describes the quality of details in an image. Humans perceive images with low sharpness as having softer edges and not detailed features. The lack of sharpness in an image can be attributed to different effects, and to identify the correct causation, each one must be independently analyzed.

A high-sharpness image captures very distinct and clear details of a scene. When it comes to visual odometry, a sharper image preserves the punctual nature of fine details, contributing to a finer displacement estimation.

**Figure 11 – A scene represented with two sharpness values. (a) is sharper than (b).**

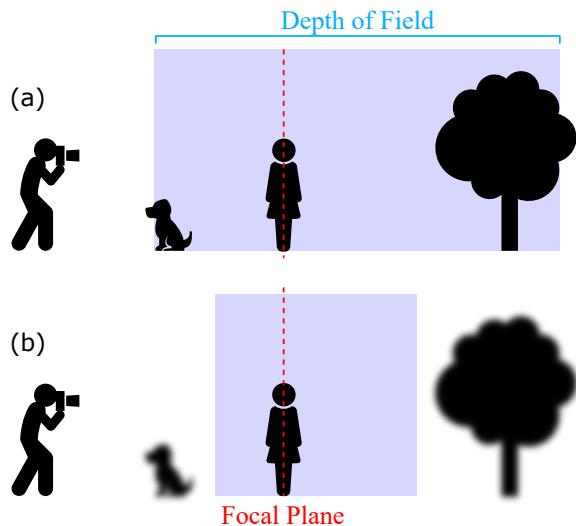


**Source: Personal Collection.**

#### 2.3.2.1 Focus

Focus is the sharpest area of an image. The distance between the object and the lens and between the lens and the sensor defines the focal plane, where all objects within are in focus. There is a tolerance region around the focal plane where objects are still in focus, the *Depth-Of-Field* (Figure 12).

**Figure 12 – Illustration of the Depth-Of-Field.** Optimal focus is achieved in the focal plane, but objects within the *Depth-Of-Field* are still in focus. (a) represents a deep DOF, while (b) a shallow DOF.



Source: Adapted from (MIKERUN, 2019).

The Depth-Of-Field is mainly defined by the lens aperture. A lower aperture leads to a bigger region where more objects are in focus, i.e., a deeper *Depth-Of-Field*. The problem arises because the aperture also affects the exposure; therefore, one must carefully consider the trade-off when adjusting it.

There are two ways to focus an object: changing the distance between the observer and the subject or adjusting the focal length. The lens's ability to shift focal length is vital, especially in situations where no relative change in distance is possible. Algorithms that implement autofocus rely primarily on focal-length micro adjustments through the embedded step motors in the lens.

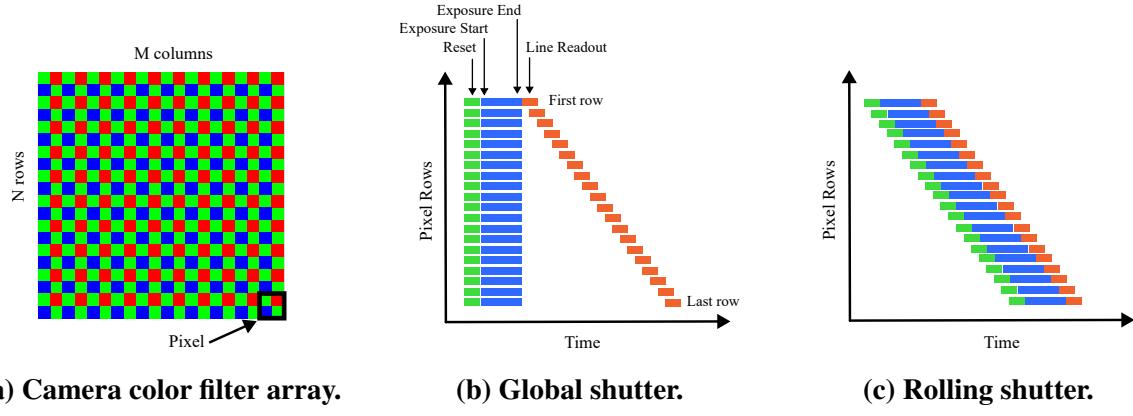
### 2.3.2.2 Motion-related Distortions

Even an object within the Focal Plane can be captured in a distorted way. If there is relative speed between the observer and the subject, the captured image might be distorted or blurry- a phenomenon called the rolling shutter effect and motion blur, respectively. These effects are closely related to how the sensor captures the image. There are two main ways of exposing the camera sensor to light: a global shutter and a rolling shutter.

In a global shutter, all the sensor pixels are simultaneously exposed and hidden from light by a "curtain" called the camera shutter. After the closure, it is read the stored brightness from the pixel array (13a) sequentially, row by row. The time between the shutter opening and closure is the shutter time or speed. The final image will likely be sharp if the photo subject is in the same position when the shutter opens and closes. But, if the object changes position

while the shutter is opened, the pixels will capture variable information over time, resulting in a blurry image (Figure 14b).

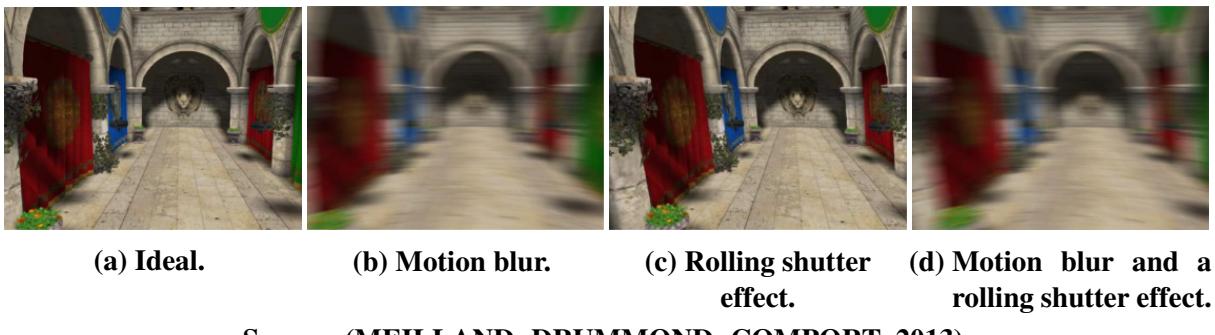
**Figure 13 – Operational difference between global and rolling shutter architecture.**



Source: Personal Collection.

Unlike the global shutter, the sensor in a rolling shutter is permanently exposed to light; thus, there is no physical shutter. Instead, the sensor pixels act as an accumulator and increase its brightness value as time passes. Periodically, the pixels are read and reset each row at a time, with a period value called the exposure time or speed. This reset pattern acts as an "electronic shutter". The motion-related artifact displays a distinct pattern since the operating principle is fundamentally different. Because the pixels are permanently exposed but periodically read, it results in a distinct visual distortion, an effect called the *rolling shutter effect* (Figure 14c). Although the rolling shutter effect does not manifest on images captured by a global shutter, both motion-related artifacts do appear on images captured by a rolling shutter (MEILLAND; DRUMMOND; COMPORT, 2013), shown in Figure 14d.

**Figure 14 – Artifacts occurred by movement of the observer.**



Source: (MEILLAND; DRUMMOND; COMPORT, 2013).

To minimize motion-related distortions, exposure/shutter time must be minimal. But, as previously noted in Section 2.3.1, a decrease in exposure time leads to a reduction in exposure, raising a potential problem with image quality.

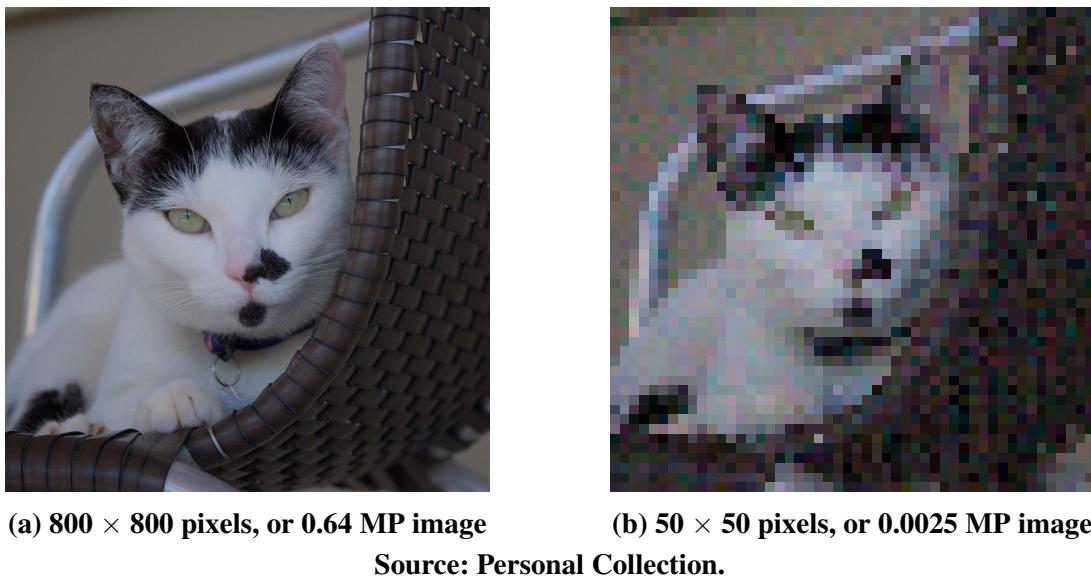
### 2.3.3 Amount of Details

Even if a scene is appropriately shot regarding exposure and sharpness, the image must have enough spatial information to allow the visual odometry algorithm to work effectively. The amount of spatial information (or details) in an image is dictated by the number of pixels and resolution.

#### 2.3.3.1 Number of Pixels

The number of pixels is the number of rows and columns of an image, mistakenly (and commonly) called "resolution". Two images of a subject with the same size but different numbers of pixels represent different levels of detail ; as long as the size is kept constant, a higher number of pixels represent a more detailed image (Figure 15a).

**Figure 15 – The same scene is represented by different numbers of pixels. Image (a) has 256 times more pixels than (b), thus capturing more details.**



**Source: Personal Collection.**

The total number of pixels (number of rows  $\times$  number of columns) of a camera sensor is commonly expressed in Millions of Pixels or MP. For instance, a 10 MP sensor will be more suitable to represent a subject than a 5 MP sensor, assuming similar capturing conditions. It is possible to decrease the number of pixels of an image through down-sampling algorithms. This class of algorithm performs different types of interpolation to keep the original image aspect ratio while reducing the number of pixels (CASTLEMAN, 1996), as illustrated in the transition from Figure 15a to 15b. It proves useful in conditions where a higher number of pixels might not contribute significantly to a better representation of an object but could slow down image processing algorithms.

## 2.4 Displacement Estimation

The heart of a visual odometry system is the displacement estimation algorithm. The algorithm's main goal is to estimate the displacement performed by the observer based on captured consecutive images. One way to compare is through image registration methods—a process of overlaying two (or more) images of the same scene taken at different times, from different perspectives, and/or by different sensors (ZITOVA; FLUSSER, 2003).

The image registration methods are generally divided into two categories: feature-based methods and area-based methods (TONG *et al.*, 2019).

Feature-based methods detect image features (e.g., corners, edges, regions) and estimate geometric correspondence between pairs by matching the location of the features in each image. This category is vastly used in powerful modern autonomous systems, mainly due to the great success of convolutional neural network-based methods (HE *et al.*, 2020).

Area-based methods compare image pairs by computing a measure based on pixel intensity. This category has low computational overhead, as it does not require extra processing to identify features when compared to feature-based methods (TONG *et al.*, 2019); therefore it is specially suitable for low performance CPU. Through simple mathematical operations on the similarity measure, it is possible to estimate displacement between image pairs. Some measures examples are: cross-correlation (LEWIS *et al.*, 1995), the sum of squared differences (KYBIC; UNSER, 2003), and mutual information (CHEN; VARSHNEY; ARORA, 2003).

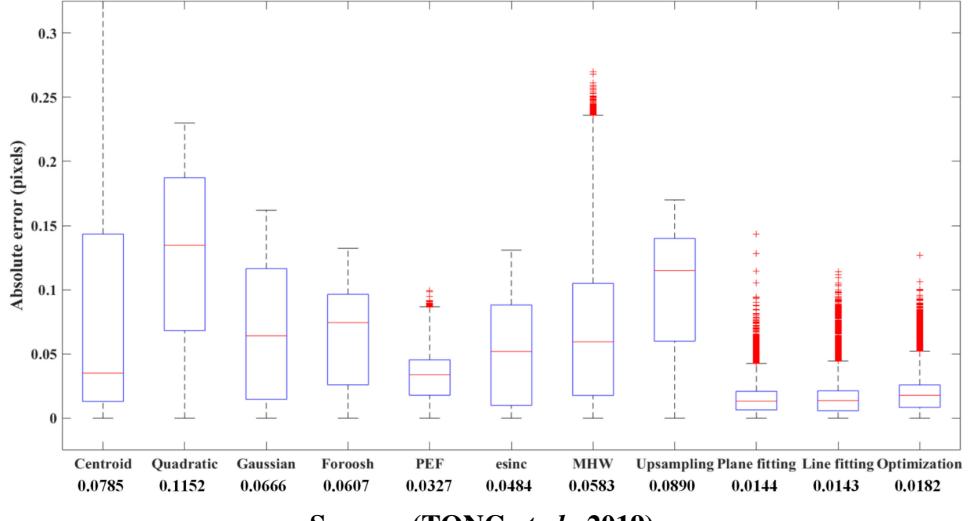
Classical area-based methods use the image representation directly from the spatial domain. Although it is the most intuitive way of estimating displacement, Fourier-based methods have been proven to offer high theoretical accuracy, high computational efficiency, and insensitivity to frequency-dependent noise and intensity contrast (TONG *et al.*, 2019).

A comprehensive review by Tong *et al.* (2019) compared Fourier-based image registration methods divided into five categories: plane fitting (STONE *et al.*, 2001; MALCOLM *et al.*, 2002), line fitting (HOGE, 2003; TONG *et al.*, 2015), sawtooth estimation (BALCI; FOROSH, 2006), frequency estimation (XU; VARSHNEY; NIU, 2006) and optimization (PUYMBROECK *et al.*, 2000).

The average displacement error estimated by methods from each category was summarized in Figure 16. The best average performance was achieved by methods from the line fitting category. Amisdt this category, a method proposed by Hoge (2003) is straightforward and allows 2D subpixel displacement estimation. It applies an *Singular Value Decomposition* (SVD) followed by a rank-one truncation to the cross-power spectrum, which results in noise reduction and computational complexity improvement.

For the above-mentioned reasons, the visual odometry system will be based on the SVD-based method and the following section will do a deep dive into each step of it.

**Figure 16 – Boxplots of the absolute errors for different subpixel methods.**



Source: (TONG *et al.*, 2019).

#### 2.4.1 Phase correlation

To simplify the problem without significant compromises, some premises are taken:

1. The observer faces the floor directly and continuously captures images of it;
2. The observer's movement lies in a plane parallel to the floor;
3. There is no rotation of the observer around any axis during the movement;
4. The captured images are in grayscale.

Let  $f_n(x,y)$  be the N-th image and  $f_{n-1}(x,y)$  the previous. Under the previous assumptions, we can state that

$$f_n(x,y) = f_{n-1}(x - a_n, y - b_n) \quad (14)$$

where  $a_n$  and  $b_n$  are the n-th horizontal and vertical shifts between the images measured in pixels. Thus, the odometry goal of discovering the observer's movement is closely related to the estimation of  $a_n$  and  $b_n$ .

One way of achieving it is through an operation called *cross-power spectrum*:

$$Q_n = Q(u,v) = \frac{F_{n-1}(u,v) \odot F_n^*(u,v)}{|F_{n-1}(u,v) \odot F_n^*(u,v)|}, \quad (15)$$

where  $\mathcal{F}\{\cdot\}$  is the 2D Fourier Transform,  $F_n(u,v) = \mathcal{F}\{f_n(x,y)\}$  is the spectra of  $f_n$ ,  $F_n^*(u,v)$  the complex conjugate of  $F_n(u,v)$  and  $\odot$  the Hadamard product (element-wise product) and  $u,v \in \mathbb{R}$ .

For our particular case where one image is the perfect shifted copy of the other, according to the Fourier Shift theorem, the spectrum of  $f_n(x,y)$  is

$$F_n(u,v) = F_{n-1}(u,v) \exp \{-j(ua_n + vb_n)\}. \quad (16)$$

Substituting Eq. 16 into Eq. 15 results in

$$Q_n = \exp \{-j(ua_n + vb_n)\}, \quad (17)$$

which is possible to decompose into

$$Q_n = \exp\{-jua_n\} \exp\{-jvb_n\} = q_u(u,v)q_v(u,v), \quad (18)$$

where  $q_u(u,v) = \exp\{-jua_n\}$  varies only in the  $u$  direction ( $q_u(u,v) = q_u(u)$ ), and  $q_v(u,v) = \exp\{jvb_n\}$  varies only in the  $v$  direction ( $q_v(u,v) = q_v(v)$ ). Since  $q_u(u)$  and  $q_v(v)$  are orthogonal functions (their inner product is zero), it is valid to say that  $Q_n$  is a rank one function.

In practice,  $f_n(x,y)$  and  $f_{n-1}(x,y)$  are digital images  $I_n[x,y]$  and  $I_{n-1}[x,y]$  with  $M$  columns and  $N$  rows, i.e. they are discrete functions with limited domain.

This transition implies that the continuous function  $Q_n = Q_n(u,v)$  under the new domain will be a matrix  $\mathbf{Q}_n \in \mathbb{C}^{M \times N}$  with  $M$  rows and  $N$  columns. Analogously, since  $Q_n$  is composed by a product of two orthogonal functions,  $\mathbf{Q}_n$  is a rank-one matrix that can be factorized as

$$\mathbf{Q}_n = \mathbf{q}_u \mathbf{q}_v^H, \quad (19)$$

where  $\mathbf{q}_u \in \mathbb{C}^M$  and  $\mathbf{q}_v \in \mathbb{C}^N$  are orthogonal column-vectors corresponding to  $q_u(u,v)$  and  $q_v(u,v)$  respectively, and  $\mathbf{q}_v^H$  is the Hermitian transpose of  $\mathbf{q}_v$  (HOGE, 2003).

#### 2.4.1.1 SVD

The Singular Value Decomposition (SVD) is a factorization of a real or complex-valued matrix. The SVD of an  $m \times n$  complex-valued matrix  $\mathbf{A}$  is (STRANG, 2022)

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^H \quad (20)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal matrices, and  $\Sigma$  is a diagonal matrix. The entries of  $\Sigma$  are the eigenvalues of  $\mathbf{A}$ , such that  $\sigma_i = \Sigma_{ii}$  and  $\sigma_i > \sigma_{i-1} > \dots > 0$ . The columns of  $\mathbf{U}$  and  $\mathbf{V}$  are called left and right-singular vectors of  $\mathbf{A}$ , respectively. These singular vectors are the eigenvectors of  $\mathbf{A}\mathbf{A}^H$  and  $\mathbf{A}^H\mathbf{A}$ , respectively. Since  $\mathbf{A}\mathbf{A}^H$  and  $\mathbf{A}^H\mathbf{A}$  are square matrices, the singular vectors form two sets of orthonormal basis  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$  and  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ .

The SVD factorization is widely used in many applications, as it decomposes the original matrix  $A$  into the fewest optimal components. In our case, we know that  $\mathbf{Q}_n$  is rank-one, therefore theoretically  $\sigma_1 \neq 0$  and the rest  $\sigma_i = 0$  for all  $i = 2, 3, \dots, \min(n,m)$ . In practice, due

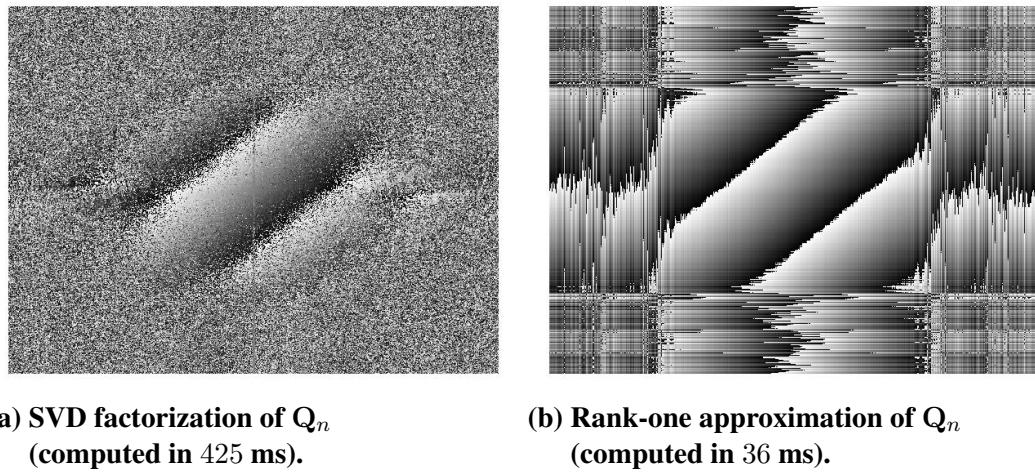
to unexpected effects such as noise, lens distortion, and numerical errors, the first eigenvalue ( $\sigma_1$ ) is dominant, but the rest are non-zero.

One way of exploiting the rank-one nature of  $\mathbf{Q}_n$  is through the Eckart-Young theorem (ECKART; YOUNG, 1936). This theorem allows us to estimate the SVD matrix only by computing the first eigenvalue, left and right singular vectors:

$$\tilde{\mathbf{A}} = \mathbf{u}_1 \sigma_1 \mathbf{v}_1^H. \quad (21)$$

Comparing Eq. 19 and Eq. 21, it is clear that we can compute  $\mathbf{q}_u$  and  $\mathbf{q}_v$  through an SVD decomposition followed by a rank-one approximation—an widely known and solvable problem (KUMAR; SCHNEIDER, 2017). When put side by side  $\mathbf{Q}_n$ , its reconstructed version through only SVD and SVD followed by rank-one approximation, we can see a significant improvement not only in runtime but also in noise reduction (Figure 17).

**Figure 17 – Phase of the cross power spectrum matrix of two shifted 2D signals. (a) is reconstructed from an SVD factorization, and (b) is reconstructed from a rank-one approximation.**



**Source: Personal Collection.**

Finally, we can find the space-domain displacements  $a_n$  and  $b_n$ , which are encoded in the *phases* of the complex exponential contained in the estimated vectors  $\mathbf{q}_u$  and  $\mathbf{q}_v$ , respectively. We extract the phases with the entry-wise *angle* operator  $\angle$ :

$$\mathbf{p}_u = \text{unwrap}(\angle \mathbf{q}_u) \quad (22a)$$

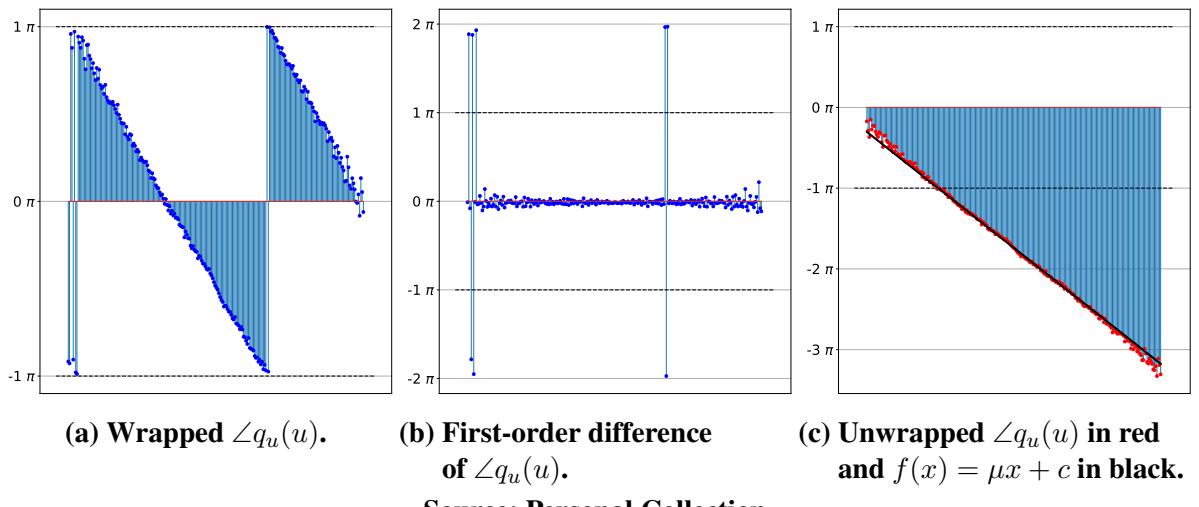
$$\mathbf{p}_v = \text{unwrap}(\angle \mathbf{q}_v) \quad (22b)$$

#### 2.4.1.2 Phase unwrapping

The function  $\text{unwrap}(\cdot)$  performs the *phase unwrapping* based on a method proposed by Itoh (1982). The method is composed of the following steps:

1. Compute the first-order difference, i.e.  $d[k] = x[k] - x[k - 1]$ , of the wrapped signal (Figure 18b);
2. Identify points where the difference is close to  $\pm 2\pi$ . These are the points where phase wrapping occurs (outliers in Figure 18b);
3. Remove the phase wrapping by subtracting or adding  $2\pi$  to the transition points where wrapping is identified;
4. Calculate the unwrapped signal by a cumulative sum operation:  $c[k] = \sum_{i=1}^k d[i]$  (Figure 18c).

**Figure 18 – Steps of phase unwrapping algorithm by Itoh (1982). Wrapping is expected when first-order finite difference is close to  $\pm 2\pi$ .**



The unwrapping is necessary since  $\angle \mathbf{q}_u$  and  $\angle \mathbf{q}_v$  are restricted to the interval  $[-\pi, \pi)$  (Figure 18a). After phase unwrapping, the entries of  $\mathbf{p}_u$  and  $\mathbf{p}_v$  should lie over lines with slopes  $\mu$  (assuming  $f(x) = \mu x + c$ ) proportional to the displacements  $a_n$  and  $b_n$  (Figure 18c). The line coefficients are obtained via least squares, then converted to linear shift through

$$a_n = \frac{\mu p_u N}{2\pi} \quad (23)$$

and

$$b_n = \frac{\mu p_v M}{2\pi}. \quad (24)$$

#### 2.4.1.3 Three-dimensional trajectory

Computed the n-th 2D displacement  $\delta_n = [\hat{a}_n, \hat{b}_n, 0]^T$  we can convert to a 3D displacement through rotation. If the observer orientation is obtained from an IMU as a quaternion  $\mathbf{q}_n$ , as defined in Eq. 7, the rotation is achieved by

$$\Delta_n = \mathbf{q} \otimes \boldsymbol{\delta}_n \otimes \mathbf{q}^*, \quad (25)$$

and the cumulative 3D displacement computed through

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \Delta_n. \quad (26)$$

#### 2.4.1.4 Windowing and practical effects

In signal processing, windowing is the act of "cutting" the signal to its relevant region. Every acquired signal is implicitly windowed, as there is no infinite memory to represent infinite signals. The windowing operation is achieved by multiplying the signal  $s(t)$  by its window  $w(t)$ , i.e.:

$$s_w(t) = s(t)w(t). \quad (27)$$

The new multiplicative term  $w(t)$  implies that the spectrum of  $s_w(t)$  will be

$$S_w(\omega) = S(\omega) * W(\omega), \quad (28)$$

where  $S(\omega)$  and  $W(\omega)$  are the spectra of  $s(t)$  and  $w(t)$ , respectively.

If  $w(t)$  is a rectangular window,  $S(\omega)$  will be the sinc function. Assuming the window has limited length, the windowed signal  $s_w(t)$  will have a distorted spectrum when compared to  $s(t)$  spectrum. This distortion manifests as a broadening in frequency peaks and, thus, is called *spectral leakage* (Figure 19).

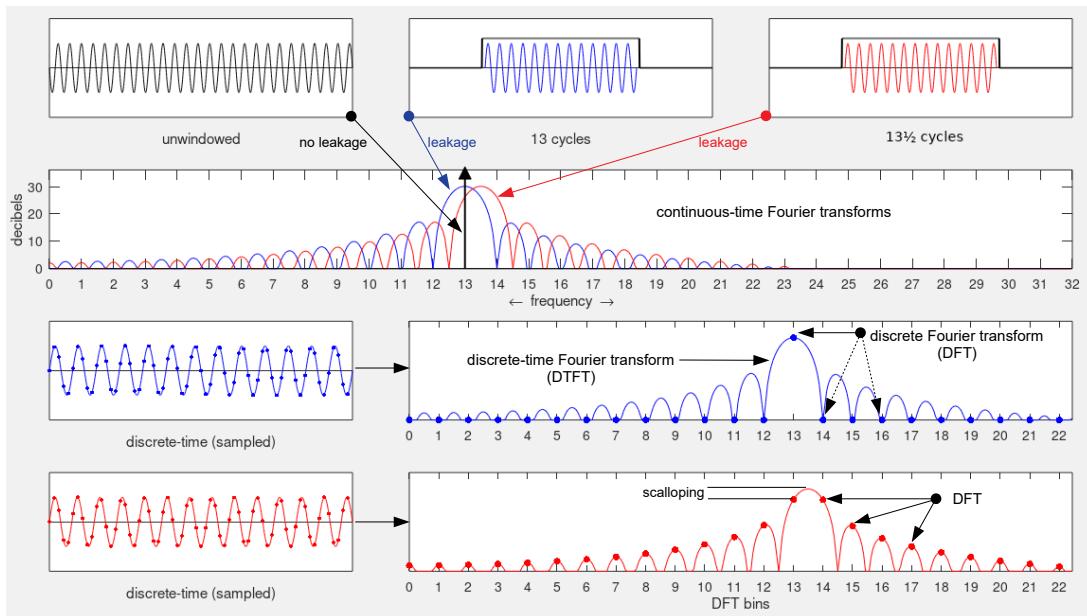
Besides spectral leakage, if the window length equals a non-integer multiple of the signal period, the discrete version of  $s(t)$  is affected by artifacts related to *edge effect*. This effect manifests by distorted frequency representation near the signal's edges and unnatural high frequencies (scalloping in Figure 19).

Other types of windows may be employed (MANOLAKIS; INGLE, 2011) to mitigate these phenomena. Each window has a different impulse response and, thus, might be more adequate for applications with different necessities.

Two-dimensional signals, such as images, are also prone to spectral leakage and edge effects. To reduce these, they undergo space-domain windowing before the 2D FT (Figure 20). Specifically, the Blackman-Harris proved to be the most suitable to enhance the proposed method's precision and stability (STONE *et al.*, 2001).

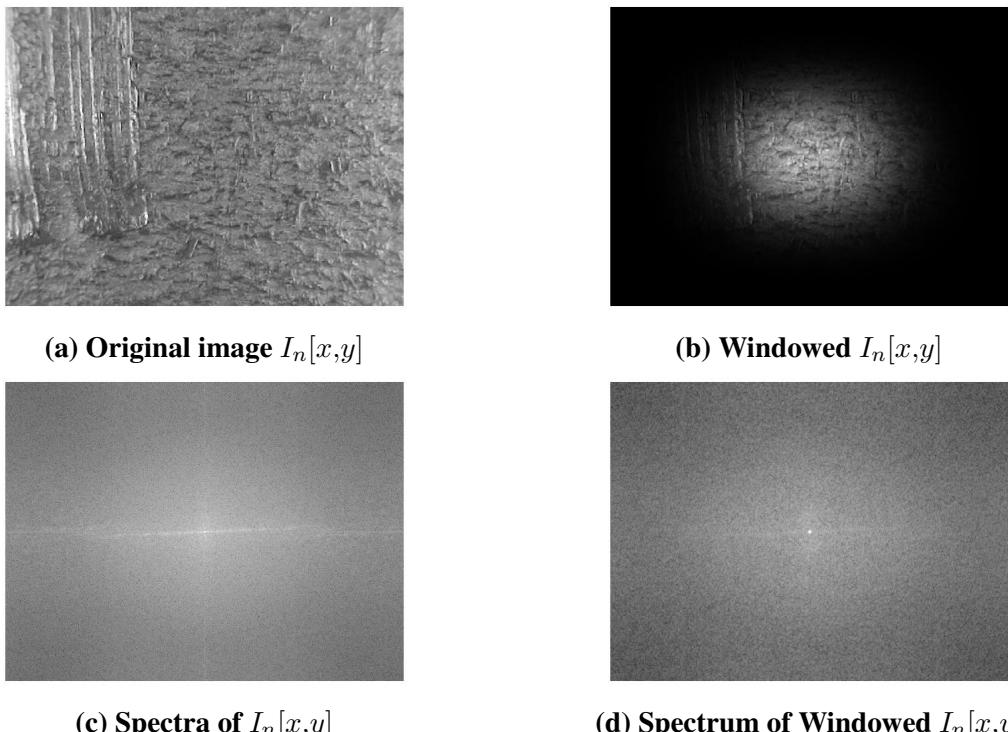
Also, after the unit-rank factorization of Eq. (19) through SVD, a rectangular windowing of  $\mathbf{Q}_n$  is applied. This aims to exclude high-frequency components that tend to be corrupted (see the peripheral region of Figure 17b) (HOGE, 2003).

**Figure 19 – Spectral leakage caused by rectangular windowing.**



Source: Adapted from (K, 2018).

**Figure 20 – Example of spatial windowing and its result on the spectra.**



Source: Personal Collection.

## 2.5 Commentaries

This chapter laid down the basic concepts to understand visual odometry, its importance in NDT, and the algorithms that solve this problem. Among the NDT methods, US inspection was explained in depth. To understand the advantages of visual odometry, traditional mechanical

odometry was explained. Then, important aspects of image capturing were introduced to aid decisions regarding improvements in image capturing and analysis of results. Finally, every step of the displacement estimation algorithm was explained, along with supplementary concepts of linear algebra and signal processing used.

### 3 MATERIALS AND METHODS

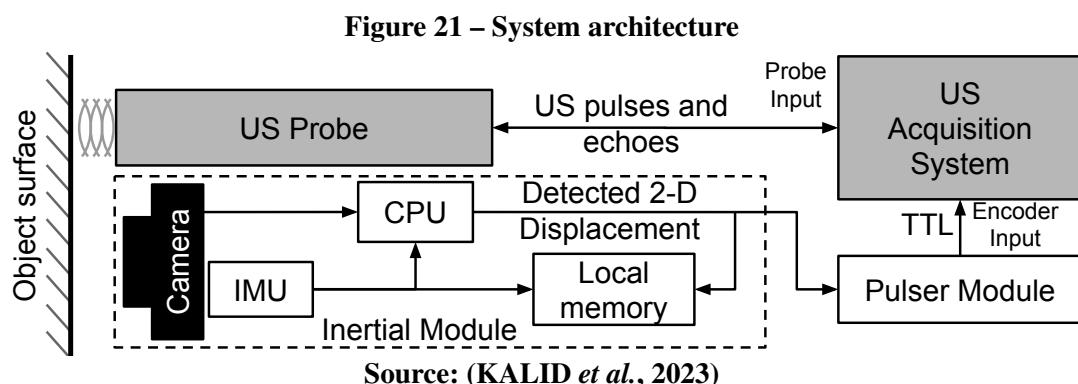
This chapter is divided into three sections. The first will describe the general architecture of the proposed system in terms of components and how they interact with each other. In the second section, the electronic hardware and mechanical parts will be described. The last section will dissect the steps involved in estimating displacement between image pairs.

#### 3.1 System Architecture

The ensemble required for an US acquisition using a visual odometer is composed by three major parts:

1. Inertial Module: Continuously captures images of the surface and sends them to Central Processing Unit (CPU). Once an image pair is available, the CPU performs the phase correlation of frames  $n - 1$  and  $n$  which estimates the  $n$ -th 2D displacement. The displacement, represented as a pair of floats, is locally stored and transmitted through Ethernet to the pulser module. In parallel, the instantaneous orientation quaternion for instant  $n$ , provided by the IMU, is stored as well;
2. Pulser module: Upon reception of the displacements, the Pulser Module immediately emulates the Transistor-Transistor Logic (TTL) pulses that would be generated by a conventional wheel encoders for the same  $n$ -th 2D displacement. This is step mandatory, as modern US acquisition systems only supports this type of input;
3. US transducer and acquisition system: The US acquisition system receives the TTL pulses through the encoder digital port. It interprets as if they were generated by a wheel encoders fixed at the transducer. Finally, it fires the transducer according to user-defined configuration.

The received US echoes don't affect the 2D displacement estimation. The camera and IMU are mechanically fixed to the US probe. The architecture is summarized in Figure 21.



## 3.2 Hardware

### 3.2.1 Inertial Module

The camera was a Raspberry Pi NoIR V2 8MP (PI, 2016), which contains an IMX219 Sony sensor. It captures video at 1080p 30 fps, 720p 60 fps or 640x480p 90 fps. It doesn't have auto-focus, therefore focus is achieved only by changing the distance between subject and camera. A macro lens was placed in front of the camera to enhance the ability to capture micro detail of the surface and minimize focusing distance, maximizing image sharpness. Additional lighting was done by four high power LEDs.

It was used a Raspberry Pi (RPi) 4 single board computers for CPU implementation. The IMU was a BNO055 connected to the RPi through its General Purpose Input/Output (GPIO) ports.

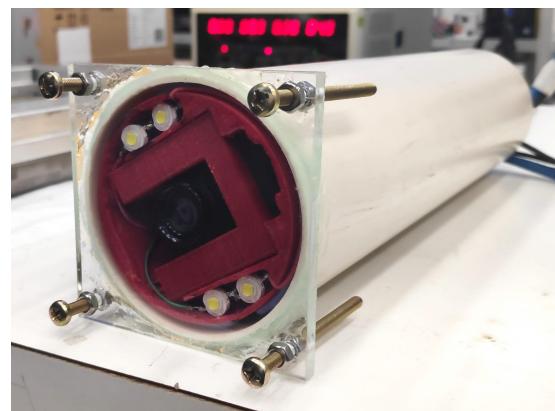
The electronic parts were mounted on a 3D printed support, which ensures a fixed relative position between the components (Figure 22a).

The prototype was encased in a sealed container with a transparent acrylic front (Figure 22b). The transparent window allows the camera and lighting system to be kept inside the dry environment while the module stays submerged. Adjustable bolts were used to maintain optimal distance between the camera and the object surface for appropriate focus, as the Picamera v2 native lens has a fixed focal length.

**Figure 22 – Visual odometer hardware.**



**(a) Electronic components mounted on 3D printed support.**



**(b) Prototype encased in a sealed container.**

Source: (JUNIOR, 2023); Personal Collection.

### 3.2.2 Pulser Module

The pulser module was implemented using a RPi 4B. It was directly connected to the Inertial Module through an Ethernet cable. The displacements were transmitted using TCP pro-

tocol (PETERSON; DAVIE, 2007), and were based on the work presented by Everton *et al.* (2016).

The relation between a number of pulses and the value of displacement is defined by the user a priori (e.g., 10 pulses/mm).

### 3.2.3 Ultrasound acquisition

The ultrasound acquisition system was a *Eddyfi Panther*<sup>TM</sup>. Among the available ports, two were used: the encoder-dedicated 3-channel port and the US transducer port.

The ultrasound transducer was a 64 element 5 MHz transducer by Imasonic (Figure 23a). The detailed specifications are presented in Table 1.

**Table 1 – Specifications of the linear phased array from Imasonic (IMASONIC SAS, France).**

Specification	Value
Number of elements	64
Central Frequency	5 MHz
Fractional bandwidth	40 %
Element width	0.5 mm
Inter element width	0.6 mm
Element depth	10 mm

**Source: Personal Collection**

The relation between received pulses and a US acquisition is defined by the user (e.g., one acquisition every ten pulses) and depends directly on the speed during the inspection. If faster movements are expected, a faster acquisition rate is recommended (e.g., one acquisition every two pulses). Both values were set based on the best empirically obtained results for the given inspection routine.

The transducer and inertial module were mechanically coupled through a 3D printed outer support (Figure 23b).

**Figure 23 – Ultrasound transducer working alongside the visual odometer.**



**(a) Ultrasound transducer connected to the acquisition system.**

**(b) Transducer and Inertial Module mechanically coupled.**

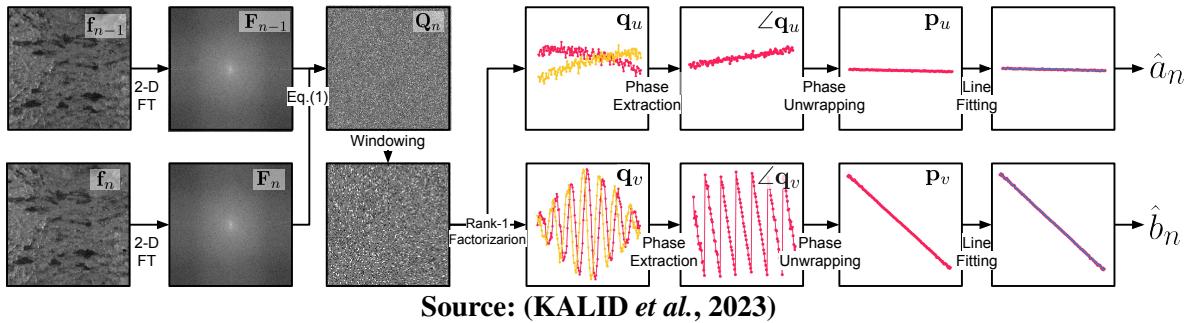
**Source: Personal Collection.**

### 3.3 Software

The implementation is written in Python language and available at <https://github.com/thiagokalid/Virtual-Encoder-ECNDT-2023.git>.

Figure 24 shows an illustrative chart of the main steps of the displacement estimation method.

**Figure 24 – Steps performed to compute 2D displacement ( $\hat{a}_n, \hat{b}_n$ ) from two consecutive frames  $f_{n-1}$  and  $f_n$ .**



The Picamera was set to the lowest resolution configuration ( $640 \times 480$ ) as it allows maximum capturing rate (90 fps stable or 200 fps unstable). It has a rolling shutter set to the lowest exposure time possible,  $20\mu\text{s}$  approximately. Since the estimation algorithm complexity depends on the image size, the  $640 \times 480$  image were down-sampled by a factor of two through bi-cubic interpolation (FADNAVIS, 2014) resulting in a size of  $320 \times 240$ . Finally, a Blackman-Harris spatial windowing was applied to decrease frequency leakage as described in Section 2.4.1.4 (omitted from Figure 24, but shown in Figure 20b).

Then, the 2D spectra are computed through 2D FFT, followed by frequency windowing (second column in Figure 24). The frequency window is a rectangular type that extracts the  $N/2$  central pixels of the spectra, where  $N = \min(640/2, 480/2)$  (STONE *et al.*, 2001).

Sequential pairs of spectra were used for computing the cross-power spectrum using Eq. 15 (first figure on the third column of Figure 24), followed by rank 1 SVD approximation. The SVD was computed through Scipy *svds(.)* implementation (VIRTANEN *et al.*, 2020). Only the 50 central pixels are kept for the following steps (second figure in third row of Figure 24), as they keep the relevant signal phase information, illustrated in Figure 17b. These criteria were found through empirical experimentation.

The result of the SVD are two complex-valued vectors  $\mathbf{q}_u$  and  $\mathbf{q}_v$  whose real and imaginary components are represented by two different colors in Figure 24 fourth column.

The phases of  $\mathbf{q}_u$  and  $\mathbf{q}_v$  are extracted but display the wrapping pattern, compensated through the unwrapping algorithm explained in Section 2.4.1.2. Point are compensated when they are above 80 % of  $\pm 2\pi$ . After unwrapping, it is done a linear regression (subtle blue line overlapping points on the seventh column of Figure 24) and the displacement estimates ( $\hat{a}_n, \hat{b}_n$ ) are computed through Eq. 23 and 24.

### 3.4 Commentaries

In this chapter, the proposed solution was decomposed into its main parts. Firstly, the overall architecture is exposed and the purpose of each component is explained, as well as how they interact with each other. Then, the physical parts of the system are shown, and some remarks regarding electronic hardware are made. Finally, the steps of displacement estimation are summarized, and the parametrization of each step is exposed.

## 4 RESULTS AND DISCUSSION

This chapter contains the experimental results of the displacement estimation methods presented in Section 2 and 3. Two sets of experiments were made: one to compute quantitative metrics regarding the method accuracy and another to simulate in-situ operation during a US inspection.

Section 4.1 introduces aspects of the experimental setup, for example: specimens used, medium considered and trajectories performed. Section 4.2 explains the adjustments methodology necessary to convert a pixel-based displacement to an SI unit, such as a millimeter. Section 4.3 exposes the results and important remarks regarding noticed patterns, experimental difficulties and limitations.

### 4.1 Experimental Setup

For the adjustments procedure and first experiment, two steel specimens were used to obtain quantitative results: a planar and a cylindrical shaped. Each experiment was reproduced in a dry environment and underwater. The two mediums (air and water) aimed to replicate the contact and immersion ultrasound inspection. Aluminum frames were installed on the specimen perimeter to restrict the movement of the visual odometer. The visual odometer was moved following the aluminum templates to obtain a more precise and controlled movement. Figure 26a and 27a show the planar specimen in a dry environment and the cylindrical specimen underwater, respectively.

The last experiment aims to test the entire ensemble composed by the visual odometer and US acquisition system in conditions close to real-life inspections. Therefore, a freehand trajectory was performed; the specimen was a half-pipe with discontinuities, and the operation was underwater.

### 4.2 Adjustments

Before the experiments, a set of adjustments were made to enhance the image quality and adjust the system displacement estimate.

#### 4.2.1 Focus

The Picamera v2 lens has a fixed focus, as discussed in Section 3.2.1. To achieve good focusing, the bolts and screws of the inertial module were adjusted. This effectively altered the camera's position, simulating a change in focal length. Based on a qualitative assessment, the

setup was adjusted until the captured image's sharpness and focus were good enough. The final distance between the acrylic window and a planar surface was 20.00 mm.

#### 4.2.2 Displacement

The 2D displacement estimates are shown in Section 2.4 and 3.3 are in a number of pixels. To convert these pixel-based displacements to SI units (e.g., millimeters), the pixel size must be determined.

Two independent linear trajectories were performed to estimate the pixel size. They were aligned with the x and y axes in the 2D plane and consisted of a single unidirectional trajectory with known length. The pixel size was calculated by dividing the total cumulative estimated displacement (in pixels), obtained using the proposed method, by the corresponding total measured distance (in millimeters).

Additionally, the monocular camera might be rotated around its axis (clear in Figure 22b). To compensate for this effect, this rotation was measured during the prior to the experiment to be later compensated. Table 2 contains the pixel size estimates.

**Table 2 – Pixel size estimation results.**

Specimen	Medium	Displacement / [pixels]		Measurements / [mm]		Pixel size / [mm]	
		x-axis	y-axis	Width	Height	Width	Height
Planar	air	16640	7228	358.00	200.33	0.022	0.028
	water	17588	7788			0.020	0.026
Cylindrical	air	13186	8847	298.33	216.67	0.023	0.024
	water	17588	7788			0.023	0.024

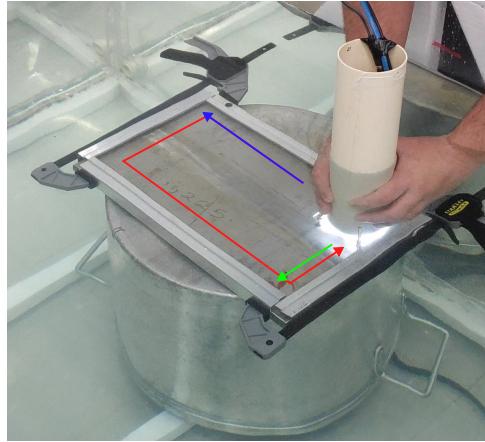
**Source: Adapted from (JUNIOR, 2023)**

### 4.3 Experiments

#### 4.3.1 Controlled trajectories

To validate the proposed method, experiments were conducted involving the movement of the visual odometer over planar or cylindrical specimens along controlled trajectories. In each experiment, a different type of trajectory was developed: single-x, single-y, and closed-loop (Figure 25). Single-x and single-y are unidirectional trajectories along the x or y axes. The higher movement speed was the only difference between this type of trajectory and the one done during the adjustment procedure. Closed-loop was a trajectory along the specimen perimeter with the ending point coinciding with the starting point. These experiments aimed to compare the estimated 2D displacement with the actual path measurements.

**Figure 25 – Planar specimen underwater and the trajectories performed: closed-loop (red), single-x (blue) and single-y (green).**

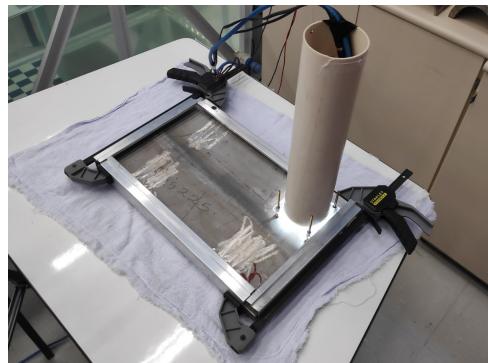


Source: Personal collection.

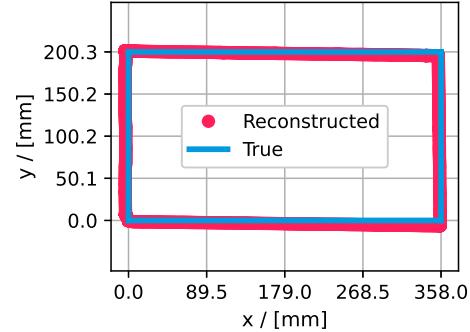
The aluminum frames along the perimeter of the cylindrical specimen allowed a 298.33 mm radial and 216.67 longitudinal displacement. Considering the specimen's theoretical radius is 85.00 mm, this corresponds to moving along 99.60° of the half-pipe circumference.

The results are summarized in Table 1.

**Figure 26 – Experimental setup and result for a planar specimen in dry environment.**



(a) Planar template



(b) Planar path (2D)

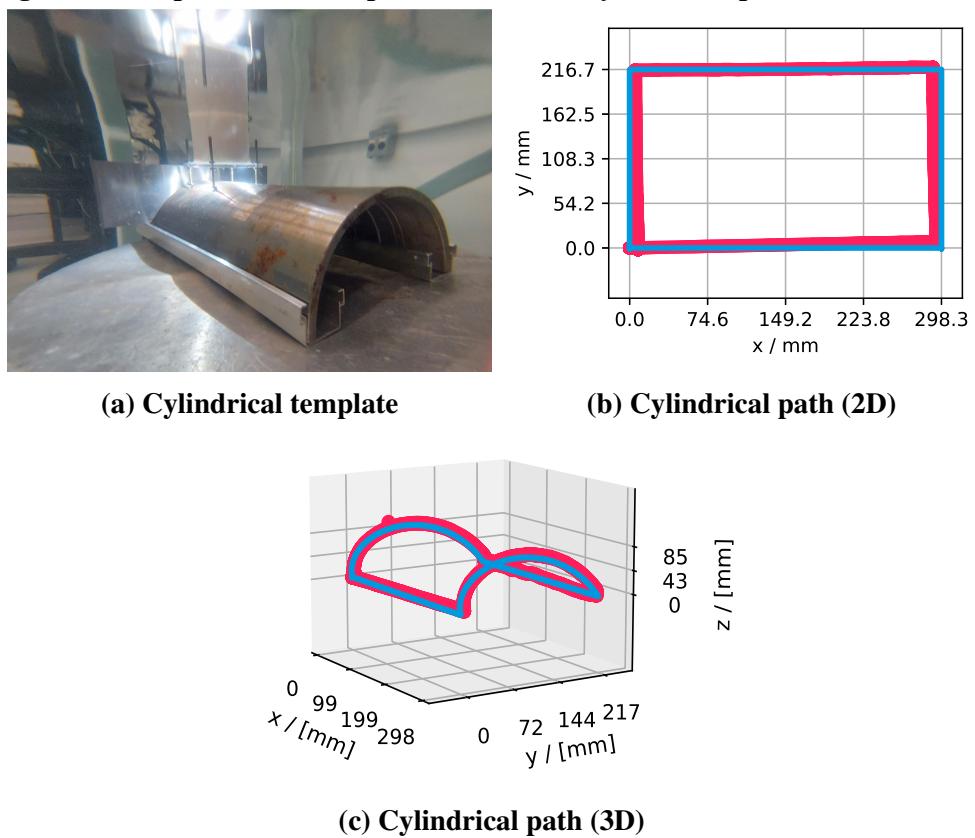
Source: Adapted from (KALID *et al.*, 2023)

**Table 3 – Cumulative displacements, in millimeters, measured for the 2D trajectories reconstructed from the data provided by the visual odometer.**

Test type	Axis	Planar			Cylindrical		
		Expected	Air	Water	Expected	Air	Water
Single-x	x	358.0	358.9	357.8	298.3	299.4	293.0
	y	0.0	-4.7	-7.1	0.0	0.4	-8.4
Single-y	x	0.0	-2.6	-2.5	0.0	1.6	-0.3
	y	200.3	200.7	200.6	216.7	216.5	210.5
Closed loop	x	0.0	-3.4	-3.8	0.0	-1.6	8.2
	y	0.0	7.2	5.5	0.0	5.6	12.2

Source: Adapted from (KALID *et al.*, 2023)

**Figure 27 – Experimental setup and result for a cylindrical specimen underwater.**

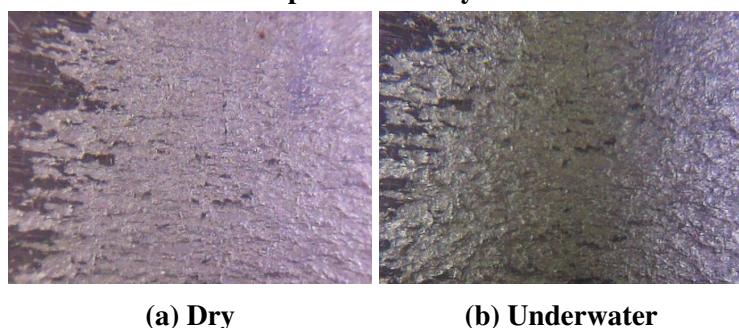


Source: Adapted from (KALID *et al.*, 2023)

The proposed method results in accurate estimates, as shown in Table 1. The cumulative displacement presented a relative error between 0.05 % (single-x, planar and underwater) up to 2.80 % (closed-loop, cylindrical and underwater) of the traveled distance.

An unexpected pattern occurred: x-axis estimates were overall more accurate than y-axis. This phenomenon might be related to the lightning asymmetry, accentuated when the visual odometer was submerged. The difference is clear when comparing images from each medium side by side (Figure 28).

**Figure 28 – A similar frame captured in a dry environment and underwater.**

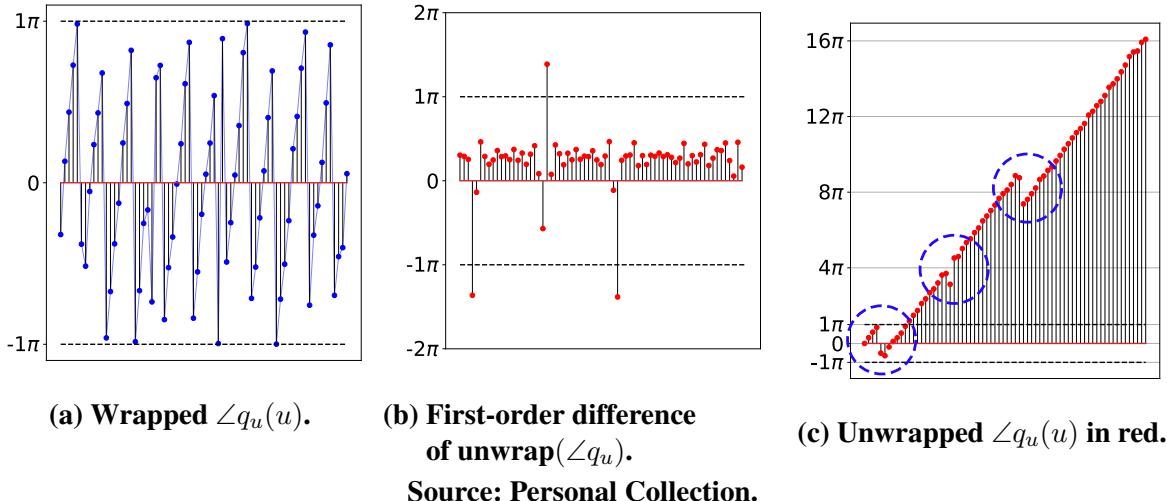


Source: Personal Collection.

Also, the employed phase unwrapping algorithms did contribute to estimation error, as it rarely yielded incorrect results in some frames. In the results presented in Figure 29, the

algorithm was able to work properly in most unwrapping, but in occurrences (highlighted in red), it was not. This effect is probably associated with the unwrapping occurring at the same time as samples with high noise, and since the algorithm proposed by Itoh uses a hard threshold, it is unable to adapt to this type of scenario.

**Figure 29 – Steps of phase unwrapping algorithm by Itoh (1982) yielding estimation error. The algorithm was unable to unwrap data in some places (highlighted with blue circles). All points showed in 29b are supposed to be within  $[-\pi, \pi]$  if the operation was successful.**



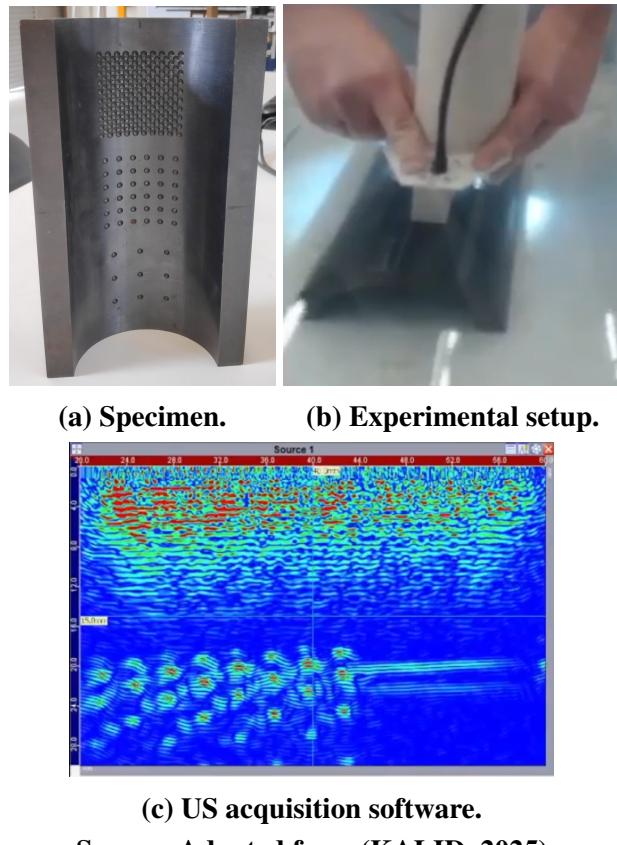
Overall, estimates for a dry environment were more accurate. This pattern might be related to lower image distortion—as the picamera and macro lens were designed for dry operation—, fewer suspended particles, more uniform scene illumination (Figure 28), and better maneuverability due to higher friction between the bolts and the specimen.

#### 4.3.2 Operation along US acquisition system

A final experiment, using the visual odometer and US acquisition system, was conducted to test the viability of the proposed solution. The goal was to show the visual odometry working side by side with the US acquisition system, properly triggering acquisition. Figure 30 illustrates an example of the real-time operation; the transducer is over a set of discontinuities of the specimen that are clearly shown in the US acquisition software. A qualitative analysis of the inspection shows a low response time for the entire ensemble, i.e., any movement performed was quickly noticed in the US acquisition system.

The Raspberry Pi 5 was able to estimate displacements at about 20 FPS, which, depending on the movement speed during the inspection, represents a real-time operation. The processing bottleneck was the displacement estimation algorithm, namely the SVD rank one decomposition; supplementary services, such as capturing images, communication, etc., did not require much processing capabilities.

**Figure 30 – An instant of the real-time operation. A video of the entire operation is available on YouTube.**



Source: Adapted from (KALID, 2025).

The real-time operation was generally smooth, but only under controlled conditions, such as: carefully controlled movement (slow speed, low acceleration, and low vibration), short operation periods due to LED overheating, and the use of carefully chosen specimens (some specimens had a reflective finish that led to saturated regions in the captured image).

#### 4.4 Commentaries

In this chapter, experimental results were presented by applying the methodology proposed in Chapter 2 and 3. Applying the adjustments proposed by the adjustment procedure, accurate displacement estimates were obtained with an error upper bound equal to 2.80 %. Additionally, the methodology to reconstruct 3D cylindrical trajectory by incorporating IMU data into a 2D displacement was successful. However, the lightning and optical problems in the scene were shown to possibly interfere more than expected in the final estimate, inducing a disparity between the x-axis and y-axis displacement accuracy.

When assembled together with the US acquisition system and visual odometer, the ensemble worked together. The US data was properly acquired, but the visual odometer was too sensitive to environmental and operational factors.

## 5 CONCLUSIONS

In this work, a method for visual odometry was proposed. As most NDT routines involve transducer movement during inspection, an odometry system is a crucial part of the acquisition system. The proposed visual odometry system offers an alternative to the mechanical odometry counterparts, especially for underwater inspection that involves 2D or 3D movements.

The algorithm underlying the visual odometer was based on the method proposed by (HOGE, 2003) and compared pairs of images captured by a monocular camera to estimate a displacement 2D between them. Unlike most visual odometry methods, the one proposed in this work employs a ground-facing camera, which, for most inspections, is a suitable decision as it makes the lighting and estimation process easier.

A set of experiments were conducted to test the proposed system's validity. The proposed solution could estimate 2D trajectories with errors up to 2.80 % of the total displacement considering different experimental setups. After fusing the 2D displacement with IMU data, a 3D trajectory was successfully obtained, proving that the system could be, in the future, able to estimate in real-time 3D trajectories as well.

Furthermore, the visual odometer was demonstrated to work as part of a US inspection ensemble. Properly firing the US acquisition system, the visual odometry system showed good responsiveness. However, when freehand trajectories were performed under an experimental setup much closer to what a field inspection would look like, the proposed visual odometry system showed potential improvements points to make the system more robust, reliable, and accurate.

Therefore, although the visual odometry system showed some practical limitations, it proved to be a valid alternative to conventional mechanical odometers in underwater US inspections.

### 5.1 Future Works

Throughout the development of this work, some improvements were identified that fall outside the current scope but would significantly enhance the reliability, accuracy, limitations, and metrological validity of the results. Notable improvements include:

- Selecting an optical apparatus, i.e., camera and lenses, suitable for underwater use;
- Designing a mechanical housing capable of withstanding high-pressure environments for subsea inspections;
- Enhancing the lighting system to dynamically adjust based on the captured image;
- Implementing an autofocus mechanism (possibly with new electronic and mechanical parts) to increase the operational range between the camera and specimen;

- Integrating an additional monocular camera to estimate depth, i.e., the distance between the acrylic window and the specimen. This could enhance the system's robustness to unpredictable changes in the operational distance;
- Implement noise robustness feature to the phase unwrapping algorithm. Instead of using a hard-threshold approach, it is possible to identify wrapping points through outlier detection algorithms, such as RANdom SAmple Consensus (RANSAC));
- Developing a standardized procedure to evaluate system repeatability. This would include, for instance, an automated robotic arm to move the visual odometer along the trajectories;
- And along with the last point, evaluating the visual odometer's performance as a measurement tool from a metrological standpoint. This would include calibrating the system using a standardized method (in line with guidelines from institutions like the International Bureau of Weights and Measures [BIPM]); assessing displacement uncertainty; and identifying system components that contribute most to measurement uncertainty.

## REFERENCES

- AQEL, M. O. *et al.* Review of visual odometry: types, approaches, challenges, and applications. **SpringerPlus**, Springer, v. 5, p. 1–26, 2016.
- BALCI, M.; FOROOSH, H. Subpixel estimation of shifts directly in the fourier domain. **IEEE Transactions on Image Processing**, IEEE, v. 15, n. 7, p. 1965–1972, 2006.
- BERNARDES, E.; VIOLET, S. Quaternion to euler angles conversion: A direct, general and computationally efficient method. **Plos one**, Public Library of Science San Francisco, CA USA, v. 17, n. 11, p. e0276302, 2022.
- CARVALHO, A. *et al.* Reliability of non-destructive test techniques in the inspection of pipelines used in the oil industry. **International journal of pressure vessels and piping**, Elsevier, v. 85, n. 11, p. 745–751, 2008.
- CASTLEMAN, K. R. **Digital image processing**. [S.l.]: Prentice Hall Press, 1996.
- CHEN, H.-M.; VARSHNEY, P. K.; ARORA, M. K. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. **IEEE transactions on Geoscience and Remote Sensing**, IEEE, v. 41, n. 11, p. 2445–2454, 2003.
- Dezhi Zheng *et al.* A Capacitive Rotary Encoder Based on Quadrature Modulation and Demodulation. **IEEE Transactions on Instrumentation and Measurement**, v. 64, n. 1, p. 143–153, jan. 2015. ISSN 0018-9456, 1557-9662. Available from Internet: <http://ieeexplore.ieee.org/document/6840345/>.
- DIGITAL, U. **What is Quadrature?** 2020. <https://www.usdigital.com/blog/what-is-quadrature/> [Accessed: 15/06/2024].
- ECKART, C.; YOUNG, G. The approximation of one matrix by another of lower rank. **Psychometrika**, Springer, v. 1, n. 3, p. 211–218, 1936.
- EVERTON, S. K. *et al.* Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing. **Materials & Design**, Elsevier, v. 95, p. 431–445, 2016.
- FADNAVIS, S. Image interpolation techniques in digital image processing: an overview. **International Journal of Engineering Research and Applications**, v. 4, n. 10, p. 70–73, 2014.
- GEIGER, A.; LENZ, P.; URTASUN, R. Are we ready for autonomous driving? the KITTI vision benchmark suite. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 3354–3361, 2012. ISSN 10636919.
- GILLES, M.; IBRAHIMPASIC, S. Unsupervised deep learning based ego motion estimation with a downward facing camera. **Visual Computer**, Springer Science and Business Media Deutschland GmbH, v. 39, n. 3, p. 785–798, 3 2021. ISSN 01782789. Available from Internet: <https://link.springer.com/article/10.1007/s00371-021-02345-6>.
- GLORIA, J. R. E. **Runway detection from map, video and aircraft navigational data**. 2016. Thesis (PhD) — Monterey, California: Naval Postgraduate School, 2016.
- GRAY, E. **What Is Focal Length in Photography? A Beginner's Guide**. 2023. <https://photographylife.com/what-is-focal-length-in-photography> [Accessed: 12/06/2024].

- HE, M. *et al.* A review of monocular visual odometry. **The Visual Computer**, Springer, v. 36, n. 5, p. 1053–1065, 2020.
- HELLIER, C. *et al.* **Handbook of nondestructive evaluation.** [S.l.], 2003.
- HO, M. *et al.* Inspection and monitoring systems subsea pipelines: A review paper. **Structural Health Monitoring**, SAGE Publications Sage UK: London, England, v. 19, n. 2, p. 606–645, 2020.
- HOGE, W. A subspace identification extension to the phase correlation method [mri application]. **IEEE transactions on medical imaging**, IEEE, v. 22, n. 2, p. 277–280, 2003.
- ITOH, K. Analysis of the phase unwrapping algorithm. **Applied optics**, Optica Publishing Group, v. 21, n. 14, p. 2470–2470, 1982.
- JUNIOR, E. T. **Contributions to Visual Odometry Applied to Non-Destructive Testing.** 2023. Thesis (Masters) — Universidade Tecnológica Federal do Paraná, 2023.
- K, B. **Spectral leakage caused by windowing.** 2018. [https://commons.wikimedia.org/wiki/File:Spectral\\_leakage\\_caused\\_by\\_%22windowing%22.svg](https://commons.wikimedia.org/wiki/File:Spectral_leakage_caused_by_%22windowing%22.svg) [Accessed 23-10-2024].
- KALID, T. E. **Ultrasound inspection with visual odometer.** 2025. <https://youtu.be/NhYWLDUWTBM> [Accessed 12-01-2025].
- KALID, T. E. *et al.* Virtual encoder: a two-dimension visual odometer for ndt. **in] Proceedings of the 13th ECNDT**, v. 2023, 2023.
- KAZIK, T.; GOKTOGAN, A. H. Visual odometry based on the Fourier-Mellin transform for a rover using a monocular ground-facing camera. **2011 IEEE International Conference on Mechatronics, ICM 2011 - Proceedings**, p. 469–474, 2011.
- KIRKPATRICK, L.; FRANCIS, G. **Physics: A World View.** Cengage Learning, 2006. ISBN 9780495010883. Available from Internet: <https://books.google.com.br/books?id=T8XFwAEACAAJ>.
- KOEPPIK. **Different apertures of a lens.** 2019. [https://en.wikipedia.org/wiki/Aperture#/media/File:Lenses\\_with\\_different\\_apertures.jpg](https://en.wikipedia.org/wiki/Aperture#/media/File:Lenses_with_different_apertures.jpg) [Accessed 15-06-2024].
- KUMAR, N. K.; SCHNEIDER, J. Literature survey on low rank approximation of matrices. **Linear and Multilinear Algebra**, Taylor & Francis, v. 65, n. 11, p. 2212–2244, 2017.
- KYBIC, J.; UNSER, M. Fast parametric elastic image registration. **IEEE transactions on image processing**, IEEE, v. 12, n. 11, p. 1427–1442, 2003.
- LEI, H. *et al.* Ultrasonic pig for submarine oil pipeline corrosion inspection. **Russian Journal of Nondestructive Testing**, Springer, v. 45, p. 285–291, 2009.
- LEWIS, J. P. *et al.* Fast template matching. *In:* QUEBEC CITY, QC, CANADA. **Vision interface.** [S.l.], 1995. v. 95, n. 120123, p. 15–19.
- MAIERHOFER, C. Nondestructive evaluation of concrete infrastructure with ground penetrating radar. **Journal of materials in civil engineering**, American Society of Civil Engineers, v. 15, n. 3, p. 287–297, 2003.
- MAIMONE, M.; CHENG, Y.; MATTHIES, L. Two years of Visual Odometry on the Mars Exploration Rovers. **Journal of Field Robotics**, John Wiley & Sons, Ltd, v. 24, n. 3, p. 169–186, 3 2007. ISSN 1556-4967.

- MALCOLM, D. *et al.* Strain measurement in biaxially loaded inhomogeneous, anisotropic elastic membranes. **Biomechanics and Modeling in Mechanobiology**, Springer, v. 1, p. 197–210, 2002.
- MANOLAKIS, D. G.; INGLE, V. K. **Applied digital signal processing: theory and practice.** [S.l.]: Cambridge university press, 2011.
- MANSUR, V. *et al.* Deploying complementary filter to avert gimbal lock in drones using quaternion angles. In: IEEE. **2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON).** [S.l.], 2020. p. 751–756.
- MEILLAND, M.; DRUMMOND, T.; COMPORT, A. I. A unified rolling shutter and motion blur model for 3d visual registration. In: **Proceedings of the IEEE International Conference on Computer Vision.** [S.l.: s.n.], 2013. p. 2016–2023.
- MIKERUN. **Depth-of-field.** 2019. <https://commons.wikimedia.org/wiki/File:Depth-of-field.svg> [Accessed 23-06-2024].
- NISTÉR, D.; NARODITSKY, O.; BERGEN, J. Visual odometry. In: IEEE. **Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.** [S.l.], 2004. v. 1, p. I–I.
- NISTÉR, D.; NARODITSKY, O.; BERGEN, J. Visual odometry for ground vehicle applications. **Journal of Field Robotics**, Wiley Online Library, v. 23, n. 1, p. 3–20, 2006.
- NIVEN, I. The roots of a quaternion. **The American Mathematical Monthly**, Taylor & Francis, v. 49, n. 6, p. 386–388, 1942.
- NOURANI-VATANI, N.; BORGES, P. V. K. Correlation-based visual odometry for ground vehicles. **Journal of Field Robotics**, John Wiley & Sons, Ltd, v. 28, n. 5, p. 742–768, 9 2011. ISSN 1556-4967. Available from Internet: <https://doi.org/10.1002/rob.20407>.
- OLYMPUS. **Scanners and Accessories.** 2014. [https://www.olympus-ims.com/pt/scanners/mini-wheel/#!cms\[focus\]=cmsContent10296](https://www.olympus-ims.com/pt/scanners/mini-wheel/#!cms[focus]=cmsContent10296) [Accessed: 10/06/2024].
- PETERSON, L. L.; DAVIE, B. S. **Computer networks: a systems approach.** [S.l.]: Morgan Kaufmann, 2007.
- PI, R. **Raspberry Pi Camera Module 2 NoIR.** 2016. <https://www.raspberrypi.com/products/pi-noir-camera-v2/> [Accessed 12-01-2025].
- PUYMBROECK, N. V. *et al.* Measuring earthquakes from optical satellite images. **Applied Optics**, Optica Publishing Group, v. 39, n. 20, p. 3486–3494, 2000.
- REINA, G. *et al.* Wheel slippage and sinkage detection for planetary rovers. **IEEE/Asme Transactions on Mechatronics**, IEEE, v. 11, n. 2, p. 185–195, 2006.
- RIEDER, H. *et al.* Online monitoring of additive manufacturing processes using ultrasound. In: **Proceedings of the 11th European Conference on Non-destructive testing.** [S.l.: s.n.], 2014. v. 1, p. 2194–2201.
- SCHMERR, L. W. **Fundamentals of ultrasonic nondestructive evaluation.** [S.l.]: Springer, 2016. v. 122.
- STONE, H. S. *et al.* A fast direct fourier-based algorithm for subpixel registration of images. **IEEE Transactions on geoscience and remote sensing**, IEEE, v. 39, n. 10, p. 2235–2243, 2001.

- STRANG, G. **Introduction to linear algebra.** [S.l.]: SIAM, 2022.
- TONG, X. *et al.* A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 53, n. 8, p. 4143–4156, 2015.
- TONG, X. *et al.* Image registration with fourier-based image correlation: A comprehensive review of developments and applications. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, IEEE, v. 12, n. 10, p. 4062–4081, 2019.
- VIRTANEN, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. **Nature Methods**, v. 17, p. 261–272, 2020.
- XIA, S. *et al.* A Survey on Human Performance Capture and Animation. **Journal of Computer Science and Technology**, Springer New York LLC, v. 32, n. 3, p. 536–554, 5 2017. ISSN 18604749.
- XU, M.; VARSHNEY, P. K.; NIU, R. A subspace method for fourier based image registration. In: IEEE. **2006 Fortieth Asilomar Conference on Signals, Systems and Computers.** [S.l.], 2006. p. 425–429.
- ZETEC. **Manual 2D Encoded Scanner Designed for Both Flat and Curved Surfaces.** 2019. <https://www.zetec.com/wp-content/uploads/2019/08/NDT-Sweeper-Datasheet.pdf> [Accessed: 12/06/2024].
- ZHOU, J. *et al.* Underwater image enhancement method with light scattering characteristics. **Computers and Electrical Engineering**, Pergamon, v. 100, p. 107898, 5 2022. ISSN 0045-7906.
- ZITOVA, B.; FLUSSER, J. Image registration methods: a survey. **Image and vision computing**, Elsevier, v. 21, n. 11, p. 977–1000, 2003.