Development of non-intrusive solutions for identification (THID) and authentication in the THz domain

Dragos-Florin Nastasiu

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE SAVOIE MONT BLANC

préparée dans le cadre d'une cotutelle entre l'*Université Savoie Mont Blanc* et l'Académie Technique Militaire "Ferdinand I" de Bucarest

Spécialité : **Signal, Image, Parole, Télécoms**

Arrêté ministériel : le 25 mai 2016

Présentée par

Dragoș NASTASIU

Thèse dirigée par **Frédéric GARET**, **Cornel IOANA** et **Alexandru ȘERBĂNESCU**
préparée au sein des **Centre de Radiofréquences, Optique et Micro-nanoélectronique des Alpes (CROMA), Grenoble Images Parole Signal Automatique (GIPSA)** et **Département des Communications et des Technologies de l'Information**
dans **les Écoles Doctorales d'Electronique, Electrotechnique, Automatique et Traitement du Signal, respectivement Ingénierie des Systèmes de Défense et de Sécurité**

Développement de solutions pour l'identification (THID) et l'authentification par des approches non intrusives dans le domaine THz

Thèse soutenue publiquement le **29/05/2024**,
devant le jury composé de :

Frédéric GARET
Professor HDR, Université Savoie Mont Blanc, France (Directeur de thèse)
Cornel IOANA
Maitre de conférences HDR, Université Grenoble Alpes, France (Directeur de thèse)
Alexandru ȘERBĂNESCU
Professeur, Académie Technique Militaire "Ferdinand I", Roumanie (Directeur de thèse)
Angela DIGULESCU-POPESCU
Maitre de conférences, Académie Technique Militaire "Ferdinand I", Roumanie (Rapporteur)
Andrei ANGHEL
Professeur, University of Science and Technology POLITEHNICA, Roumanie (Rapporteur)
Patrick Mounaix
Rapporteur, IMS Bordeaux University (Rapporteur)
Emmanuel TROUVE
Professeur, LISTIC – Polytech, France (Examinateur)
Sri Krishnan
Professeur, Faculty of Engineering & Architectural Science Toronto, Canada (Examinateur)
Alexandre Locquet
Adjunct Assistant Professor, Georgia Tech Lorraine-CNRS Laboratory, France (Examinateur)
Maxime BERRIER
Maitre de conférences, Université Savoie Mont Blanc, France (Examinateur)

T H E S I S

Development of novel non-invasive methods for identification and authentication in THz domain

Presented and defended by

Dragoș NASTASIU

for jointly obtaining the

DOCTORATE DEGREE

of University of Savoie Mont Blanc
Doctoral School for Electronics, Power Systems, Automatic Control and Signal Processing
Specialization: Signal, Image, Speech, Telecommunications

and the
PHD DEGREE IN TECHNICAL SCIENCES

of Military Technical Academy "Ferdinand I" of Bucharest
Faculty of Communications and Electronic Systems for Defense and Security
Specialization: Electronic Engineering, Telecommunications and Information Technologies

Thesis directed by Frédéric GARET, Cornel IOANA and Alexandru ȘERBĂNESCU.

Prepared in Centre de Radiofréquences, Optique et Micro-nanoélectronique des Alpes (CROMA), Grenoble Image Parole Signal Automatique laboratory (GIPSA-lab) and at the Faculty of Military Electronic and Informatics Systems.

Defended in Chambery, on 29.05.2024, in front of the jury:

Reviewers:	*Professor* Patrick MOUNAIX
	Professor Andrei ANGHEL
Examiners:	*Professor* Frédéric GARET
	Professor Alexandru ȘERBĂNESCU
	Professor Emmanuel TROUVE
	Professor Sri KRISHNAN
	Professor Alexandre LOCQUET
	Associate Professor Cornel IOANA
	Associate Professor Maxime BERNIER
Invited:	*Associated Professor* Angela DIGULESCU

Acknowledgements

I would like to express my deepest gratitude to the esteemed members of my thesis jury for their invaluable contributions and support. My heartfelt thanks go to the reviewers, Professor Patrick Mounaix and Professor Andrei Anghel, whose expert evaluations and feedback have greatly enhanced the quality of my work.

I am also profoundly grateful to my thesis directors, Professor Frédéric Garet, Professor Alexandru Șerbănescu and Associate Professor Cornel Ioana. Your guidance, insights and unwavering support have been essential in shaping my research and bringing this project to fruition.

Additionally, I would like to thank the examiners, Professor Emmanuel Trouve, Professor Sri Krishnan, Professor Alexandre Locquet and Associate Professor Maxime Bernier, for their constructive critiques. Your detailed feedback and constructive comments have been crucial in refining and improving my research.

I am thankful to Associate Professor Angela Digulescu for her participation as an invited committee member and for her contributing valuable perspectives to this project. Your expertise and support have been instrumental in the completion of my research and I am deeply appreciative of your time and effort.

I want to express my profound appreciation and love to Mădălina, who stood by me through the toughest times, offering me endless support and appreciation. Your care and constant love gave me the strength to keep going and inspired me to find solutions to even the most difficult problems.

I also want to thank my family, who has continuously encouraged and believed in me even in my most challenging times.

For my colleagues and professors from MTA, I thank you for your unwavering support and productive discussion which transposed into meaningful and interesting ideas: Popescu Florin, Despina Cristina, Stănescu Denis, Rîncu Iulian and Enache Florin.

I would also like to extend my sincere appreciation to the CROMA laboratory for their warm welcome. I am particularly thankful to Jean François Roux, Pierre-Baptiste Vigneron, and Emilie Herault, who made me feel like one of their own.

Lastly, I want address special thanks to all my close friends who stayed with me late-nights to discuss and helped me take a break from everyday worries: Scripcaru Răzvan, Antonescu Răzvan, Stan Dragoș, and Cîrîc Iulian.

In conclusion, I am deeply grateful to everyone who has supported me throughout this journey. Your encouragement and assistance have been invaluable. Thank you for being a part of my story.

Contents

List of figures

List of tables

List of acronyms and abbreviations

ACT – Activation
AGG – Aggregation
CPU – Central processing unit
CWI – Continuous Wave imaging
DC – Direct current
DCT – Discrete Cosine transform
DEKM – Deep embedded k-mean
DL – Deep Learning
DTW – Dynamic time warping
DWT – Discrete wavelet transform
EM – Electromagnetic
FFI – Fast Fourier image
FFT – Fast Fourier transform
FNN – False nearest neighbor
FSDI – Frequency Slide Difference Image
FSII – Frequency Slice Integration Image
GGC – General graph convolution
GNN – Graph neural network
GPU – Graphical processing unit
IFFT – Inverse fast Fourier transform
IR – Infrared
k-NN – k-nearest neighbor
LTG-GaAs - low-temperature grown gallium arsenide
MF – Match filter
ML - Machine Learning
MLP – Multilayer perceptron
MPI – Maximum Peak Image
mPI – Minimum Peak Image
MRA – Multi resolution analysis
MSE – Mean squared error
MW – Microwave
NN – Neural network
ODL – Opical delay lines
PD-MFI – Phase diagram match filter image
PD-DCI – Phase diagram dynamic characterized image
PD-CMI – Phase diagram convolution magnitude image
PD-IWI – Phase diagram independent warping image
PD-DWI – Phase diagram dependent warping image
pSNR – Peak signal-to-noise ratio
pReLU – Parametric rectified linear un
ReLu – Rectified linear unit
it
RI-WPD – Rotation invariant wavelet packet decomposition
ROI – Region of interest
SNR – Signal-to-noise ratio
SSIM – Structural similarity index measure
SVM – Support vector machine
THz – Terahertz
THz-TDS – Terahertz Time domain spectroscopy
TOF - Time-of-flight
TPI – Terahertz pulsed imaging
fs – Femtosecond
TSI – Time Slice Image
TSDI – Time Slice Difference Image
TSI – Time Slice Image
TSII – Time Slice Integration Image
TOFI – Time-of-flight Image
TI-WPD – Translation invariant wavelet packet decomposition
WT – Wavelet transform
WPD – Wavelet packet decomposition

Chapter 1. Introduction

In many imaging systems, various research directions can be pursued to enhance the overall system's outcomes. In the relatively new THz technology, we can identify several challenges associated with the preprocessing step, but the focus is on image displaying methods. In this thesis, significant attention is dedicated to investigating amplitude estimation techniques within ideal and noisy conditions [1]. In all imaging systems, regardless of their technology, when the estimation of the received signal is poor, the resulting image has low quality. Thus, novel image displaying methods are introduced to increase the capacity to estimate the amplitude at reception.

On the other hand, concerning computer vision tasks such as image characterization and classification, the subjects addressed in this thesis envision image classification under view-point variation constraints. Features derived from an image subject to translation or rotation differ from those originating from the original image. The highly probable consequence of this aspect is an erroneous classification of the patterns. In this context, the concept aims to establish a singular decomposition method for images sharing the same pattern but undergoing rotation or translation. Afterwards, an essential step, developed in this thesis, is to introduce novel discriminative feature vectors and to train a variety of machine learning (ML) and deep learning (DL) classifiers to analyze their capacity in classification tasks.

1.1 THz technology

The aim of this section is to underscore the significance of the increasingly emerging and advancing THz technology. [2]. Terahertz (THz) imaging marks a novel advancement in diagnostic and sensing technologies, rapidly expanding in scope owing to its benefits and recent progress in the efficient generation, manipulation, and detection of THz signals [3]. THz waves occupy the quasi-optical domain, situated between the microwave (MW) and infrared (IR) regions within the electromagnetic (EM) spectrum, typically around 1 THz (10^{12} Hz). Given the absence of a rigorously defined frequency band, numerous researchers opt to specify the range based on the specific properties of the samples and materials analyzed via THz radiation. As it is depicted in Figure 1-1, the frequency band is limited within 0.3 – 3 THz, but can be extended in both sides without exceeding 0.1 – 10 THz [4].

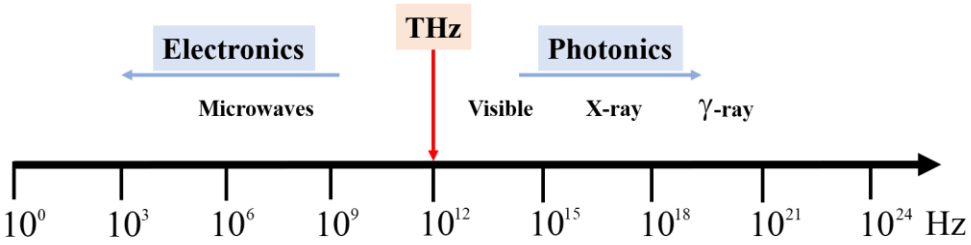


Figure 1-1. The electromagnetic spectrum

THz radiation offers a variety of unique spectroscopic characteristics that motivates the actual scientific interest and development of the THz sensing as a powerful and complementary technique of inspection [3]. It allows for a multidirectional analysis by providing complementary information to other sensing technologies such as X-ray computed tomography scans, ultrasounds, microwaves, etc. Using THz radiation to create an image brings a variety of advantages which are

described in the following paragraph. Due to the shorter wavelength compared with waves at lower frequencies (ultrasounds, microwaves, etc.), a THz image has greater spatial resolution. Many common materials such as paper, fabrics, cardboard, ceramics, and plastic composites are relatively transparent to THz radiation, motivating its use as a contactless and nonintrusive diagnostic instrument in many applications [5], [6], [7], [8], [9], [10]. Another important property of THz radiation is its sensitivity to chemical and structural information [11]. Moreover, THz waves interact with materials in several ways, including reflection, transmission, absorption, and scattering. These interactions depend on the material's dielectric properties, which are the refractive index and absorption coefficient. As THz waves are sensitive to changes in these properties, it is possible to detect subtle changes in the chemical composition and structure of a material. Typically, THz applications leverage fundamental characteristics of THz radiation, including its non-ionizing nature, the adequate transparency of materials, and the distinct spectral signatures exhibited by many chemical compounds within the THz range. These characteristics offer significant contrast in images featuring materials with high electrical conductivity (e.g., metals) or substantial static dipoles (e.g., water), as a result of the pronounced absorption or reflection of radiation in these mediums. THz imaging is versatile, enabling the generation of diverse images from the same measured data, including panchromatic-absorption images, refractive-index images, and time-of-flight (TOF) images.

The specific contrast of THz radiation, associated with the increased spatial resolution, opens new perspectives in the imaging framework. In the biological diagnosis domain, the strong absorption of THz in water molecules plays a major role by limiting the penetration depth of the THz rays. Taking this into account, the research focuses on surface inspection, such as skin conditions and cancer diagnosis [12], [13], [14], [15], teeth imagery and identification of dental caries [16]. The THz radiation has a great potential in assuring a higher quality control of production lines by allowing for an in-depth analysis of the integrity and properties of an object: paint thickness analysis and drying time [17], impact control at the surface of steel sheets [18], on site measurement of additives in polymers [19]; Another sector where THz technologies are successfully applied is represented by the preventive maintenance. The possibility of a volumetric and nonintrusive analysis lead to a lot of opportunities in this area: detection of air bubbles in foams [20], detection of paint corrosion points for vehicles and airplanes [21], detection of fissures under a plaster [22]. Applications of security and defense can also take advantage of the features of the THz radiation such as the security of the control points (check-points) in airports, where the detection of a dangerous individual is imperative to be done with ease [23]; detection of explosives, and drugs [24]. Another interesting branch for THz radiation is the field of telecommunications where it can take advantage of both optical and electronic characteristics of the waves. An important application is the usage of higher frequency carrier signals to increase the transfer rate of two close terminals up to 20 GB/s [25], [26].

1.2 THz imaging

The THz domain remains one of the least explored areas within the electromagnetic spectrum, thus the knowledge gap is even more reflected in the imaging domain. Depending on the technology of the THz sources and detectors, we can divide the imaging techniques based on the type of generated signal. Therefore, a THz system can create an image based on pulses or a continuous wave:

- THz pulsed imaging (TPI)
- Continuous wave THz imaging (CWI)

A continuous wave approach detects the intensity distribution of the scattering effect on the edges of the motives on the samples by using a narrow frequency band, thus a CW system

generally provides information about the transmitted energy [27]. On the other hand, a pulsed imaging system considers the interactions between the emitted signal and the sample through which it propagates. These interactions envision pulse reflections, amplitude attenuation, phase and waveform deformations, etc., thus providing more information about the analyzed sample such as: electric field magnitude, depth of scattering centers, optical path length (transmission time), absorption spectrum, phase and pulse shape [27]. On average, in terms of power, a continuous wave system usually outputs higher energy compared to the pulsed imaging technique though, there are pulse sources that can generate THz signals having a power higher than 14 mW [28].

Furthermore, due to the commonly adopted electro-optic sampling method [29] and photoconductive antenna technology [30] used to detect THz radiation in TPI systems, their structures are complex, but with advantages such as: improved signal-to-noise ratio (SNR) and larger bandwidth of waveforms, usually up to 5 THz [31], [32]. However, CW THz imaging systems are compact and can provide real time analysis capabilities without additional pump optical detection elements [33].

1.3 THz pulsed imaging

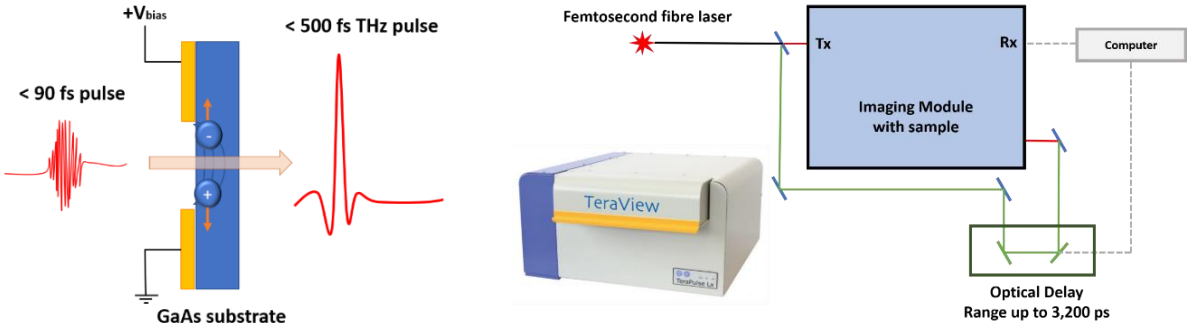
In this thesis, THz pulsed imaging is our choice of image creation. The measurements are conducted with a TeraPulse Lx system from TeraView, Ltd., a bench top THz transmission spectrometer that consist of the following components: an integrated PC to control the system; an external laptop which enables the user to control the TeraPulse Lx equipment (remote heads) and manipulates the THz analysis data via the TeraPulse Software; an enclosed optical system, an enclosed electronics system and a THz focused beam.

The generation of THz pulsed radiation is based on a photoconductive switch [34], [35], [36]. These photoconductive emitters rely on creating THz pulses using the femtosecond laser pulse (below 90 fs) to excite a biased low-temperature grown gallium arsenide substrate (LTG-GaAs). The process of generating a THz pulse is presented in Figure 1-2 (a). The emitted pulse (below 500 fs) has the power distributed over a frequency range spreading from 0.1 THz and up to 6 THz. THz pulse is radiated when the electrical dipole moment of electron/hole pairs, photogenerated in LTG-GaAs by the femtosecond laser pulse is modulated on sub picosecond timescales due to bias voltage applied in-between electrodes. The change in current density, and therefore photocurrent, comes from two processes:

1. The femtosecond laser illumination that results in a rapid change of the carrier density

2. The acceleration of photo-generated carriers under the influence of an external electric fields

Coherent detection of the THz radiation uses a similar photoconductive antenna circuit. With the help of the femtosecond pulse synchronized to the THz emission, the photoconductive gap is gated and a current proportional to the THz electric field is measured. The time domain of the THz signal can be sampled by varying the optical path length to the receiver. In this manner, both the amplitude and the phase of the incident pulse is obtained with a dynamic range (DR) greater than 95 dB.



(a) (b)

Figure 1-2. (a) THz pulse generation; **(b)** THz system with all its components: femtosecond laser, imaging module with sample, optical delay line and computer.

In general, most of the THz systems work in transmission or reflection mode [37]. The TeraPulse Lx system has a Transmission and Reflection Imaging Module (TRIM) that allows to use both types of imaging techniques by interchanging elements. The diagram of the system and its modularity feature are presented in Figure 1-2 (b). During the thesis, the transmission mode is used to acquire the raw dataset from which the image will be constructed. The working principle of the transmission mode is shown in Figure 1-3. The object to be scanned is placed on a 2D movable holder at the focal plane of the incident THz pulse. After traveling through the sample, the transmitted signal is acquired by the receiver. Both the transmitter and the receiver are required to be positioned perpendicularly to the sample under test in order to increase the sensitivity at reception.

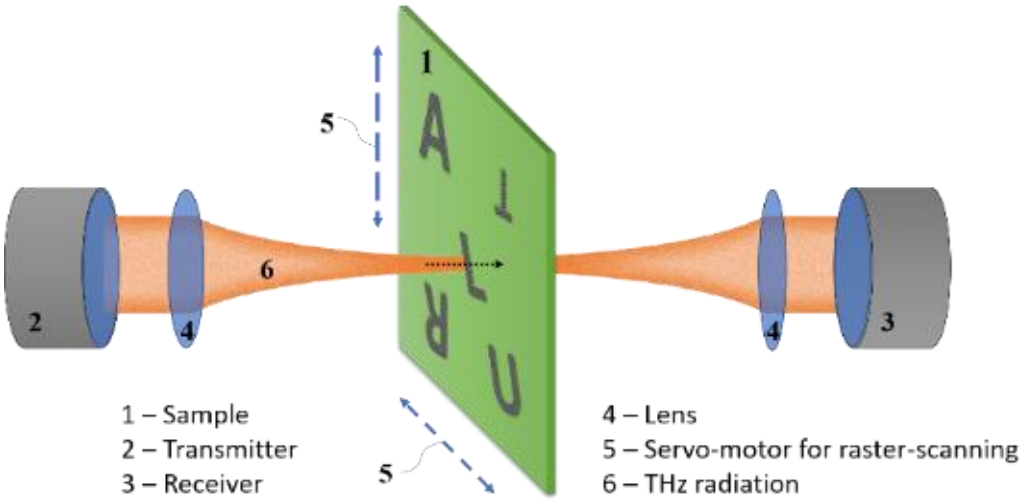


Figure 1-3. Transmission Mode acquisition principle

The samples consist of a polyethylene or paper substrate with a variety of metallic ink shapes (stars, squares, circles and hexagons) at different scales printed on it. The TeraPulse Lx system is configured to raster-scan in transmission mode an area of 300 mm-by-300 mm of the sample with a spatial resolution of 0.3 mm or 0.1 mm. The spatial resolution is the capacity to distinguish two close patterns. In far-field, the THz wavelength is approximately 300 micrometers at 1 THz. The physical distance between the THz emitter and receiver is 10 cm, adjusted to have a maximum amplitude signal when measured with no sample in-between. After the scan, the

acquired measurements are recorded in a 3D matrix, where the first two dimensions corresponds to the coordinates of each pixel of the image. The time-domain waveform of each of these pixels is stored in the third dimension over 2048 temporal points corresponding to 50ps. Figure 1-4 shows some of the samples analyzed and used in this thesis. The THz images of the samples are displayed using 1 THz frequency.

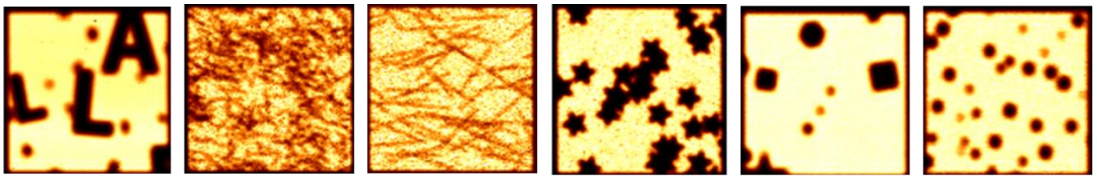


Figure 1-4. Overview of some samples used in our study and imaged using THz pulsed imaging systems at 1 THz.

1.4 Challenges and limitations in THz imagery

This part presents the challenges tackled in this thesis by providing brief descriptions about their importance and why is it necessary to overcome them.

1.4.1 Image reconstruction under ideal and noise conditions

Image reconstruction from 3-dimensional raw data is a challenging problem in many fields including medical imaging or material science [38], [39], [40]. The reconstruction process uses mathematical algorithms to convert the acquired THz data into an image of an object or scene. In most of the cases, regardless the technology of the imaging system (acoustic, radar, X-ray, THz, etc.), the challenge is to estimate the amplitude of the received signal and to use that value to reconstruct the image [41]. These estimates allow for a representation of the intensity of an image, as it corresponds to the brightness of the sample being imaged. Accurate amplitude estimation is essential as it enables high-quality images with high contrast, otherwise the resulting images may be too dark or too bright, making it difficult to discern details or identify objects within the image. Secondly, the amplitude estimation is critical for the detection and analysis of subtle changes in the image. This is particularly important in medical imaging, where even small changes in tissue density or structure can indicate disease or injury [42], [43]. Lastly, a robust amplitude estimation is important to calibrate and normalize the imaging system to ensure consistent measurements across different images and imaging sessions.

1.4.2 View-point variation in Computer Vision-related tasks

View-point variation is a challenge in computer vision-related tasks, such as object recognition, detection, and pose estimation [44], [45], [46], [47]. It refers to the changes in the position, orientation, or perspective of an object or scene relative to the sensors, which can significantly affect the appearance of the object in images. As a result, the task of recognizing or detecting the object becomes difficult for computer vision algorithms, as the same object may have different shapes if observed from different viewpoints. This variation of the observation angle can be caused by various factors, such as the relative positions of the object and sensors, the angle of the camera or sensor, the lighting conditions, etc. In addition, the complexity of the object's shape, texture, and color can also affect the degree of view-point variation. In THz imaging, view-point

variation can be caused by changes in the angle of incidence of the THz radiation or changes in the relative positions of the object and the THz source and detector.

Figure 1-5 shows how three different images of the same pattern, but which is rotated with 0, 45 and 180 degrees, have distinctive feature vectors. In this case, addressing view-point variation is a critical challenge in computer vision as it is essential for real-world applications. For instance, autonomous vehicles must be able to recognize and track other vehicles and pedestrians from different viewpoints to ensure safe driving. Similarly, object recognition and tracking are essential for medical, surveillance, robotics, and augmented reality applications. In THz imagery, researchers have developed several methods to address the view-point variation challenge, including data augmentation [48] and feature-based approaches [49]. However, this challenge remains an ongoing area of research, and there is still much work to be done to improve the robustness and accuracy of computer vision algorithms under varying viewpoints.

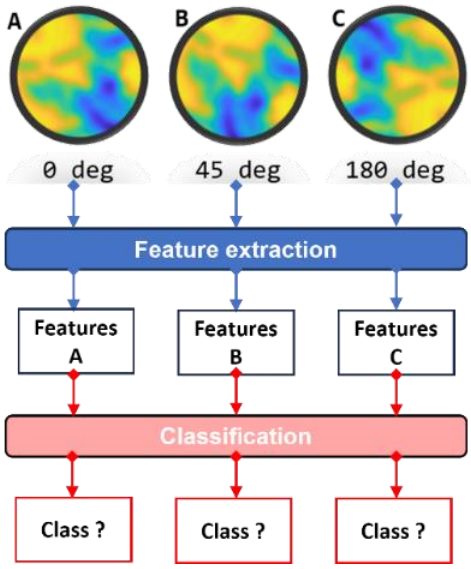


Figure 1-5. View-point variation challenge in computer vision

1.4.3 Thesis outline

Chapter 2 presents the state-of-the art techniques used for THz image reconstruction, their limitations and therefore, the need for new techniques to display the image.

Chapter 3 focuses on a modern signal representation space, the phase diagram, in order to define novel methods which, improve the quality of the reconstructed image. The quality assessment is achieved using different full- and no- reference quality metrics.

Chapter 4 describes the means of tackling characterization and classification tasks under invariance constraints. Before detailing the novelties, the chapter focuses on presenting some state-of-the-art techniques used in nowadays imaging systems and how they can be used as a starting point to develop two feature extraction techniques. These techniques are based on invariant Wavelet Packet Decomposition (WPD) and entropy as the mean to characterize the images.

Chapter 5 focuses on the classification task by testing the proposed methods from Chapter 3 on publicly available image datasets, but also on a dataset of THz images created in the laboratory. The chapter also presents a comparison between different approaches from the literature and assesses our methods in this regard.

Finally, Chapter 6 concludes the thesis by synthesizing the main contributions and the perspectives for future research directions.

Chapter 2. State-of-the-art image reconstruction

2.1 Introduction

In the THz imaging framework, the acquisition setup has the most important role. Its features, limitations, and a priori configuration, determine the quality of the measured data. In this context, a critical aspect concerns the data processing methods employed for reconstructing the image from raw data. These factors collectively contribute to the overall precision and integrity of the THz imaging system, consequently influencing the ultimate diagnostic outcome. In most of pulsed THz imaging systems, the resulting image pixels contain time-domain waveforms recovered by a raster-scanning (point-by-point) method.

The problem of THz image reconstruction is to estimate the amplitude of the received signal [50]. A primary constraint of THz imaging pertains to the signal-to-noise ratio (SNR), underscoring the significance of this factor when utilizing THz systems. The image's quality significantly relies on the performance of the THz system and the operational conditions, which are not always perfectly controllable, unlike in a research laboratory setting. Consequently, certain disturbances, such as those induced by low-performance imaging systems or critical measurement conditions, require correction through signal post-processing. In recent years, numerous techniques have been introduced for THz signal denoising, aiming to substantially enhance the signal-to-noise ratio (SNR) using methods like wavelets [51], shrinkage [52], etc. These approaches are primarily formulated under the assumption of additive noise, often regarded as constant regardless of the signal's strength. This problem is classically tackled by using either a robust pre-processing method that enhances and exploits the strongest characteristics in the 3D data structure, or more common methods such as post-processing algorithms.

2.2 Problem definition

Suppose the imaging system is under the influence of a slow-varying physical perturbation having random amplitude, random phase, and random frequency up to 100 Hz. One thing to mention is that the randomness of perturbation’s parameters is of gaussian type. This noise models the instability of the hardware setup and its sensitivity to any slow physical move that might appear while the system is in use (e.g., temperature drift). To be more precise, we exemplify the following possible perturbations: the temperature effect on laser; delay line motion faults (rolling, pitching) in translation stage; and noise coming from the THz detection chain which is also proportional to the signal [53]. We can image scenarios within identification and authentication applications in airport security systems, where achieving precise standoff detection of weapons, including steel, firearms, and explosive items, is imperative. Some of the challenges that standoff detection is facing, are:

i) extending the distance range at which effective identification can occur;

ii) and improving signal detection and estimation over atmospheric and environmental noise and perturbation.

Under the influence of the described perturbation the THz image will degrade, loose contour information and will have a deformed contrast, making the further processing difficult and therefore having an unreliable result at the end of the processing chain.

The various perturbations can be more precisely modeled as a multiplication [54], [55], [56] between the received waves and an analytic signal that describes the instantaneous amplitude and phase interactions between the emitted wave and various disturbing factors.

Under the assumption of the physical perturbation, we consider the emitting THz pulse defined, in numerical form, by:

$$s[n] = A[n]e^{j\phi[n]}, \tag{2.1}$$

where $A[n]$ is the envelope of the transmitted signal and $\phi[n]$ is the instantaneous phase. For a monochromatic pulse, $\phi[n]$ can be expressed as:

$$\phi[n] = 2\pi f_n n, \tag{2.2}$$

where $f_n = \frac{f_0}{f_s}$, f_0 is the signal's frequency and f_s is the sampling frequency.

Considering the nature of the perturbation, for example the slow temperature drift, and the raster-scanning approach in pulsed THz imaging, we model the perturbation for each individual received THz pulse:

$$p_{i,j}[n] = \alpha_{i,j}[n]e^{j2\pi f_{i,j}n+\varphi_{i,j}} \tag{2.3}$$

where $\alpha_{i,j}$, $f_{i,j}$ and $\varphi_{i,j}$ are the envelope, frequency and phase of the perturbation signal.

The received pulse is the product between the two exponential signal and can be described as in:

$$r_{i,j}[n] = s[n] \cdot p_{i,j}[n], \tag{2.4}$$

where the real part of the signal is:

$$\begin{aligned} \Re e\{r[n]\} = \alpha_{i,j}[n]A[n] \cdot \Big(&\cos[2\pi f_n n] \cdot \cos\left[2\pi f_{i,j} n + \varphi_{i,j}\right] + \\ &+ \sin[2\pi f_n n] \sin\left[2\pi f_{i,j} n + \varphi_{i,j}\right]\Big). \end{aligned} \tag{2.5}$$

However, this model of the perturbation can be extended to real-time THz imaging, which implies a wider THz beam that illuminates the whole sample. In contrast, a raster-scanning approach scans the sample point by point with a THz beam radius smaller than 1mm. For THz imaging with wider beam, as in radar imaging, diffraction and scattering effects have to be added to slow-varying perturbations. [57]. The scattering effect in raster-scan imaging has neglectable effects as the beam width is small enough to minimize the encountering of unexpected scattering points. Additionally, even if we suppose there are scatterers inside the illuminated zone, the distance between them is small compared to the whole THz pulse that they can only affect small parts of pulse.

Thus, In THz imaging with a wider illuminating zone, the received THz beam has the following expression:

$$r[n]=\sum_{i=1}^{M}\sum_{j=1}^{N}\alpha_{i,j}A\left[n-\frac{d_{i,j}}{c}f_s\right]e^{j\phi\left[n-\frac{d_{i,j}}{c}f_s\right]}, \tag{2.6}$$

where M,N represent the number of scatter points along each axis, $\alpha_{i,j}$ represents the amplitude attenuation coefficients, $d_{i,j}$ is the distance between each scatter points and the receiver and c is the speed of light.

If we consider that $\phi[n]$ is the phase of a monochromatic pulse, the received THz pulse can be written as:

$$r[n]=\sum_{i=1}^{M}\sum_{j=1}^{N}\alpha_{i,j}A\left[n-\frac{d_{i,j}}{c}f_s\right]\cdot e^{-j2\pi\frac{d_{i,j}}{c}\cdot f_n n_0}\cdot e^{j2\pi f_n n}, \tag{2.7}$$

where the real part of the signal is:

$$\begin{aligned}\Re e\{r[n]\}=\sum_{i=1}^{M}\sum_{j=1}^{N}\alpha_{i,j}\cdot A\left[n-\frac{d_{i,j}}{c}f_s\right]\cdot\Bigg(&\cos[2\pi f_n n]\cdot\cos\left[2\pi f_n\frac{d_{i,j}}{c}n\right]-\\&-\sin[2\pi f_n n]\sin\left[2\pi f_n\frac{d_{i,j}}{c}n\right]\Bigg).\end{aligned} \tag{2.8}$$

As it can be observed from (2.5) and (2.8), the two types of perturbation are different in nature, but are similarly mathematically modeled for the two types of THz imaging techniques: raster-scanning and real-time systems. Having a similar model, the methods we proposed in this thesis for image reconstruction can be easily extended to both scenarios, but we focus only on raster-scanning as it is the acquisition technique used to obtain our images.

As (2.5) and (2.8) show, the propagation effect multiplicatively influences the amplitude, requiring careful consideration to develop an accurate method for synthesizing images. THz image reconstruction happens to be a tough problem, especially in the context of physical perturbations and environmental interference. Classic time- or frequency-domain methods usually fail to distinguish between interference and the informational signal, thus resulting in an improper estimation of the amplitude and a low-quality image with contour information loss and deteriorated contrast. The effects of the described perturbation are presented in Figure 2-1. The THz images are reconstructed using the maximum amplitude value of the received pulses in both cases. The perturbated THz image is the result of numerically adding the multiplicative noise to the original THz image. In the following perturbed image, we add multiplicative noise with random amplitude between 0.5 - 1, frequency between 50 - 100 Hz, and phase between 0 to π.

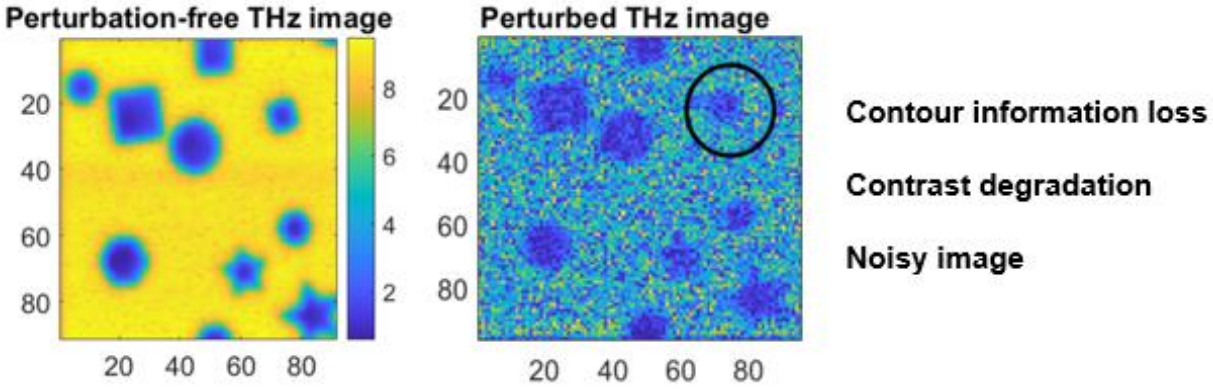


Figure 2-1. The effects of the slow-varying perturbation on the THz image reconstruction.

2.3 Mathematical modeling of the THz image

Before diving into the purpose of the chapter, we should first define the means of representing 3-D data acquired by our THz system after imaging a sample. The volumetric data can be expressed as a discrete function, based on the received THz signals defined in (2.8) :

$$g[x,y,n] = \mathrm{Re}\{r_{x,y}[n]\} \;\; | \;\; n \in \mathbb{N}^{+} \tag{2.9}$$

where $x = \overline{1...W}, y = \overline{1...H}$ iterates through the width and the height of the imaged sample on each scanning point (x,y); n represents the sample number; $\mathrm{Re}\{r_{x,y}[n]\}$ are the discrete real part of the complex THz signals of length N that propagated through the sample and were received by a THz detector. With regards with the photoswitch sampling technique used in TPI systems, we consider the equivalent sampling frequency, f_s and therefore, the sample number is expressed as $n = t \cdot f_s$, where t represents the timestamps of the THz pulse.

Some of the state-of-the-art methods to reconstruct images consider either the time or frequency-domain representation of the THz pulses [58]. Thus, it is necessary to define the frequency volumetric image by applying a Fast Fourier Transform (FFT) on the 3rd dimension of the discrete image function, g :

$$G[x,y,f_k] = \sum_{n=1}^{N} g[x,y,n] \cdot e^{-\frac{j2\pi}{N} f_k n} \tag{2.10}$$

where f_k is the frequency starting from 0 and going up to the Nyquist frequency, $\frac{f_s}{2}$.

In Figure 2-2 we depict an illustration of the data format presenting a 3-D image, where each pixel is a THz waveform represented in time and frequency domain, respectively.

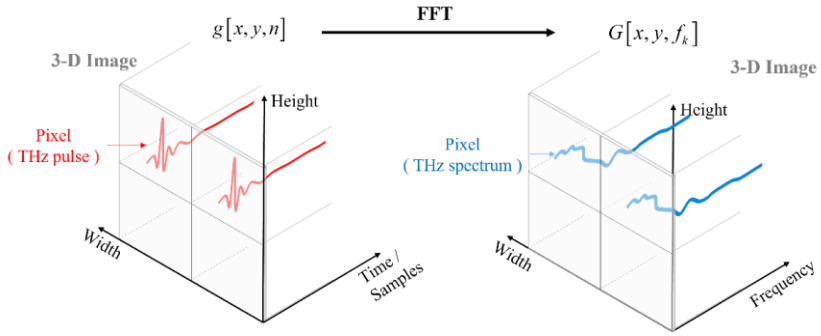


Figure 2-2. Illustration of 3-D data format in time and frequency domain

2.4 State-of-the art image reconstruction methods in THz imagery

In pre-processing, the problem of how to display the image arises. In this part, we detail state-of-the-art reconstruction methods based on time and frequency domain representation of the signals. For illustration purposes we consider images that have 100 by 100 pixels corresponding to a physical image of 3 by 3cm with a spatial step of 0.3mm.

The objective is to study the state-of-the-art methods to reconstruct images while observing their ability to remove noise and artefacts. All methods are compared in both ideal and noisy context. We suppose the system is under the influence of a physical perturbation with random amplitude, phase and frequency. In all cases, the noise has the following parameters: the amplitude is a number randomly picked from the range 0.5 - 1, the random frequency range is 50 - 100 Hz, and the phase is randomly chosen from 0 to π. The multiplicative perturbation conditions remain consistent across methods to highlight their impact on the information within the image and facilitate comparison among state-of-the-art approaches employed in contemporary THz systems. The effect of this specific noise is visible through contour information loss for dominant, wide motives and even total loss of information in case of small details. Other important aspects are contrast degradation and additional background noise. In terms of information, the images in noise conditions are highly corrupted with little to no easily detectable details. The contrast is degraded and can be analyzed within image histograms, as their shape is related to image appearance and information. In most scenarios, depending on the information, for dark images, the majority of the histogram's bins are densely clustered towards the lower, darker end of the intensity scale. Similarly, for light images, the higher end of the scale is populated. For a low-contrast image, the histogram tends to be centered around the middle of the scale, whereas for a high-contrast image, the histogram spans a broad range of the intensity scale. THz images are presented along with corresponding histograms so that an overview of changes in pixel intensities is provided.

2.4.1 Time-domain methods

The widely used reconstruction space is the time-domain [59], [60]. One commonly utilized approach involves presenting the maximum amplitude of the signal for each pixel in the image. The *max* function reduces the dimensionality of the volumetric data by discarding all amplitude values except the maximum. The image resulted from this procedure is denoted as Maximum Peak Image (MPI) and can be expressed as:

$$MPI[x,y] = \max\{g[x,y,n]\}, \tag{2.11}$$

where the $\max\{.\}$ function operates on each received signal at (x,y) scanning point.

Figure 2-3 shows an example of MPI reconstruction method in ideal conditions and under the influence of the multiplicative perturbation. Due to the random distortion defined in (2.8), the reconstructed image is corrupted and information deformed, aspects also visible in the histogram, where pixel values are spread among all domain.

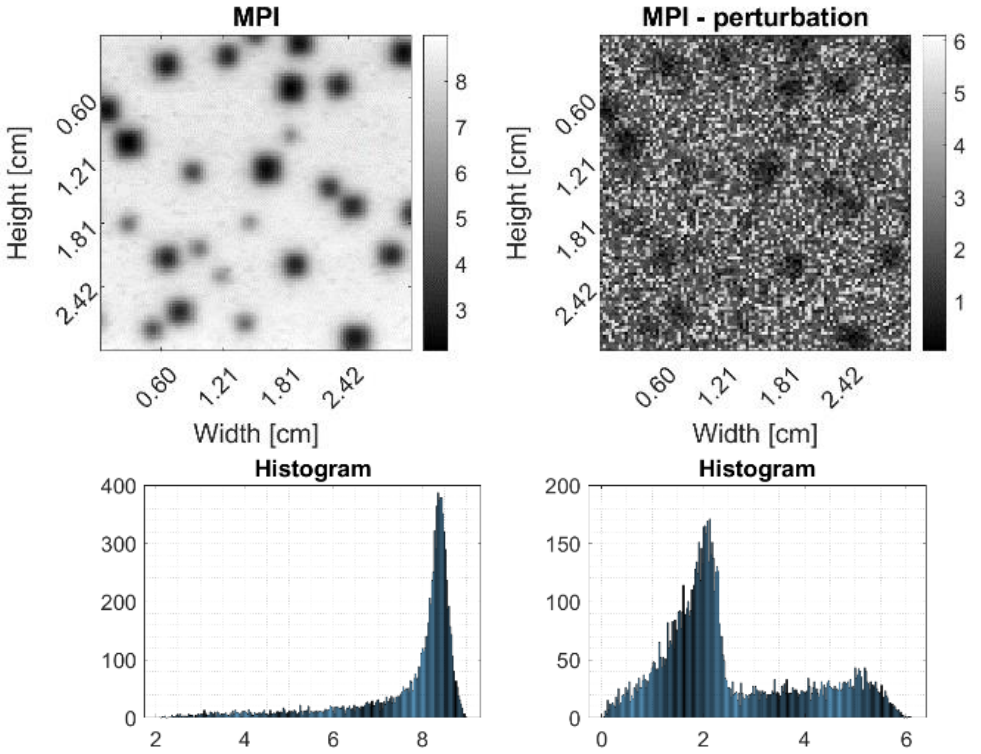


Figure 2-3. Maximum peak value images in ideal and noisy context

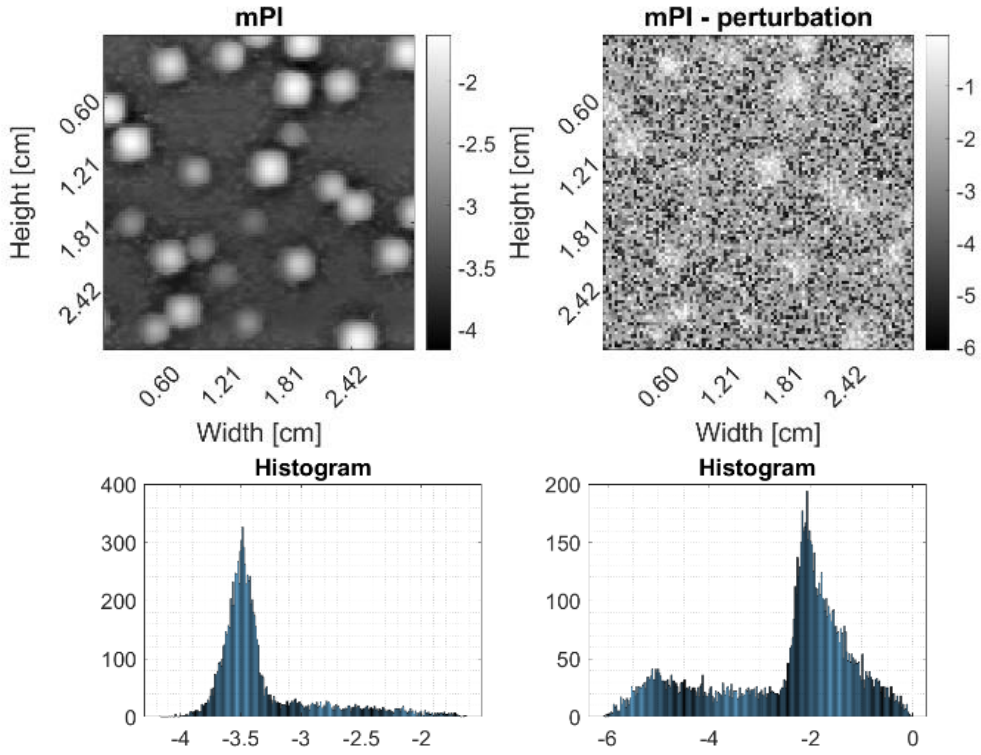


Figure 2-4. Minimum peak value images in ideal and noisy context

At the same time, one can display the image using the signal's minimum amplitude. This image is further denoted as Minimum Peak Image (mPI) and is expressed as:

$$mPI\left[x,y\right]=\min\left\{g\left[x,y,n\right]\right\}, \tag{2.12}$$

where the $\min\left\{*\right\}$ function operates on each received signal at $\left(x,y\right)$ scanning point.

Figure 2-4 shows an example of this case in both ideal and noisy conditions. Similarly, to MPI, mPI images prove to have the same characteristics, mentioning only that the image intensities and histogram are reversed as the image was reconstructed using the minimum amplitude values.

The two aforementioned display methods derive from a broader approach, wherein the image can be rendered using the amplitude at a particular temporal position, usually a timestamp where the signal is prominent, or where the reflections of the THz pulse are visible. The image is further denoted as Time Slice Image (TSI) [61] and is defined as:

$$TSI\left[x,y\right]=g\left[x,y,n_{slice}\right], \tag{2.13}$$

where $n_{slice}=f_s\cdot t_{slice}$ is the timestamp used to create the image.

In Figure 2-5, we depict the reference THz pulse, $s\left[n\right]$, initially defined in (2.1) and the values used to create a TSI image and subsequent methods based on the same reconstruction principle. The reference pulse is the signal measured without a sample and it is used to measure all samples presented in this thesis. The THz reference pulse is considered constant for each measuring point.

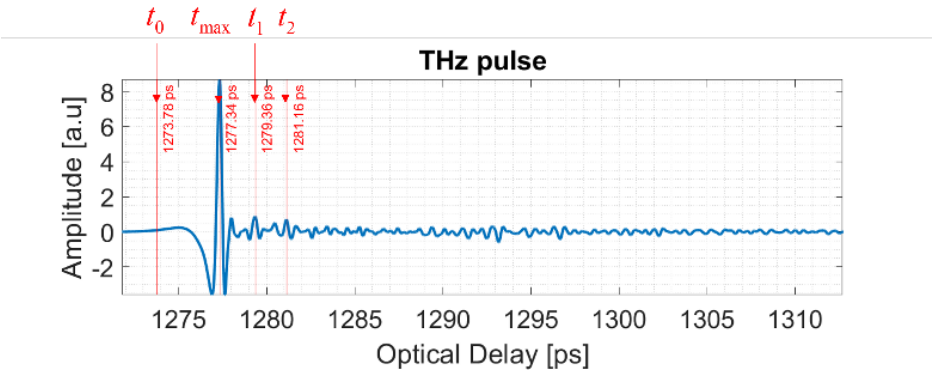
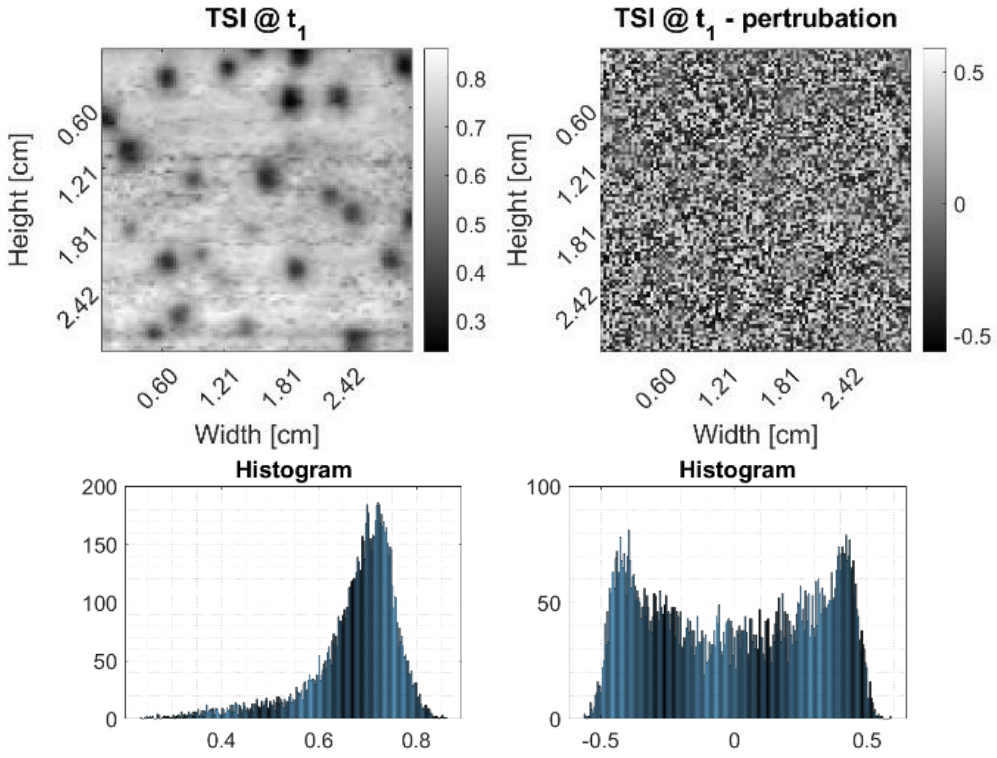


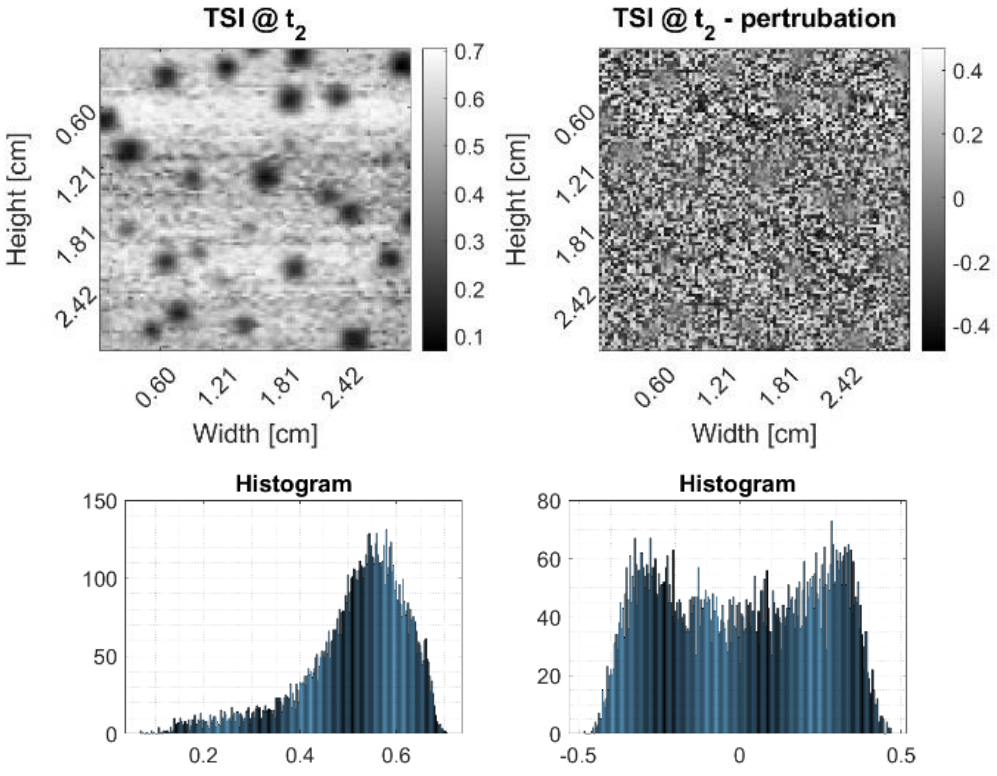
Figure 2-5. THz pulse and the corresponding timestamps of its reflections

As TPI systems use optical delay lines (ODLs) to sample the electric field in the THz pulse, the *Ox* axis in Figure 2-5 has the optical delay values used to probe the THz electrical field. The timestamps are used in subsequent methods in this section and correspond to the start of the pulse, $t_0=1273.78$ ps, the first, $t_1=1279.36\,ps$, and second, $t_2=1281.16\,ps$ reflection of the THz pulse, respectively.

We chose these timestamps as their amplitudes are larger than others, thus having higher probability to correctly reconstruct the image. Usually, the choice of other values yields poor results as they have low energy and are sensitive to noise. Additionally, the timestamp of the maximum peak value, $t_{max}=1277.34\,ps$, is denoted in the same figure as it is used in later methods to reconstruct images.

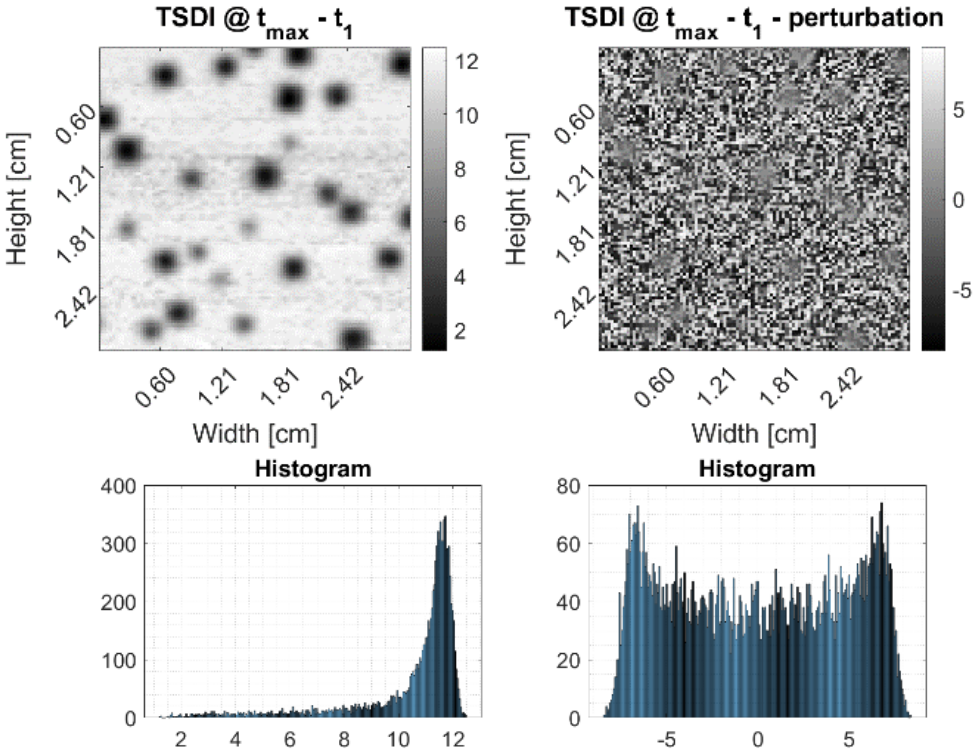


(a)

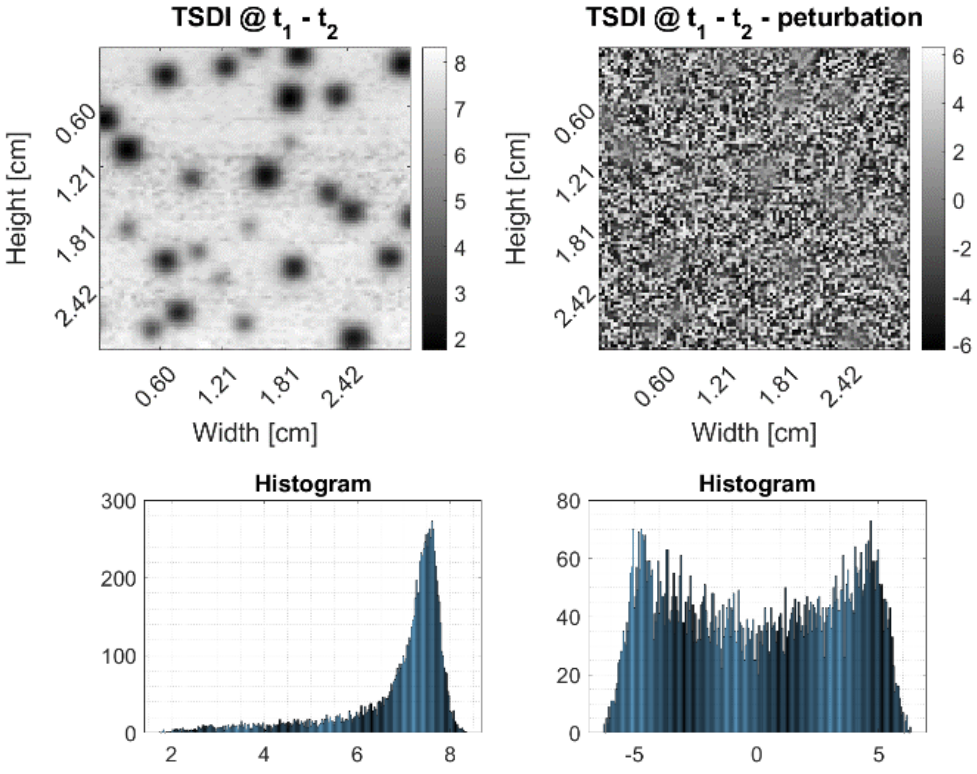


(b)

Figure 2-6. Time Slice Images in ideal and noisy context at: **(a)** 1279.36 ps; **(b)** 1281.16 ps with perturbation;



(a)



(b)

Figure 2-7. Time slices difference images in ideal and noisy context: **(a)** between maximum peak and first reflection; **(b)** between first and second reflection;

Considering the two reflections' timestamps depicted in Figure 2-5, we reconstruct the corresponding TSI images presented in Figure 2-6 in ideal context (a) and noisy conditions (b). As the reflections' energy is low compared to the noise, the image is highly sensitive in terms of pixel intensities, resulting in images and histograms resembling noise.

Another method uses the difference between the signal's amplitudes at two specific timestamps positions and to display the image. Thus, the Time Slice Difference Image (TSDI) [61] is defined as:

$$TSDI\left[x,y\right]=g\left[x,y,n_i\right]-g\left[x,y,n_j\right], \tag{2.14}$$

where $n_i = f_s \cdot t_i$ and $n_j = f_s \cdot t_j$ represent two different time slices.

Figure 2-7 (a) presents THz images reconstructed using the difference between the amplitude values at $t_{\max}$ and t_1, while Figure 2-7 (b) shows images reconstructed with the amplitude difference between t_1 and t_2. Considering the phase and frequency deformation as consequences of noise's presence, the THz pulses and their individual reflections' positions are arbitrary shifted at each measuring point, (x,y). It is worth mentioning that $t_{\max}$ was chosen on the reference pulse and will not correspond to the maximum amplitude value on the received THz signals. Thus, the major drawback of this method regards the fixed timestamps which do not always correspond to the exploited signals' components. Additionally, there is no obvious mean to select optimal signal's components and to find their position within the measured signal. The consequence is a highly corrupted image with most of the information hidden in noise.

In the same manner as TSDI, we can integrate between two positions to reconstruct the image. A Time Slice Integration Image (TSII) [61] is therefore expressed by:

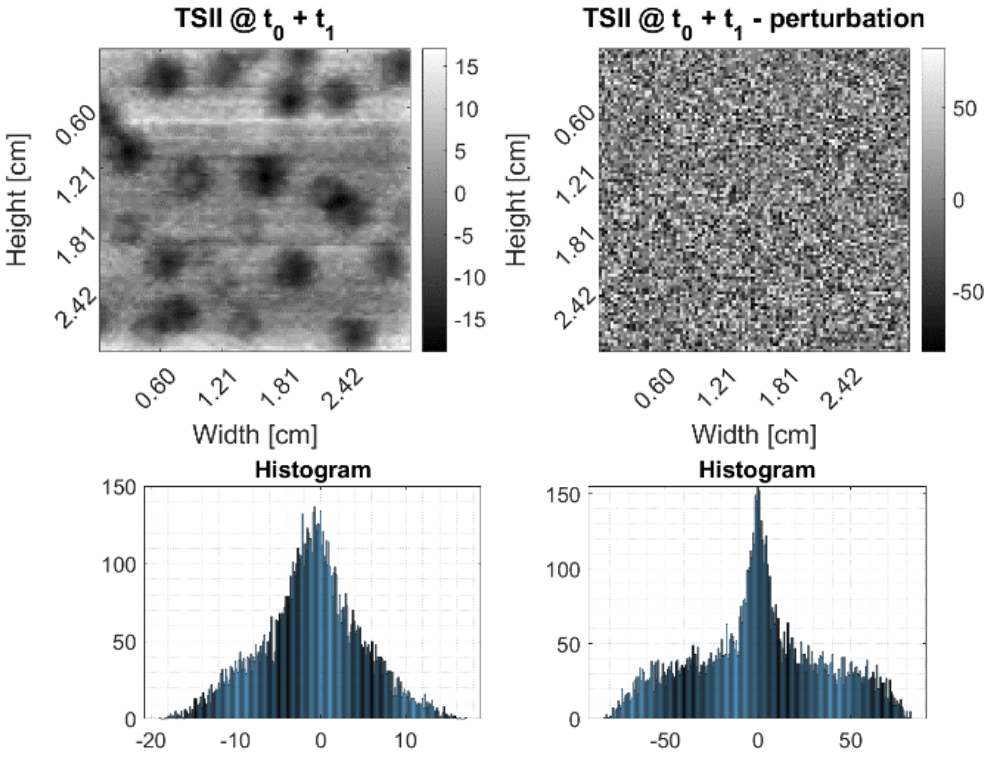
$$TSII\left[x,y\right]=\sum_{k=n_i}^{n_j} g\left[x,y,k\right], \tag{2.15}$$

where the summation iterates through all timestamps from $n_i = f_s \cdot t_i$ to $n_j = f_s \cdot t_j$. Employing identical timestamps as in the preceding methods, TSI and TSDI, in Figure 2-8 (a) the images are reconstructed by integrating all values between t_0 and t_1. In the second scenario, in Figure 2-8 (b), the images are reconstructed by integrating from t_0 to t_2. TSII images exhibit similar limitations as the preceding methods, with the most significant being the lack of information.

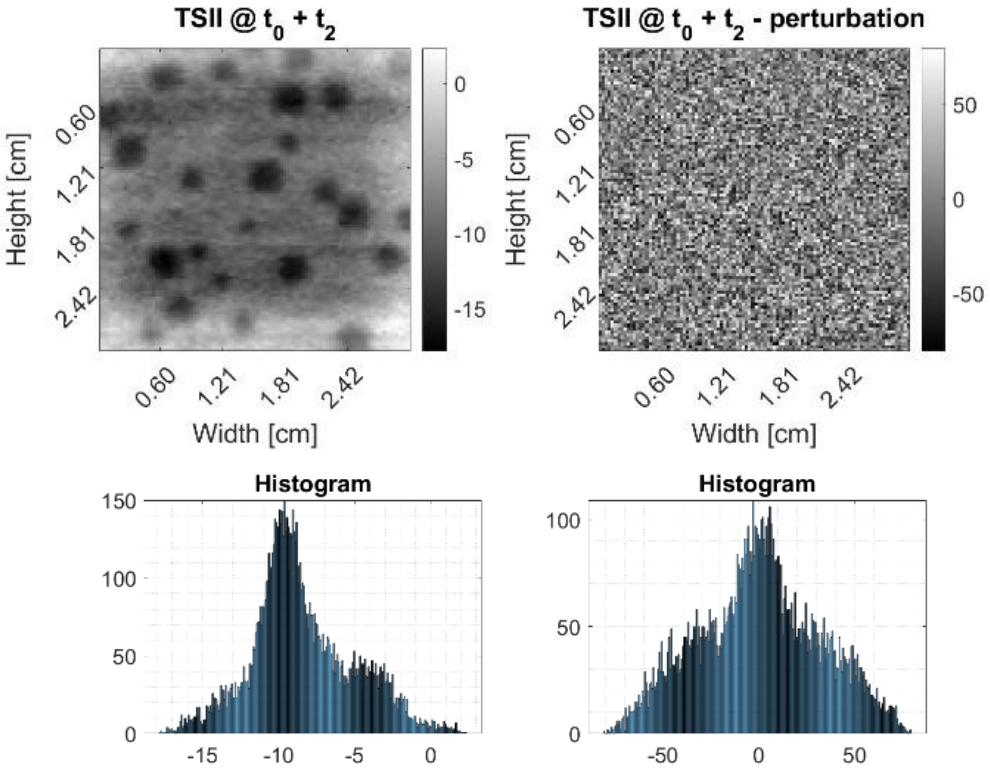
Another approach is based on determining the time of flight (TOF) related to the time shift of the THz pulse, and to display it for each pixel. To ascertain the time shift or lag, we calculate the position of the maximum value in the cross-correlation between the emission and reception, also known as Matching Filtering (MF). To display Time of Flight Images (TOFI) [62], we first compute the cross-correlation between $s\left[n\right]$, initially defined in (2.1), and all measured THz signals. The pixels values are in fact the positions of the cross-correlation maxima determined as in:

$$TOFI\left[x,y\right]=\operatorname{argmax}\left\{\left\langle g\left[x,y,n\right],s\left[n\right]\right\rangle\right\}, \tag{2.16}$$

where $\langle . \rangle$ denotes the correlation operator.

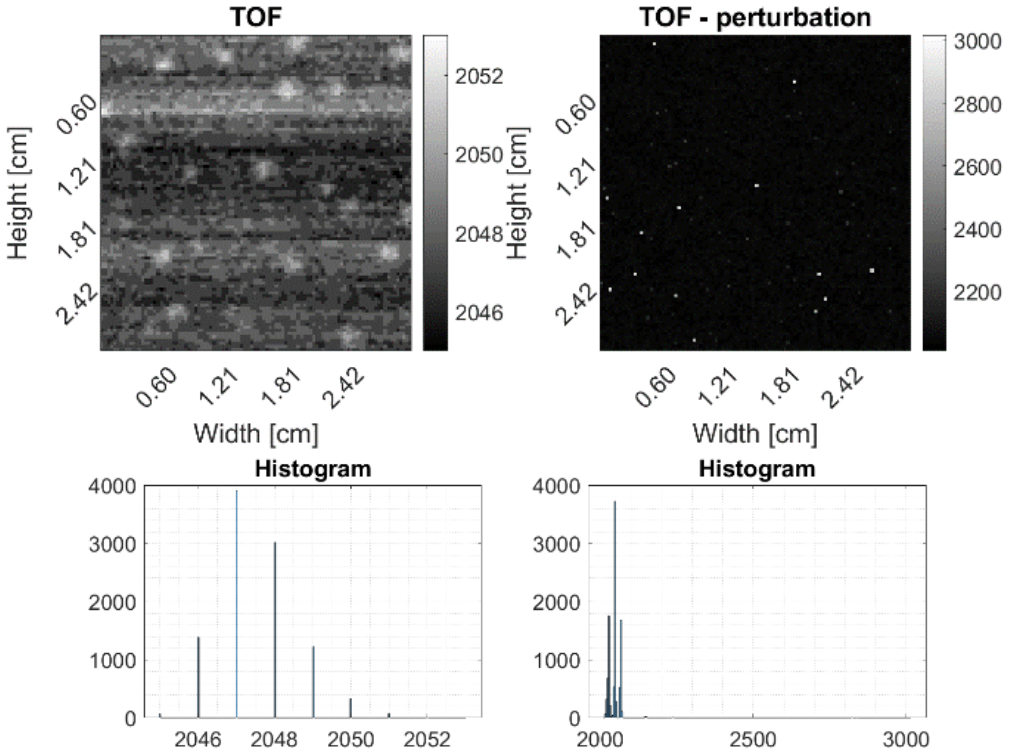


(a)

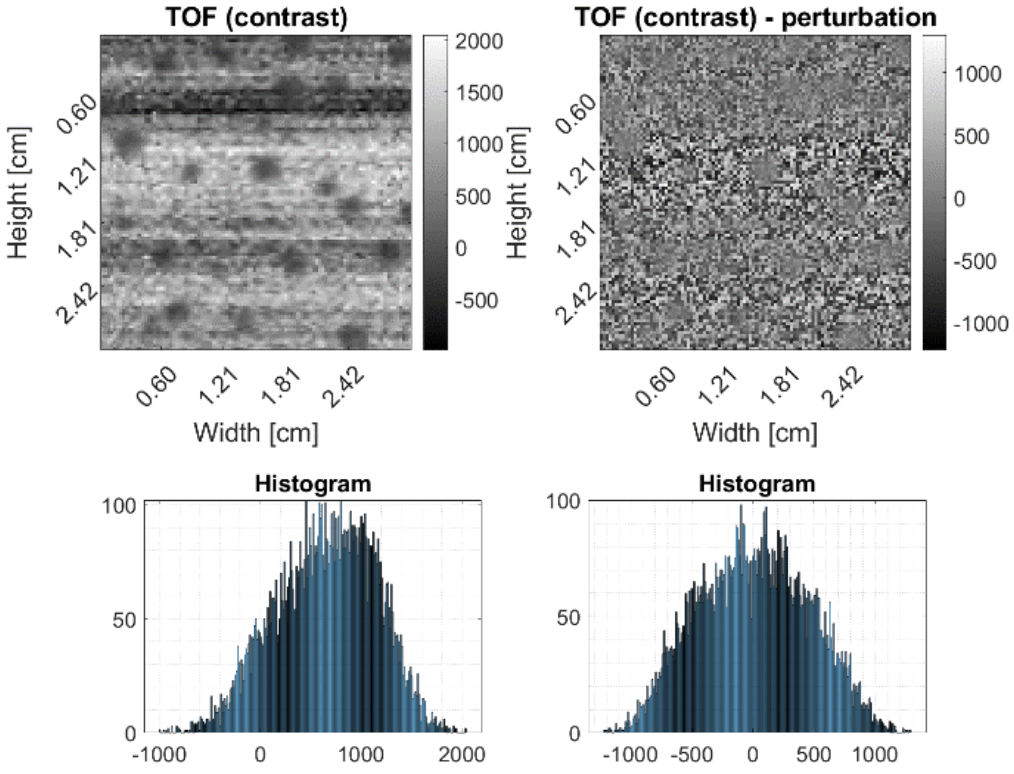


(b)

Figure 2-8. Time slices integration images in ideal and noisy context: **(a)** between maximum peak and first reflection; **(b)** between first and second reflection



(a)



(b)

Figure 2-9. Time-of-flight images in ideal and noisy context: **(a)** without contrast enhancement; **(b)** with contrast enhancement

In the common case of channels exhibiting convolutional effects [63], the signal transmitted to a receiver can take multiple paths to reach the destination. These paths may involve reflections, diffractions, and scattering from various objects in the environment. As a result, the transmitted signal may experience multiple delays and attenuations. Using MF, the signal of interest can be optimally detected in this scenario. However, the model of the multiplicative perturbations tackled in this thesis showcases more complex deformations resulting in an efficient MF approach. Frequently, when employing TOFI, the signal reflected from the front surface of the scanned object emerges as the strongest signal and is identified as the maximum value. If the surface remains flat, the TOF will remain approximately constant across the entire surface, yielding a uniform image that is sensitive to either the flatness or the optical thickness of the sample being measured. To magnify the contrast of such TOF-based image, one can also compute the first derivative of the THz pulse at a certain temporal position set as working point. In such imaging process the slight temporal shift is multiplied by the slope of the THz pulse, which is maximum if the working point is the inflection one, as in:

$$TOFI\left[x,y\right]=\frac{g\left[x,y,n_2\right]-g\left[x,y,n_1\right]}{n_2-n_1}\cdot\operatorname{argmax}\left\{\left\langle g\left[x,y,n\right],s\left[n\right]\right\rangle\right\}, \tag{2.17}$$

where n_1 and n_2 are two consecutive positions. Figure 2-9 (a) presents TOFI images, and (b) depicts the contrast enhanced TOFI images. Amplitude randomness and phase deformations introduced by multiplicative perturbations makes this approach unusable as the computed lags are randomly varying losing information within the image. Moreover, as the measured sample is not multi-layered, TOF values are close to one another and the image contrast is low.

2.4.2 Frequency-domain methods

Frequency-domain methods exploit the spectral information of the THz pulses [64], [65]. Figure 2-10 presents the spectrum of the reference THz pulse, $s\left[n\right]$, emphasizing the spectral components, $f_i=i\text{ THz}\mid i\in\left\{1,2,3,4,5\right\}$, which are used to reconstruct the images in the following methods. These spectral components are chosen from the frequency pool of the THz imaging system. From its technical description manual, the spectral range varies from 0.06 THz and up to 5 THz. Thus, frequency values higher than 5 THz are not used as the signal-to-noise ratio is below 1.

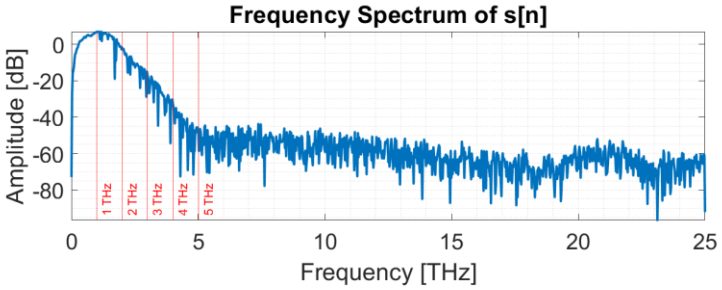
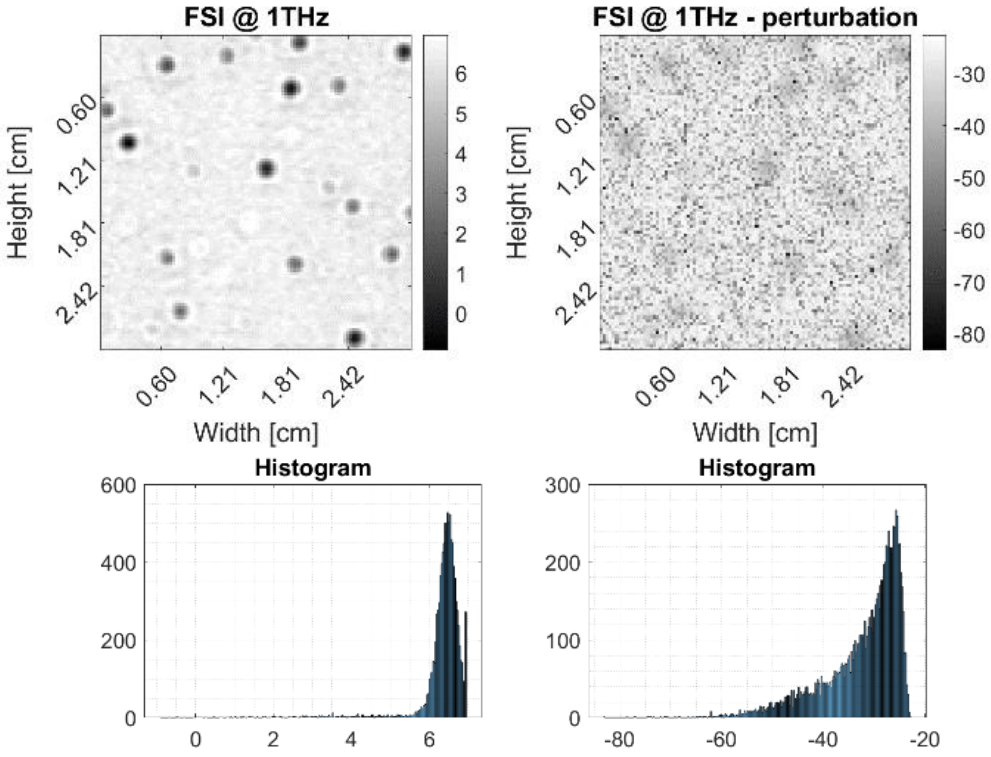


Figure 2-10. Frequency spectrum of the THz reference signal

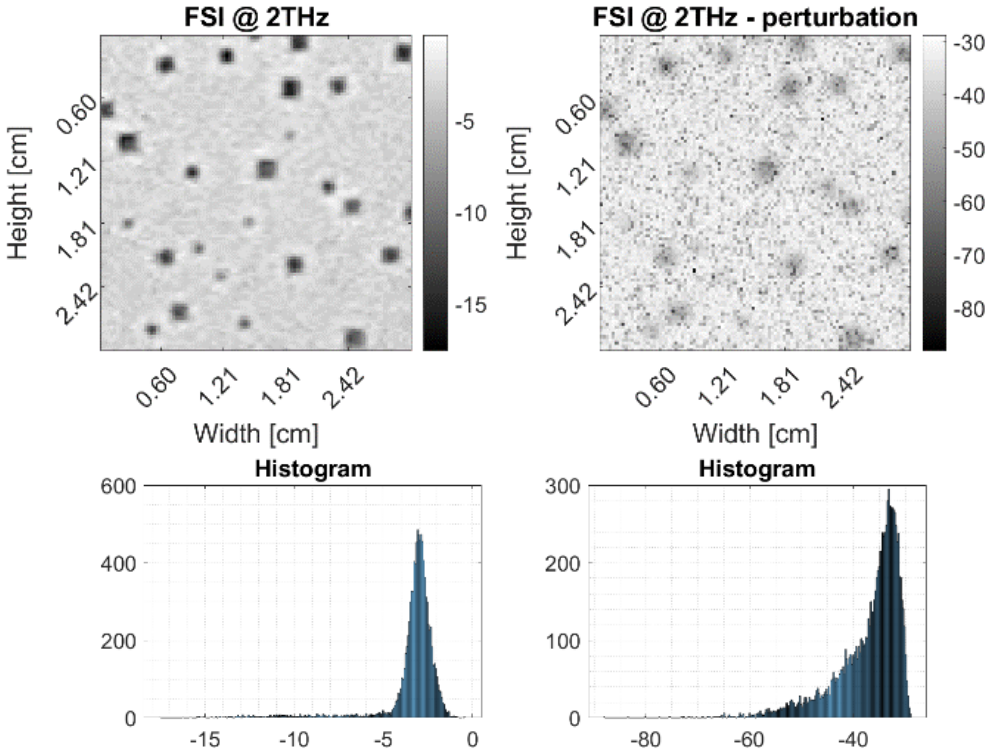
One of the most straightforward approaches uses the amplitude of a specific frequency component as pixel value. The Frequency Slice Image (FSI) [58] can be expressed as:

$$FSI\left[x,y\right]=G\left[x,y,f_k\right], \tag{2.18}$$

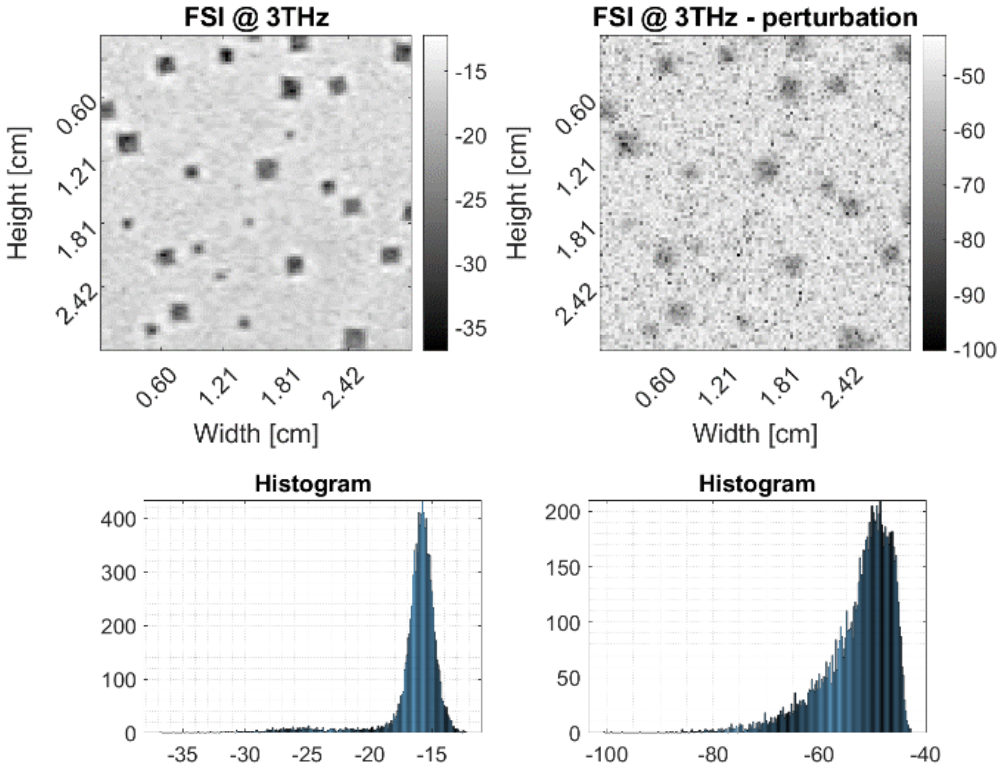
where f_k is the chosen frequency. In Figure 2-11 we depicted THz images reconstructed using the amplitude value of the frequency components starting with 1 THz and going up to 5 THz.



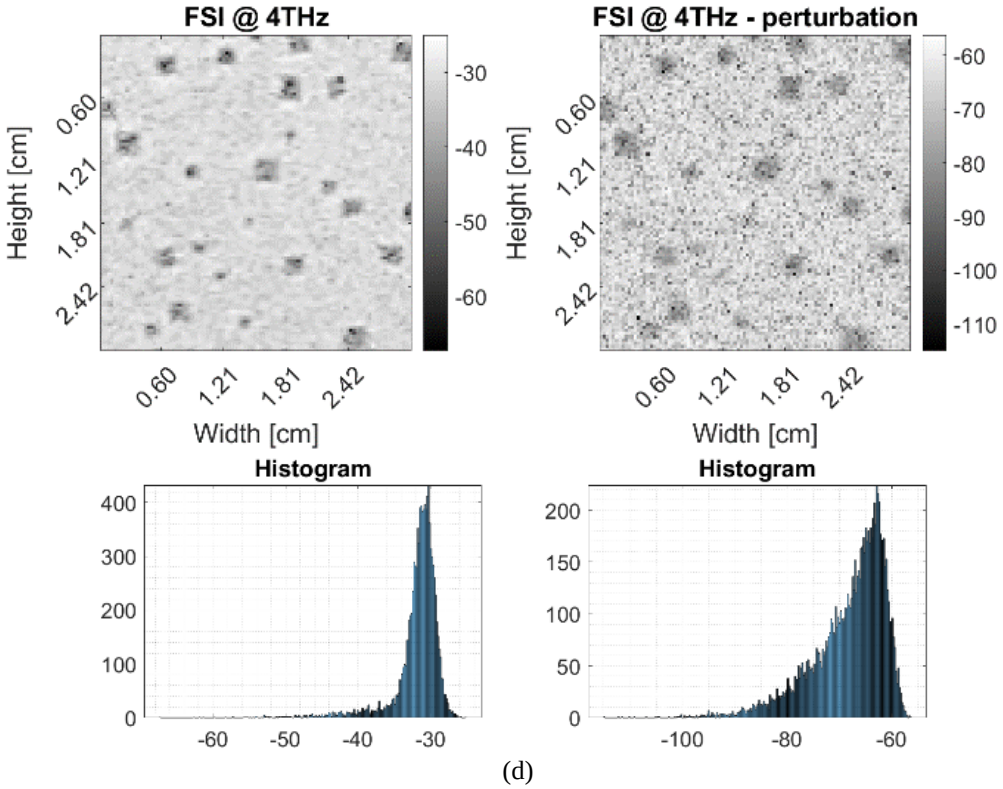
(a)



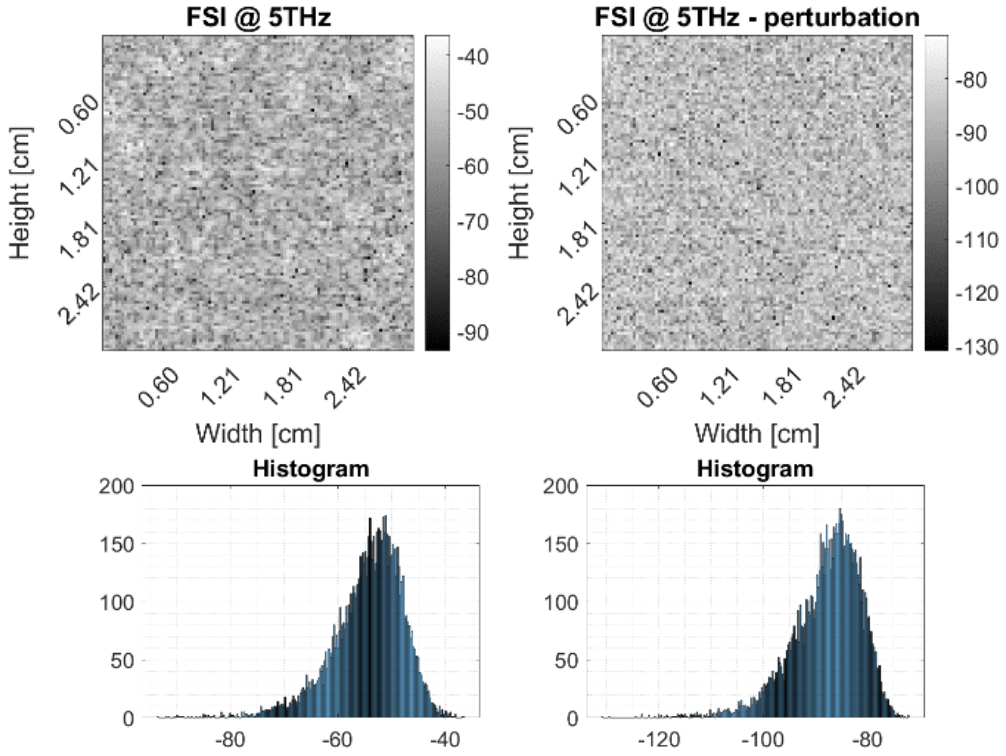
(b)


FSI @ 3THz
FSI @ 3THz - perturbation
Height [cm]
Width [cm]
0.60
1.21
1.81
2.42
-15
-20
-25
-30
-35
-50
-60
-70
-80
-90
-100
Histogram
400
300
200
100
0
-35
-30
-25
-20
-15
150
50
-100
-80
-60
-40

(c)


FSI @ 4THz
FSI @ 4THz - perturbation
Height [cm]
Width [cm]
0.60
1.21
1.81
2.42
-30
-40
-50
-60
-70
-80
-90
-100
-110
Histogram
400
300
200
100
0
-60
-50
-40
-30
150
50
-100
-80
-60

(d)



(e)

Figure 2-11. Frequency slice images in ideal and noisy context at: **(a)** 1 THz; **(b)** 2 THz; **(c)** 3 THz; **(d)** 4 THz and **(e)** 5 THz.

In general, low frequency values reconstructs images with blurred information and high contrast, while higher frequencies generate images with higher spatial resolution, but with lower contrast. The spatial resolution is limited by the diffraction limit. The imaging resolution is limited to half of the wavelength, λ, modified by the refractive index n_i of the medium and the angle θ of the cone of focused beam:

$$d_{\text{limit}} = \frac{\lambda}{2n_i \sin\theta} \tag{2.19}$$

As the multiplicative noise is present in all effective bandwidth, it cannot be removed by selecting a specific frequency to reconstruct the image. Based on the noise characteristics (arbitrary amplitude, frequency and phase), the spectral components are randomly affected and thus, obstructing an efficient image reconstruction.

Similarly, as observed in the time-domain, we can create an image considering the amplitude difference between two distinct frequencies. We denote this image as FSDI (Frequency Slice Difference Image) and define it as in:

$$FSDI[x,y] = G[x,y,f_1] - G[x,y,f_2], \tag{2.20}$$

where f_1 and f_2 are the chosen frequencies to create the image.

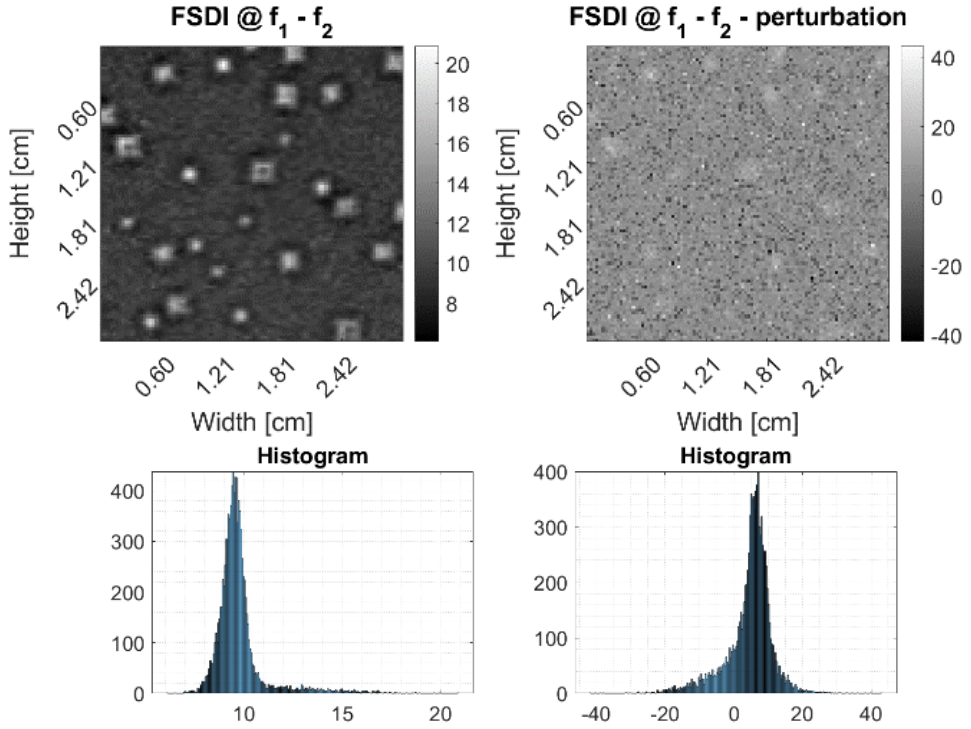
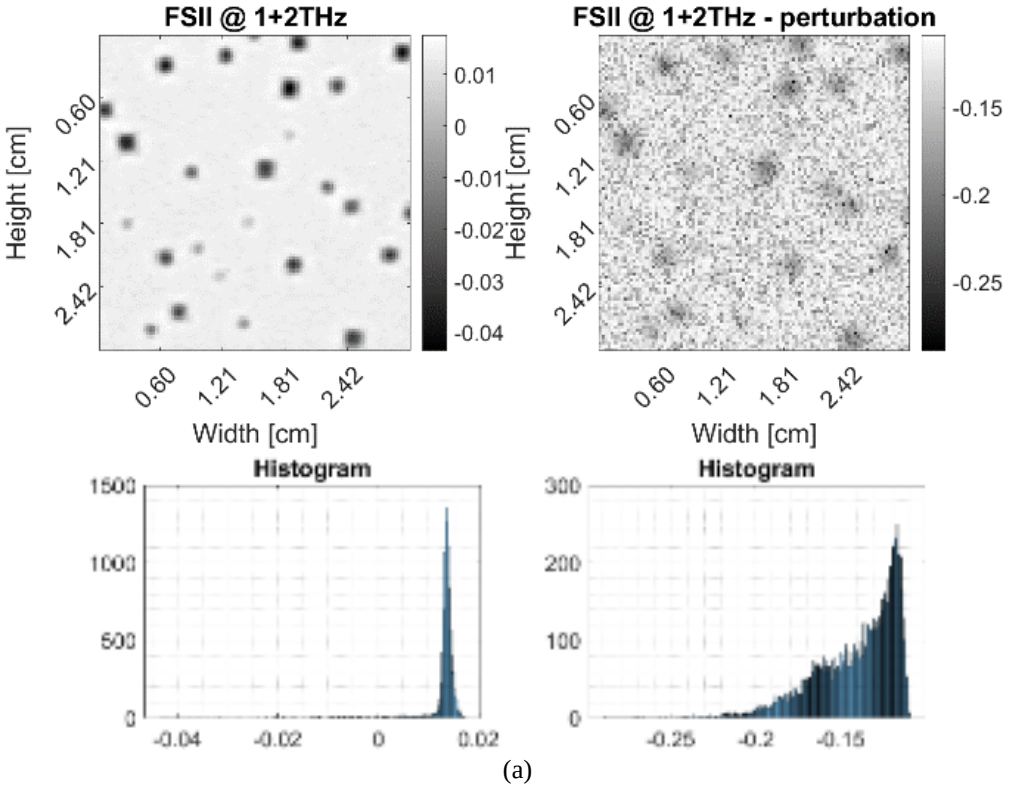
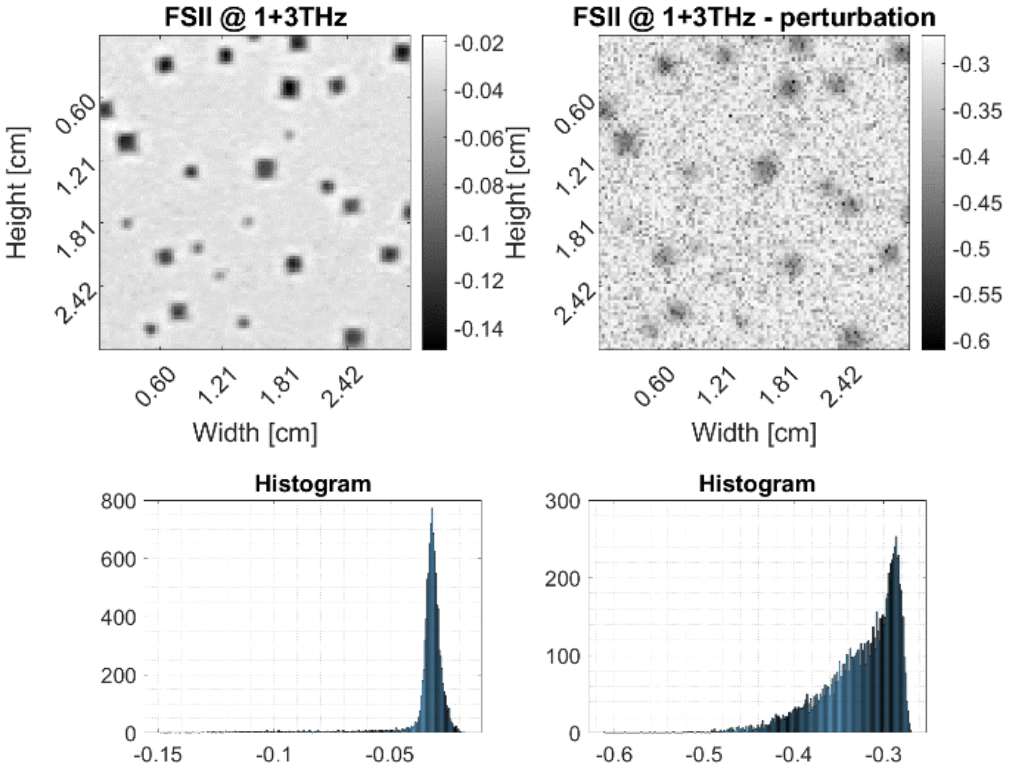


Figure 2-12. Frequency Slice Difference in ideal and noisy context: **(a)** 1 THz - 2 THz; **(b)** 1 THz – 3 THz.



(a)



(b)

Figure 2-13. Frequency Slice Integration Image: **(a)** from 1 THz to 2 THz; **(b)** 1 THz to 3 THz

Examples of this method are presented in Figure 2-12 (a) where we use the amplitude difference between frequency components at $f_1 = 1\ THz$, $f_2 = 2\ THz$ and Figure 2-12 (b) where the considered frequency are $f_1 = 1\ THz$ and $f_3 = 3\ THz$.

The images loose almost all information with mentioned parameters. In this case, fine-tuning is a complex process with little rate of success as the efficiency of the method is strictly related to noise and its influence on spectral information: the greater the influence of noise, the higher the likelihood of failure to recover information.

Consequently, we can also integrate the amplitudes between two distinct frequencies. The image displayed using this method is further denoted as FSII (Frequency Slice Integration Image) [65] and is defined as in:

$$FSII[x,y] = \sum_{k=f_1}^{f_2} G[x,y,k], \tag{2.21}$$

where the summation iterates through frequencies from f_1 to f_2.

Figure 2-13 (a) and (b) shows examples of FSII reconstruction by integrating between 1-2 THz and 1-3 THz, respectively. However, one important disadvantage is that there are no efficient bandwidth selection algorithms which can automatically adapt from measure to measure to reconstruct the image in a way that reduces noise while maximizing information. Thus, FSII, but also FSI and FSDI are rather overwhelming approaches as they need lot of fine-tuning and the results are not always expected to be of high quality.

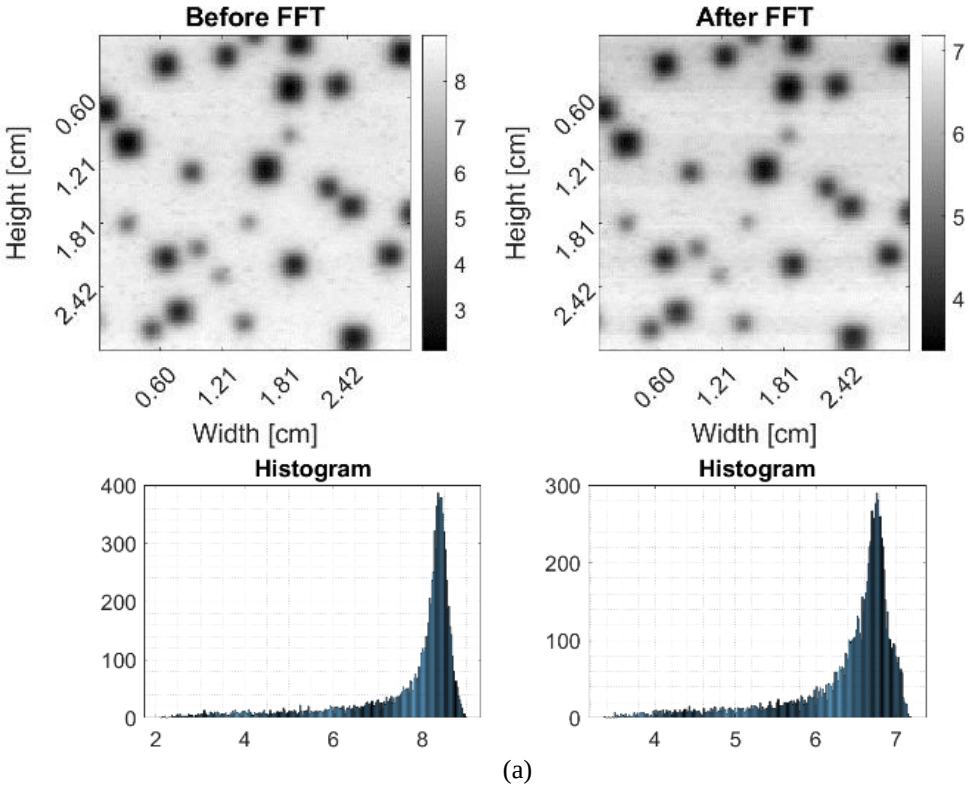
Filtering procedures applied in electromagnetic sensing are used to enhance the THz image reconstruction. A first example is the Fourier Filter method which exploits the part of the signal's spectrum where the amplitudes are not lower than a threshold [66]. The waveforms in the time domain, taken along a line or column from the image, are collected into a vector. The Fourier Transform is then applied to obtain their spectra. Subsequently, these spectra are normalized to their maximum value and converted into a logarithmic scale. The frequencies whose amplitude are lower than 3dB are discarded and the Inverse Fourier Transform is applied to retrieve the time-domain signals.

The FFI (Fourier Filtered Image) can be briefly described as in:

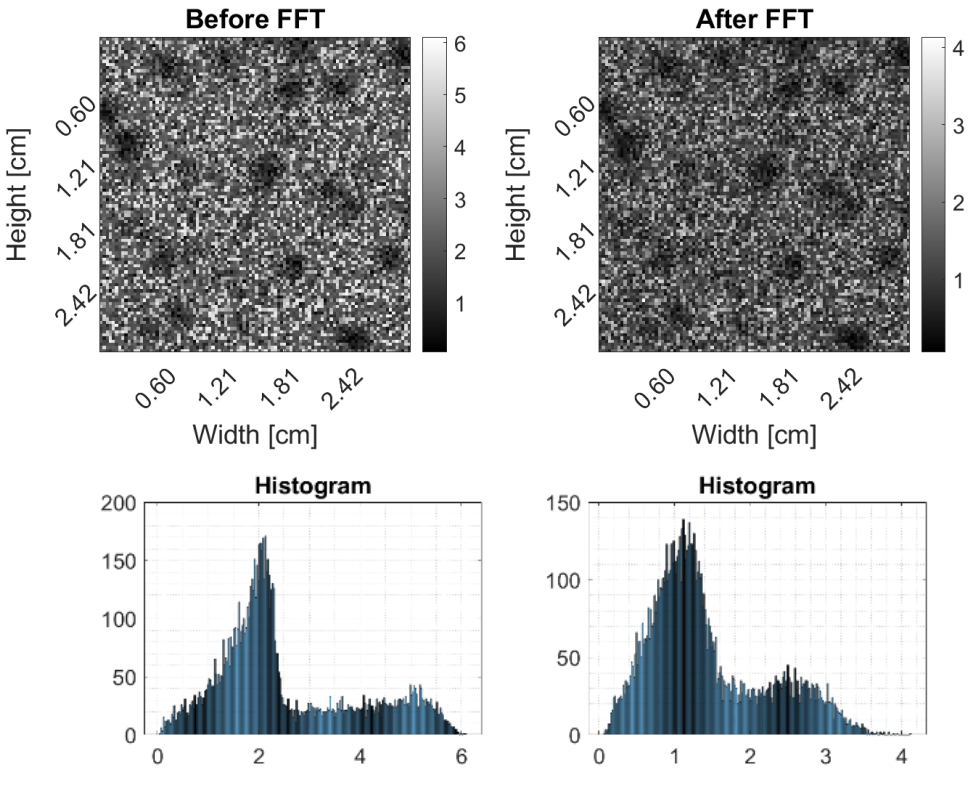
$$FFI[x,y] = \max\left\{ IFFT\left\{ \Pi\left\{ \sum_{n=0}^{N-1} g[x,y,n] \cdot e^{-j2\pi\frac{k}{N}n} \right\} \right\} \right\} \tag{2.22}$$

where IFFT is the Inverse Fast Fourier Transform, Π is the bandpass filter that selects the effective spectrum of the THz pulses, and *max* is the function which reduces the dimensionality from 3-D to 2-D by selecting the maximum amplitude value after the filtering procedure.

Figure 2-14 shows a THz image in both ideal and noisy conditions before and after filtering. In both cases, the filtered image does not greatly improve the original image. Moreover, it lowers the contrast as the multiplicative noise is present in the whole spectrum.



(a)



(b)

Figure 2-14. FFT filtering procedure in: **(a)** ideal conditions; **(b)** noisy context

2.4.3 Subsections concluding remark

As until now, the presented methods were individually applied on each THz signal. In this case, we refer them as pre-processing methods as they operate on the physical level of the image, which is the electric field of the THz radiation. While a part of the pre-processing methods ensures a roughly good image reconstruction, they are sensitive to noise, which deforms the contours of the shapes, blurs the image and adds unwanted artifacts. Some methods have potential in recreating images, but they require either fine-tuning or aid from other tools in order to achieve a quality result.

In the following subsection, we briefly present post-processing methods, as they are applied after the image was already reconstructed in order to enhance its quality and to remove any additional noise.

2.4.4 Post-processing methods

To illustrate the post-processing methods, we're utilizing the noisy images introduced in the preceding section. The objective of this sub-section is to analyze different widely used methods and to observe their robustness and capacity to remove noise and enhance the image. We chose to analyze the FSII image from Figure 2-13 (b) as it has a good quality.

A simple widely form of post-processing method is the histogram equalization [67], which is used to enhance the contrast by spreading the most frequent intensity values. Adaptive histogram equalization [68] calculates numerous histograms corresponding to distinct regions of the image. These local histograms are exploited to reallocate the intensity values among the image. Figure 2-15 shows how both methods perform on a THz image. Indeed, the contrast is increased, but the noise is visible in the background and contour information is lost in both cases.

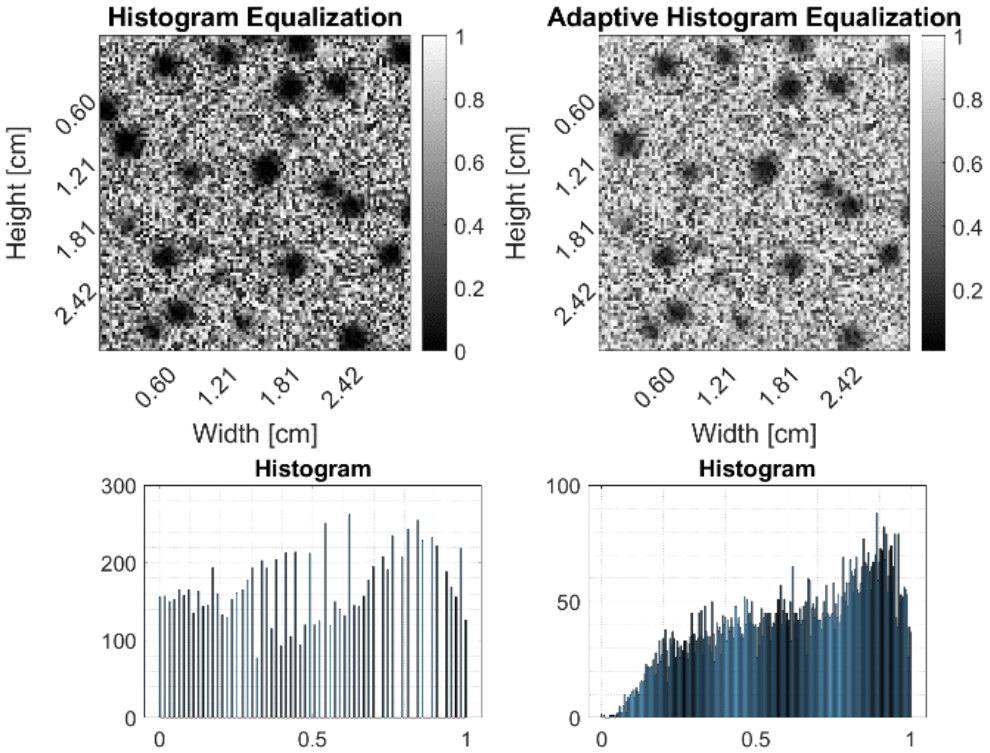


Figure 2-15. THz images obtained after histogram equalization (left) and **a**daptive histogram equalization (right) post-processing

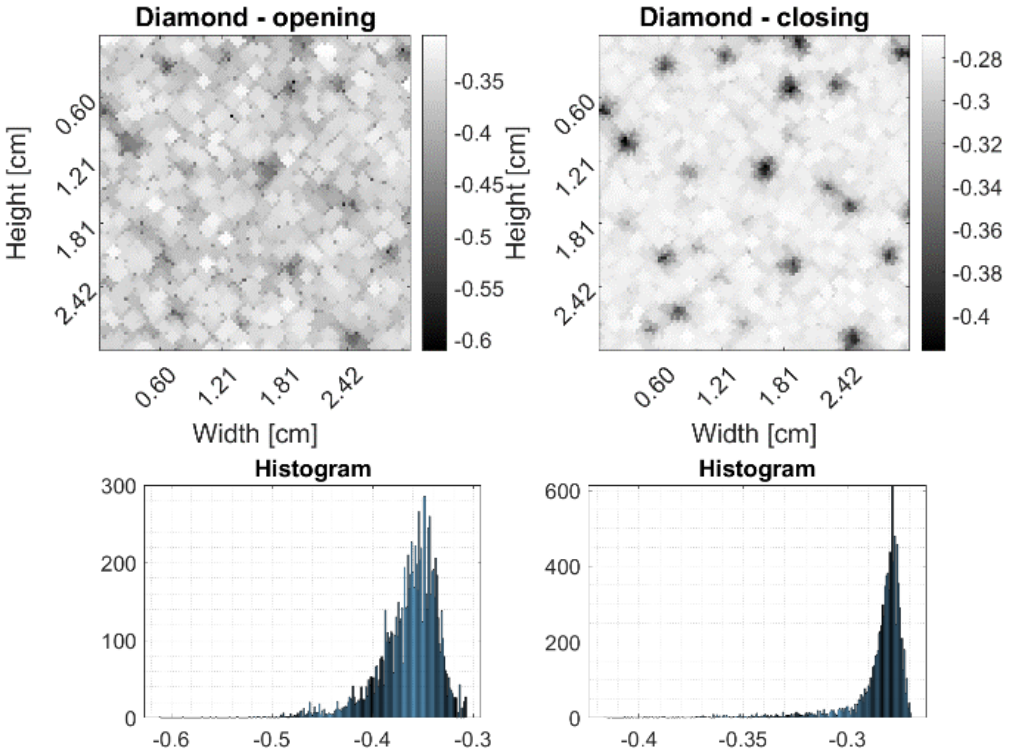
Another post-processing method is the morphological filtering [69] which is based on the erosion and dilation operators [70]. At the core of morphological image processing lie two fundamental operations: erosion and dilation. Erosion acts as a digital "shrinking" mechanism, diminishing the extent of information within the image, while dilation serves as its counterpart, expansively "growing" the image's content. Building upon these foundational concepts, two additional morphological operations emerge: opening and closing. Broadly speaking, the application of an opening operation on an image yields a smoothing effect on its borders and contours, effectively eliminating fine details such as small, narrow lines and connectors between distinct components. Conversely, closing an image entails the amalgamation of fragmented regions, effectively bridging narrow gaps or interruptions, eradicating diminutive holes, and filling in the spatial voids along the contour. These operations, intricately interwoven, offer versatile tools for manipulating image structure and enhancing visual clarity, playing a pivotal role in a myriad of image processing tasks across diverse domains.

Figure 2-16 shows examples of multiple opening and closing procedures with different structural elements such as: (a) diamond-shape with a 2-pixel distance from its center to its points; (b) disk-shape with a 2-pixel radius; (c) octagon-shape with a 3-pixel distance between its center to its sides; (d) a horizontal line of 2-pixel length; (e) a rectangle-shape with 3x2 pixels; and (f) a 3x3 pixel square. Usually, increasing the width, height or radius of the structural element, the erosion or dilation effect on image information is even more visible. By using morphological filtering procedures, the image is slightly enhanced but the most important drawback of this approach is the fact that the initial information is deformed and additional artefacts, unrelated to the original image, are added to the image.

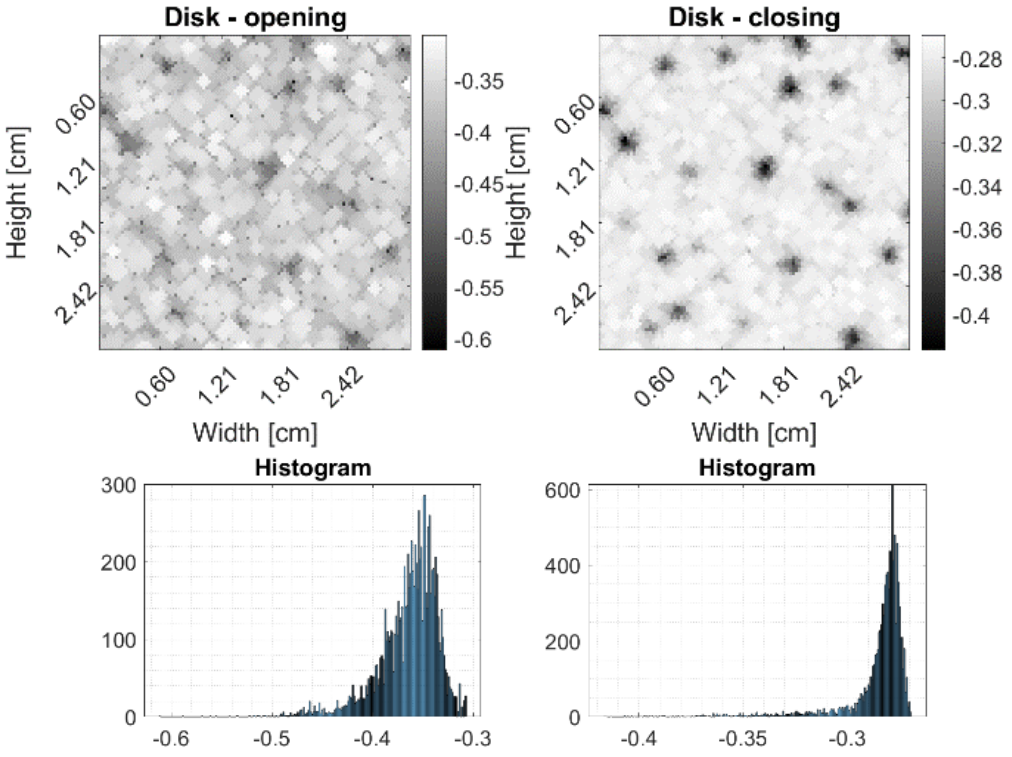
Other drawback of morphological filtering is the computational complexity, particularly for large images. Moreover, as seen in Figure 2-16 morphological operations can inadvertently lead to a loss of fine-grained details within images, potentially impacting the fidelity of processed results. Sensitivity to noise further compounds these challenges, necessitating careful consideration during parameter selection to balance noise suppression with preservation of essential image features. Additionally, the subjectivity in parameter tuning and the potential for boundary effects underscore the nuanced nature of morphological filtering, demanding expertise and iterative refinement for optimal outcomes.

We can mention other techniques such as noise removal using Wiener filters [71], linear contrast adjustment [72], median filtering [73] and unsharp mask filtering [74], which yield similarly results as the ones presented.

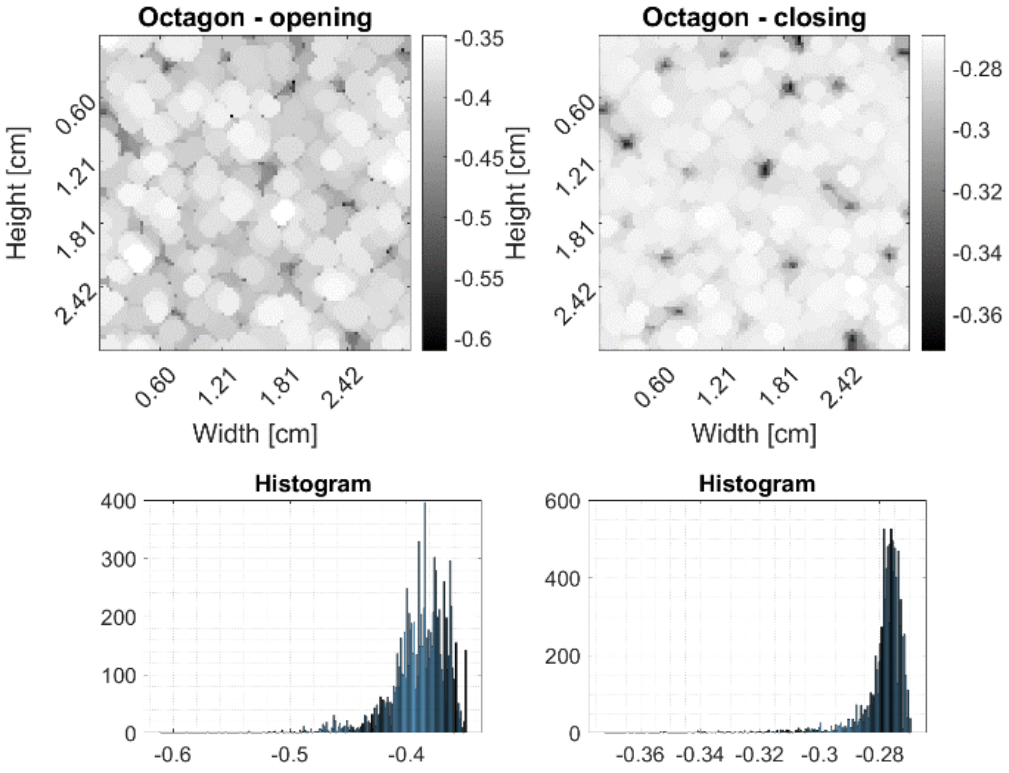
While all the methods are easy to apply and provide decent individual results in some cases, the post-processing chain consists of several enhancement techniques combined. This results in a time-consuming image reconstruction with a requirement of additional computational resources. Another problem is related to the careful algorithm adjustment that is needed to achieve the best outcome. However, even with a thorough preparation of the post-processing chain, the results are not guaranteed to be of high quality, thus enforcing the idea of a new method that can improve the overall quality of the image at the physical level, reducing noise influence and the complexity of the imaging system in terms of processing.


Diamond - opening
Diamond - closing
Height [cm]
Width [cm]
0.60
1.21
1.81
2.42
-0.35
-0.4
-0.45
-0.5
-0.55
-0.6
-0.28
-0.3
-0.32
-0.34
-0.36
-0.38
-0.4
Histogram
300
200
100
0
600
400
-0.6
-0.5
-0.4
-0.3
-0.35

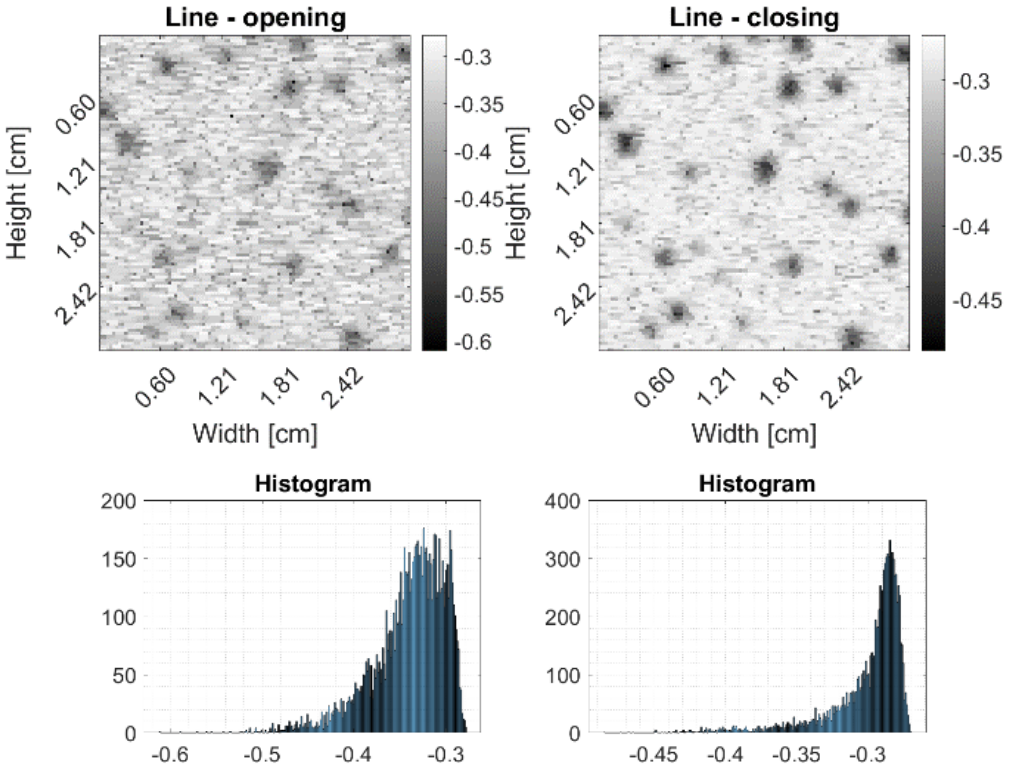
(a)


Disk - opening
Disk - closing
Height [cm]
Width [cm]
0.60
1.21
1.81
2.42
-0.35
-0.4
-0.45
-0.5
-0.55
-0.6
-0.28
-0.3
-0.32
-0.34
-0.36
-0.38
-0.4
Histogram
300
200
100
0
600
400
-0.6
-0.5
-0.4
-0.3
-0.35

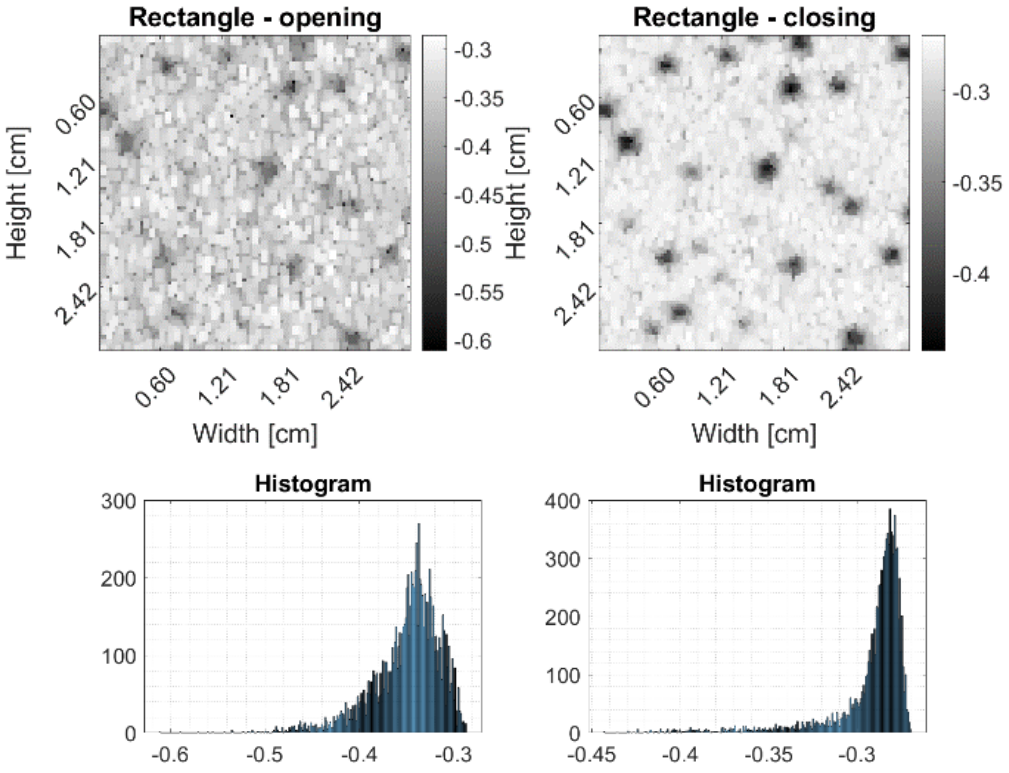
(b)


Octagon - opening
Octagon - closing
Height [cm]
Width [cm]
0.60
1.21
1.81
2.42
-0.35
-0.4
-0.45
-0.5
-0.55
-0.6
-0.28
-0.3
-0.32
-0.34
-0.36
Histogram
400
300
200
100
0
600
-0.6
-0.5
-0.4
-0.36
-0.34
-0.32
-0.28

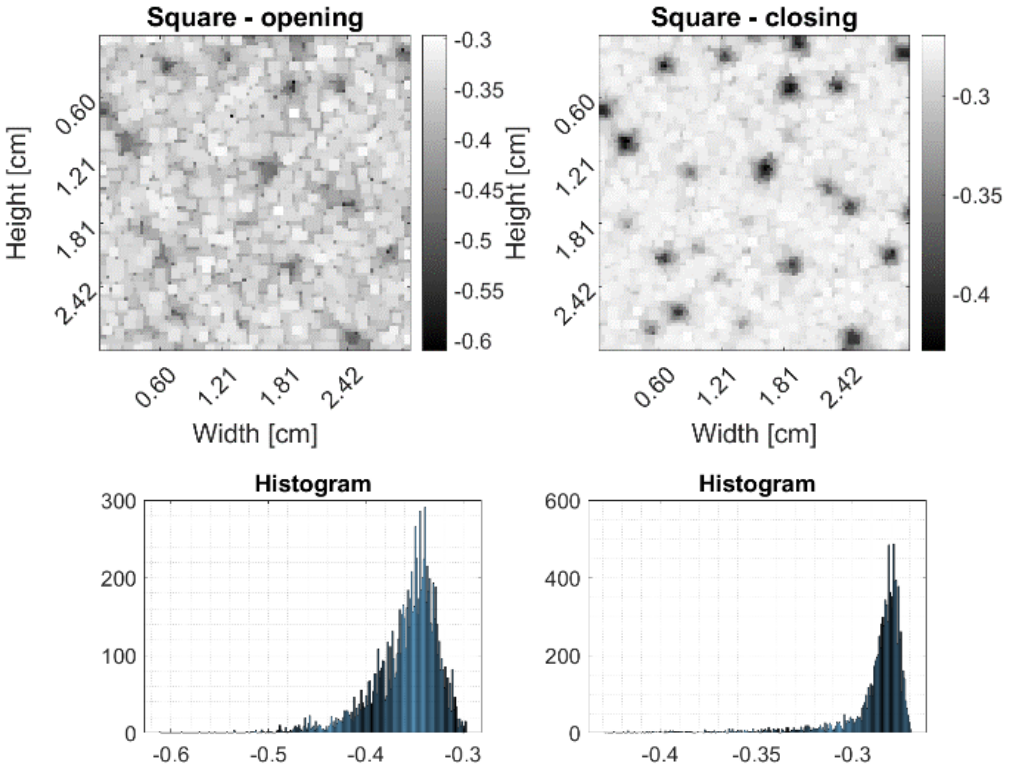
(c)


Line - opening
Line - closing
Height [cm]
Width [cm]
0.60
1.21
1.81
2.42
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
-0.6
Histogram
200
150
100
50
0
400
300
-0.6
-0.5
-0.4
-0.3
-0.45
-0.35

(d)



(e)



(f)

Figure 2-16. Multiple opening and closing procedures with different structural elements: **(a)** diamond-shape; **(b)** disk-shape; **(c)** octagon-shape; **(d)** a horizontal line; **(e)** a rectangle-shape; **(f)** square-shape

2.4.5 Subsection concluding remark

Finally, our conclusion is that the post-processing methods can off course improve the quality of THz images but the quality of the initial formed THz image is naturally crucial. Table 1 presents an overview of the advantages and disadvantages of the state-of-the-art. This is the reason why we have oriented our researches in the sense of proposing a new methodology to construct the THz images.

In the following chapter, we introduce the concept of the phase diagram, followed by the development of five innovative image reconstruction methods.

Method		Advantage	Disadvantage
Pre-processing	**Time-domain methods**	• Easy to implement • Lower computational resources • For TSI, TSII and TSDI - control over the timestamps • TOFI – possibility of imaging in layers • Perform well in laboratory contexts	• The choice of the timestamps for TSI, TSII and TSDI. • Sensitive to multiplicative perturbations like temperature drifts and delay of the optical line.
	Frequency-domain methods	• Exploit the spectrum • Controlled resolution of image by using different frequencies • Possibility to slightly improve the image under multiplicative perturbations • The multiplicative perturbation cannot be removed as it deforms the amplitude, frequency and phase	• The choice of frequency stamps for FSI, FSII, FSDI. • Filtering is not optimal as the multiplicative noise is present in the whole spectrum. • Sensitive to multiplicative perturbations like temperature drifts and delay of the optical line.
Post-processing	**Hist. Eq. Morphological filtering**	• Can improve the contrast of images in controlled environment	• Dependent on the image reconstruction algorithm • Deforms the information present in the image

Table 1. Advantages and disadvantages of the state-of-the-art methods

Chapter 3. Phase-Diagram method for image reconstruction

3.1 Introduction

A common observation arising from the examination of the state-of-the-art is the recognition of certain inherent disturbances that cannot be effectively eliminated through conventional pre-processing and post-processing methods. These disturbances typically stem from noise components or the laser pointing effect caused by yaw in the delay line of THz-TDS systems [75]. Therefore, our approach aims to mitigate the impact of low-frequency multiplicative perturbations on the quality of THz images by employing the concept of phase diagram. [76].

3.2 Introduction of Phase Diagram Representation

The representation of phase diagrams is a concept typically utilized within the domain of nonlinear dynamic system analysis. [77]. Recently, its primary advantage as a data-driven technique, which doesn't necessitate any model for data analysis, has broadened its application scope, leading to successful implementation across various fields such as: telecommunications to identify wide band modulations [78], system monitoring in order to estimate instantaneous frequency laws of signals [76], unmanned aerial vehicles' movement characterization [79], etc. However, in the context of THz imagery, the phase diagram provides a robust representation of transient signals and, therefore, a potentially better amplitude estimation and capacity to generate high resolutions images [80].

To gain a clearer comprehension of this representational space in the context of THz imagery, it is essential to define this technique based on a THz signal measured when imaging a sample. We consider the real part of the reference THz pulse defined in (2.1) of length N as follows:

$$s = \left\{ s[n] \mid n = \overline{1, N} \right\} \tag{3.1}$$

To transpose the time-series into its phase diagram representation, we define state vectors as in:

$$\overrightarrow{P_n^m} = \sum_{k=1}^{m} s[n + (k-1)d] \cdot \overrightarrow{e_p} \tag{3.2}$$

where m is the embedding dimension of the phase space, d is the delay between the signal's samples, $\overrightarrow{e_p}$ is the unit vector of the axis that defines the phase space and $M = N - (m-1)d$ is the number of points in the trajectory. The delay choice is important as it must ensure the balance between redundancy and irrelevancy in phase space [80].

If the delay is chosen too small, the state vectors will be similar to one another and their distribution defined around the principal diagonal of the space, thus a redundant trajectory is generated. On the other hand, if the delay is too large, the state vectors will be uncorrelated and an initially simple trajectory can turn into a complicated and irrelevant space representation. Determining the optimal delay estimate poses a challenge since we lack prior knowledge about the characteristics of the original phase space. In practice, the delay is chosen by computing the mean mutual information [81]. It is worth noting that there is no generally acceptable technique for delay

estimation and that the estimate is only a recommendation, generally considered as a starting point from where it is fine-tuned to offer the best representation in the phase space.

The embedding dimension, m, is calculated using a false nearest neighbor algorithm (FNN) [81]. FNN is based on the observation that when m has a small value, the real trajectory is projected onto a lower dimensional space, resulting in an increase of the false neighbors due to the folding of the trajectory. If m has a large value, the number of the false neighbors decreases due to the unfolding of the trajectory.

The algorithm of phase space reconstruction for a 3-dimensional case is depicted in Figure 3-1 where the coordinates of the trajectory are represented by delayed values of the time series.

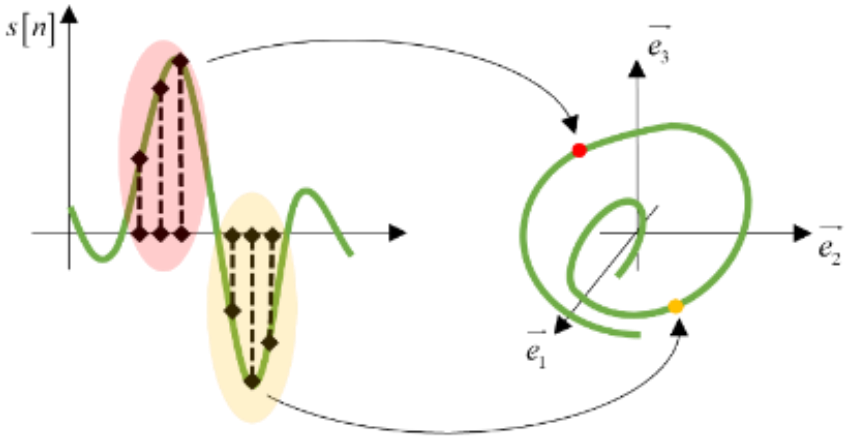


Figure 3-1. Phase Diagram reconstruction concept

All the proposed algorithms can be explained based on the following preliminary notation. Starting from the time series of length N defined as in (3.1), we can express its phase space trajectory in a matrix form as in:

$$\gamma[p,q] = s\left[q + (p-1)d\right] \tag{3.3}$$

where p runs over all dimensions up to m, q iterates through all M points in the phase space, and finally, d and m are the phase space reconstruction parameters. In other words, the matrix form contains on each column the m-dimensional cartesian coordinates which construct the phase trajectory.

We started the research by looking into and extrapolating the match-filter (MF) concept from the time-domain. As we have already described in State-of-the art image reconstruction methods in THz imagery, the MF is not optimal when the convolutive transmission channel is affected by the multiplicative perturbation.

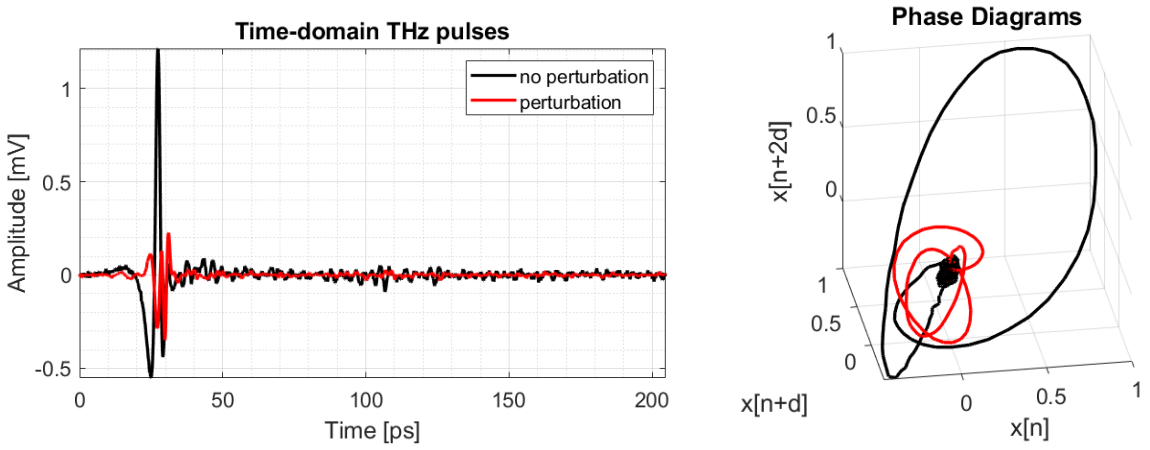
However, by generalizing the MF in order to be applied to phase diagrams, we could be able to better emphasize the information and simultaneously, to remove the effects of the multiplicative noise. The phase diagram approaches are considered time-domain reconstruction methods, as the phase trajectories are constructed from the time-domain signals.

3.2.1 Algorithm 1: Phase-diagram Matching

Our first approach uses the MF concept [82] in the phase diagram domain to fit a reference trajectory to our data. In the one-dimensional case, but also in higher dimensionality, MF is the optimal linear filter that maximizes SNR in an additive stochastic noise environment. Even though,

our context is under the assumption of multiplicative noise, MF is considered as a starting point in our search for an efficient reconstruction algorithm.

To illustrate the approaches, we consider two time-domain THz pulses as shown in Figure 3-2 (a), one being the received pulse under ideal conditions and the other is the same but under synthetically introduced multiplicative perturbation. The deformations of the multiplicative noise are more complex than the amplitude attenuation and space shifting, as we observe them in the time domain and the phase space in Figure 3-2 (b). The perturbed signal most challenging characteristic is the phase deformation which is better emphasized in the phase space in Figure 3-2 (b). Despite the observable attenuation of the THz pulse, the phase diagrams are scaled to 1 to illustrate the multiplicative noise impact on the phase.



(a) (b)

Figure 3-2. Two time-domain THz pulses and their phase diagram representation.

A phase diagram matching procedure can be defined as the convolution between the reference phase diagram, γ_{ref}, and a reversed one, $\gamma_{x,y}$, corresponding to an (x,y) pixel location in an image as in:

$$C_{x,y}[j,k]=\sum_{p=1}^{m}\sum_{q=1}^{M}\gamma_{ref}[p,q]\gamma_{x,y}[j-p+1,k-q+1] \tag{3.4}$$

where $j=\overline{1,2m-1}$ and $k=\overline{1,2M-1}$.

Therefore, a Phase Diagram Match Filter Image (PD-MFI) can be defined as the product between the highest convolution values per dimension as in:

$$\text{PD-MFI}[x,y]=\prod_{p=1}^{m}\max\left\{C_{x,y}^{p}\right\}, \tag{3.5}$$

where $C_{x,y}^{p}$ denotes convolution coefficients on the p^{th} dimension when processing the pixel information at (x,y).

An example of phase diagram MF is presented in Figure 3-3, where we use the two phase diagrams presented in Figure 3-3 to compute the convolution in (3.4). The result is represented in a three-dimensional phase space, the same space as the original diagrams. In this scenario, the

product between the maximum convolution value on each dimension is used as a pixel intensity value.

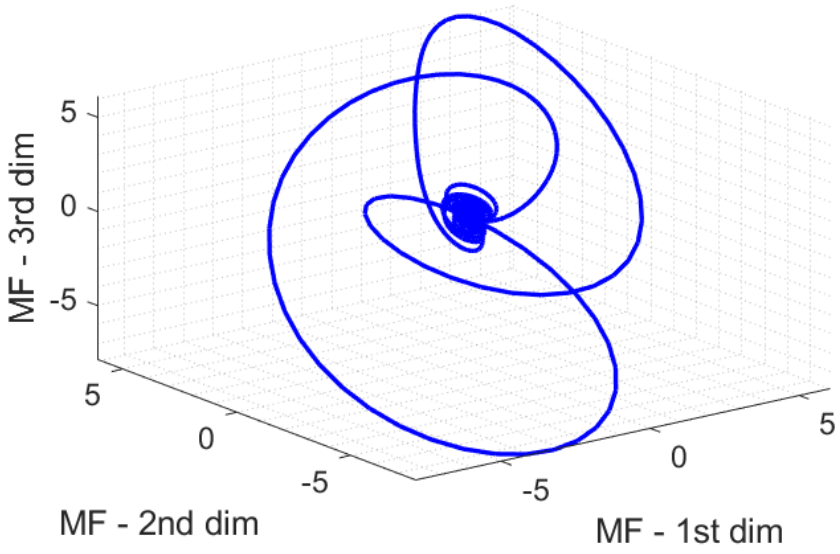


Figure 3-3. Phase Diagram MF result represented in the phase space

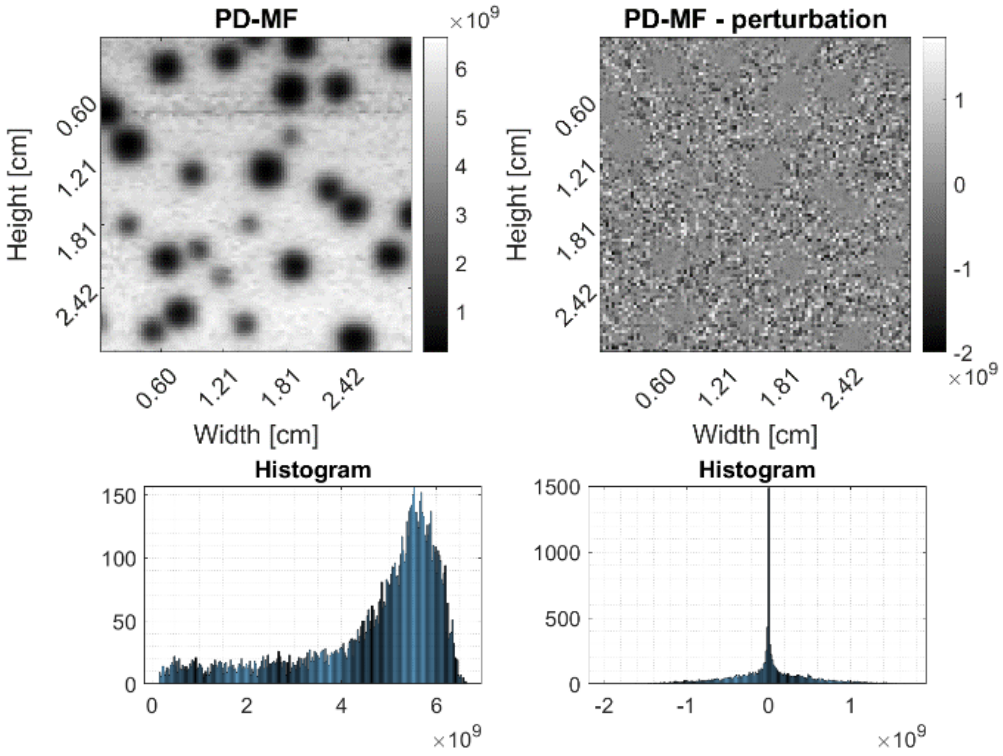


Figure 3-4. PD-MF reconstructed image.

From a physical perspective, match filtering is used to detect the most similar phase diagram information with respect to the reference diagram. The reference diagram is the diagram reconstructed from the THz reference pulse which is used as the emitted waveform when imaging samples with our system. The similarities are subsequently transcribed by (3.5) in PD-MFI. Figure

3-4 presents a PD-MFI in ideal and noisy context. In the first scenario the reconstructed image has high-contrast and conspicuous motives indicating that the information is overestimated. The contour information is not accurately reconstructed as MF tends to dilate the motives in the image. The reason behind this comes from propagation-related phenomena such as reflection, diffraction and absorption, all which are not characteristics of a convolutive channel where MF is optimal. When THz waves pass from one medium to another (from air to metallic ink in our case) the pulse deforms. The more it deforms, the lower the energy detected by MF as the information differs from the template.

In the noisy case in Figure 3-4, due to the multiplicative nature of the perturbation, MF fails to reconstruct the image as all pulses have random phase deformation, oppressing the capacity to detect the information at physical level. The resulting image is mostly noise rendering the fact that PD-MFI is not optimal in the proposed noise context.

Another idea is to express the phase diagram convolution in (3.4) using state vectors of q length, considering $C_{x,y}^{p,q}$ as cartesian coordinates on the p^{th} dimension:

$$\overrightarrow{PC}_{x,y}^{p} = \sum_{p=1}^{m} C_{x,y}^{p} \cdot \overrightarrow{e_p} \tag{3.6}$$

and $\overrightarrow{e_p}$ are unit vectors.

A Phase Diagram Convolution Magnitude Image (PD-CMI) is expressed considering the maximum magnitude values of (3.6) per dimension as in:

$$PD\text{-}CMI\left[x,y\right] = \max\left\{\left(\sum_{p=1}^{m} C_{x,y}^{p,q}\right)^{2} \mid q = \overline{1,M}\right\}, \tag{3.7}$$

where the reconstructed image integrates the convolution information from all dimensions, as opposed to PD-MF where the information was weighted in-between dimensions by their product.

An example of PD-CMI approach is presented in Figure 3-5, where we also use the two phase diagrams presented in Figure 3-3 to compute the magnitude of the phase diagram convolution in (3.6). The maximum magnitude value is used as pixel intensity value.

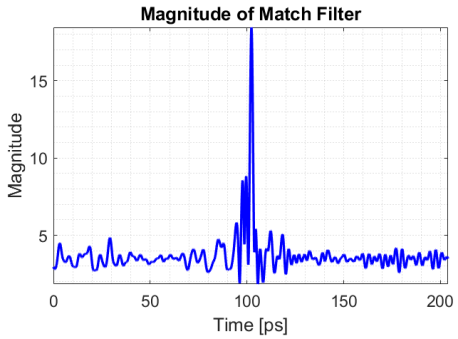


Figure 3-5. Magnitude of the phase diagram MF

Figure 3-6 presents a PD-CMI analyzed in both contexts. Compared to PD-MF, the resulting image has lower contrast in the ideal case, but it can retrieve some information in the perturbation scenario, transcribed by slightly highlighting the regions of interest.

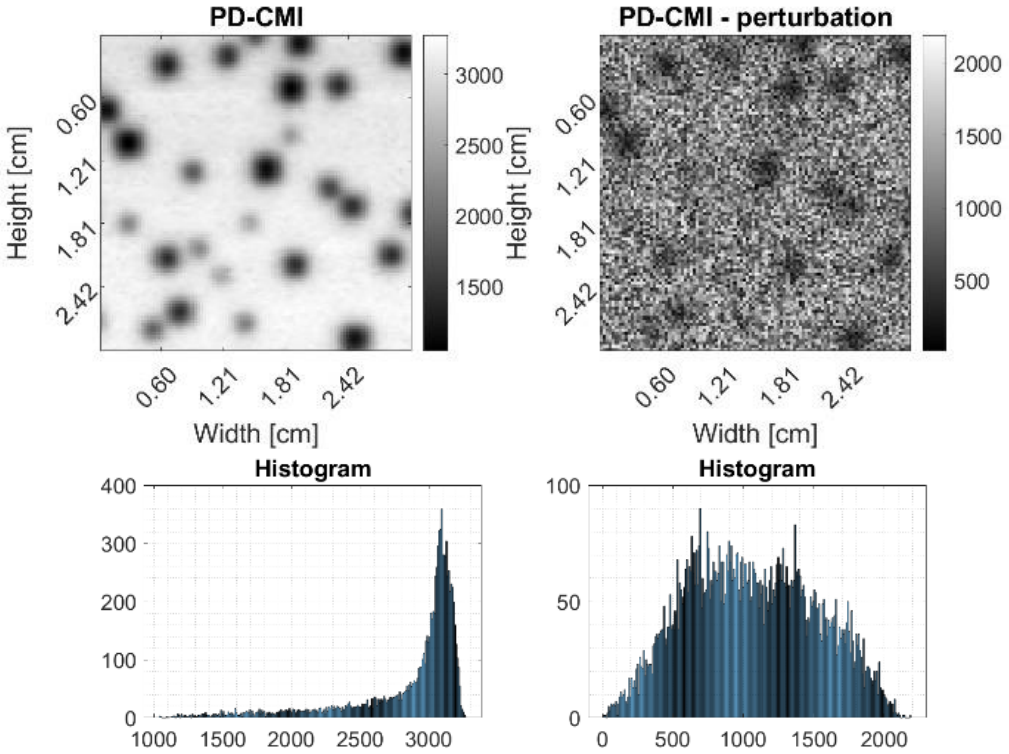


Figure 3-6. PD-CMI in ideal and noisy context

Despite the improvement of the image quality by using PD-MFI or PD-CMI, the performances are limited by the way the MF is computed. Dealing with deformations, which are more intricate than the attenuation and spatial shifts seen in convolutive transmission channels, has prompted further research to propose more robust techniques.

3.2.2 Algorithm 2: Phase-diagram Warping

In our second approach, we quantify the relative deformation of each phase trajectory with regards to the reference phase space. An optimal alignment and distance metric between two time-series can be obtain by using Dynamic Time Warping (DTW) [83]. As we exploit a m dimensional phase space the generalized version of the DTW is used [84]. The principle on which the algorithm is based comes from a physical observation that the longer the propagation path of a transient signal through a dispersive medium is, the higher its deformation is. As an example of one-dimensional case, two signals with equivalent features and components which are arranged in the same order can appear distinctive due to differences in the durations of their sections. Figure 3-7 shows an example of this case where we initially depict two signals, a monochromatic signal, Y, and a chirp signal, X. The DTW adapts one of the two signals in such way that the difference between the features is minimized and the similarities between signals are highlighted. The Euclidean distance is used to evaluate the similarity of the signals, where a distance close to 0 means the signals are similar, while a high value means the signals are distinctive.

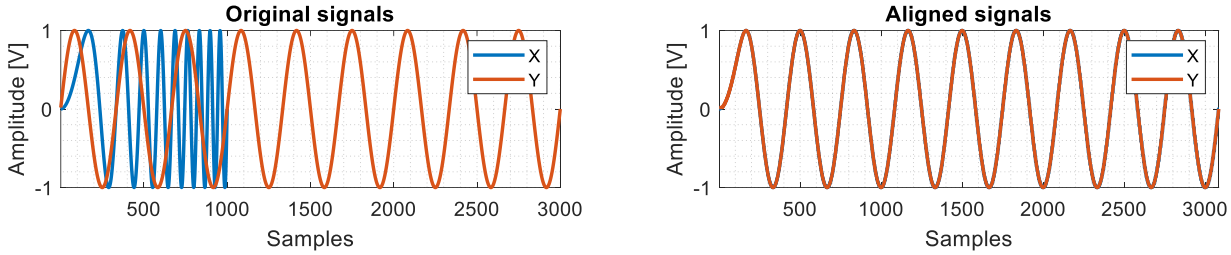


Figure 3-7. Example of DTW on two time series depicting similar components that differ in duration

In our multi-dimensional context, DTW will deform one phase space trajectory in order to match the reference one by minimizing the distance between them. Compared to MF, DTW is more versatile and can better fit the template on the actual data.

Considering the one-dimensional case, the warping cost between two time-series of length N can be expressed as:

$$DTW\left[s_1, s_2\right] = \min\left\{\sqrt{\sum_{k=1}^{K} p_k}\right\} \tag{3.8}$$

where p_k is the k^{th} warping path's point defining the alignment between two time series and $N \le K \le 2N-1$. The constraint is a consequence of constructing the distance matrix whose elements, $D\left[i, j\right]$ are the Euclidean distances between i^{th} and j^{th} sample of the first and second time series, respectively. Typically, the warping path must originate and terminate at diagonally opposite corners of the matrix, with warping steps constrained to adjacent values and points being monotonically spaced in time. The most straightforward way to determine the path is to compute the cumulative distance or the DTW as following:

$$D_c\left[i, j\right] = \left(s_1\left[i\right] - s_2\left[j\right]\right)^2 + \min\left\{D_c\left[i-1, j-1\right], D_c\left[i-1, j\right], D_c\left[i, j-1\right]\right\}, \tag{3.9}$$

where s_1 and s_2 are two time-series.

DTW is extended to multi-dimensional data considering two strategies: one computes the cumulative distances each dimension separately, and the other integrates all dimensions by redefining the distance metric from (3.9) as the Euclidean Distance between m-dimensional data points. We denote the former as DTW_I and the latter as DTW_D.

Considering the definition of one-dimensional DTW from (3.9), a Phase Diagram Independent Warping Image (PD-IWI) is constructed using DTW_I approach on the reference trajectory, γ_{ref} and the trajectory of a corresponding pixel location, $\gamma_{x,y}$ as in:

$$PD\text{-}IWI\left[x, y\right] = DTW_I\left[\gamma_{ref}, \gamma_{x,y}\right] = \sum_{p=1}^{m} DTW\left[\gamma_{ref}^{p}, \gamma_{x,y}^{p}\right], \tag{3.10}$$

where $\gamma_{ref}^{p}, \gamma_{x,y}^{p}$ denote the phase diagram's p^{th} dimension.

A Phase Diagram Dependent Warping Image (PD-DWI) is calculated using DTW_D concept by adjusting (3.9) to integrate information from all dimensions as in:

$$D_c\left[i, j\right] = \sum_{p=1}^{m}\left(\gamma_{ref}^{p,i} - \gamma_{x,y}^{p,j}\right)^2 + \min\left\{D_c\left[i-1, j-1\right], D_c\left[i-1, j\right], D_c\left[i, j-1\right]\right\}. \tag{3.11}$$

Using DTW_D's formalization in (3.11), the PD-DWI can be written as:

$$PD\text{-}DWI\left[x,y\right] = DTW_D\left[\gamma_{ref},\gamma_{x,y}\right]. \tag{3.12}$$

An illustration of the DTW_D is presented in Figure 3-8 (a) where we depict the phase diagrams' dependent alignment based on the monochromatic and chirp signals from Figure 3-7. In this scenario, DTW_D aligns concomitantly all dimensions of the original chirp's diagram in such way that it resembles the phase diagram of the sinusoid. This is better emphasized in Figure 3-8 (b) where we present the aligned phase diagrams for both signals. Figure 3-9 (a) shows the process of DTW_I, the independent phase diagram alignment for the same two signals. The superimposed phase diagrams are presented in Figure 3-9 (b).

When analyzing the two approaches, we note that for DTW_I, the result is better due to individual alignment of dimensions of the phase diagram. In this case, the aligned X-phase diagram is matched on the Y-phase diagram. For DTW_D, the alignment is not exact, as this algorithm performs on all dimensions simultaneously, making it more complex to have a better match of the two phase diagrams.

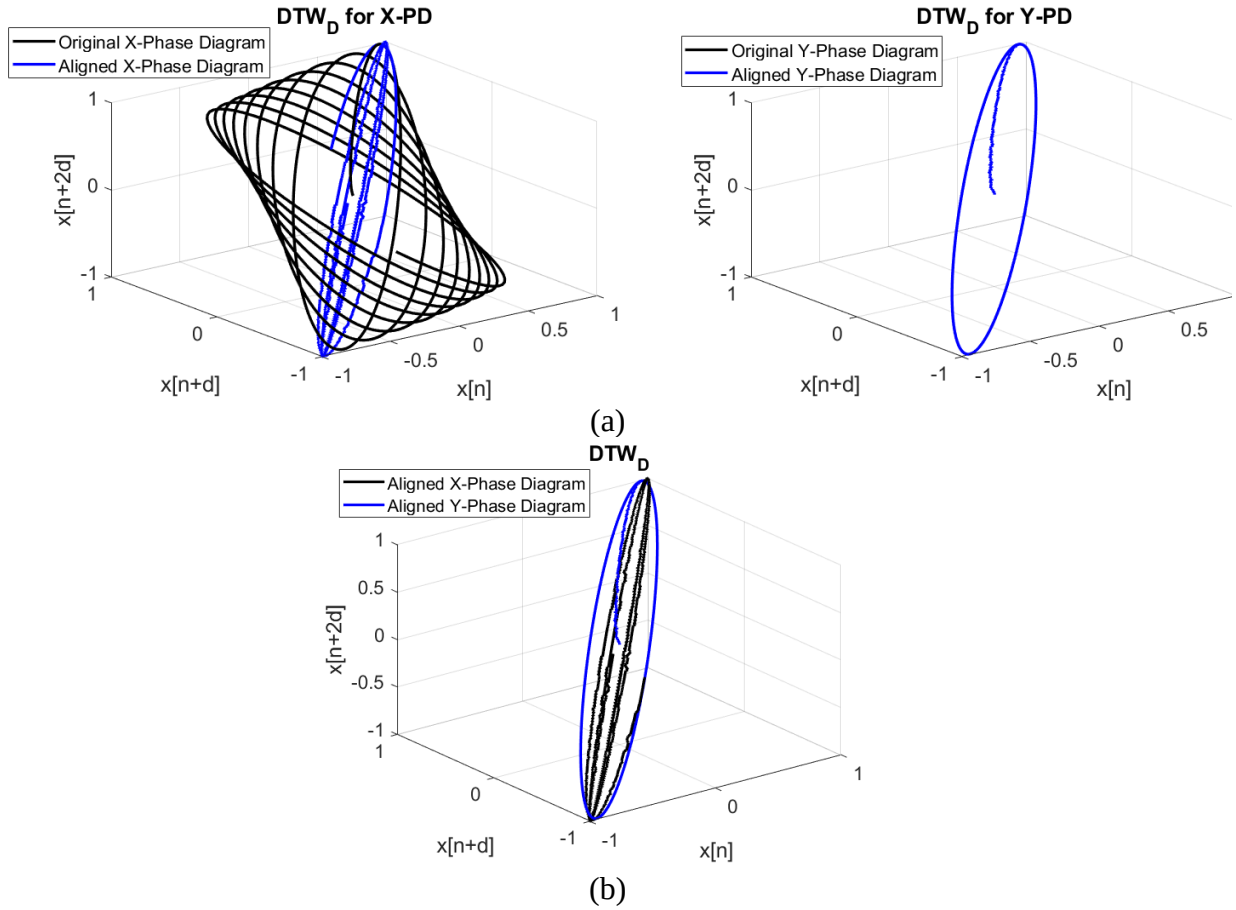


Figure 3-8. (a) DTW_D - Dependent phase diagram alignment of signals X (chirp) and Y (sinusoid); (b) Superimposed aligned phase diagrams of X and Y

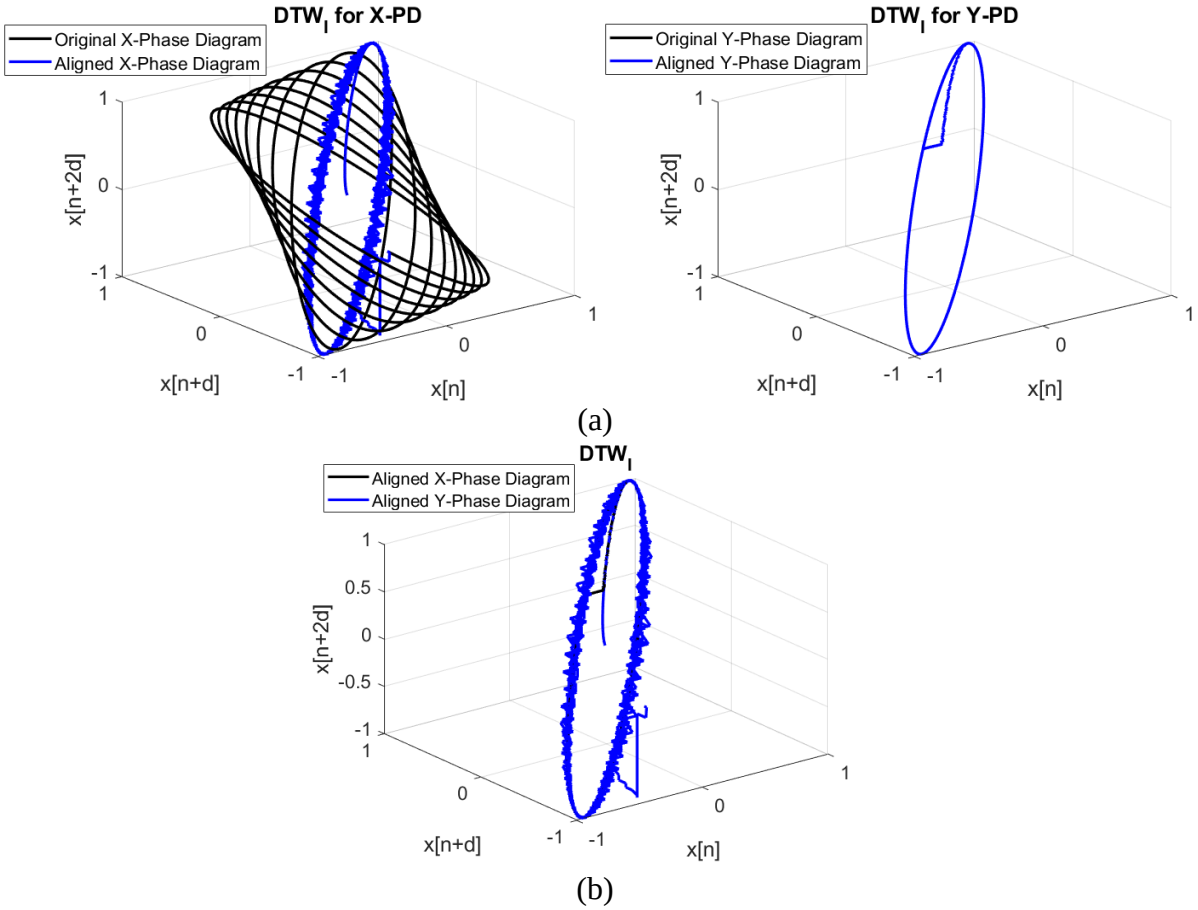


Figure 3-9. (a) DTW_I – Independent phase diagram alignment of signals X (chirp) and Y (sinusoid); (b) Superimposed aligned phase diagrams of X and Y

The resulting images using the two proposed approaches are presented in Figure 3-10 and Figure 3-11, respectively. In both approaches, in the ideal context, the reconstructed images have high quality. They successfully emphasized the information by having a constant low background level and high intensity motives which are interpreted as the image having high contrast. The consequence is the prominent separation between image components which certainly aids all further processing that an image can be subjected to. Between PD-IWI and PD-DWI, the former has a slight advantage due to the independent alignment of dimensions as it allows for better template matching between the reference and the analyzed trajectory. The reconstruction quality edge of the PD-IWI is also observed in the histograms where in the ideal case, the information is represented with a smaller intensity range than the other.

Finally, compared to MF, the additional degree of freedom introduced by DTW improves the reconstructed quality of the THz images in ideal environment. Conversely, when the multiplicative perturbation is present, the results are slightly better than with a MF approach, but not what it is expected from a reconstruction algorithm faced with a noisy environment. Some of the information is reconstructed, but the remanent noise still needs to be further processed to be eliminated. The histograms better reflect the quality of the reconstruction algorithm, where at least for PD-IWI, the contrast is higher.

Due to the observed limitations of MF and DTW-based methods, we continue our research by exploiting the dynamic evolution of the phase trajectory.

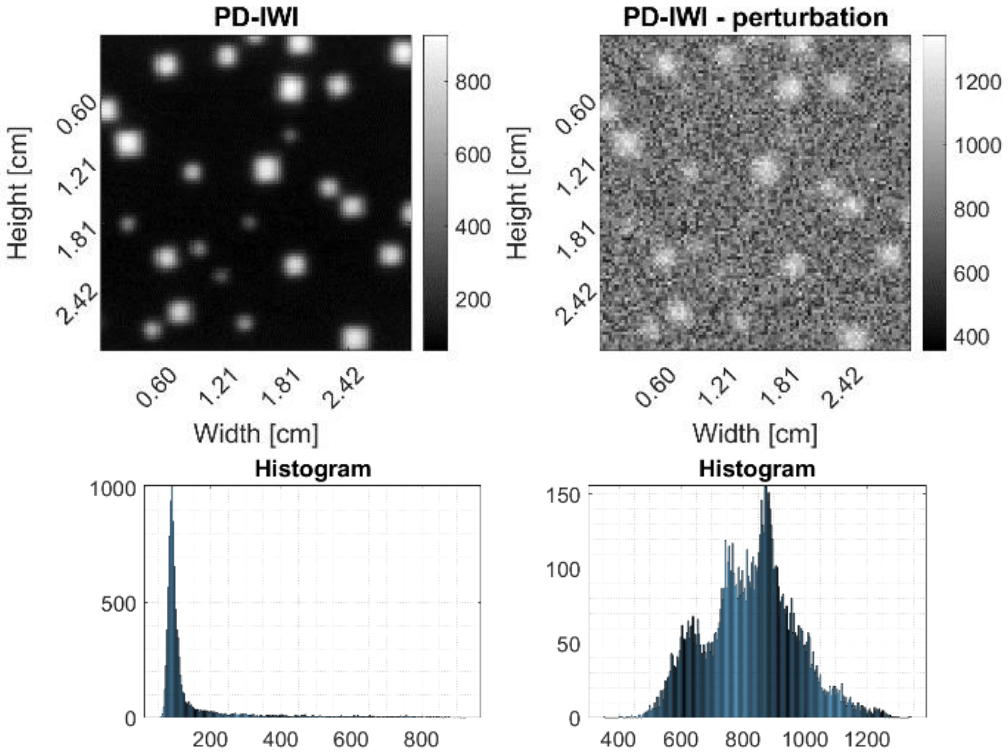


Figure 3-10. PD-IWI in ideal and noisy context

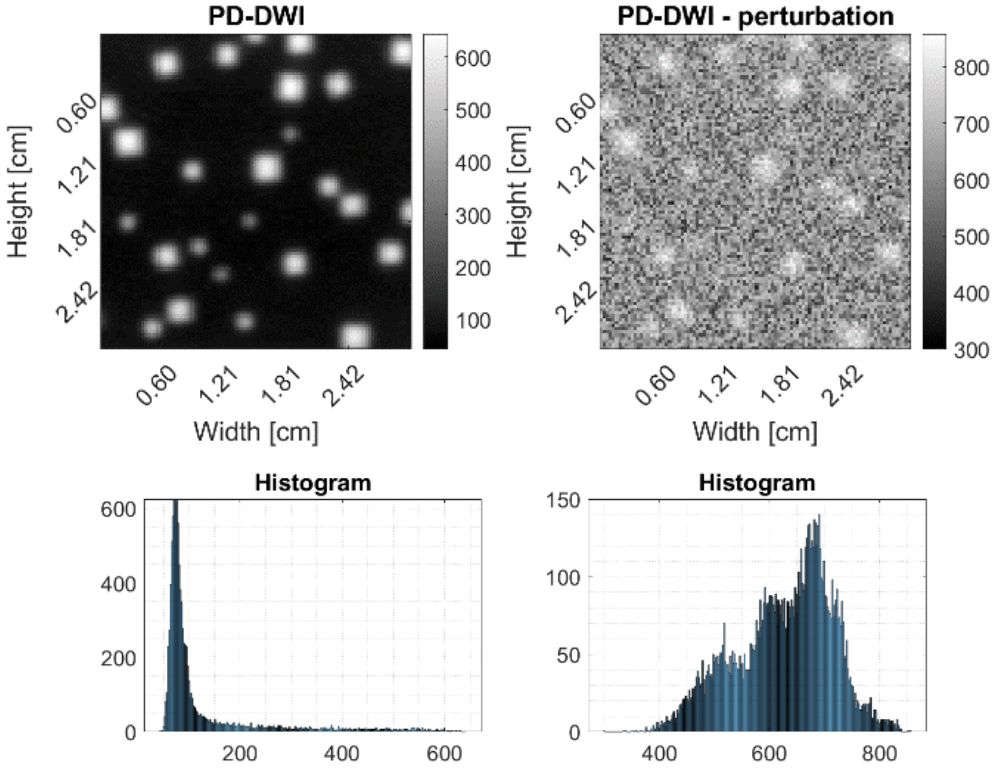


Figure 3-11. PD-DWI in ideal and noisy context

3.2.3 Algorithm 3: Frenet-Serret frame of a space curve

Our third approach is based on a characterizing the dynamics of the phase diagram using the mathematical concept of curvatures. The algorithm is depicted in Figure 3-12 and it is iteratively applied for every THz pulse by transposing the time-domain signals into their phase space. The next steps refer to a phase diagram characterization and quantization in order to estimate an intensity value corresponding to a pixel.

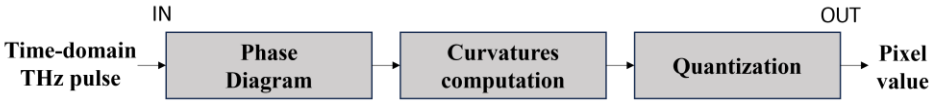


Figure 3-12. Quantization of phase diagram based on curvatures

Considering that our phase space is basically an evolving trajectory in a m- Euclidian space $\mathbb{R}^m$, it seems natural and con $B_q = T_q \times N_q = \frac{\gamma_q^{'} \times \gamma_q^{''}}{\left\| \gamma_q^{'} \times \gamma_q^{''} \right\|}$ venient to model its motion using the Frenet-Serret frame [85]. As the proposed method transposes the time-series into a three-dimensional phase space, where $m = 3$, the emphasis of the mathematical modeling is defined using this space constraint. Thus, this frame describes the trajectory using the first derivatives of tangent (T), normal (N) and binormal (B) unit vectors as follows:

$$\frac{dT}{d\delta} = \kappa N \tag{3.13}$$

$$\frac{dN}{d\delta} = -\kappa T + \tau B \tag{3.14}$$

$$\frac{dB}{d\delta} = -\tau N \tag{3.15}$$

where $\frac{d}{d\delta}$ represents the derivative with respect to the arclength δ, the parameters κ and τ represent the curvature and respectively the torsion of the curve. An example of Frenet-Serret phase diagram characterization is presented in Figure 3-13, where we emphasize the three vectors for each point of the trajectory.

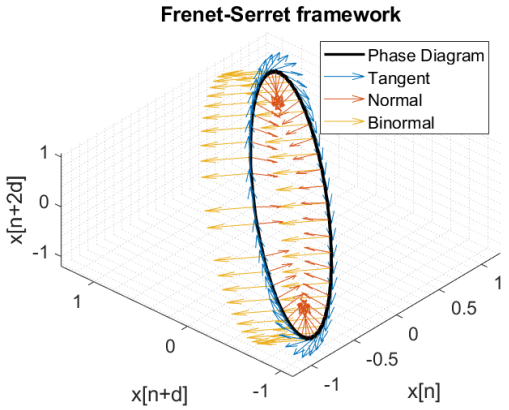


Figure 3-13. Frenet-Serret characterization of a phase diagram

The Frenet-Serret formulas given for the unit vectors depend on the curve being defined in terms of arclength parameter. In Euclidean geometry, this definition comes as a natural assumption as the arclength is a Euclidean invariant of the curve. Nonetheless, this parametrization it is not suitable to work with in practice due to the numerical complexity of the arclength's computation. In order to reduce the complexity, in practice, an alternative definition for the curvature it is usually considered.

Defining a three-dimensional phase diagram from (3.3) the Frenet-Serret frame can be rewritten as following:

$$T_q = \frac{\gamma_q^{'}}{\left\|\gamma_q^{'}\right\|} \tag{3.16}$$

$$N_q = \frac{T_q^{'}}{\left\|T_q^{'}\right\|} = \frac{\gamma_q^{'} \times \left(\gamma_q^{''} \times \gamma_q^{'}\right)}{\left\|\gamma_q^{'}\right\| \cdot \left\|\left(\gamma_q^{''} \times \gamma_q^{'}\right)\right\|} \tag{3.17}$$

$$\tag{3.18}$$

where $\times$ is the cross product operator, $\|.\|$ denotes the magnitude defined in an Euclidean sense, T_q, N_q and B_q are the unit tangent, normal and binormal vectors at the q^{th} point of the phase space $\left(q = \overline{1, M}\right)$, and $\gamma^{'}$, $\gamma^{''}$ represent the first and second order derivatives of the phase trajectory. Following the frame alternative definition, the next step is to introduce the curvature, a parameter that dynamically describes our trajectory.

Considering our phase space trajectory, the curvature can be formulated as:

$$\kappa[q] = \frac{\left\|\gamma_q^{'} \times \gamma_q^{''}\right\|}{\left\|\gamma_q^{'}\right\|^3} \tag{3.19}$$

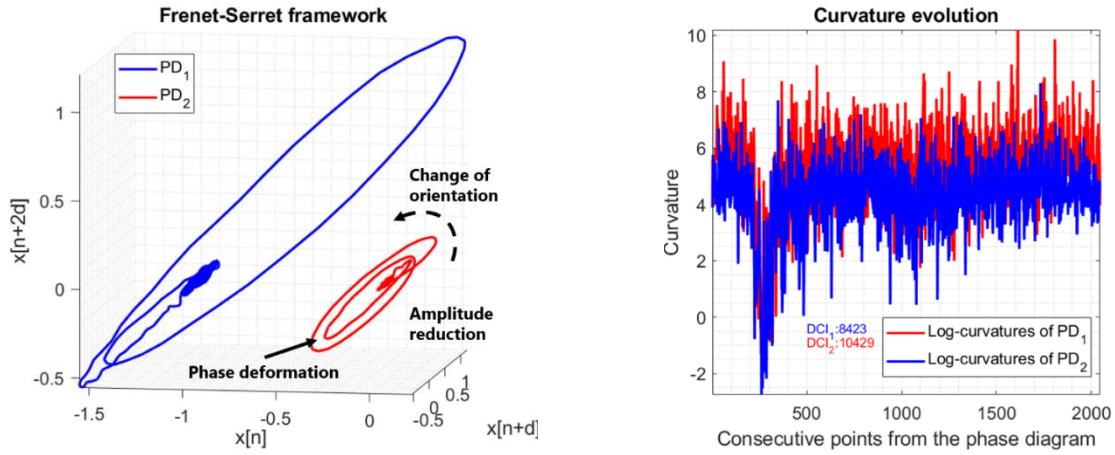
The interest in curvatures comes from the fact that they describe a three-dimensional curve intrinsically, without considering the embedding dimension and its extrinsic properties. In our case, an intrinsic property, such as the curvature, is invariant regardless the embedding dimension, m, of the phase diagram.

Generally, properties of curve which depend on the coordinate space in which it is embedded, are referred to as extrinsic. On the other hand, intrinsic properties are measured within the curve itself without any reference to another space. For an extrinsic property example, we consider the slope of a tangent line as it depends on the coordinate system in which it is defined. We expect to have highly variable curvatures due to derivatives being sensitive to phase changes. However, it does not affect our processing as the curvatures are proportional to deformations of a trajectory. Thus, a phase diagram, the more deformed is, the greater the values of its curvatures.

Finally, we can define the Phase Diagram - Dynamic Characterized Image (PD-DCI) as the mean of all logarithmical scaled curvature values:

$$DCI[x, y] = \frac{1}{M} \sum_{q=1}^{M} \log\left(\kappa_{x,y}[q]\right) \tag{3.20}$$

where $\kappa_{x,y}[q]$ represents the curvature value at the q^{th} point of a M points long trajectory corresponding to a (x, y) pixel location.



(a) (b)

Figure 3-14. (a) Example of two phase diagrams, one exhibiting multiple deformations; and (b) the evolution of log-curvatures for both phase diagrams

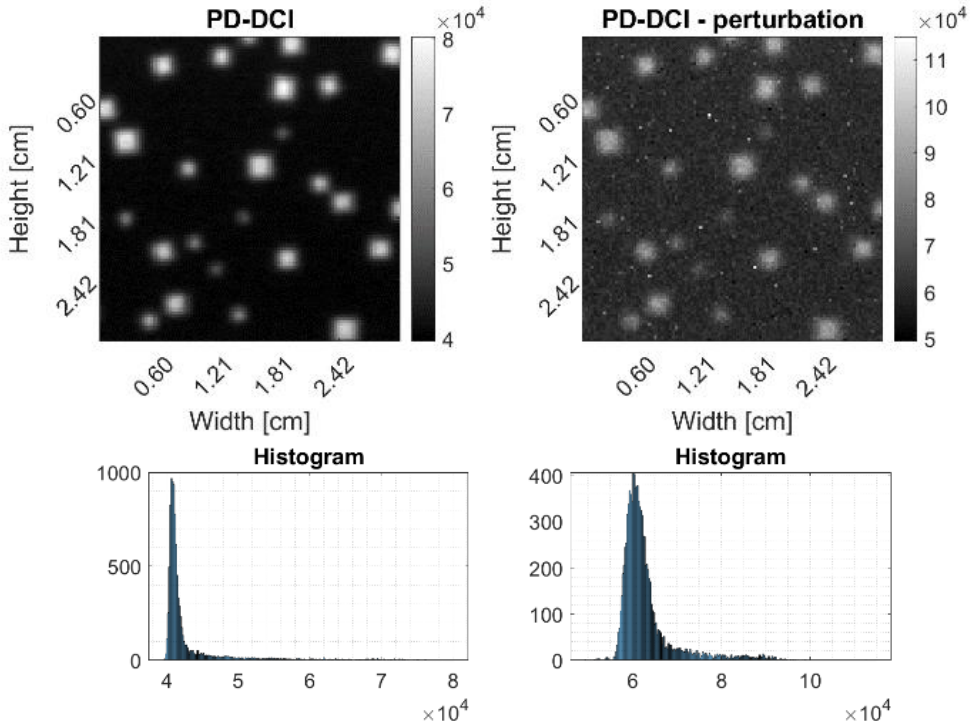


Figure 3-15. PD-DCI image in ideal and noisy context

An example of two phase diagrams is presented in Figure 3-14 (a), where one exhibits multiplicative noise patterns such as phase deformations, change in orientation and amplitude reduction. Figure 3-14 (b) shows the log-curvature evolution of the phase diagrams and the results of computing (3.20) on both trajectories. Averaging over all curvatures offers an accurate estimation of the phase trajectory deformation and noise. The logarithm is used in order to equilibrate the initial high difference between curvatures at different points, weighting noise and

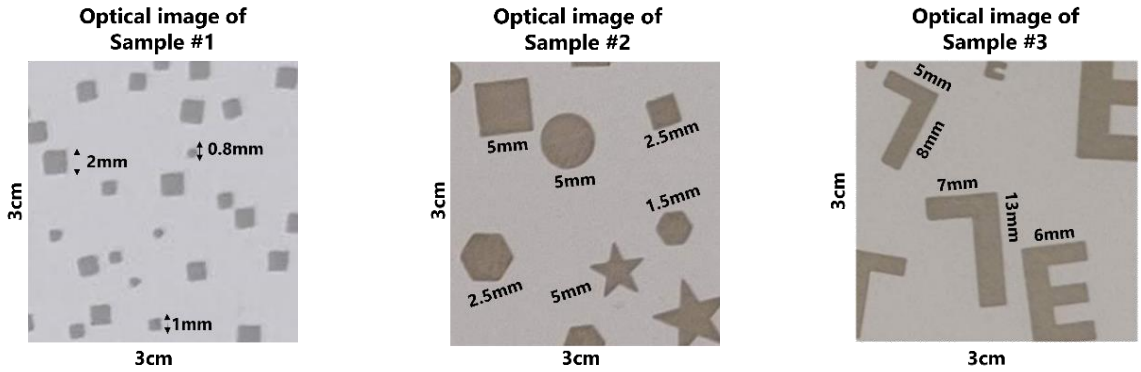
information in the phase diagram and eliminating the risk of information being overwhelmed by noise.

Figure 3-15 presents the result of our third approach. We observe that the ideal case is efficiently reconstructed recovering all contour information. Moreover, the method successfully reduced the multiplicative noise. The remaining background noise is minimal compared to other approaches. The motives are clearly defined and the naturally occurring smoothing effect on the borders is minimized. The results are reinforced by the peaks in the histograms indicating that the information is represented with comparable intensities. In the next subsection, we provide a comparative analysis between methods using image quality metrics and other samples with different information.

3.3 Comparative analysis of phase diagram methods

To cross-validate the proposed methods, we tested them on additional samples with different printed patterns. We kept the measurement and noise conditions constant across all image acquisitions and perturbation modeling to ensure consistent evaluations. To assess image quality, we applied a simple thresholding method, as referenced in [86], to segment the background from the information. This approach helps identify images with higher contrast and clearer information. Additionally, for a correctly reconstructed image, the shapes' contours should be coherent. In this context, coherence means having clearly defined contours and a distinct separation between motifs. This ensures that the shapes are accurately represented without significant distortions or artifacts. By using this thresholding method, we can determine which images exhibit the best quality in terms of contrast and information prominence. This process verifies the robustness and reliability of our methods across different samples, confirming their effectiveness in various scenarios.

For comparison purposes, the optical images of the 3cm-by-3cm samples used in this section are presented in Figure 3-16 (a), (b) and (c). The first sample in Figure 3-16 (a) is composed of polyethylene substrate with metallic ink square patterns printed on it. The patterns have their size varying from 0.8mm up to 2mm. The second sample in Figure 3-16 (b) has multiple geometric figures such as squares, circles, hexagons and stars with their size varying from 2.5mm and up to 5mm. The substrate is polyethylene and the patterns are printed with metallic ink. The last sample in Figure 3-16 (c) has also polyethylene substrate with letters from metallic ink. The size of the patterns from the image varies between 5mm up to 13mm.



(a) (b) (c)

Figure 3-16. Optical images of the samples used for methods' comparison

Firstly, we start by analyzing the first sample (Sample #1) used to define the state-of-the-art and our methods in Section 2.4 and Chapter 3. As illustrated in Figure 3-17, applying a thresholding method in the ideal scenario allows all methods to separate the information from the background, though with some noticeable differences. For instance, with MPI, PD-MF, PD-CMI, and PD-DWI, some of the shapes are linked to each other, which is generally undesirable.

In contrast, when observing FSI and PD-DCI, these methods show similarities in terms of the information they convey, but there is a trade-off between linking information and achieving contrast. Specifically, in the case of this sample, FSI provides better separation between motifs as it considers a single, higher frequency when reconstructing the image. This means that the higher the frequency used, the lower the contrast of the image, as the energy of higher frequencies is more susceptible to propagation phenomena.

On the other hand, PD-DCI reconstructs the image by taking into account all frequencies in the phase diagram. Consequently, we expect to see a slightly narrower separation between shapes but with higher contrast and a consistent background level. The ultra-wide bandwidth of the THz pulse, approximately 5 THz, makes our method robust against propagation phenomena and various types of perturbations.

The most significant advantage of PD-DCI over all other methods becomes evident under the influence of noise. Among all the methods, PD-DCI is unique in its ability to reconstruct an image from noise and achieve overall high quality with consistent images in different contexts. This robustness highlights the superiority of PD-DCI, particularly in noisy environments, making it a reliable choice for image reconstruction tasks that require both high contrast and stability against perturbations.

In Figure 3-18, we analyzed a different sample (Sample #2) under the same conditions as the first. This sample presents larger shapes of various types, including stars, squares, hexagons, and circles. The diversity and size of these shapes result in a larger absorption area, posing a significant challenge for some classic methods due to pulse deformation during propagation.

In an ideal scenario, methods such as MPI, PD-CMI, PD-IWI, PD-DWI, and PD-DCI exhibit superior reconstruction capabilities. The shapes within the binarized images produced by these methods are distinctly defined, demonstrating their effectiveness. However, a notable observation pertains to the star shapes: in all cases except PD-DCI, the contours of the star shapes are deformed, indicating a limitation in these methods' ability to maintain shape integrity under certain conditions.

Furthermore, the PD-MF approach proves unsuitable for this sample, as it causes the shapes within the image to link together, thereby failing to preserve their individuality. Additionally, the FSI method is ineffective in reconstructing the image due to the attenuation of the selected frequency by the metallic ink, which absorbs the frequency and diminishes the image quality.

These findings underscore the importance of selecting an appropriate method for image reconstruction, particularly when dealing with samples that have complex shapes and varying absorption properties. PD-DCI stands out as the most robust method in this context, maintaining the integrity of the shapes and providing high-quality reconstruction despite the challenges posed by the sample's characteristics. This highlights PD-DCI's potential as a reliable tool for applications requiring precise and accurate image reconstruction.

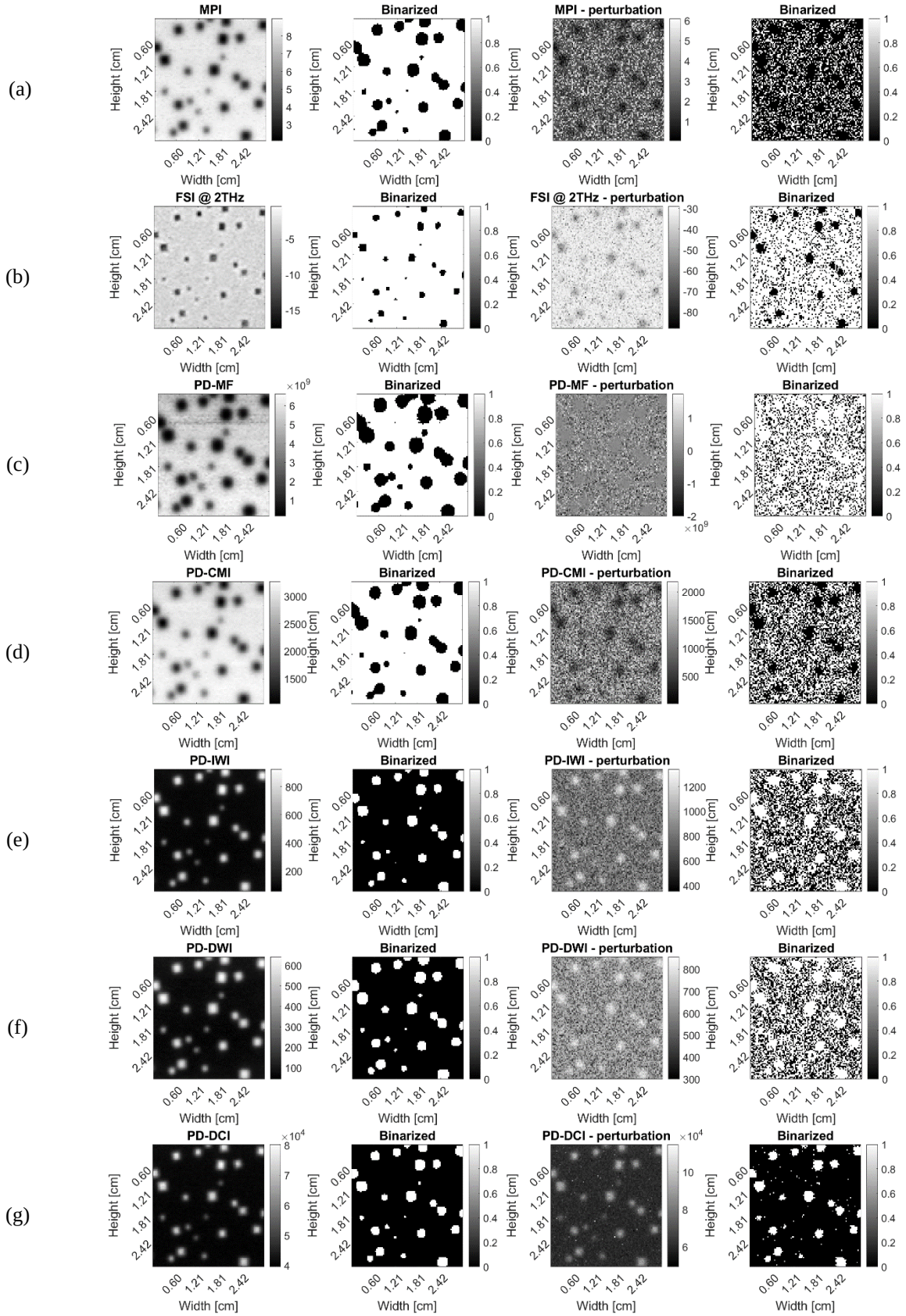


Figure 3-17. Segmentation results for a selection of methods including ours for an ideal and noisy context for sample #1: (a)MPI; (b) FSI; (c) PD-MF; (d) PD-CMI; (e) PD-IWI; (f) PD-DWI; (g) PD-DCI

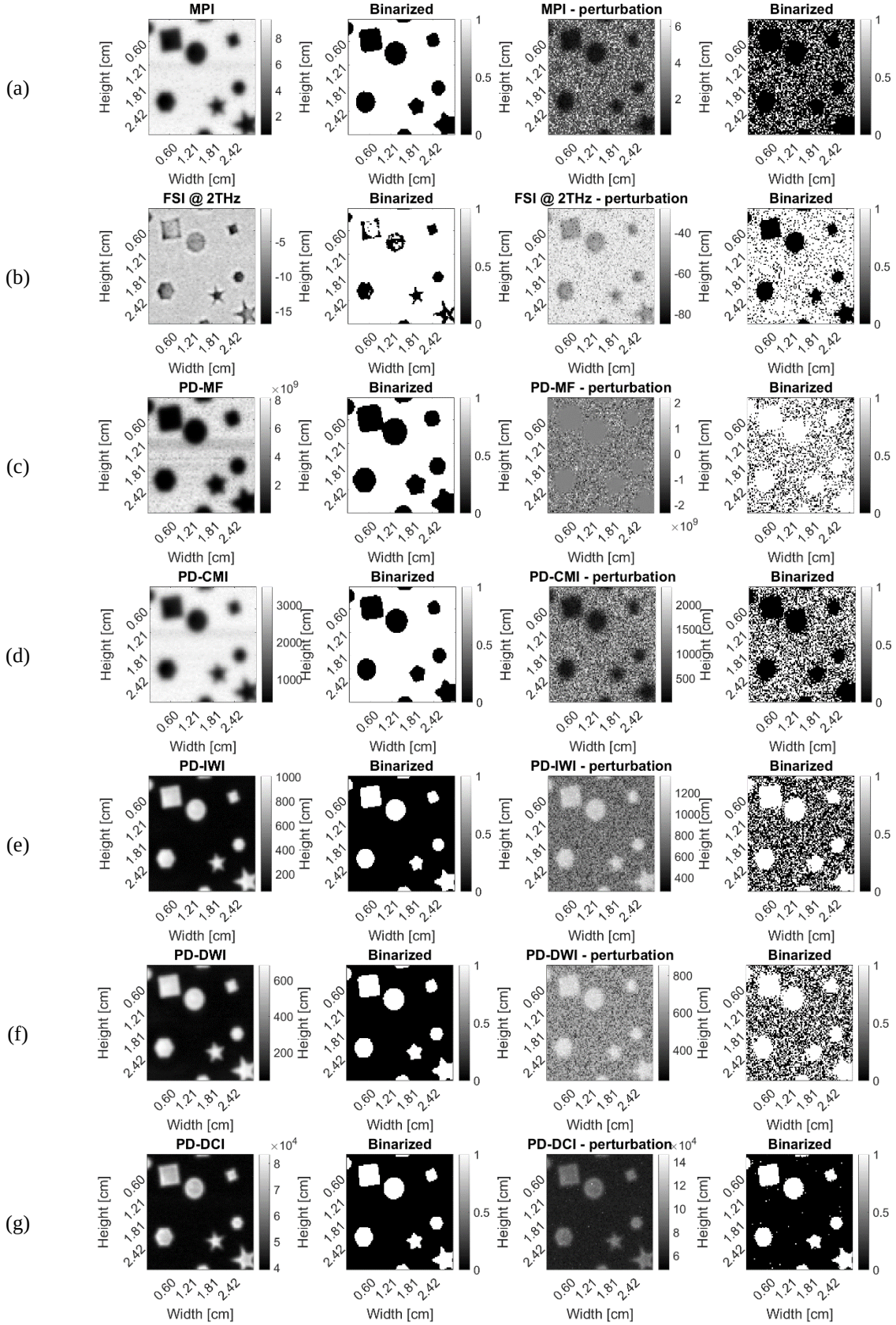


Figure 3-18. Segmentation results for a selection of methods including ours for an ideal and noisy context for sample #2: (a)MPI; (b) FSI; (c) PD-MF; (d) PD-CMI; (e) PD-IWI; (f) PD-DWI; (g) PD-DCI

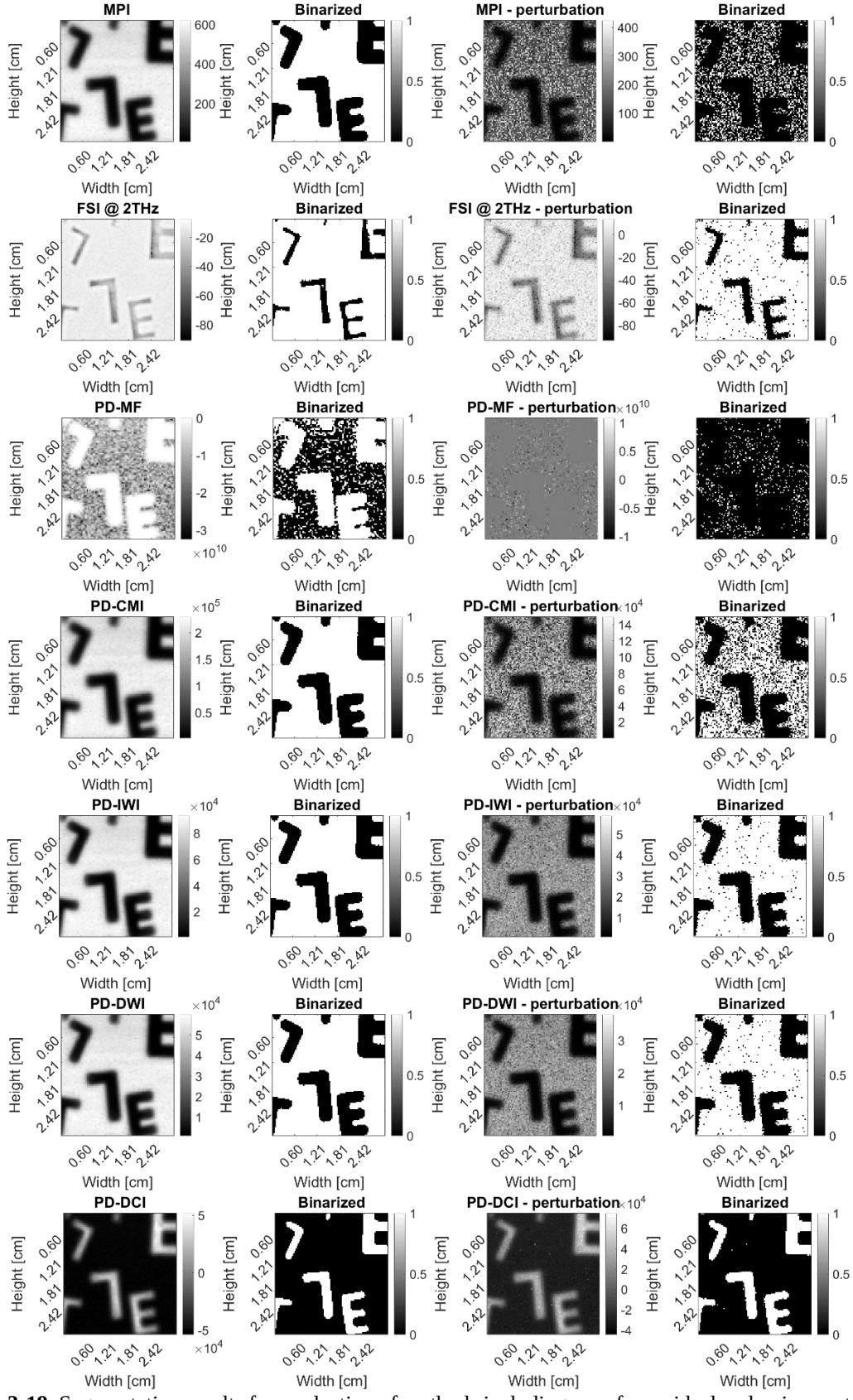


Figure 3-19. Segmentation results for a selection of methods including ours for an ideal and noisy context for sample #3: (a)MPI; (b) FSI; (c) PD-MF; (d) PD-CMI; (e) PD-IWI; (f) PD-DWI; (g) PD-DCI

Our next sample (Sample #3) is composed of the same polyethylene substrate but with another pattern printed on in. The sample presents two types of metallic-ink letters with random distribution. As we observe in Figure 3-19 but also with previous patterns, one of the main debates regards FSI and PD-DCI reconstruction methods. While all the other yield decent results in an ideal case, FSI and PD-DCI are by far two of the most promising to display an image. Nevertheless, due to the low contrast of FSI and larger absorption areas on the sample, parts of the letters are missing when segmenting FSI. In this situation, PD-DCI presents a more stable version of the image. In a noise context, FSI has remaining artefacts and incoherent letter borders. On the opposite side, PD-DCI shows little to almost no noise artefacts with a steady definition of the letters.

As for the moment, considering the experiments conducted with different samples, PD-DCI can be assessed as one of the most appropriate methods to reconstruct images. From the proposed methods, we mention that PD-CMI, PD-IWI and PD-DWI can also be used to an extent considering the tendency to overestimate the information and enlarge the motives. They can be successfully applied in order to determine regions of interest (ROIs) in images as they are able to detect the presence of information. When the objective is to eliminate artefacts as a result of multiplicative perturbations affecting the imaging system, PD-IWI, PD-DWI and PD-DCI are the methods to be applied.

Not all applications require image segmentation, therefore, a brief review of Figure 3-17, Figure 3-18 and Figure 3-19 shows us that our methods are more suitable to be used in image reconstruction. One of their primary benefits is their ability to offer enhanced contrast in the final image with a fairly constant background. Even in the perturbation context, PD-DCI has one of the best qualities compared to the others.

Further next, due to the subjective perspective of the previous analysis, we provide the results of full-reference image quality metrics such as Mean-squared error (MSE), Peak signal-to-noise ratio (pSNR) and Structural Similarity Index (SSIM). This analysis is required to objectively assess the quality of reconstructed images.

MSE is a measure of the average squared difference between an image and the ideal (or reference) image. A lower MSE value indicates a higher degree of similarity between the analyzed image and the reference image. It's important to note that while MSE is straightforward to compute, it may not necessarily correlate well with human perception of image quality. The formula of MSE for two images is defined as in:

$$MSE\left[I_1, I_{ref} \right] = \sum_{x=1}^{W} \sum_{y=1}^{H} \left(I_1[x,y] - I_{ref}[x,y] \right)^2 \tag{3.21}$$

The second metric, pSNR, is derived from MSE and represents the ratio of the maximum pixel intensity to the power of the distortion. A higher pSNR value indicates a greater similarity between the two analyzed images. The mathematical definition of pSNR is:

$$PSNR\left[I_1, I_{ref} \right] = 10 \log \left(\frac{peak^2}{MSE\left[I_1, I_{ref} \right]} \right), \tag{3.22}$$

where $peak$ is the maximum value taken from the range of image data. In our case $peak = 1$ as all the images are rescaled between $[0,1]$ range.

In contrast, SSIM combines local image structure, luminance, and contrast into a unified local quality score. With SSIM, structures and patterns are essentially discernible as variations in pixel intensities between adjacent pixels. Given the human visual system's capability to perceive such structures, the SSIM quality index is more pertinent, offering a score that correlates more closely with subjective quality perception. The closer the SSIM value approaches 1, the greater the similarity between the images. The overall score is calculated as the product between the three terms as in:

$$SSIM\left[I, I_{ref}\right] = l\left(I, I_{ref}\right)^{\alpha} \cdot c\left(I, I_{ref}\right)^{\beta} \cdot s\left(I, I_{ref}\right)^{\gamma}, \tag{3.23}$$

where

$$l\left(I, I_{ref}\right) = \frac{2\mu_I \mu_{I_{ref}} + C_1}{\mu_I^2 + \mu_{I_{ref}}^2 + C_1} \tag{3.24}$$

$$c\left(I, I_{ref}\right) = \frac{2\sigma_I \sigma_{I_{ref}} + C_2}{\sigma_I^2 + \sigma_{I_{ref}}^2 + C_2} \tag{3.25}$$

$$s\left(I, I_{ref}\right) = \frac{\sigma_{I\,I_{ref}} + C_3}{\sigma_I \sigma_{I_{ref}} + C_3} \tag{3.26}$$

where μ_I, $\mu_{I_{ref}}$, σ_I, $\sigma_{I_{ref}}$ and $\sigma_{I\,I_{ref}}$ are the local means, standard deviations and cross-covariance for images I and I_{ref}. The exponents α, β and γ are chosen to be equal to 1 to give each component the same weight when computing the SSIM index.

The results of applying the quality metrics for the three cases presented earlier in Figure 3-17, Figure 3-18 and Figure 3-19 are further discussed below. In order to use these quality metrics, we present in Figure 3-20 three reference images used for comparison. The reference images are in fact segmented optical images of the three samples. Before segmentation, the optical images are sub-sampled in order to have 100 by 100 pixels. As the reconstructed images can report different scales, they are commonly scaled in order to match a black background (0s) and white patterns (1s) for comparison purposes.

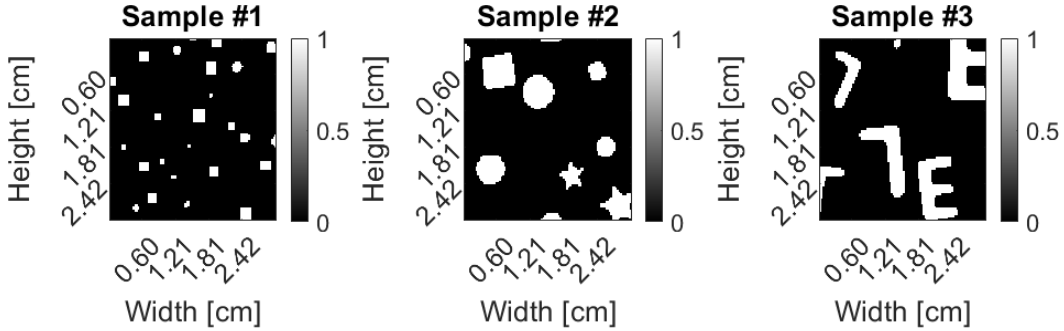
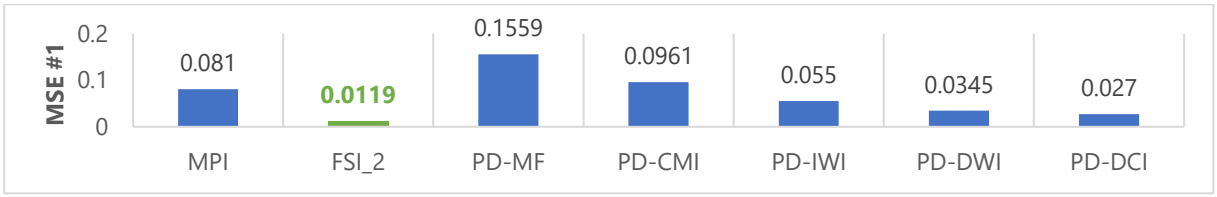


Figure 3-20. Reference images used with image quality metrics

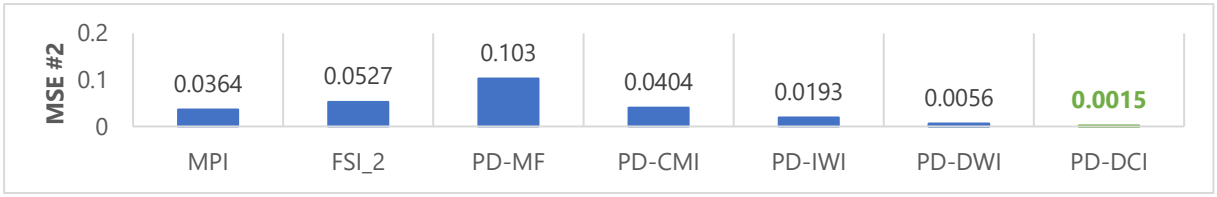
Figure 3-21 presents the MSE for all three samples and for all compared methods. For sample #1 in Figure 3-21 (a), the FSI reconstruction method has the lowest outcome, thus it has a better alignment to the original image compared to the other proposed methods. Considering the

results of sample #2 and sample #3 from Figure 3-21 (b) and (c), the lowest MSE is achieved by PD-DCI.

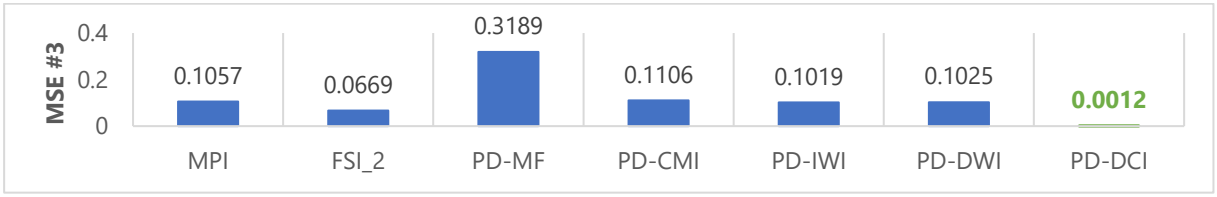
Figure 3-22 shows the pSNR evaluation of reconstruction methods. As pSNR is strictly related to MSE, the outcomes are anticipated to be alike. Thus, for sample #1 in Figure 3-22 (a), the FSI reconstruction method has the highest pSNR, while for sample #2 and sample #3 in Figure 3-22 (b) and (c), PD-DCI has the highest score.



(a)

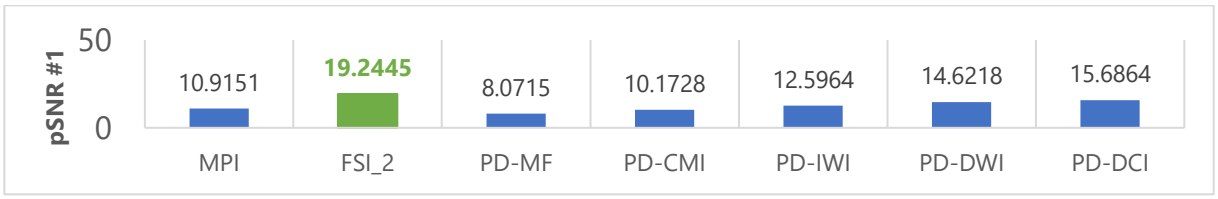


(b)

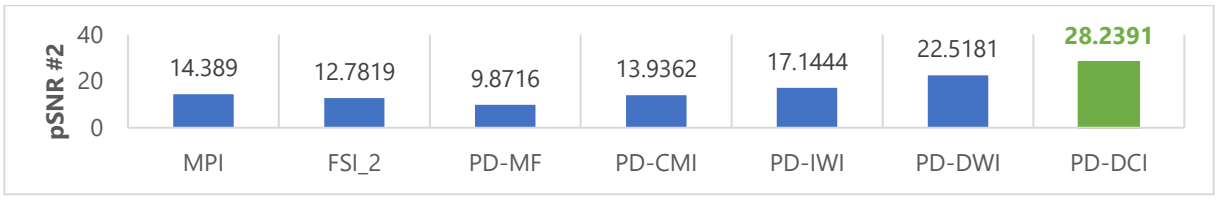


(c)

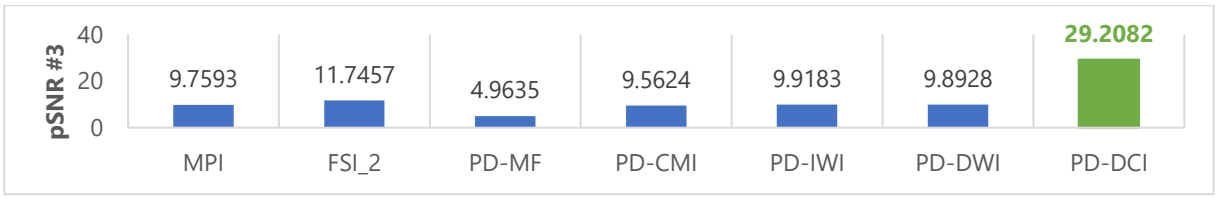
Figure 3-21. MSE for (a) Sample #1; (b) (a) Sample #2; and (c) (a) Sample #3



(a)

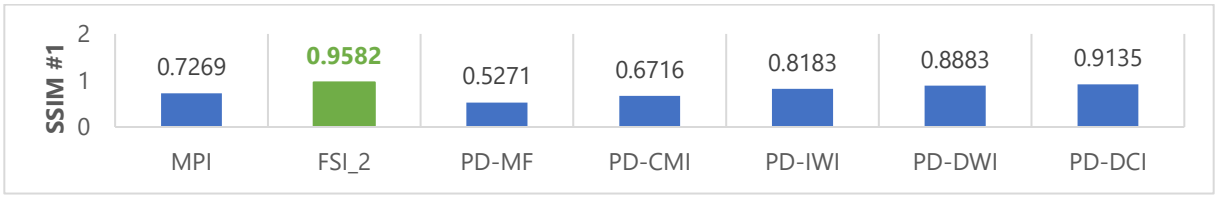


(b)

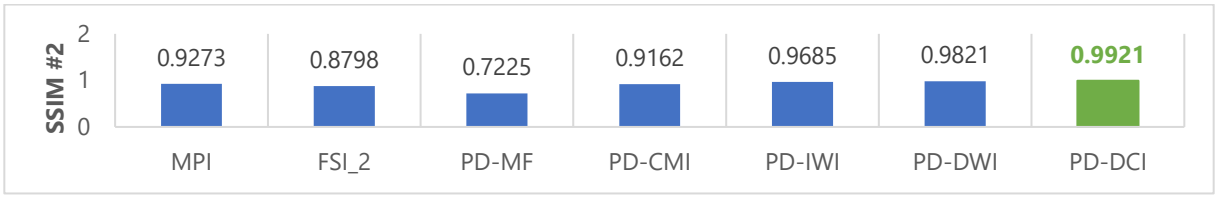


(c)

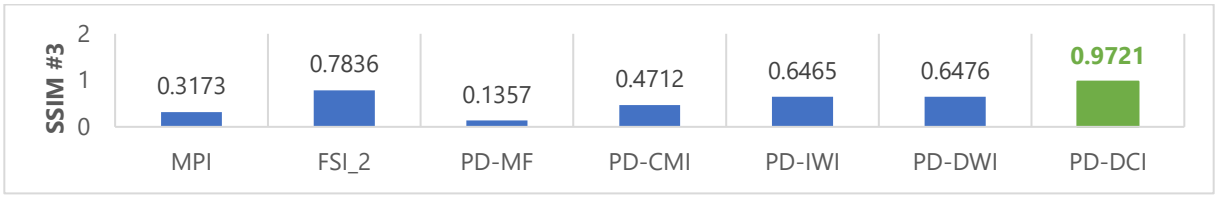
Figure 3-22. pSNR for (a) Sample #1; (b) (a) Sample #2; and (c) (a) Sample #3



(a)



(b)



(c)

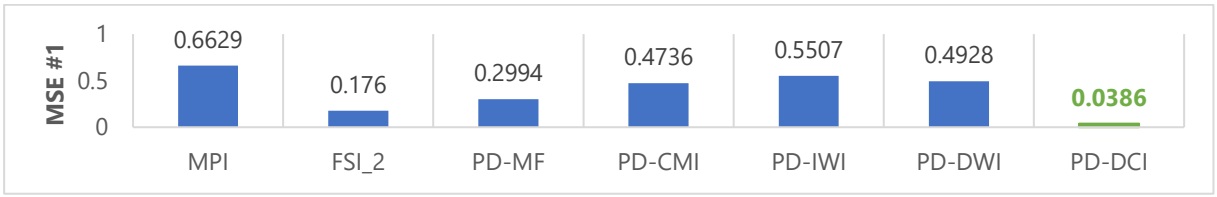
Figure 3-23. SSIM for (a) Sample #1; (b) (a) Sample #2; and (c) (a) Sample #3

In Figure 3-23 we evaluate the final SSIM results, which conclude that FSI can better reconstruct sample #1 among the rest of the methods, while sample #2 and #3 have greater SSIM for PD-DCI.

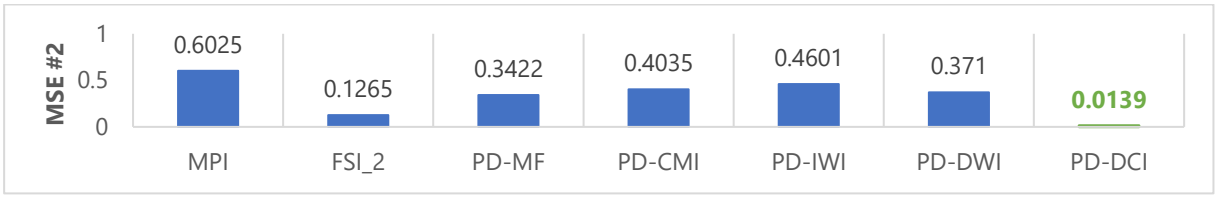
We present the same analysis but in a multiplicative perturbation context for MSE in Figure 3-24, pSNR in Figure 3-25 and SSIM in Figure 3-26. The binarized noisy images are compared with their references. As it is expected considering the initial results and discussions about the proposed methods in the start of this sub-section, with all three samples, the reconstructed images using PD-DCI yield the best quality. Most of the artefacts and deformations as consequences of the applied noise are eliminated and the information is emphasized.

In conclusion of this section, certain observations can be made. The proposed methods have an increased complexity when it comes to an algorithmic implementation. However, despite their complexity, the reconstructed images can have better quality in terms of contour information and structural similarity compared to the reference image in an ideal context. Nevertheless, there are some situations when classic methods fail such as when the absorption surface is larger and a FSI approach is unsuitable due to a possible spectral frequency attenuation. In this case, PD-DCI is recommended as it considers the whole spectrum when reconstructing the image. Nevertheless, PD-DCI seems to be the second choice when the motives in the sample are small enough. In this case, despite detecting all the information, the contour is deformed compared to a FSI approach.

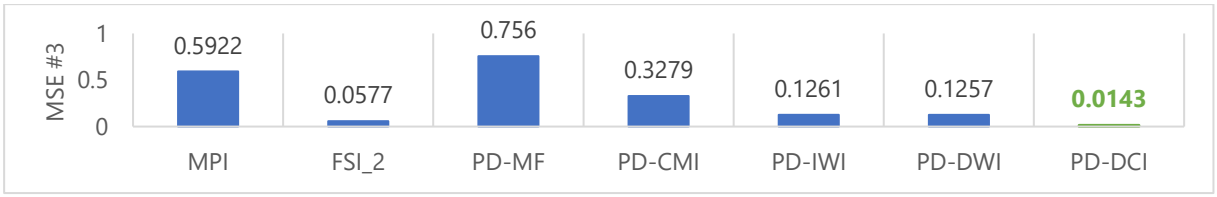
Another important mention is when the multiplicative noise is present. In this scenario, the results shown until now prove the increased capacity of the methods to remove artefacts and to reconstruct a close to the ground truth image. PD-DCI has demonstrated superior performance compared to all other methods, making it the recommended choice when the system is affected by slow-varying perturbations.



(a)

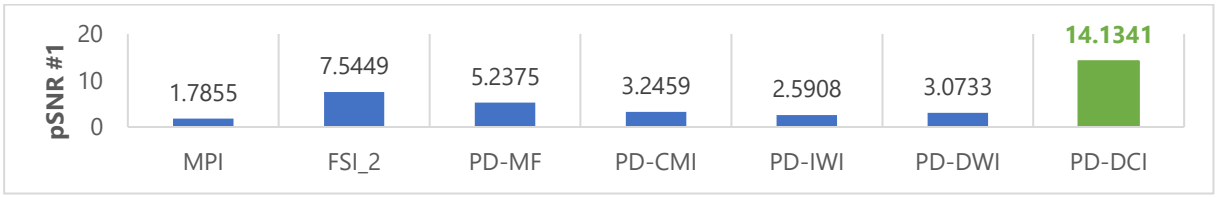


(b)

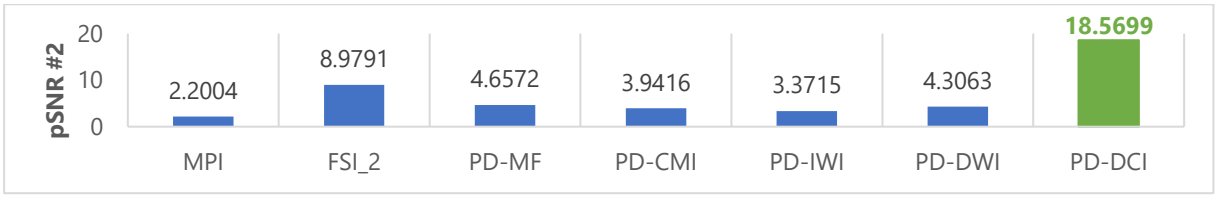


(c)

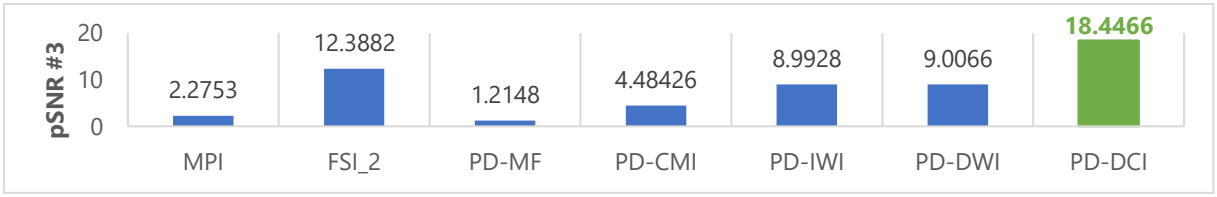
Figure 3-24. MSE for (a) Sample #1; (b) (a) Sample #2; and (c) (a) Sample #3



(a)

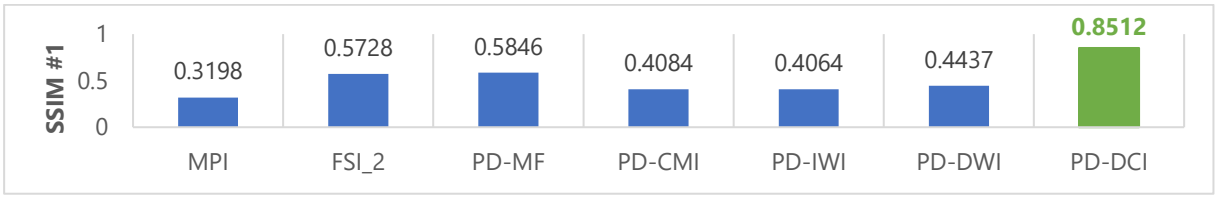


(b)

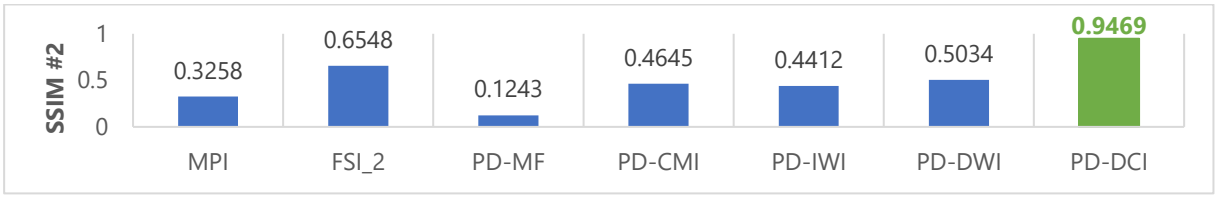


(c)

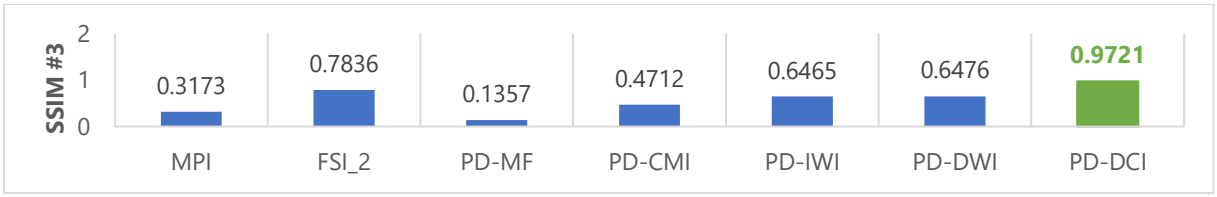
Figure 3-25. pSNR for (a) Sample #1; (b) (a) Sample #2; and (c) (a) Sample #3



(a)



(b)



(c)

Figure 3-26. SSIM for (a) Sample #1; (b) (a) Sample #2; and (c) (a) Sample #3

3.4 Chapter summary

This chapter five new reconstruction methods are introduced and tested in an ideal and noisy context, providing a performance comparison in terms of image quality, contrast and segmentation. The proposed methods are based on the phase diagram representation of time-domain signals. In most of the cases, their ability to reconstruct information surpass the classic methods. However, there are situations with limitations and challenges which force-open new research paths revolving around a phase-diagram approach.

In the next chapter, we discuss classification approaches considering view-point variation constraints such as translations and rotations. As the perspective of an object in an image is not constant in everyday applications, we propose a couple of entropy-based features computed from an invariant wavelet decomposition that increases the classification accuracy in a ML context.

Chapter 4. THz image analysis

4.1 Introduction

As in most imaging technologies performing in a variety of frequency bands, THz image processing pipelines encounter analogous challenges in pattern recognition: sensing instabilities such as view-point variation and feature extraction. After a THz image has been reconstructed with a method described in the previous chapter, the naturally subsequent step in a classification framework is image analysis and feature extraction.

Nowadays, Deep Learning (DL) approaches are the most common choice in computer vision tasks due to rapid advances in computer memory, CPU and GPU, corroborated with DL models' capacity to optimally extract features in highly-hierarchical manner [87]. However, in THz imaging framework, DL is still in its early stage. For example, THz DL-based computer vision requires large datasets which are, to our best knowledge, very few publicly available and task-specific such as an image restoration benchmark dataset [88] and a dataset prepared to evaluate object detection algorithms [89].

One of the main obstacles to create large THz image datasets is the time-consuming image acquisition process. For an actual THz TDS system, the acquisition process can take from a few minutes to hours depending on the size of the sample and the a priori configuration of the system. For example, each 100 by 100-pixel image presented in Chapter 2 took approximately fifteen minutes to be created with the systems' parameters set as follows: step size is 0.1mm and there are 10 averaging procedures when measuring one pixel. The obvious trade-off is between the acquisition time and the higher quality of the image in terms of spatial step. Rapid acquired images lack information and resolution, while the others are information enhanced, but wearisome. Thus, considering the absence of available data and the difficulties of creating large datasets, a Machine Learning (ML) approach is more suitable in this case.

The majority of ML classification methodologies employ a hybrid approach, utilizing diverse complementary tools and paradigms to construct an accurate discriminatory system. This system should be robust and efficiently handle translations or rotations, two of the most common deformations when it comes to image patterns and real-world applications. Bearing in mind the observation point uncertainty in a real-world THz system, we firstly review state-of-the-art feature extraction methods based on image transforms. Afterwards, we propose a scheme that offers a unique sub-band decomposition regardless of image translations or rotations, followed by entropy-based feature vectors extractions that are further engineered into suitable formats and fed to a large pool of classifiers.

The focus of this chapter is on researching, evaluating and assessing novel features with a THz image dataset, but also with public benchmark datasets for classification purposes.

4.2 State-of-the-art image analysis for ML

Before diving into details, we briefly review the notions needed to describe image analysis methods. These methods essentially act as transforms that break down images into weighted combinations of orthogonal or biorthogonal basis functions, whose weights can be subsequently examined and converted into features. From this standpoint, we view images as vectors within a vector space containing all other images, and the basis functions as the primary features that convey the essence of the transformation. It is important to note that image transforms are equivalent in terms of information and energy, the difference being only about their distribution among the resulting coefficients.

All transforms presented in subsequent sections of this chapter use a specific transformation kernel. These transformation kernel represent the basis or the functions used to decompose the data. To extend the one-dimensional transforms to higher dimensions, the separability of the basis functions is essential from an implementation perspective.

For example, a two-dimensional transformation kernel, $k[x, y, a, b]$, is separable if it can be rewritten as:

$$k[x, y, a, b] = k_1[x, a] k_2[y, b] \tag{4.1}$$

where x, y, a and b are variables traversing the original and transform domains. Moreover, if k_1 and k_2 are functionally equal, the transformation kernel is also symmetric. The consequence of kernel separability transposes into a simple implementation of transforms in case of images. The complexity of the initial two-dimensional case is reduced to multiple one-dimensional transforms.

This means that an image transform can be computed by applying the one-dimensional transform on each row, then on each column, or the conversely without any changes in the final result due to symmetry.

In the following subsection, we discuss some of the most common image transforms and formally describe each feature extraction technique individually.

4.2.1 Discrete Cosine Transform

One of the most frequently employed transforms in compression and data processing is Discrete Cosine Transform (DCT) [90]. These applications are suitable for DCT as it concentrates the majority of the image's visual information into a small number of coefficients. For DCT, the set of basis functions, or kernels, are represented by a collection of cosines with different amplitudes and frequencies.

In a two-dimensional context, the DCT of an image, $f[x, y]$, is defined as:

$$C_{DCT}[a, b] = \alpha_a \alpha_b \sum_{x=1}^{W} \sum_{y=1}^{H} f[x, y] \cos \frac{\pi(2x+1)a}{2W} \cos \frac{\pi(2y+1)b}{2H}, \tag{4.2}$$

where $x = \overline{1...W}$ and $y = \overline{1...H}$ iterate through the width and the height of the image; a and b are the transform variables with the same range as image variables. The two scaling coefficients before summations are expanded as in:

$$\alpha_a = \begin{cases} \frac{1}{\sqrt{W}}, & a = 1 \\ \sqrt{\frac{2}{W}}, & 2 \le a \le W \end{cases}, \qquad \alpha_b = \begin{cases} \frac{1}{\sqrt{H}}, & b = 1 \\ \sqrt{\frac{2}{H}}, & 2 \le b \le H \end{cases}. \tag{4.3}$$

For simplicity, let's consider that the image has its width equal to its height so we can define DCT transformation matrix. This matrix allows us to visualize the basis functions and to redefine (4.2) in matrix form.

Thus, the DCT transformation matrix for an W-by-W image can be expressed by:

$$T_{DCT}^{ab} = \begin{cases} \dfrac{1}{\sqrt{W}} & a = 1 \quad 1 \le b \le W \\ \sqrt{\dfrac{2}{W}} \cos \dfrac{\pi (2b+1) a}{2W} & 1 \le a \le W \quad 1 \le b \le W \end{cases} . \tag{4.4}$$

Therefore, DCT becomes:

$$C_{DCT} = T_{DCT} \cdot f \cdot T_{DCT}^{'} . \tag{4.5}$$

where T_{DCT} and are the original and transposed transformation matrix.

Figure 4-1 (a) shows an example THz image of 128-by-128 pixels which will be used as input to DCT in (4.5). The image will be used to exemplify all the following feature extraction methods presented in this chapter. The DCT matrix is shown in Figure 4-1 (b), where we observe the frequency variation of the basis. Horizontally, the frequencies increase from left to right, while vertically frequencies increase from top to bottom. The coefficient $C_{DCT}[0,0]$ is often called Direct Current (DC) coefficient representing the "zero" frequency component. Figure 4-1 (c) shows the logarithmically scaled DCT coefficients and emphasizes the concentration property of DCT where most of the energy is distributed in the top-left corner, corresponding to lower frequencies.

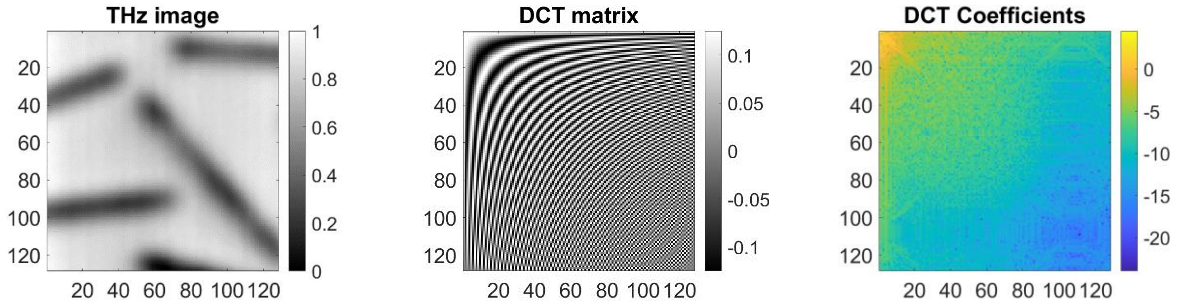


Figure 4-1. DCT transform example: (a) THz image; (b) DCT matrix transform; and (c) DCT coefficients.

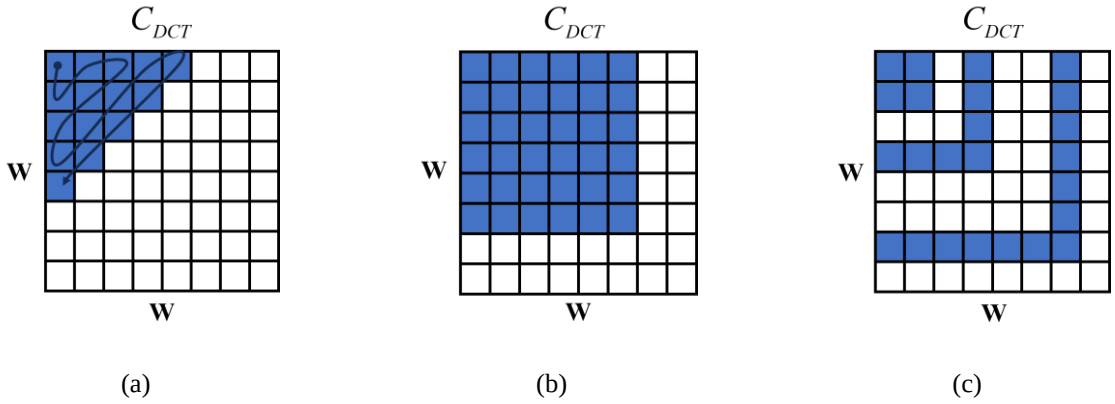


Figure 4-2. DCT feature selection methods: (a) zig-zag masking; (b) zonal masking; and (c) Jing's bands

In most cases, when decomposing an image with DCT, the features which can be extracted are in fact the full DCT coefficients. Nevertheless, there exist alternatives for reducing the dimensionality of the feature vector and to keep only the most important coefficients [91]. One is to make the selection with a zig-zag mask as it is depicted in Figure 4-2 (a) with blue. Other ideas are to mask an entire region as in Figure 4-2 (b) or to use Jing's bands as it is depicted in Figure 4-2 (c).

The features presented earlier can be formalized as a collection of values, in the order of their introduction, as in:

$$DCT_{full} = \left\{ C_{DCT}\left[a,b\right] \right\}, \tag{4.6}$$

$$DCT_{zig-zag} = mask_{zig-zag}\left\{ C_{DCT}\left[a,b\right] \right\}, \tag{4.7}$$

$$DCT_{reg} = mask_{reg}\left\{ C_{DCT}\left[a,b\right] \right\}, \tag{4.8}$$

$$DCT_{Jing} = mask_{Jing}\left\{ C_{DCT}\left[a,b\right] \right\}, \tag{4.9}$$

where DCT_{full}, DCT_{zz}, DCT_{reg} and DCT_{Jing} represents the whole DCT decomposition, the zig-zag approach, the masked region and features extracted using Jing's band masking, respectively.

4.2.2 Discrete Wavelet Transform

Wavelet Transforms (WTs) are powerful mathematical frameworks for analyzing data with features varying over multiple scales. As an example, for a time-series, features can be represented by transient phenomena or highly- or slowly-varying frequencies over a time frame, etc. In two-dimensions, WTs specialize in detecting features such as edges and textures.

Wavelet transformations implementation based on multiresolution analysis (MRA) [92] concept is an important signal processing approach which includes three major concepts: sub-band coding, quadrature mirror filtering and pyramidal image decomposition. As signals often are composed of multiple physically meaningful elements, it is useful to analyze each element in spectrally separated sub-bands, but on the same time-scale as the original data. Intuitively, MRA is decomposing data into parts which can later be added back together to reconstruct the original input. To represent and analyze data at multiple resolutions, scaling and wavelet functions are concomitantly used. The former creates approximations of data with a difference in resolution by a factor of 2, while the latter encodes the information gap between two neighbor approximations.

A Discrete Wavelet Transform (DWT) expresses data as a linear combination of orthonormal or biorthonormal basis functions, which are the scaling and wavelet functions. Under the fundamental conditions of MRA [92], the scaling function, φ, can be defined as a linear combination of double-resolution versions of itself:

$$\varphi\left[x\right] = \sum_{k\in\mathbb{Z}} \sqrt{2}\cdot h_{\varphi}\left[k\right]\varphi\left[2x-k\right], \tag{4.10}$$

expression called the refinement equation, where the integer, k, indicates the position of the scaling function on the $x-axis$ and $h_{\varphi}\left[k\right]$ are the scaling function coefficients.

Considering the scaling function from (4.10), the MRA theory [92] states that there exist a wavelet function, ψ, which can cover the differences between adjacent scaling spaces as:

$$\psi[x]=\sum_{k\in\mathbb{Z}}\sqrt{2}\cdot h_{\psi}[k]\varphi[2x-k] \tag{4.11}$$

where $h_{\psi}[k]$ are the wavelet function coefficients.

Usually, the two functio $T_{DCT}^{'}$ ns, $\varphi[x]$ and $\psi[x]$, are referred to as the approximation (or father) wavelet and detail (or mother) wavelet. Their recursive nature allows to describe the approximation and detail coefficients which determines a filter-based approach when defining DWT as in:

$$C_{\varphi}[j,k]=\sum_{x}h_{\varphi}[x-2k]C_{\varphi}[j+1,k], \tag{4.12}$$

$$C_{\psi}[j,k]=\sum_{x}h_{\psi}[x-2k]C_{\varphi}[j+1,k], \tag{4.13}$$

where $C_{\varphi}[0,k]=f[k]$ is the original signal.

As depicted in Figure 4-3 (a) [93], the filters decompose the signal's spectral content into two equally-sized sub-bands, one which is the approximation and the other which contains the details. Equations (4.12) and (4.13) can be interpreted as low-pass and high-pass filters followed by a 2-factor sub-sampling. $H_{\varphi}[\omega]$ and $H_{\psi}[\omega]$ are the transfer functions of the two filters. Figure 4-3 (b) shows a two-stage filter bank where each stage yields transform coefficients at scales $j+1$ and $j+2$. One important observation is that at the highest scale, j, the transform coefficients are the function or image itself, $f[x]$. There are equivalently transposed in signal processing as low- and high-pass filters which split the spectrum into halves.

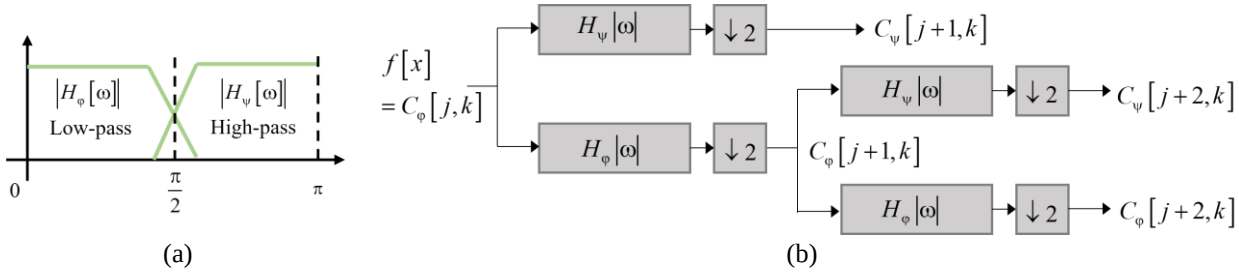


Figure 4-3. (a) Frequency spectrum of the filter bank and; (b) a two-level wavelet analysis

From this point, a two-dimensional implementation of DWT requires a two-dimensional scaling function $\varphi[x,y]$ and three wavelet functions, $\psi^{W}[x,y]$, $\psi^{V}[x,y]$ and $\psi^{D}[x,y]$. These wavelets have the natural ability to emphasize directional information as in columns, rows and diagonals. The directionality of the two-dimensional wavelet functions is the direct consequence of their definition as the product between two one-dimensional and separable wavelet functions [92].

The two-dimensional DWT is implemented employing the same approach as in one-dimension. The image, $f[x,y]=C_{\varphi}[0,k,l]$, is initially low-pass and high-pass filtered and down sampled column-wise. The low-pass component characterizes the low-frequency and vertical

information from the original image, while the high-pass component emphasizes the high-frequency and vertical details. The same procedures are applied row-wise to each of the intermediary sub-images resulting in 4 quartered-sized sub-images or wavelet coefficients, three which are directionally sensitive: $C_{\psi}^{V}[j+1,k,l]$ (vertically) , $C_{\psi}^{H}[j+1,k,l]$ (horizontally), $C_{\psi}^{D}[j+1,k,l]$ (diagonally) and one approximation, $C_{\varphi}[j+1,k,l]$ as in:

$$C_{\psi}^{D}[j+1,k,l]=\sum_{x}\sum_{y}h_{\psi}[y-2k]h_{\psi}[x-2k]C_{\varphi}[j,k,l], \tag{4.14}$$

$$C_{\psi}^{V}[j+1,k,l]=\sum_{x}\sum_{y}h_{\psi}[y-2k]h_{\varphi}[x-2k]C_{\varphi}[j,k,l], \tag{4.15}$$

$$C_{\psi}^{H}[j+1,k,l]=\sum_{x}\sum_{y}h_{\varphi}[y-2k]h_{\psi}[x-2k]C_{\varphi}[j,k,l], \tag{4.16}$$

$$C_{\varphi}[j+1,k,l]=\sum_{x}\sum_{y}h_{\varphi}[y-2k]h_{\varphi}[x-2k]C_{\varphi}[j,k,l]. \tag{4.17}$$

where j is the scale, k and l are transform variables. A block diagram of the process and a visual representation of a one-level decomposition are depicted in Figure 4-4 (a) and (b) [93], respectively.

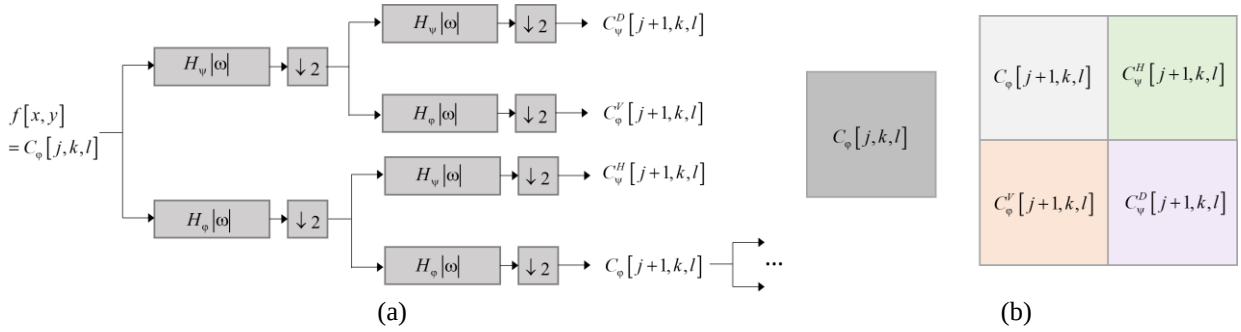


Figure 4-4. (a) One-level DWT of an image; (b) Visual representation of a one-level DWT.

We discussed the implementation of wavelet coefficients in DWT and further we present two of the most used wavelet functions. Figure 4-5 depicts some of these two-dimensional bases: (a) the ‘Daubechies’; and (b) ‘Coiflet’ families, respectively. The 'Daubechies' family comprises orthogonal wavelets distinguished by the greatest number of vanishing moments within a specified support width. The 'Coiflet' family, like the 'Daubechies' family, consists of compactly supported orthogonal wavelets with the highest number of vanishing moments for both the scaling and wavelet functions. Vanishing moments indicate the smoothness and regularity of functions, representing the number of moments that integrate to zero. In practical terms, this implies that a wavelet family with a higher number of vanishing moments is better equipped to distinguish finer details and variations, rendering it more suitable for applications where such details are crucial.

The compact support of a wavelet function is a property that allows for efficient detection and localization of features in signals or images. A wavelet function is compactly supported if the waveform is non-zero inside and zero outside a predefined interval.

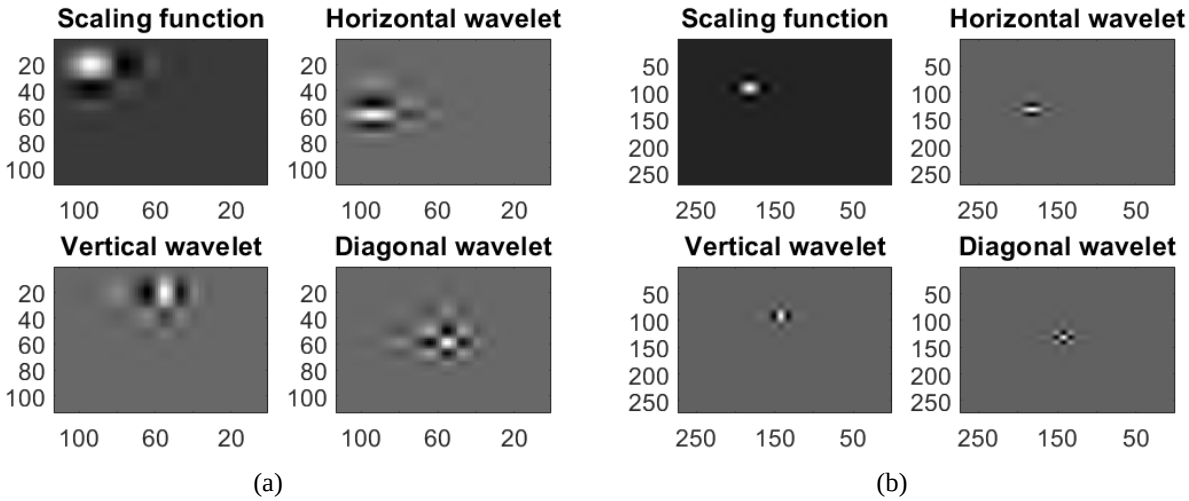


Figure 4-5. Representative scaling and wavelet functions: (a) 'Daubechies' family; (b) 'Coiflet' family.

To better visualize DWT, we present in Figure 4-6 an example of two-level decomposition for a THz image with Daubechies wavelet family. It can be observed that each sub-image specializes in characterizing information based on directions at different scales. The directional nature of each sub-band image is described, where HL stands for horizontal, LH for vertical, HH for diagonal details and LL for the approximation. This notation represents the order of the low-pass and high-pass filtering procedures when generating each sub-image.

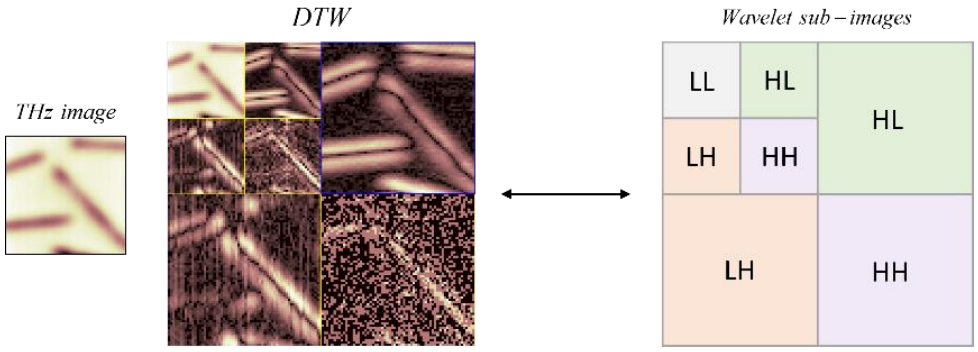
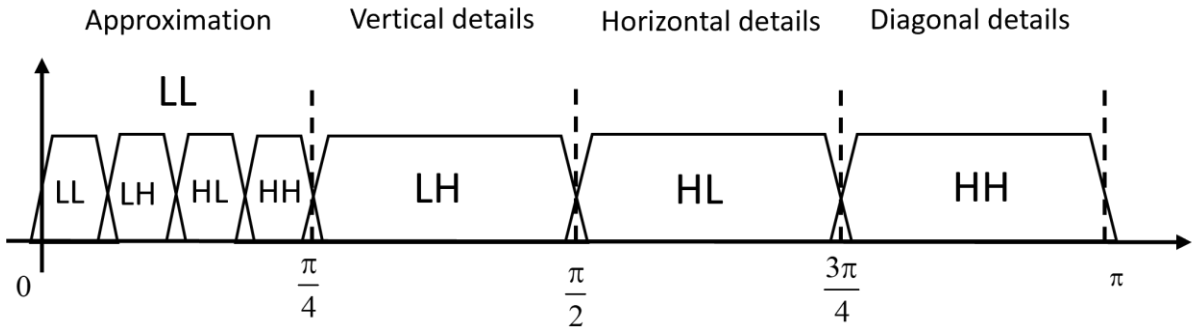
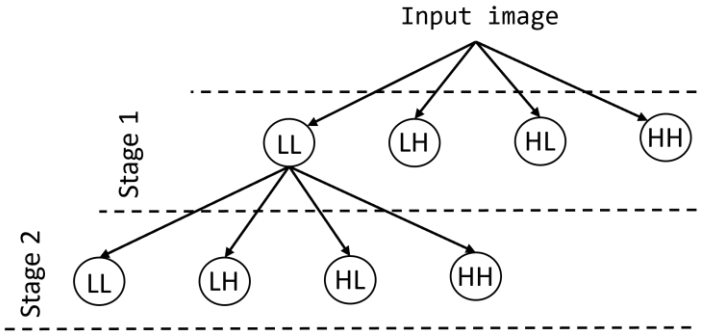


Figure 4-6. A two-level DTW example on a THz image with 'db2' wavelet family.

As the entire DWT is implemented using filtering procedures, we can visualize the frequency spectrum partitioning for a two-level decomposition in Figure 4-7 (a) [93] or the equivalent DWT decomposition tree in Figure 4-7 (b). The initial root node is assigned the approximation of the image, while the child nodes of the first level contain the approximation (LL) and detail components (LH, HL and HH). For the second level, the low-frequency content (LL) from level 1 is further decomposed. This process can have several stopping criteria such as: a predefined number of levels of decomposition, a criterion based on energy or entropy measurements, or a limitation coming from data dimensions. As a trivial example, an 8 by 8-pixel image can be decomposed a maximum of 2 levels. The first level of decomposition yields four 4 by 4-pixel sub-images and the second level yields sixteen 2 by 2-pixel images. Thus, another level of decomposition cannot exist as 1 by 1 sub-images are most of the time irrelevant and feature extraction methods usually do not apply in this case. In usual applications, the latter stopping criterion is never used as for larger images, the process will be tedious and inefficient.



(a)



(b)

Figure 4-7. (a) DWT frequency partitioning for a two-level decomposition and (b) its analysis tree representation equivalent.

In the case of DWT some of the most common features include statistical properties [94] such as the mean of wavelet sub-bands; energy [95] computed as the sum of squared DWT coefficients which reveals the energy distribution; or the Shannon-entropy [96] of the wavelet sub-bands which offers information about their complexity.

The feature vector composed from the means of all DWT coefficients, DWT_{mean}, can be formalized as a collection of values as follows:

$$DWT_{mean} = \left\{ \frac{1}{N_j} \sum_k \sum_l C_{DWT}\left[j,k,l\right] \mid j \text{ is the scale} \right\}, \tag{4.18}$$

where C_{DWT} represent all DWT coefficients (approximation and details) and N_j is the number of elements at scale j either for the approximation or detail coefficients.

The energy-based feature vector, DWT_E, is defined as a collection of wavelet sub-band energy values which can be concisely expressed as in:

$$DWT_E = \left\{ \frac{1}{N_j} \sum_k \sum_l C_{DWT}\left[j,k,l\right]^2 \mid j \text{ is the scale} \right\}. \tag{4.19}$$

The entropy-based feature vector, DWT_{Ent}, uses the Shannon definition [97] and it is created from all the entropy values of the DWT coefficients as in:

$$DWT_{Ent} = \left\{ \sum_k \sum_l C_{DWT}\left[j,k,l\right]^2 \log_2\left(C_{DWT}\left[j,k,l\right]^2\right) \mid j \text{ is the scale} \right\}. \tag{4.20}$$

4.2.3 Wavelet Packet Decomposition

Considering the frequency partitioning in Figure 4-7, we observe that details are represented with wider bandwidths compared to the low-frequency content. To have an increased control over the frequency spectrum, DWT can be generalized to achieve a more flexible decomposition, called Wavelet Packet Decomposition (WPD).

WPD provides a richer representation of images allowing for an in-depth analysis and possible selection of sub-bands to have an optimal decomposition in a sense defined by different criteria. A two-dimensional WPD of an image up to level $j+1$ can be recursively expressed as:

$$C_{4k}\left[j+1,k,l\right] = \sum_x \sum_y h_\psi\left[y-2k\right] h_\psi\left[x-2k\right] C_k\left[j,k,l\right] \tag{4.21}$$

$$C_{4k+1}\left[j+1,k,l\right] = \sum_x \sum_y h_\psi\left[y-2k\right] h_\varphi\left[x-2k\right] C_k\left[j,k,l\right] \tag{4.22}$$

$$C_{4k+2}\left[j+1,k,l\right] = \sum_x \sum_y h_\varphi\left[y-2k\right] h_\psi\left[x-2k\right] C_k\left[j,k,l\right] \tag{4.23}$$

$$C_{4k+3}\left[j+1,k,l\right] = \sum_x \sum_y h_\varphi\left[y-2k\right] h_\varphi\left[x-2k\right] C_k\left[j,k,l\right] \tag{4.24}$$

where $f\left[x,y\right] = C_0\left[0,k,l\right]$ is the original image.

The recursive process of WPD decomposes the image at scale j, $C_k\left[j\right]$, into four sub-images of quarter-size $C_{4k}\left[j+1\right]$, $C_{4k+1}\left[j+1\right]$, $C_{4k+2}\left[j+1\right]$ and $C_{4k+3}\left[j+1\right]$. WPD algorithm is illustrated in Figure 4-8 [93] and it is similar to DWT to a point. In contrast to DWT, WPD decomposes every sub-band, not only the low frequency content. The same approach is applied where each column of the image at level j is filtered using the wavelet and scaling functions and every other column is kept. Afterwards, the rows of the resulting sub-images undergo additional filtering, with every alternate row being preserved.

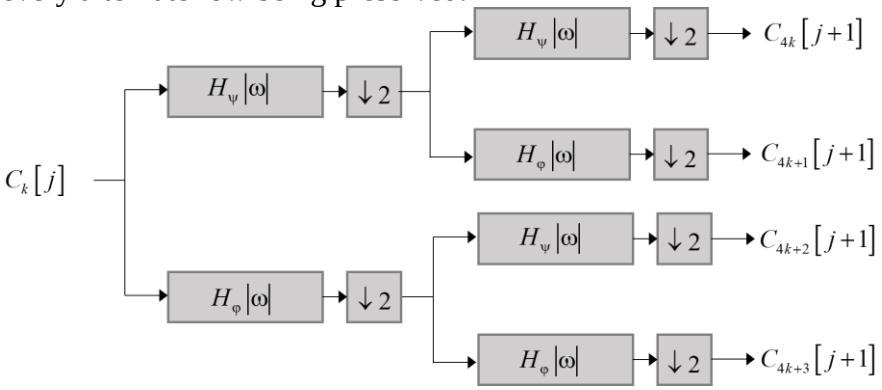


Figure 4-8. WPD decomposition process

The difference between DWT and WPD is even more visible when partitioning the frequency spectrum. Figure 4-9 (a), (b) shows the differences between decompositions. As we can observe, the WPD it is the overcomplete version of DWT by having all spectrum analyzed with smaller bandwidths compared to DWT, offering a more flexible and powerful approach when characterizing images.

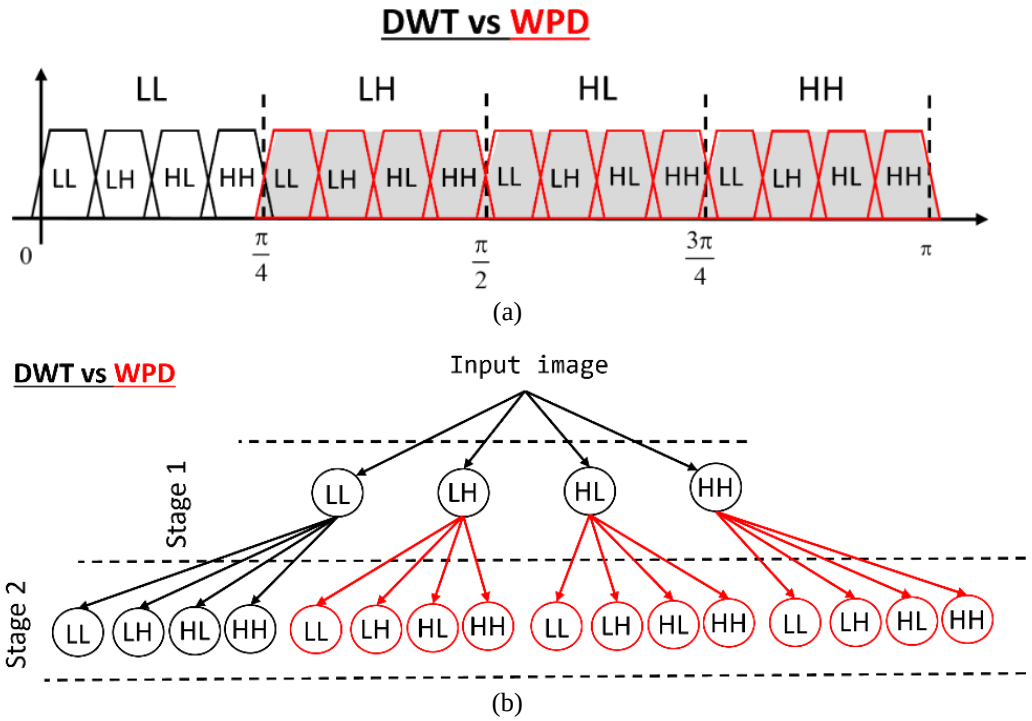


Figure 4-9. (a) WPD frequency partitioning vs. DWT and (b) the WPD analysis tree in contrast with DWT

However, the overcomplete WPD is computationally expensive to work with in practice, therefore a selection of most important sub-bands is desirable. Optimal decompositions are usual based on entropy- or energy-based cost functions [98]. The energy criterium is a reasonable criterion used in image compression [99], while the entropy is usually found in pattern recognition systems [100], [101]. Considering one of the thesis objectives to classify THz patterns, we further detail the entropy-based algorithm of sub-band selection and use this optimization technique in every experiment. The process of searching for a space of minimum entropy can be explained in two steps:

i) Firstly, we compute the Shannon-entropy for the parent node, E_p and its four child nodes, E_A, E_V, E_H and E_D.

ii) Then, if E_p is greater than the sum of its child nodes ($E_A + E_V + E_H + E_D$), retain the child nodes as being part of the "best" basis, otherwise retain the parent node.

The nonnormalized Shannon-entropy formula for an image is expressed as follows:

$$E\{f[x,y]\} = \sum_x \sum_y f^2[x,y] \log_2 \left[f^2[x,y] \right]. \tag{4.25}$$

This optimization of the WPD tree generates a space of minimum-entropy sub-images. The concept stems from the relationship between entropy and structural information in image processing: high entropy values are associated with noise, whereas lower values suggest the presence of patterns and motifs, reflecting the complexity of the analyzed image.

From an algorithmic point of view, Figure 4-10 (a) shows the WPD complete decomposition and Figure 4-10 (b) presents the optimal base selected with regards to the entropy cost function. In Figure 4-10 (c) we observe the sub-space selection algorithm. In the first level of the decomposition, the LH, HL and HH coefficients remain in the "best" basis, meaning that their entropy is lower than the sum of their child nodes. The first level LL coefficients respect the second condition in the algorithm and thus are further decomposed considering its child nodes in the "best" basis. It is worth mentioning that the selection algorithm is also a decomposition optimization technique: we do not need the overcomplete WPD in order to select the "best" basis. The optimal sub-space is generated at each step of the decomposition decreasing memory consumption and processing time.

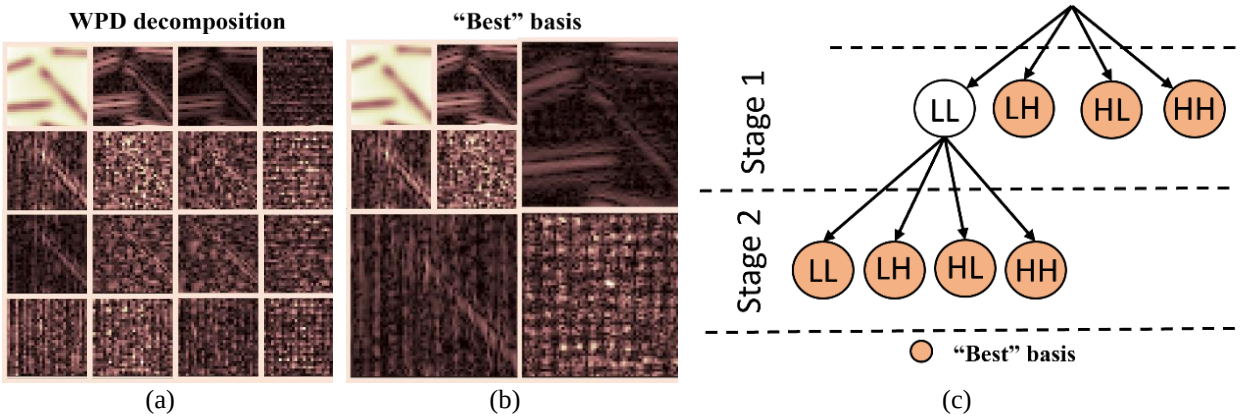


Figure 4-10. (a) a 2-level WPD; (b) "best" basis or subspace of minimum entropy; and (c) graph representation of "best" basis

As with the DWT case, some of the most common feature vectors that can be computed from WPD "best" basis are created from the mean of the wavelet coefficients, WPD_{mean} ; their energy, WPD_E ; and entropy, WPD_{Ent} . We formalize the feature vectors as a collection of values as in:

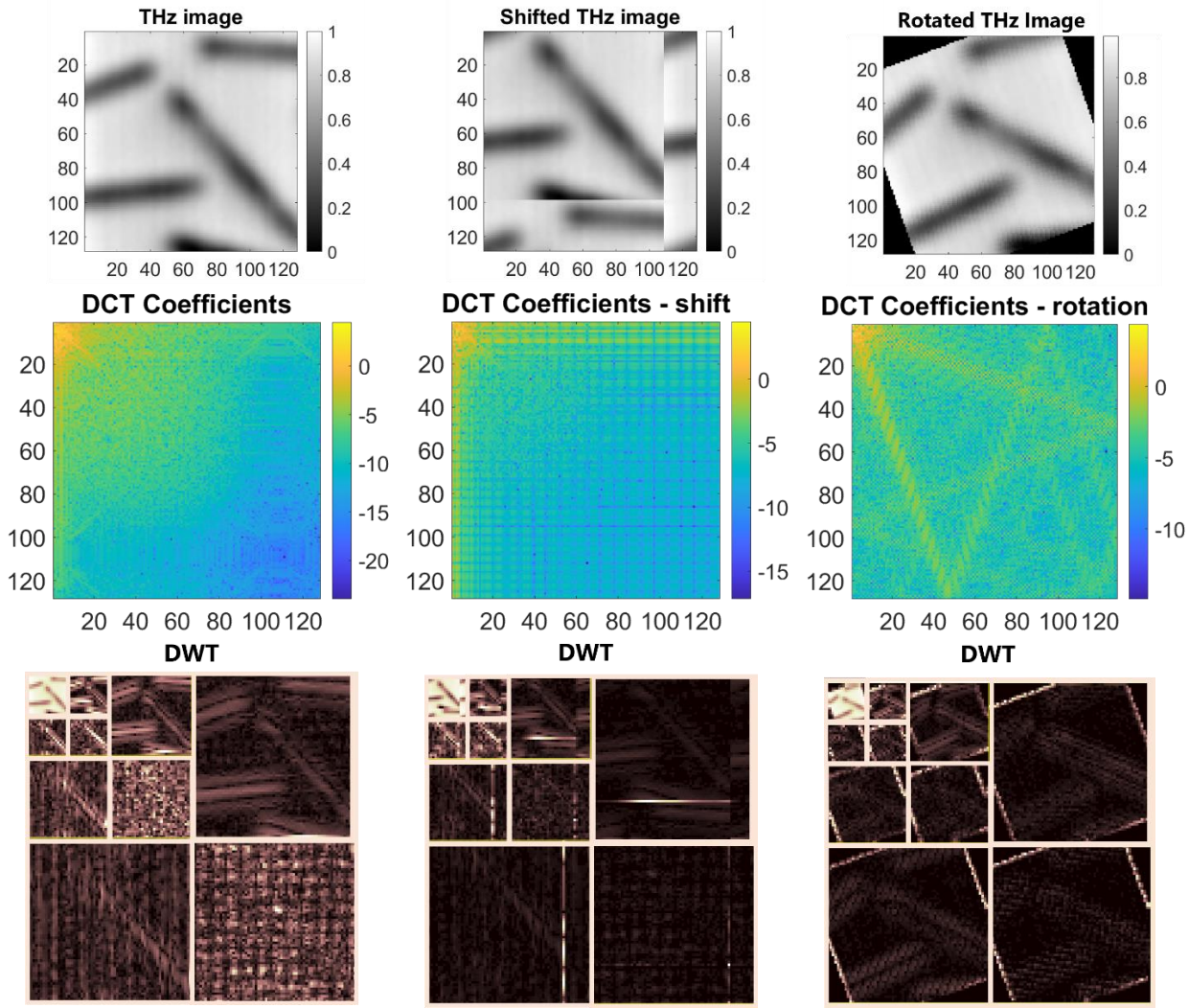
$$WPD_{mean} = \left\{ \frac{1}{N_j} \sum_k \sum_l C_{WPD}[j,k,l] \mid j \text{ is the scale} \right\}, \tag{4.26}$$

$$WPD_E = \left\{ \frac{1}{N_j} \sum_k \sum_l C_{WPD}[j,k,l]^2 \mid j \text{ is the scale} \right\}, \tag{4.27}$$

$$WPD_{Ent} = \left\{ \sum_k \sum_l C_{WPD}[j,k,l]^2 \ln\left(C_{WPD}[j,k,l]^2\right) \mid j \text{ is the scale} \right\}, \tag{4.28}$$

where C_{WPD} represent all WPD "best" basis coefficients (approximations and details), and N_j is the number of elements at scale j either for the approximation or detail coefficients.

An important drawback of DCT and DWT is the partial sub-band exploration, making image analysis unthorough and inflexible. The two transforms' coefficients are even more contrasting when the patterns are rotated or translated, as the energy distribution among decomposition coefficients is completely different from one image to another. We illustrate the comparison between the original THz image and its version that has been shifted and rotated in case of DCT and DWT in Figure 4-11.



(a) (b) (c)

Figure 4-11. Overview of DCT and DWT analysis: (a) of original image; (b) of shifted image; (c) of rotated image;

Comparatively to DCT and DWT, WPD also lacks translation invariance, thus it is sensitive to image pattern location with respect to a chosen origin. This translates into features that are discordant, meaning that in the feature space, two images exhibiting identical patterns, with one being translated, have features which are far apart, making subsequent classification step challenging. This is even more problematic for image rotations. An image is usually visualized as a grid of pixels, therefore, an image rotation of some random degree transposes pixels in-between grid points. This means that in digital image processing, an image rotation introduces intensity changes and deformations due to the grid arrangement of pixels. Moreover, opposed to image

translations, when rotating an image, part of the initial information is lost and replaced with 0's to fill the empty values. In Figure 4-12 (a) we present the THz image and a 3-level WPD "best" basis with 'db2' wavelet family. In contrast, we present in Figure 4-12 (b) and (c) the shifted and rotated THz image and their 3-level WPD "best" bases computed in the same conditions. It is obvious that even if the three images represent the same pattern, the optimal bases are different.

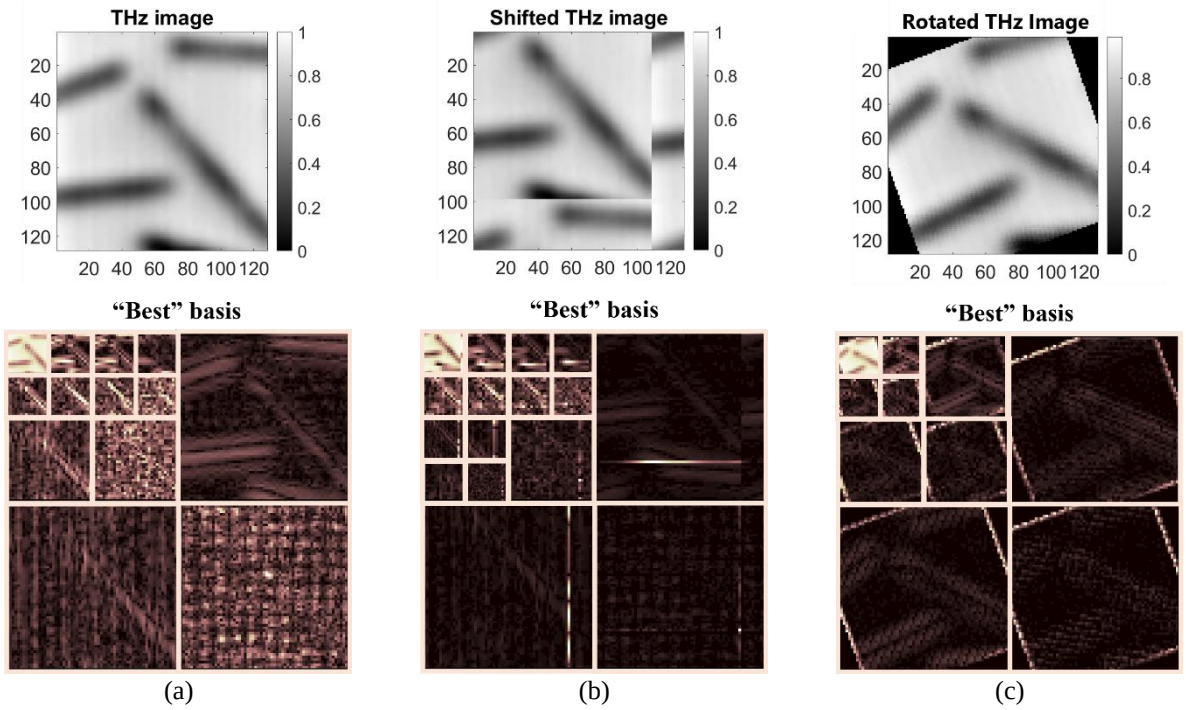


Figure 4-12. Comparison of "best" bases in three cases: (a) original THz image; (b) shifted THz imaged; and (c) rotated THz image.

Generally, when describing the same phenomenon captured from different perspectives, the goal is to minimize the variations of the extracted features, or in the WPD case, the decomposition. Therefore, it is preferable to achieve identical decomposition for the same pattern, regardless of whether it is translated or rotated relative to an origin point.

The following subsection describes our proposed method, examining every component it implies from image analysis to feature extraction frameworks, in the same manner as the state-of-the-art was presented. It provides a comprehensive analysis of the improved versions of the WPD and examples emphasizing their invariant capabilities when it comes to decomposing an image under view-point variations. In the same subsection, we introduce novel feature extraction techniques that are used to characterize image decompositions.

4.3 Invariant Wavelet Packet representation

We propose a framework that can be freely exploited when translation and rotation constraints are present in classification tasks. Figure 4-13 provides an overview of the framework. Briefly, the framework we developed decomposes an image using the invariant versions of WPD, namely the Translation Invariant WPD (TI-WPD) and Rotation Invariant WPD (RI-WPD). Each one of these two decompositions tackle the view-point variation challenge we addressed in our research. From these two distinctive decompositions we generate novel feature vectors such as N-directional entropy and N-regional entropy which are further engineered and prepared to be fed to

multiple ML classifiers such as neural networks (NNs), decision trees, support vector machine (SVM), k-nearest neighbors (K-NN) and graph neural networks (GNNs), etc.

Further on, we introduce each concept used in the framework starting with the decompositions and followed by features. Brief descriptions of the classification frameworks are presented in the next chapter.

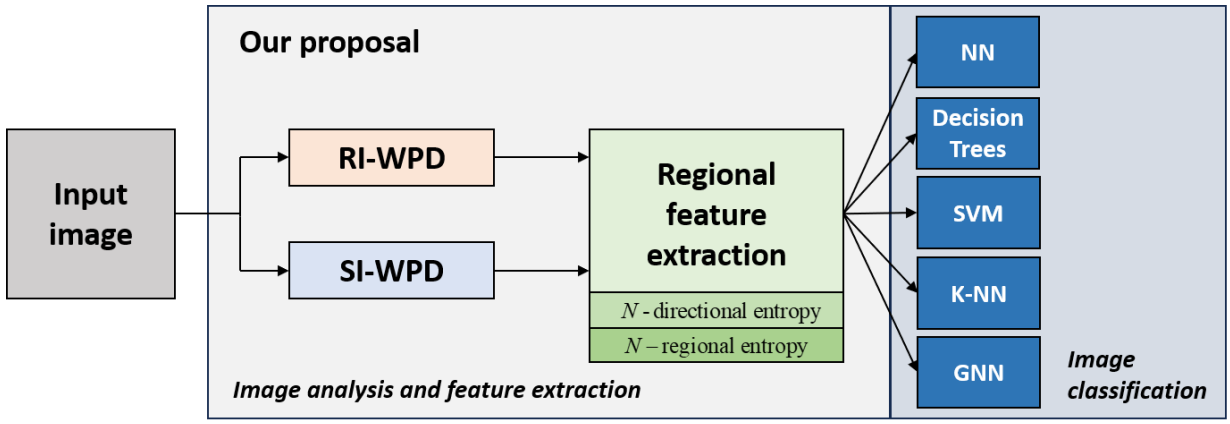


Figure 4-13. Overview of the proposed schema for image analysis and classification task.

4.3.1 Translation-Invariant Wavelet Packet Decomposition

Achieving translation invariance involves introducing an extra degree of freedom to WPD. To visualize and to offer an intuitive interpretation of the difference between WPD and TI-WPD, we provide in Figure 4-14 the comparison between the two decomposition operators. A decomposition operator is a functional block used to design and implement a complex analysis system as TI-WPD.

The WPD operator in Figure 4-14 (a) depicts the low-pass (LP) and high-pass (HP) filtering procedures, followed by down-sampling factor of 2 and the "best" basis selection algorithm. It's important to note that the optimization algorithm relies on minimizing the entropy cost function detailed in Section 4.2.3. For the WPD case, the decomposition operator is substantially the first level formalization of the filtering-based representation from Figure 4-8 with the additional optimization block.

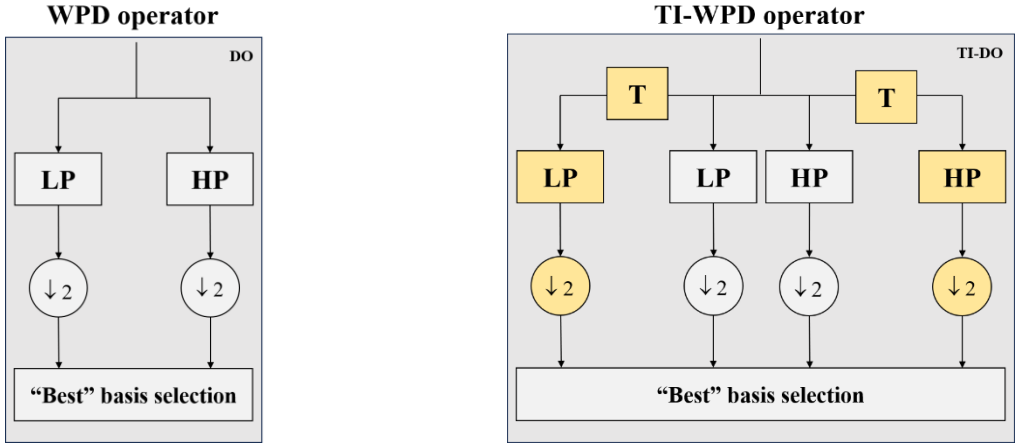
For the TI-WPD, in Figure 4-14 (b), the additional component is a translation block, where T specifies the number of pixels to be shifted. The translation block generates circularly shifted wavelet coefficients. In this scenario, compared to the WPD operator, we enlarge the generated subspace at each decomposition step by computing more wavelet coefficients and therefore, increasing the complexity of the process. However, the trade-off between the higher complexity and translation invariance, is favorable if we consider the significant property we attain as a result of exploiting the translation block.

It is proven that the translation values, T, are dyadic in nature in order to assure the invariance conditions imposed in our research [102]. Usually, as in the 2D case there are two translation axes, we imply the translations as a pair of two values, (T_V, T_H), whose indices V and H denote the shifting axes, vertical and horizontal, respectively. Therefore, for a dyadic image of $W \times W$ pixels, T_V and T_H take values of the form, 2^t, where $t \in \{0,..\log_2(W)-1\}$.

The choice of t is essential since it impacts both the chosen subspace and the minimization of the entropy cost function. The existing choice of the translation values transposes into the fact that there exists, $T_{V\min}$ and $T_{H\min}$, that generate the subspace of minimum entropy. To our

knowledge, there is no specific algorithm to compute these values, therefore, we empirically chose $T_V = T_H = 1$ as they have always generated the subspace that minimizes the entropy cost function in all of the tests and experiments on our datasets. Thus, when defining TI-WPD, the pair of translation values that generates the subspace at each decomposition step are: $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$.

The importance of having a subspace of minimum entropy stirs up from the interpretation of entropy in images, as lower values imply that there is structural information present in images, or wavelet coefficients in our case.



(a) (b)

Figure 4-14. Comparison between: (a) WPD operator; and (b) TI-WPD operator

With respect to the translation invariance conditions, for an image, $f[x,y]$, the TI-WPD can be defined as follows:

$$C_{4k}^{(T_V,T_H)}[j+1,k,l] = \sum_x \sum_y h_\psi[y-2k]\, h_\psi[x-2k]\, C_k^{(T_V,T_H)}[j,k,l] \tag{4.29}$$

$$C_{4k+1}^{(T_V,T_H)}[j+1,k,l] = \sum_x \sum_y h_\psi[y-2k]\, h_\varphi[x-2k]\, C_k^{(T_V,T_H)}[j,k,l] \tag{4.30}$$

$$C_{4k+2}^{(T_V,T_H)}[j+1,k,l] = \sum_x \sum_y h_\varphi[y-2k]\, h_\psi[x-2k]\, C_k^{(T_V,T_H)}[j,k,l] \tag{4.31}$$

$$C_{4k+3}^{(T_V,T_H)}[j+1,k,l] = \sum_x \sum_y h_\varphi[y-2k]\, h_\varphi[x-2k]\, C_k^{(T_V,T_H)}[j,k,l] \tag{4.32}$$

where $f[x,y] = C_0[0,k,l]$ and (T_V,T_H) are the pair of translations values. Equations (4.29) - (4.32) represent all the coefficients that are generated when using TI-WPD. From this collection of wavelet coefficients, we discard the ones that do not verify the entropy minimization criteria and finally, determining the "best" basis that is invariant to image translations.

Figure 4-15 shows an alternative visual representation of TI-WPD based on Mallat's pyramidal decomposition principle [92], where at each decomposition step, the TI-WPD operator is applied and generates the subspace of minimum entropy.

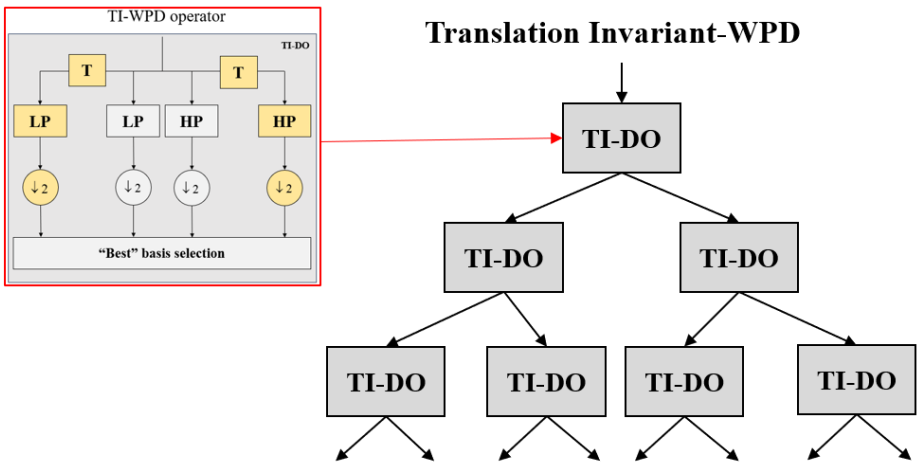
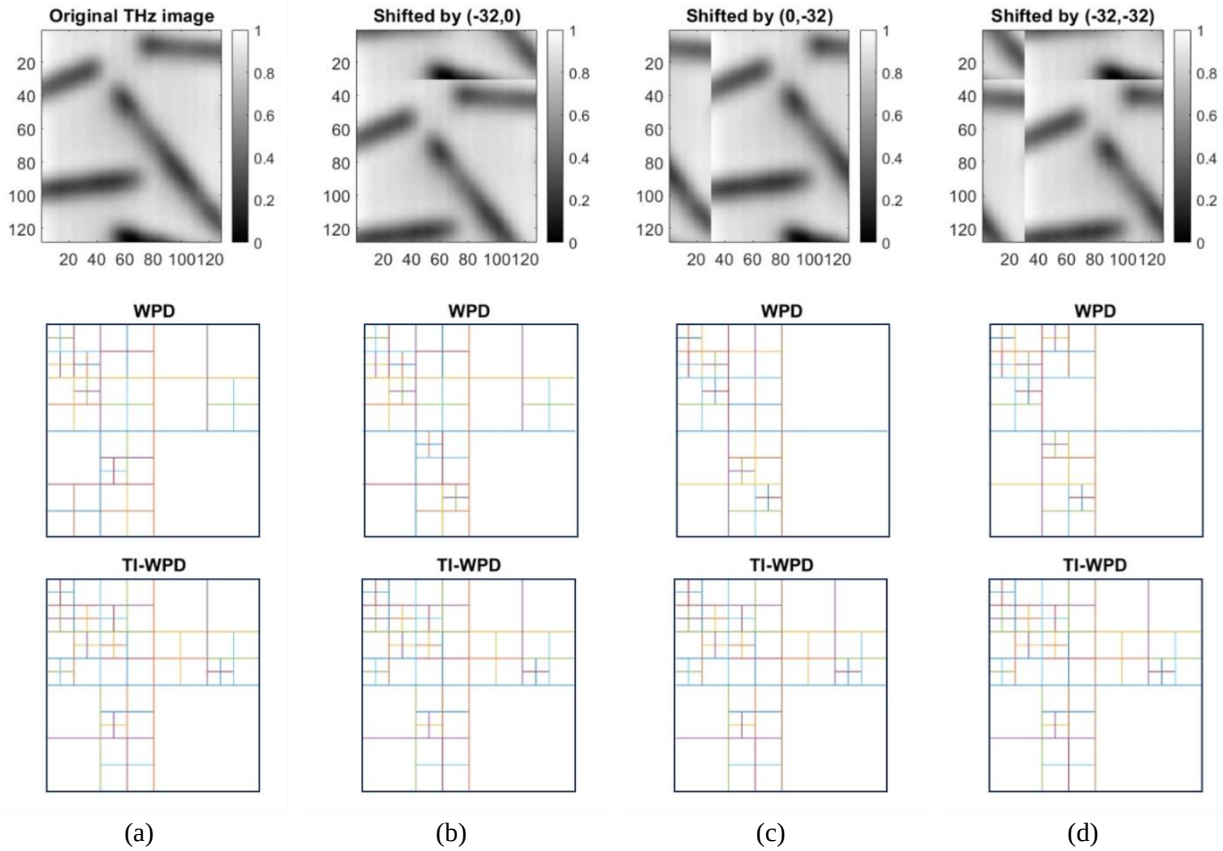


Figure 4-15. TI-WPD visualization according to Mallat's pyramidal decomposition

To observe the effectiveness of TI-WPD, a THz image is circularly translated in different directions and decomposed using WPD and TI-WPD simultaneously. It is worth noting that the four images contain the same information and no pixel value was substituted, removed or changed, apart from the applied translations.



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

Figure 4-16. Comparison between WPD and TI-WPD in 4 cases: (a) original unshifted THz image; (b) horizontally shifted with 32 pixels; (c) vertically shifted with 32 pixels; and (d) horizontally and vertically shifted with 32 pixels.

In the example from Figure 4-16, we used the "coiflet5' wavelet family and additionally, we increased the depth to a 4-level analysis to better emphasize and observe the advantages of TI-WPD over WPD. The choice of the wavelet family is strictly related to the information contained in the image and the nature of the application. Therefore, due to the fact that images we analyze on a daily basis in computer vision tasks have many patterns, contrasting characteristics and a wide variety of possible objects, one of the main challenges of wavelet analysis is the choice of the wavelet family. Therefore, in most of our tests we used two families, namely the 'Coiflet' and 'Daubechies' wavelets.

Figure 4-16 shows the decompositions results for the same THz image but which is firstly translated horizontally with 32 pixels, then vertically with 32 pixels and finally both vertically and horizontally with 32 pixels. As expected, when using WPD to analyze the provided images, we observe that in all four cases, the "best" basis selected is different as the image information is concentrated in other frequency sub-bands. In this case, a subsequent classification is hindered as the potential features are extracted from different sub-bands. On the other hand, TI-WPD assures that the selected "best" basis remains the same under any translation regardless of the direction and value.

4.3.2 Rotation-Invariant Wavelet Packet Decomposition

The other imposed constraint is more complex than the translation invariancy as rotations in digital images introduce deformations and loss of information. Nevertheless, the WPD can be adjusted to attain rotation invariance [103]. Figure 4-17 outlines a method for achieving rotation-invariant decomposition.

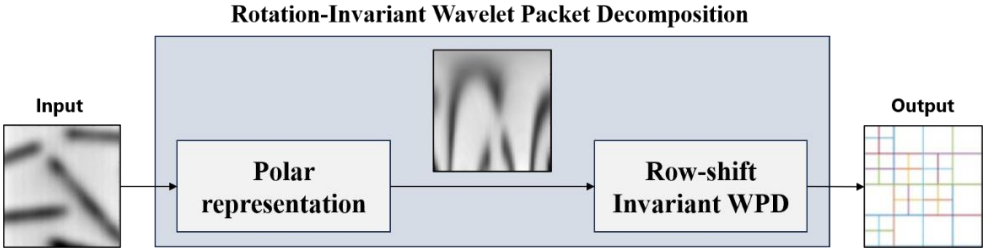


Figure 4-17. Overview of the conceptual approach of RI-WPD

To begin with, the challenging issue of rotation is simplified by converting it into a simpler translation problem using the polar representation of the image. In this representation, a translation corresponds to a rotation in the cartesian grid. The mapping from an image, $f[x,y]$, to a polar representation, $f_P[\rho,\theta]$ is defined by the following parameter relations:

$$\rho = \sqrt{(x-x_c)^2 + (y-y_c)^2}, \tag{4.33}$$

$$\theta = \tan^{-1}\frac{y-y_c}{x-x_c}, \tag{4.34}$$

Where (x_c, y_c) is the center of the image, (x,y) denotes the sampling pixel in the cartesian grid and (ρ,θ) is the radius and angular position in the polar representation.

In contrast to TI-WPD, achieving rotation invariance involves introducing an extra degree of freedom during the decomposition process, coupled with the "best" basis selection algorithm as it is shown in Figure 4-18 (a), where T_H denotes the translation only on the horizontal axis.

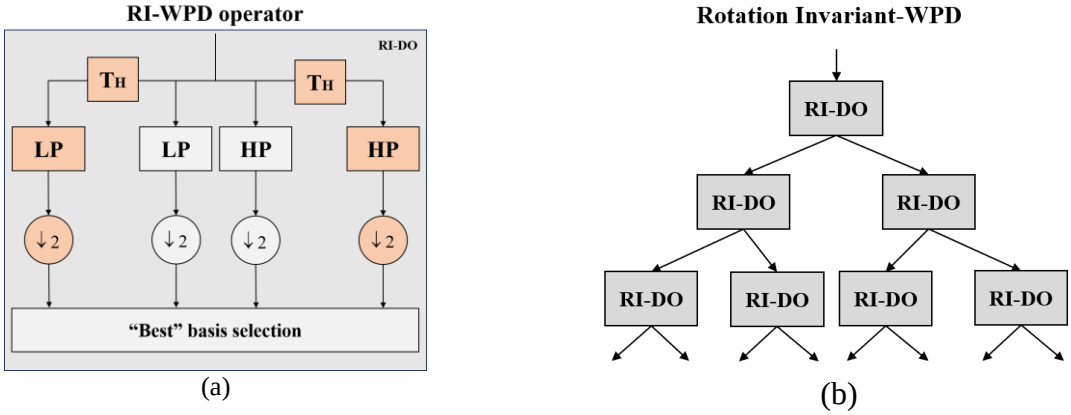


Figure 4-18. (a) RI-WPD operator; and (b) RI-WPD process according to Mallat's pyramidal decomposition

Following Mallat's pyramidal decomposition principle, at every node, we produce the subspace comprising all wavelet packet coefficients and their row-shifted versions using the RI-WPD decomposition operator. The RI-WPD process is depicted in Figure 4-18 (b). In contrast with TI-WPD, RI-WPD is implemented using only the translations on the horizontal axis, as shifting on vertical direction distorts the image when transposed back in the cartesian domain. Row-shifting indicates that the translation block, T_H, accepts only the pair of values $(0,0)$ and $(0,1)$, where the first value expresses the translation on the vertical axis and the other on the horizontal axis of the polar image. As RI-WPD is implemented in the similar conditions as the TI-WPD, the translation value of 1 is sufficient in order to achieve the desired rotation-invariancy.

It is worth noting that during the conversion of an image from the cartesian domain to the polar domain and back, there is some loss of information inherent to the polar transform. However, we are interested in only the information we lose when transposing into the polar domain, as these images are further used when generating rotation invariant wavelet coefficients. Using (4.33) and (4.34) you transform each pixel's (x, y) coordinates into polar parameters, (ρ,θ), the radius and angular position relative to the center of the image. In this scenario, the extent of information loss and distortion depends on factors such as the resolution of the polar grid and the complexity of the image such as finer details, patterns and resolution. Another important aspect is the interpolation of pixel values. In many cases, the interpolation is needed to estimate the pixel values at non-integer coordinates. This process induces differences between the original cartesian image and its polar representation. However, in order to alleviate the possible artefacts, we can consider a bilinear interpolation approach when converting from the cartesian to the polar domain.

The bilinear interpolation can be defined as a multilinear polynomial:

$$f(x,y) \approx a_{00} + a_{10}x + a_{01}y + a_{11}xy, \tag{4.35}$$

where f is a function with a point as argument (x, y) and a_{00}, a_{10}, a_{01} and a_{11} are the coefficients. It is assumed that the value of f is known at the four points $P_{11} = (x_1, y_1)$, $P_{12} = (x_1, y_2)$,

$P_{21} = (x_2, y_1)$ and $P_{22} = (x_2, y_2)$. The coefficients of the bilinear interpolation are found by solving the linear system:

$$\begin{bmatrix} 1 & x_1 & y_1 & x_1 y_1 \\ 1 & x_1 & y_1 & x_1 y_2 \\ 1 & x_2 & y_2 & x_2 y_1 \\ 1 & x_2 & y_2 & x_2 y_2 \end{bmatrix} \begin{bmatrix} a_{00} \\ a_{10} \\ a_{01} \\ a_{11} \end{bmatrix} = \begin{bmatrix} f(P_{11}) \\ f(P_{12}) \\ f(P_{21}) \\ f(P_{22}) \end{bmatrix}. \tag{4.36}$$

Therefore, the coefficients are computed as:

$$\begin{bmatrix} a_{00} \\ a_{10} \\ a_{01} \\ a_{11} \end{bmatrix} = \begin{bmatrix} x_2 y_2 & -x_2 y_1 & -x_1 y_2 & x_1 y_1 \\ -y_2 & y_1 & y_2 & -y_1 \\ -x_2 & x_2 & x_1 & -x_1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} f(P_{11}) \\ f(P_{12}) \\ f(P_{21}) \\ f(P_{22}) \end{bmatrix} \tag{4.37}$$

The bilinear interpolation uses the four nearest pixels around the location of the pixel with unknown value. The unknown pixel's value is the weighted average of those four pixels. Considering the high complexity of the wavelet transforms, one of the advantages it provides regard the straightforward implementation and efficiency. Moreover, compared to nearest-neighbor interpolation method, a bilinear interpolation offers smoother transitions between fine lines and details, aiding in preserving the information from the original image.

Figure 4-19 illustrates examples of cartesian-to-polar and their corresponding polar-to-cartesian transforms of the same THz image used to exemplify TI-WPD.

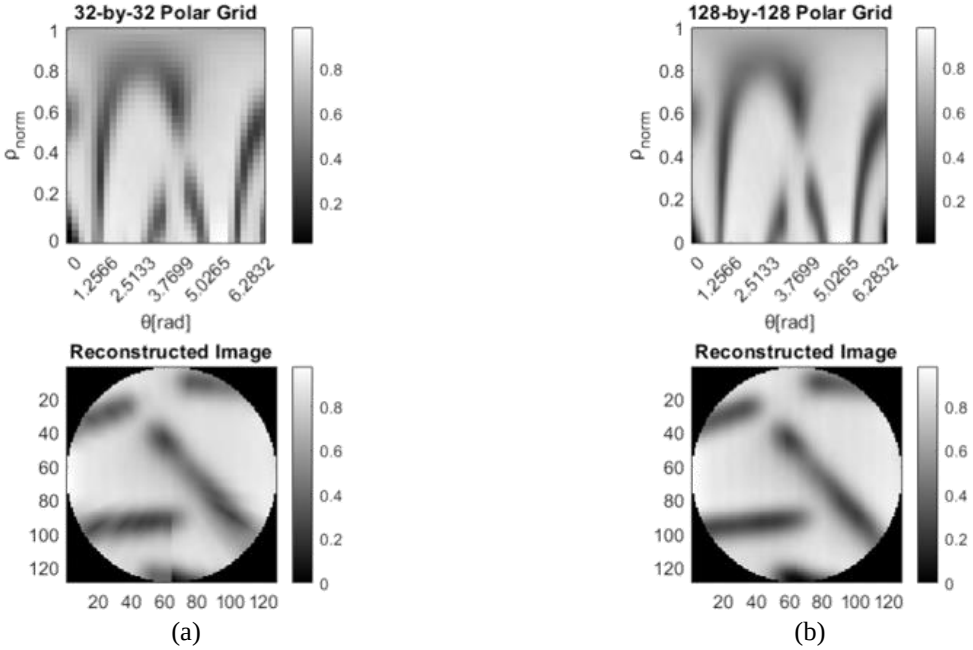


Figure 4-19. Cartesian to polar image representations: (a) for 32-by-32; and (b) 128-by-128 polar grid

Each presented case has different resolutions of the polar grid as follows: (a) 32-by-32; (b) 128-by-128. It is easily observable that the more the grid's resolution is increased, the better the

details are retained in the image. For our applications, using a high-resolution polar grid, such as 128-by-128, it is a suitable way to tackle the problem of information integrity.

Considering the row-shift invariance of a polar image, the RI-WPD can be expressed as in:

$$C_{4k}^{(0,T_H)}[j+1,k,l]=\sum_x\sum_y h_\psi[y-2k]h_\psi[x-2k]C_k^{(0,T_H)}[j,k,l], \tag{4.38}$$

$$C_{4k+1}^{(0,T_H)}[j+1,k,l]=\sum_x\sum_y h_\psi[y-2k]h_\varphi[x-2k]C_k^{(0,T_H)}[j,k,l], \tag{4.39}$$

$$C_{4k+2}^{(0,T_H)}[j+1,k,l]=\sum_x\sum_y h_\varphi[y-2k]h_\psi[x-2k]C_k^{(0,T_H)}[j,k,l], \tag{4.40}$$

$$C_{4k+3}^{(0,T_H)}[j+1,k,l]=\sum_x\sum_y h_\varphi[y-2k]h_\varphi[x-2k]C_k^{(0,T_H)}[j,k,l], \tag{4.41}$$

where $f_P[\rho,\theta]=C_0[0,k,l]$ and $(0,T_H)$ are the pair of translations values. Equations (4.38) - (4.41) represent all the coefficients that are generated throughout the RI-WPD. From all these coefficients, we select only the ones that are in accordance with the entropy minimization criteria, thus, determining the subspace of wavelet coefficients that are invariant to image rotations.

Figure 4-20 presents the decompositions results for the same THz image (a) but which is firstly rotated counter clock-wise with: (b) 25 degrees; (c) with 125 degrees; and (d) with 225 degrees. As it was presumed, when we decompose the images using WPD, we observe that in all four cases, the "best" basis selected is different.

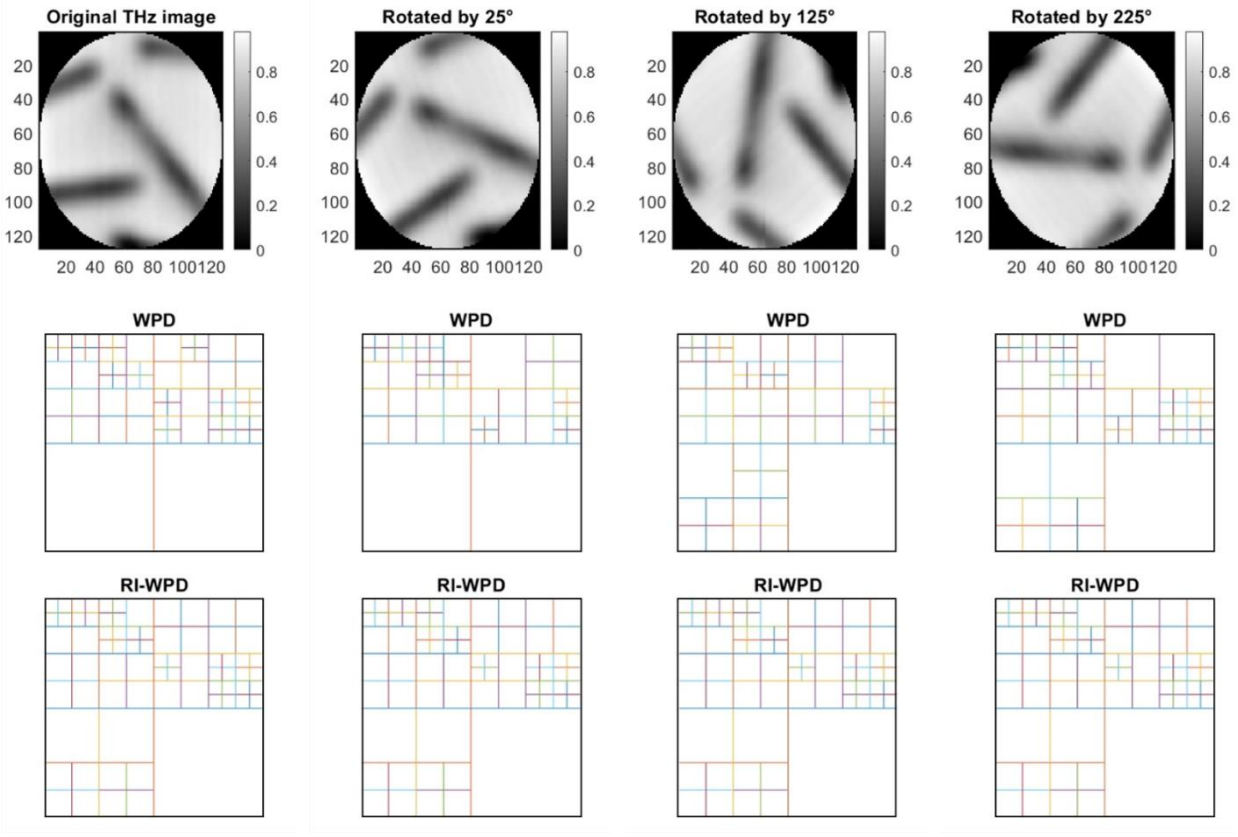


Figure 4-20. Comparison between WPD and RI-WPD in 4 cases: (a) original THz image; (b) 25° counter clock-wise rotation; (c) 125° counter clock-wise rotation; and (d) 225° counter clock-wise rotation.

However, RI-WPD assures that the selected "best" basis is the same under any rotation regardless of the direction and value. It is noteworthy that the black outer zones in the images are the result of transposing the image back to the cartesian domain. The idea is that in order to assure that the image information does not change when the image is rotated, as it is usually the case, we assume that most of the information is concentrated inside the circle which is inscribed the analyzed image. Thus, the outer corners of the image can be discarded.

As a reminder, the objectives of the two invariant versions of the WPD are to solve the view-point variations of the image patterns and to offer a unique decomposition. However, these two conditions can hardly be achieved concomitantly with a single decomposition. This is due to the fact the rotations and translations differ in nature when it comes to digital images. Therefore, our proposal is to use TI-WPD and RI-WPD in conjunction when analyzing an image, resulting in two "best" bases, one invariant to translations, and the other invariant to rotations.

In the following subsections, we introduce novel frameworks used to compute intrinsically different features, which significantly increase the robustness of classifiers on a variety of image datasets. It is worth noting that the framework we define here exploit the nature of the cost function used in the wavelet decompositions, namely, the entropy. As we started the proposal with a search of a sub-space of minimum entropy, it is intriguing to extract characteristics that further express and describe the entropy distribution among wavelet coefficients. However, we are not limited by the entropy and other features can be used to compute the features using the framework such as the: energy, mean, standard deviation, skewness, higher order statistics, etc. Before detailing the first feature extraction framework, it is important to remind that for an $W \times W$ dyadic image, the entropy is 0 when the image has constant intensity values and it reaches its maximum, $W \cdot \log_2 W$, when the probabilities of a pixel values are uniformly distributed, as seen in a random image.

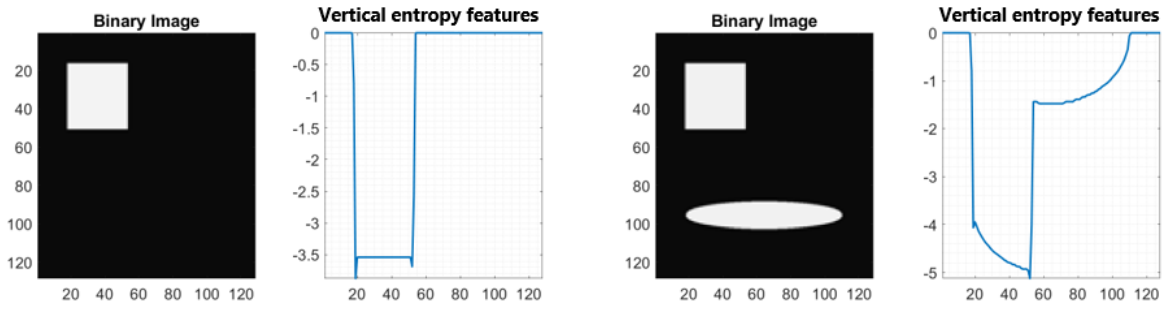
4.3.3 Vertical entropy features

The vertical feature framework consists of determining features along the vertical direction. For a dyadic image of $W \times W$, we generate a feature vector of length W by computing the entropy measure for each individual column. For the stated dyadic image, $f[x, y]$, the vertical entropy features are defined as a collection of values computed as follows:

$$V_{features} = \left\{ v_x \mid x = \overline{1, W} \right\}, \tag{4.42}$$

where $v_x = \sum_x f_y^2[x] \log_2 \left[f_y^2[x] \right]$ and f_y is the y^{th} row.

An example of this procedure is illustrated in Figure 4-21, where we can note two binary images, one containing a square pattern in the upper left corner and the other including the additional wide ellipse covering the lower part of the image. The entropy values in both scenarios in Figure 4-21 (a) and (b) determine the present of information and quantifies it in order to be used as features in subsequent computer vision tasks. For the first case presented in Figure 4-21 (b), the entropy senses the overlapping of the two objects and integrates the vertical entropy coming from the first image with the ellipse's information.



(a) (b)

Figure 4-21. Vertical entropy feature extraction for: (a) a binary image with a square; and (b) a binary image with an additional ellipse.

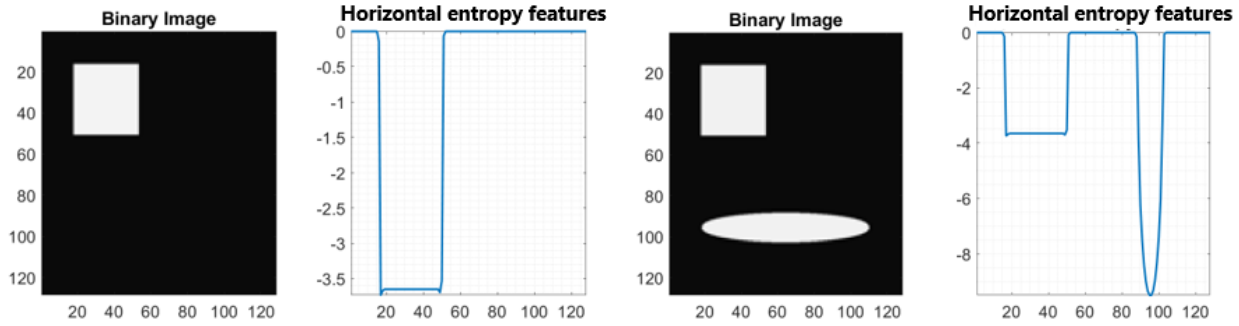
4.3.4 Horizontal entropy features

In the same manner as the vertical entropy features, we compute the entropy vector for each individual row, determining the horizontal features. For the same image, $f[x,y]$, the horizontal entropy features are formalized as in:

$$H_{features} = \left\{ h_y \mid y = \overline{1,W} \right\}, \tag{4.43}$$

where $h_x = \sum_y f_x^2[y] \log_2 \left[f_x^2[y] \right]$ and f_x is the x^{th} column.

As in the previous framework, the same two images are used in order to determine the features and evaluate the results. The extracted features are depicted in Figure 4-22 (a) for the first image; and (b) for the second. With this approach, the objects are separated and individually characterized, implying the complementary relations between the vertical and horizontal frameworks.



(a) (b)

Figure 4-22. Horizontal entropy feature extraction for: (a) a binary image with a square; and (b) a binary image with an additional ellipse.

4.3.5 N-directional entropy features

Starting from the same ideas of the two previous frameworks, we outline the information distribution in N directions or regions of interest (ROI) originating from the image center and

extending to its borders. Thus, the feature vector of an image contains N entropy values, corresponding to each region and can be defined as:

$$D_{features} = \left\{ d_n \mid n = \overline{1,N} \right\}, \tag{4.44}$$

where $d_n = \sum_x \sum_y f^2[x,y] \log_2 \left[f^2[x,y] \right]$ and $(x,y) \in ROI_n$.

The decision of N relies on the information content and spatial distribution within the image. Generally, it is advisable for the maximum value should be a power of 2 and lower than image's smallest dimension, W to prevent pixel overlap between regions. If these criteria aren't met, the analysis may still offer a satisfactory description but might involve redundant information and increased computational overhead. This is due to the fact that for a number of regions greater than W, pixels will be part of two or multiple regions at the same time, thus the redundancy occurs.

To exemplify the N-directional approach, two binary images are tested. The first image and its characterizations are presented in Figure 4-23 (a) for $N = 4$; (b) for $N = 8$; and (c) for $N = 16$ directions. The image used in those four cases depicts a single pattern as in the previous frameworks, a white square in the upper left corner. In the first scenario, for $N = 4$, we split the image into four regions annotated with numbers, denoting the directions, as shown in Figure 4-23 (a). The polar plot on the right side presents the entropy values of each region. We use a polar plot as it offers an easily interpretable representation, in accordance with the analyzed image. As we observe, only in ROI 1 information is present as it has the only non-zero entropy value, while the others are zero due to constant intensity in those ROIs. As we increase the number of directions of analysis, the polar plot becomes more detailed, providing a more precise depiction of the information distribution within the image.

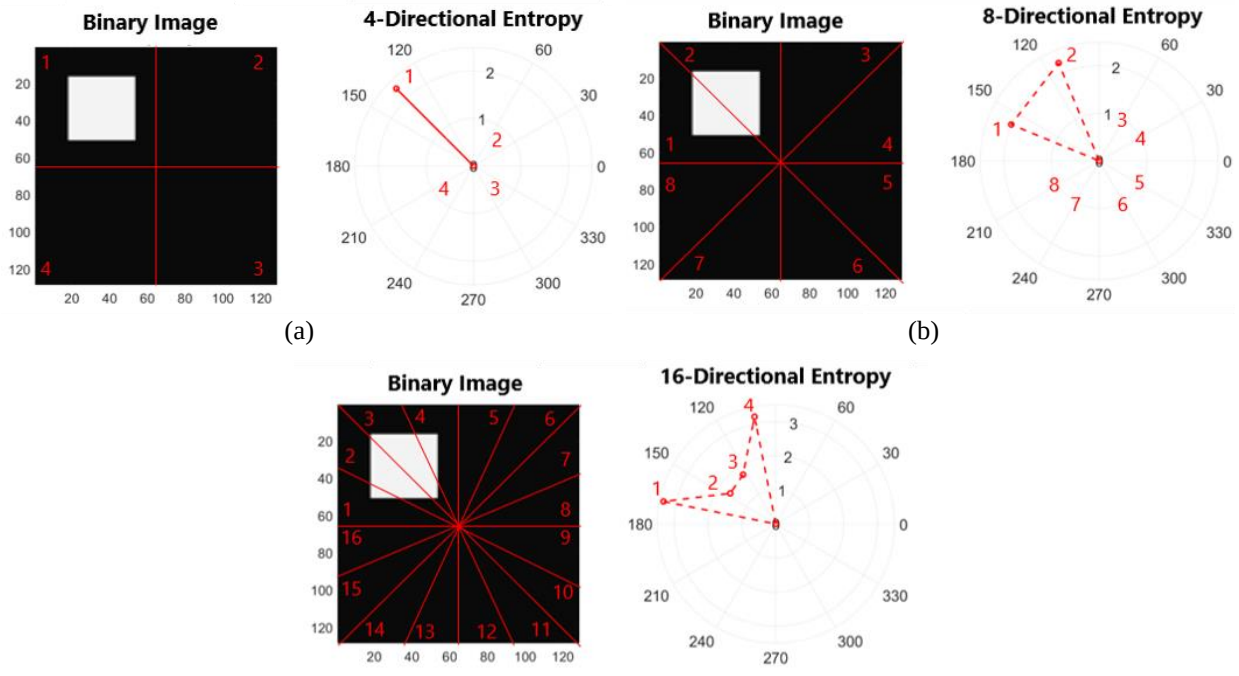


Figure 4-23. N -directional entropy characterization for a binary image with a white square for: (a) $N = 4$; (b) $N = 8$; and (c) $N = 16$;

In Figure 4-23 (b) as the number of directions became eight, the white square is characterized by two entropy values from ROIs 1 and 2 which are similar in value, but not the same. This difference comes from the imperfect alignment of the white square. Lastly, Figure 4-23 (c) shows a 16-directional analysis, where we observe the four entropy values from ROIs 1, 2, 3 and 4, each describing parts of the white square.

For the second image, with the additional wide ellipse, the N-directional analysis is presented in Figure 4-24 (a) for $N=4$; (b) for $N=8$; and (c) for $N=16$. It is worth noting that in this scenario, increasing the number of directions to $N=16$, allows us to jointly localize and quantify image information content: both the square and the ellipse are clearly characterized by their entropy values. Considering the vertical and horizontal frameworks, the N-directional entropy fuses the information coming from both in such a manner that it allows for a more detailed analysis.

The N-directional feature extraction framework can strengthen the imposed invariance conditions due to its circular nature. The input image can be slightly preprocessed to generate features quasi-invariant to image rotations only by considering that the information is mostly concentrated inside the circle inscribed in the analyzed image as it is depicted in Figure 4-24 (d). Indeed, there is a trade-off between information loss and quasi-invariant N-directional features. Nonetheless, considering the objective of the thesis, it is worth experimenting with this approach. The complete invariance is achieved when the degree of image rotation is equal to the central angle of any ROI, α_n, which is the angle formed by the two lines defining the region and that has its vertex in the center of the inscribed circle as shown in Figure 4-24 (d). In this case, the features are circularly shifted while their values remain the same as it is shown in Figure 4-25. The 8-directional entropy analysis of the two THz images yield the same results but which are rotated with 180 degrees. In every other case, the information exchanges between the ROIs will impact the generated features to some degree, thus the quasi-invariancy to rotations.

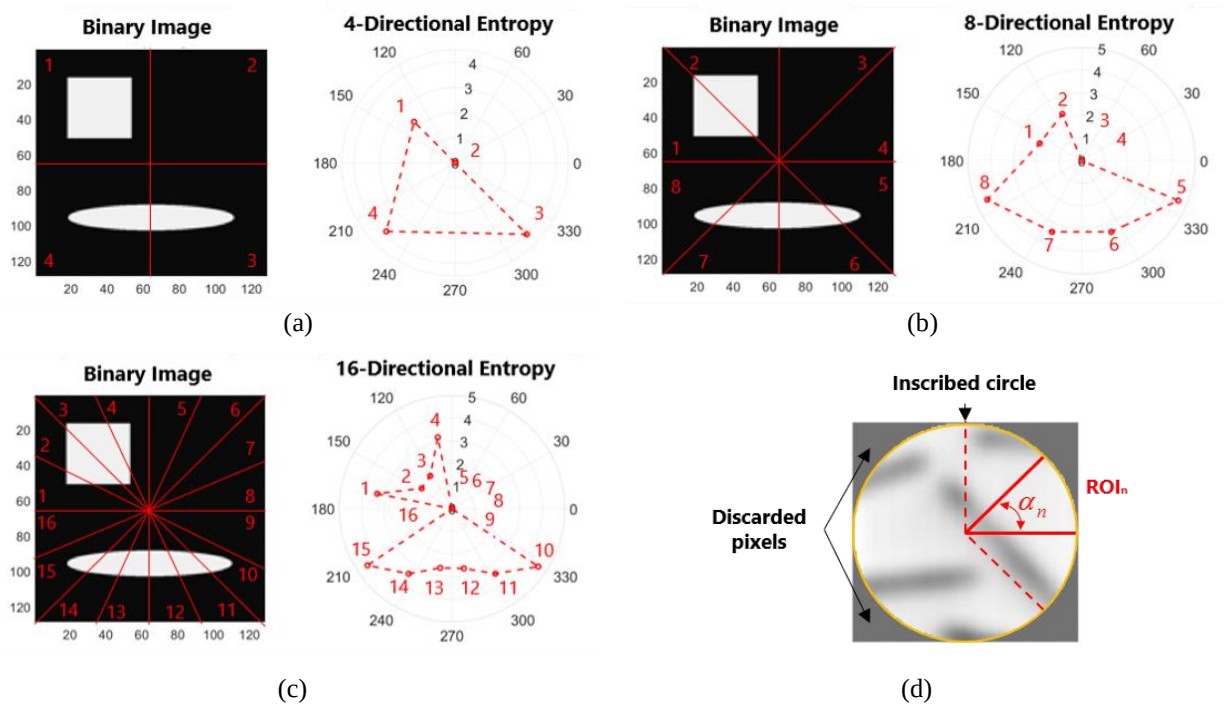


Figure 4-24. N -directional entropy characterization for a binary image with a white square and an ellipse for: (a) $N=4$; (b) $N=8$; (c) $N=16$; and (d) ROI and angle created by two lines defining a ROI

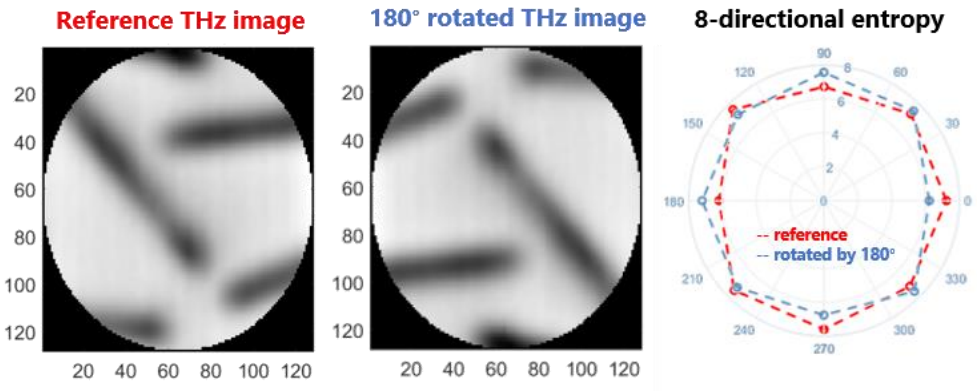


Figure 4-25. 8-directional entropy for a reference THz image and its 180° rotated version.

Due to direct relationship between the invariance and the ROI angle, we conclude that the more ROIs we use, the smaller the angle is; thus, the features are becoming resilient in front of image rotations when N has a greater value. However, as the value of N increases, the number of pixels inside the ROI decreases, resulting in the computed entropy becoming less relevant. The value of N should be corelated with the patterns and their size. In this research, we generally use 16, 32 or 64 ROIs per image when computing the directional entropy.

4.3.6 N-zonal entropy features

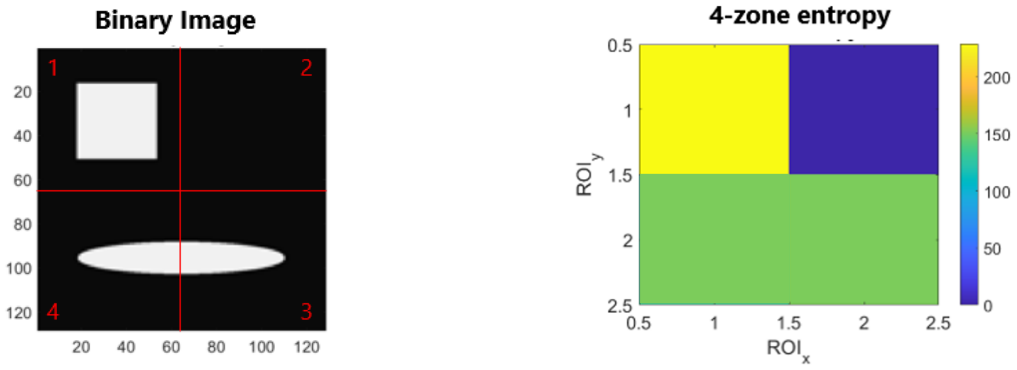
As the patterns or information present in an image are not always known, the previous framework is complemented by other types of regions to compute features. Another interesting approach is a square mask. In this context, the image is split in N square zones covering the whole image and which are also considered ROIs. The feature vector is expressed using the same principle as in the previous frameworks using a collection of values as in:

$$Z_{features} = \left\{ z_n \mid n = \overline{1,N} \right\}, \tag{4.45}$$

where $z_n = \sum_x \sum_y f^2[x,y] \log_2 \left[f^2[x,y] \right]$ and $(x,y) \in ROI_n$.

In an ideal case, the number of ROIs, N, should be generally chosen such that the dimensions of a ROI capture important parts of the pattern or motive present in the analyzed image. For a dyadic image of $W \times W$ pixels, $N = 2^r$, where $r \le \log_2 W$. As in the N-directional framework, the trade-off is between the resolution of the analysis and the information contained in each ROI. Considering the trade-off, N should not be too large, as this transposes into small area ROIs which are usually insufficient and irrelevant. For the purpose of the thesis, the usual chosen values are 16 and 64 zones.

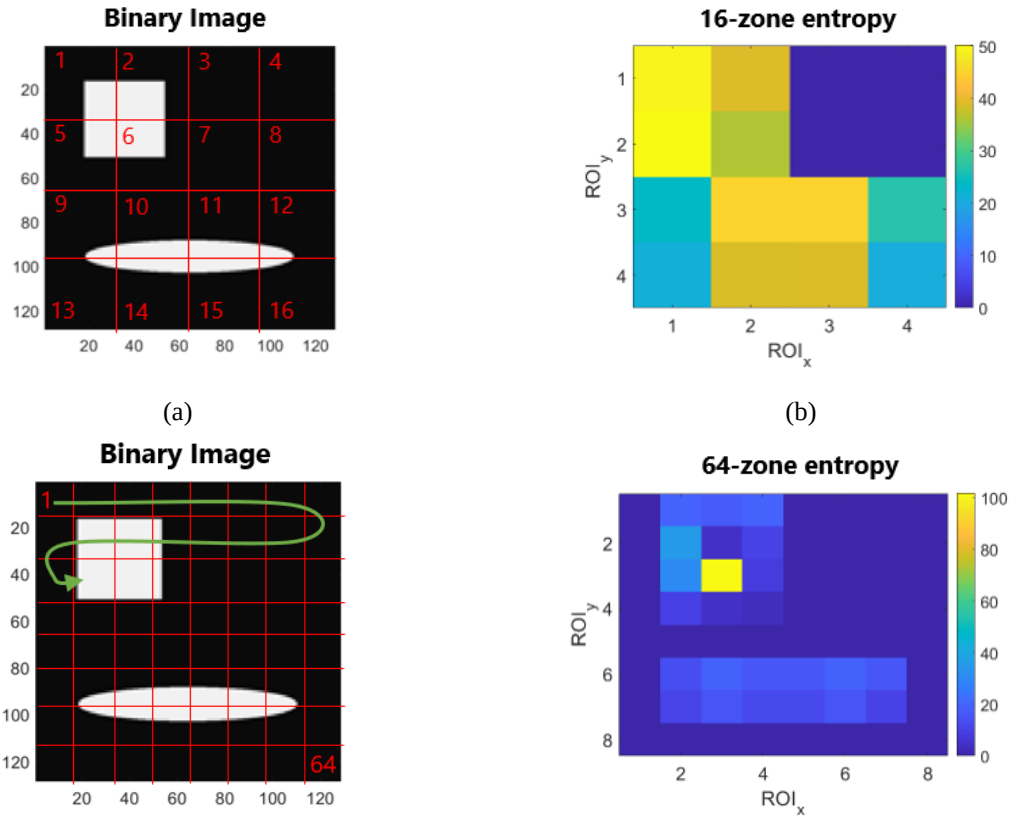
An example analysis is presented in Figure 4-26 for the same binary image with a square and an ellipse. The first case when $N = 4$ zones is analogous to a 4-directional analysis using the previous framework so we expect the same features. The representation of the 4 square ROIs is presented in Figure 4-26 (a). The corresponding and conveniently arranged entropy features are depicted in Figure 4-26 (b). In this framework, a more relevant and intuitive representation of the features is in grid form as it resembles the ROIs positions in the original image. In Figure 4-26 (b) the entropy features depict an initial low-resolution distribution of information inside the image.



(a) (b)

Figure 4-26. (a) The binary image and the corresponding ROIs for a (b) 4-zonal entropy characterization

As we increase the number of analysis zones from $N=4$ to $N=16$, we observe that the analysis starts offering insights about the distribution of information. The 16 ROIs are presented in Figure 4-27 (a) and the 16-zonal entropy characterization in Figure 4-27 (b). However, the separation of the two objects is not clear not even in this case. In the last scenario presented in Figure 4-27 (c) and (d), when $N=64$, the characterization of the image is even more detailed, the objects being clearly distinguished and characterized.



(a) (b)

(c) (d)

Figure 4-27. (a) The binary image and the corresponding 16 ROIs for a (b) 16-zonal entropy characterization; (c) The corresponding 64 ROIs for a (b) 64-zonal entropy characterization;

It is noteworthy that this framework is not invariant to rotations, but is used alongside with the N-directional analysis to enhance the discriminatory power of the features. In this manner, by using different ROIs when computing the distribution of information, we have more perspectives about the patterns and the data analyzed.

A second example analysis is presented in Figure 4-28 where we provide the 64-zonal entropy features for a THz image. The lower entropy values correspond to regions in the original image where the distribution of pixel intensities is mostly constant and organized. The higher values are correlated with noisy regions with little to no informational gain.

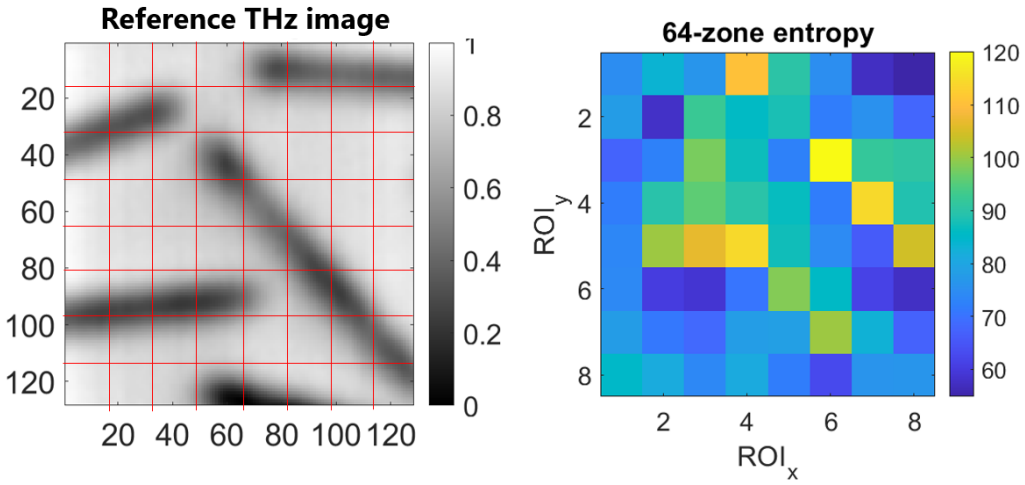


Figure 4-28. THz image and its corresponding 64-zone entropy analysis.

4.4 Chapter summary

This chapter presented various methods of image analysis, including DCT, DWT, and WPD, with an emphasis on the common features extracted using these transforms. These approaches generally perform well in straightforward scenarios where there are no variations in viewpoint. However, to address real-world cases where the image of a pattern or subject can vary significantly, we have enhanced these image decomposition techniques and computed additional features based on these upgraded versions.

In the final subsection, we developed and introduced the tools we propose to use in image analysis procedures, specifically Translation-Invariant Wavelet Packet Decomposition and Rotation-Invariant Wavelet Packet Decomposition. The features derived from our frameworks can be further engineered and organized to adapt to a variety of machine learning classifiers. This adaptability ensures that our methods can be effectively utilized in diverse application scenarios.

Additionally, these frameworks are versatile and can be integrated with numerous statistical measures, allowing the computation of complementary features and providing different perspectives on the analyzed image. By incorporating these advanced techniques, our methods offer a more robust and comprehensive approach to image analysis, capable of handling the complexities and variations encountered in real-world scenarios. This enhances the overall accuracy and reliability of image classification and pattern recognition tasks, making our frameworks valuable tools in the field of image analysis.

Chapter 5. THz image classification

5.1 Introduction

In the previous chapter, a general image analysis framework was introduced considering view-point variations constraints. Starting from the invariant wavelet decompositions, we developed feature extraction methods that are able to provide localized and complete insights about the analyzed image. The frameworks allow for complementary analysis that result in a powerful set of features that can be used as discriminating assets. From this point, the last challenge tackled by the thesis regards the image classification, a task typically regarded as one of the final stages in an image processing chain.

Considering the scarcity of THz data, ML classifiers are preferred compared to a DL approach. Therefore, this chapter provides details about novel feature assembling techniques prior to feeding them to ML algorithms. In this case, feature assembling refers to the patterns or structures in which the already computed features are transposed and can be further exploited in conjunction with ML classifiers. These assembling techniques are tailored based on the graph representation of a wavelet decomposition, but also on physical considerations such as the wavelet frequency sub-bands. In other words, starting from the vertical, horizontal, N-directional or N-zonal frameworks, the resulting features can be represented in graph form, thus reflecting the structural frequency decomposition of the TI-WPD or RI-WPD tree. The features coming from this approach are perfectly adapted to train a Graph Neural Network (GNN) for the image classification task. The second idea we propose is to aggregate the features based on their depth and type, two identifiers of wavelet sub-images. In this case, the features are fed to classic ML classifiers such as k-NN, SVM, decision trees, etc.

Finally, the chapter ends by presenting the comparative results of ML image classification using classical features presented in the previous chapter and the features coming out from the frameworks we proposed. To correctly asses our approach, we tested a large pool of ML classifiers on many proprietary THz datasets, but also on other non-THz public benchmark datasets. In the following sub-sections, we explain the assembling techniques and their corresponding ML classifiers, considering a WPD "best" basis. Nevertheless, one family of transforms on which these assembling techniques can be applied is the one that follow the multiresolution concept and can be represented through a graph. As already mentioned, one example is WPD, but TI-WPD and RI-WPD are included and also preferred in the conducted experiments due to their invariant nature.

5.2 Graph-approach in image classification

Starting from a WPD "best" basis structure, the proposed features are assembled in the same manner, retaining the frequency sub-bands selection structure. The process can be visualized in Figure 5-1 where we depict the 2-level WPD "best" basis of an image where three level-one and four level-two sub-images are considered to "best" represent the original THz image. On each sub-image we compute one or multiple features using the vertical, horizontal, N-directional, or N-zonal frameworks. To maintain the structure of the WPD basis, for sub-images not selected to be part of the "best" basis we use a feature vector of zeros with the same size as the vector computed for the other nodes.

A graph signal processing approach provides the mathematical tools to define these graph-structured features. Thus, a brief introduction of the concepts adapted to our research is needed. Generally, a graph is an object that represents relations between entities. Entities are commonly referred to as "nodes" or "vertices", while relations are the "edges" connecting the nodes. A graph

can be directed if the edges connecting the nodes point in only one direction. If the directions do not exist, or the edges are bidirectional, the graph is undirected. In our case, both types of graphs can be implied when assembling the features, however we choose to use the restricted case of directed graphs to retain the parent-child relation concept also presented in a wavelet decomposition. An example of a directed graph's nodes and edges is presented in Figure 5-1 along with the features computed on the sub-images from the "best" basis.

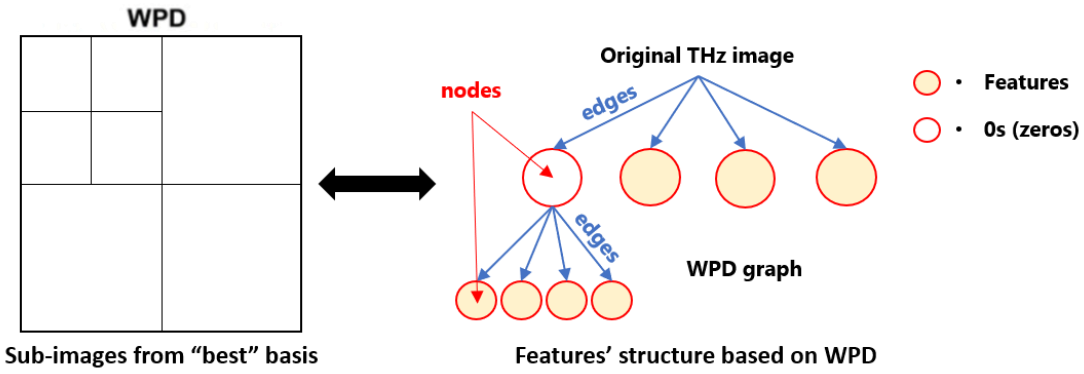


Figure 5-1. Graph-structured features starting from a WPD

A directed graph, G_f, containing the set of features and edges can be formally defined as follows:

$$G_f = (V, E) \tag{5.1}$$

where V and E represent the vertices' features and edges, respectively. It is noteworthy that the set of elements, $\left\{(v_i, e_{ij}) \in G_f \mid i, j = \overline{1, N_V}\right\}$, are ordered so that the graph is directed.

The edges of the graph, E, can be weighted if any relation between wavelet coefficients is considered, otherwise, all edges will share the same weight, equally weighting the features. In this thesis, we consider that all the wavelet sub-images and their corresponding features bring information of the same priority and importance as they are all apriori selected with a pruning algorithm. Thus, the edges' weights, e_{ij}, are equal to 1. Other approaches might consider a weighting law for the "best" basis edges, or a more exhaustive idea by exploiting the full wavelet decomposition while assigning weights to edges according to an informational criterion.

The structure of directed and unweighted graphs containing N_V vertices can be explicitly defined with an adjacency matrix, A, as in:

$$A = \left\{a_{ij} \mid i, j = \overline{1, N_V}\right\}, \tag{5.2}$$

where

$$a_{ij} = \begin{cases} 1 & \text{if there is a edge from node } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}, \tag{5.3}$$

and a_{ij} indicates the connection from node i to node j.

The conclusive idea of using graph structured features is that they can be fed to GNNs, an emerging trend in signal processing and machine learning that has the potential in increasing performances in computer vision tasks [104], [105]. In our research, GNNs take as input graph-structured features and aggregates them into new graph-representations using the general framework presented in [106]. Finally, the newly transformed features are passed to a Multi-Layer Perceptron (MLP) network for classification. A MLP is a neural network with multiple layers suitable for solving non-linear classification problems.

Using (5.1) and (5.2) we define the features, their edges and structure according to a WPD graph. One example of such graph-representation is presented in Figure 5-2 (a) where the nodes, v_i, represent the wavelet sub-images; the edges, e_{ij}, are the arrows connecting the nodes. Additional self-loop edges are also depicted. Self-loops in GNNs refer to edges connecting nodes to themselves. These self-loops are not present in a typical WPD graph, but from an algorithmic point of view they stabilize the learning process by addressing issues related to the lack of self-connections in classic convolutional operations on graphs, making the process more robust [107], [108]. They are also a simple mechanism for nodes to influence their own features during message-passing [109], [110], [111] or aggregation steps [112], [113]. This mechanism is particularly useful for refining the representation of nodes based on their own information. However, the benefits of self-loops vary depending on the specific application and the characteristics of the graph [114], [115]. Our experiments demonstrated that the inclusion of self-loops increases the convergence speed during training by facilitating the information exchange between nodes.

The adjacency matrix in Figure 5-2 (b) represents the graph from Figure 5-2 (a): with 1s we show the weighted connections between nodes and with 0s we represent the absence of connection. Each one of the nodes, v_i, contains the horizontal, vertical, N-directional, N-zonal features or a combination of these features.

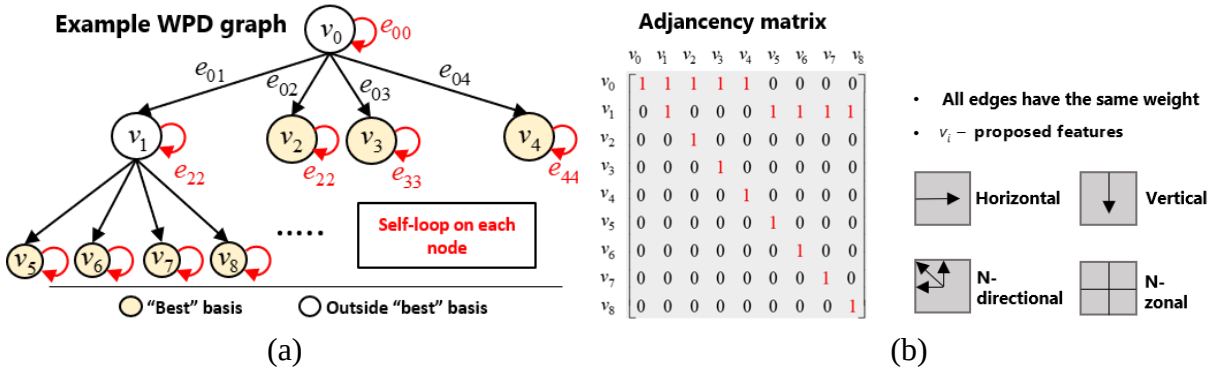


Figure 5-2. (a) Example graph with self-loops; and (b) its adjacency matrix with the proposed features corresponding to each node

Having described the structure of the features, we now introduce the GNN architecture and its general mathematical formalization. In this case, we start by presenting the intra-layer graph-processing, we continue with the inter-layer relationship and layer stacking, and finally we present the classification neural network.

If we consider the GNN's architecture presented in Figure 5-3 [106], the k-th layer, conveniently denoted as General Graph Convolution (GGC) layer, can be defined as in:

$$v_p^{(k+1)} = AGG\left(\left\{ACT\left(DROPOUT\left(BN\left(W^{(k)}v_q^{(k)} + b^{(k)}\right)\right)\right), q \in \xi(p)\right\}\right), \tag{5.4}$$

where $v_p^{(k+1)}$ is the k-th layer embedding node, $W^{(k)}$ and $b^{(k)}$ are the trainable weights and biases, and $\xi(p)$ is the local neighborhood of v_p, referring to all nodes that have direct connection to the analyzed node.

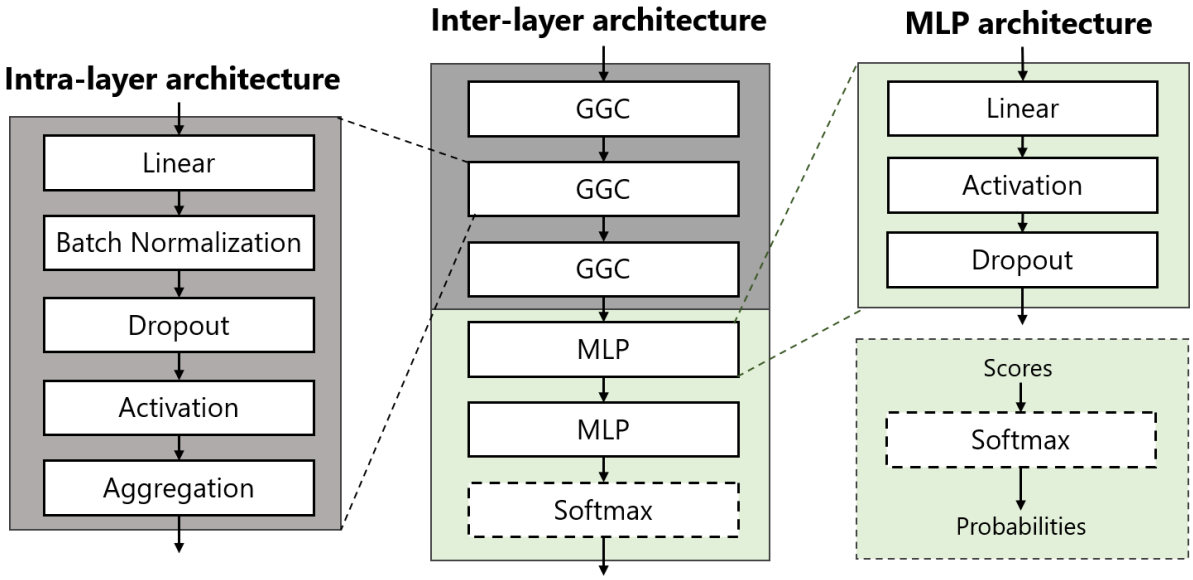


Figure 5-3. The architecture of the GNN

Considering the intra-layer architecture depicted in Figure 5-3, we further describe the role of each component step by step in the following:

a) The linear layer, often referred as the "weight" layer has an important role in transforming the aggregated neighborhood information into updated node representations. In general, this layer follows a linear mapping as shown in (5.4) using the trainable parameters, W and b, representing the weighting matrix and the bias.

b) Batch normalization is used to improve the training stability and convergence speed. In the proposed architecture, batch normalization operates independently on each node's features, considering the features of a single node as a batch [116]. For a node, v_q, with a feature vector of length N_{Feat}, the node-wise batch normalization is defined as follows:

$$BN\left(v_q\right) = \frac{v_q - \mu_q}{\sigma_q}, \tag{5.5}$$

where μ_q and σ_q are the mean and the variance of the feature vector corresponding to node v_q.

These two variables are further defined as in:

$$\mu_q = \frac{1}{N_{Feat}} \sum_{i=1}^{N_{Feat}} v_q[i], \tag{5.6}$$

$$\sigma_q = \sqrt{\frac{1}{N_{Feat}} \sum_{i=1}^{N_{Feat}} \left(v_q[i] - \mu_q\right)^2}. \tag{5.7}$$

c) The dropout, as in classic neural networks, offers a method to mitigate overfitting and enhance the model's generalization [104]. A mathematical formalization of this technique incorporates the input vector, a probability, p, representing the fraction of units to drop out and a binary mask of identical shape as the input vector, where each entry is either 0 (dropped out) or 1 (kept). Thus, the dropout process can be expressed as in:

$$DROPOUT(x) = \frac{Mask \otimes x}{1 - p}, \tag{5.8}$$

where x represents the hidden layer output. The probability, p, is a hyperparameter that is usually tuned based on the characteristics of the data and the complexity of the model. In our case, our experiments showed that a value of 0.3 best fits our data and thus, it is used throughout the presented results.

d) The activation function introduces non-linearity to the model, enabling it to learn intricate patterns within the data and facilitating information propagation in GNNs. Some of the most common activation functions in the context of GNNs are Rectified Linear Unit (ReLU), Parametric Rectified Linear Unit (PReLU) or Swish [106]. As the choice of an activation function often depends on the specific task and dataset, it is a common practice to experiment with different functions to find the best solution empirically. However, from a computational point of view, especially when working with exhaustive wavelet decompositions, we chose to use the activation with the lowest computational impact, ReLU, which is defined as in:

$$\mathrm{ReLU}(x) = \max(0, x). \tag{5.9}$$

where x is the input vector.

As already mentioned, the favorable characteristic of ReLU is that it induces sparsity in the network when the input is negative and thus, being beneficial for memory and computational efficiency. Another aspect regarding the choice of ReLU is that it has demonstrated empirical success in various machine learning applications, many state-of-the-art models using it as an activation function [104], [106], [117]. From this point of view, ReLU is an appealing first choice when experimenting and it is used throughout the comparative analysis later in this chapter.

e) The final aggregation step is a crucial component in a GNN architecture as it is in charge of combining information from neighbor nodes to update node representations. As for our

scenario, we experimented with the *sum* aggregation defined for the aggregated representation $v_q^{(k+1)}$ for node q at layer $k+1$ as in:

$$SUM: \ v_q^{(k+1)} = \sum_{i \in \xi(q)} v_i^{(k)}, \tag{5.10}$$

where $\xi(q)$ represents all the neighbors of node q.

The *sum* function preserves global information and considers the overall contribution of all neighbor nodes.

f) Having defined the GGC layer, the following step is to stack multiple layers to create a graph neural network as shown in Figure 5-3. In practice, stacking is one of the most frequent options. One of the reasons behind stacking multiple layers is that they allow the model to learn hierarchical representations of the input graph, with each layer increasing the complexity of the patterns and dependencies in the represented data [104], [111], [113], [118]. In this sense, we propose to stack three GGC layers as they are shallow enough to prevent overfitting, and deep enough to capture complex feature relationships [119].

g) The final neural network, MLP, depicted in Figure 5-3 stacks two layers, each being composed of a linear layer, with ReLU activation and a Dropout function. The two-layer feed-forward neural network is a common classifier, usually used in conjunction with a Softmax [120] loss function in computer vision classification task. The Softmax function accepts a vector of real numbers as input and transforms it into a probability distribution through normalization. Given an input vector, x, the Softmax function produces the following output:

$$\text{softmax}\left(x_i\right) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}, \tag{5.11}$$

where x_i is the i-th value of the input vector, e is the base of the natural logarithm and K is the total number of classes.

One advantage of the Softmax function is that it offers a probabilistic interpretation of the output. The output scores of each individual class are transposed into probabilistic confidence values. If the output value at position i is close to 1, it indicates high confidence that the input belongs to class i, while values close to 0 indicate low confidence.

An overview of the proposed graph-based approach in image classification is presented in Figure 5-4. Starting from a database of images, we decompose all the images using TI-WPD and RI-WPD. From all the sub-images in the "best" bases we extract the proposed features and assemble them in graph-form. The features are then split in training, validation and test sets with a ratio of 60/20/20. The described GNN is trained considering the cross-validation approach [121]. Cross-validation is a model evaluation technique wherein the dataset is divided into multiple subsets. The model is trained on some of these subsets and evaluated on the remaining ones. This method offers a more reliable estimate of the model's performance than a single train-test split. By averaging the performance across multiple folds, it minimizes the influence of randomness from a

single split, thus mitigating overfitting and decreasing the variance in performance estimates, such as those derived from the confusion matrix.

The trained weights of the GNN model are then used to assess its classification performance using the confusion matrix and the accuracy obtained on the test dataset. The schematically workflow presented in Figure 5-4 is the methodology used in all experiments we conducted.

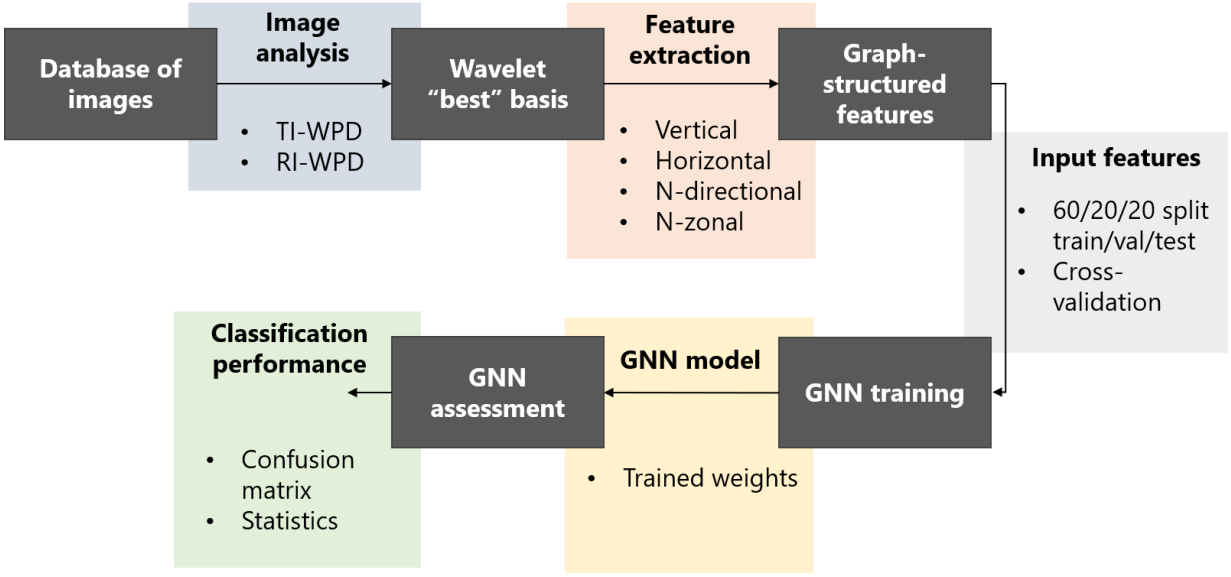


Figure 5-4. Overview of the graph-based approach in image classification

5.3 A 4-channel aggregation framework with frequency localization of wavelet sub-bands

This subsection presents the second novel feature extraction technique which is based on aggregating characteristics and including the frequency information of the "best" basis in the final feature vector. This process can be visualized in Figure 5-5, where we start by depicting an example wavelet 2 level decomposition of an image computed using either TI-WPD or RI-WPD. The 4-channel aggregation procedure implies a joint between the aggregated sub-images' features based on their level and type; and the structural frequency domain arrangement of the "best" basis expressed by the collection of sub-images' normalized central frequencies.

The 4-channel aggregation name comes from the four types of wavelet sub-images represented by the approximation and three detail coefficients. The aggregation table showed in Figure 5-5 is firstly comprised by a collection of features that can be defined as in:

$$F_{d,p} = \sum_{i=1}^{N_{d,p}} \Upsilon\left\langle C_{p,i}^{d} \right\rangle, \tag{5.12}$$

where $C_{p,i}^{d}$ are the $p-$ type coefficients at level d of decomposition, $N_{d,p}$ is the number of $p-$ type coefficients at level d, and $\Upsilon\langle\cdot\rangle$ is the formal expression of features extraction techniques. Many features extraction techniques can be used and adapted with the presented framework. However, considering that the pruning algorithm is based on the minimization of entropy, it is worth to further experiment with the same concept of measuring the informational randomness of

each wavelet coefficient. Moreover, the values of the sub-images' entropy are already computed through the pruning algorithm; thus, we also reduce the time complexity by avoiding additional computations.

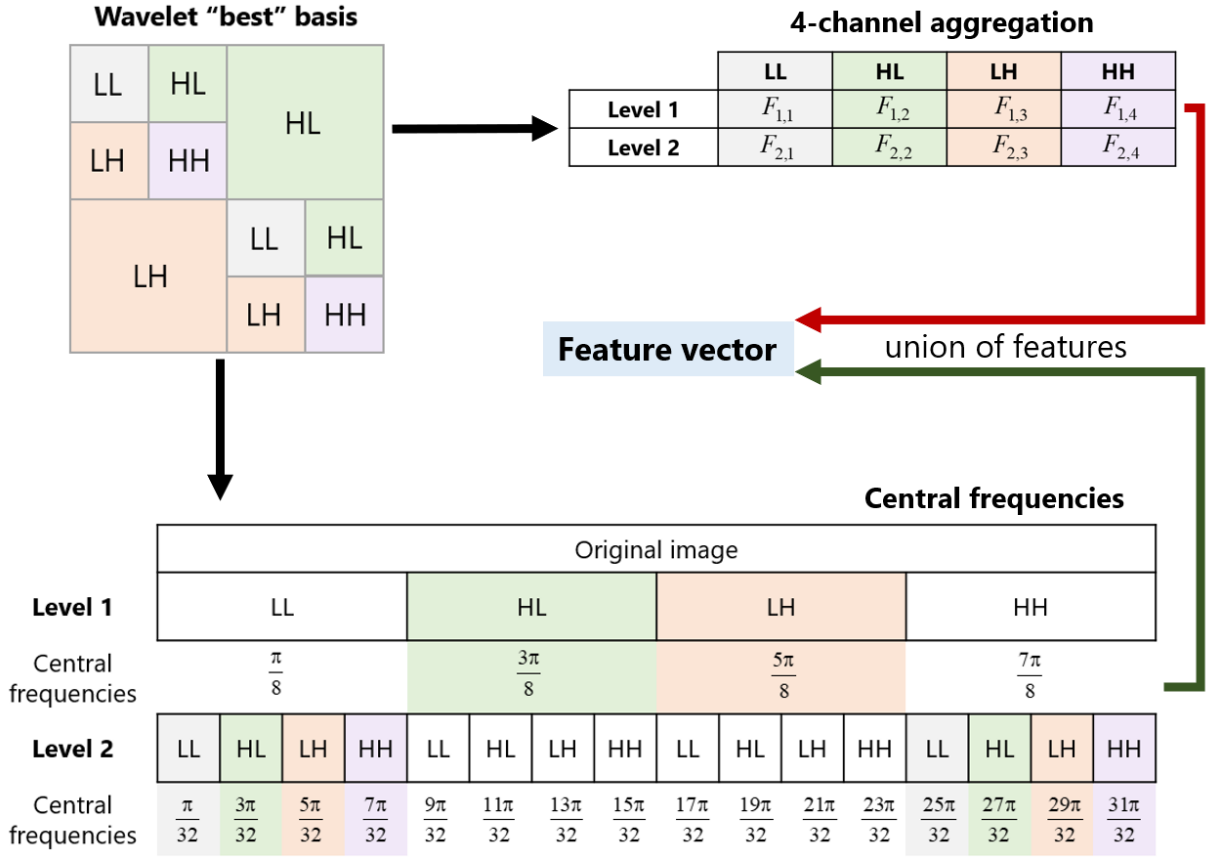


Figure 5-5. A 4-channel feature aggregation example for a wavelet "best" basis

The central frequencies of the wavelet sub-images are normalized between 0 and π. As it is presented in Figure 5-5, they are added as elements to the features defined in (5.12), thus creating the final feature vector coming from the 4-channel aggregation approach is:

$$F_{ch-agg} = \left\{F_{d,p}\right\} \cup \left\{F_{central}\left(C_{p,i}^{d}\right) \mid 1 \le d \le D,\ 1 \le p \le 4,\ 1 \le i \le N_{d,p}\right\}. \tag{5.13}$$

ML classifiers such as k-NN, SVM, NN and Decision Trees are used to evaluate the features on several datasets of images. As for the GNN approach, for 4-channel aggregation with sub-band localization we present in Figure 5-6 an overview of the image classification workflow. We start from a database of images of different classes. These images are further decomposed using TI-WPD and RI-WPD. The "best" bases are characterized using the proposed framework and the feature space is created. The initial feature space is then split in training, validation and testing sets in order to comply with the cross-validation principles. A variety of ML classifiers are simultaneously trained and tested using the confusion matrix and the accuracy inferred from it.

In the subsequent section, we showcase the outcomes of our proposed frameworks, which are tested on two proprietary THz databases and then expanding to public benchmark datasets found in image pattern recognition literature.

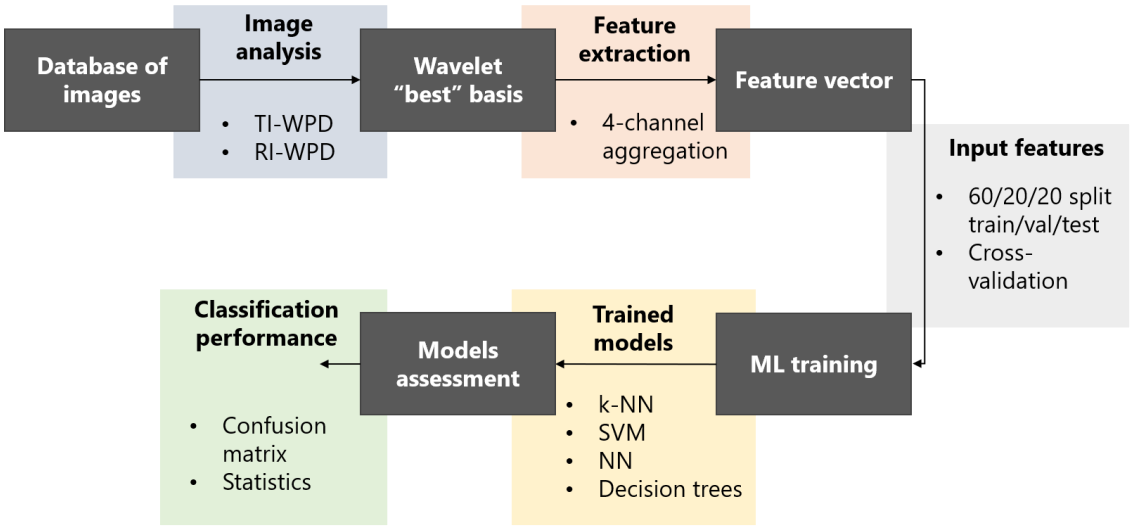


Figure 5-6. Overview of the 4-channel framework in image classification

5.4 Results and discussion

In Table 2 we present a synthesis of the image analysis decompositions along with the feature extraction techniques assessed on multiple databases and with different classifiers. In other words, Table 2 depicts all scenarios evaluated in our research and provides the general guidelines of this section. Before diving into results, we first describe the scenarios, the ML classifiers and introduce the evaluated databases. Following this section, we present more details about the databases along with the performance analysis of the classifiers.

a) In the DCT case, we test as features the full decomposition coefficients which is exhaustive due to the higher dimension of the feature vector; a zig-zag approach improving the efficiency by selecting a smaller number of coefficients that traverse the whole frequency content of the DCT; a regional mask, extracting all the high-frequency coefficients from the bottom-right region of the DCT matrix; and Jing's bands with the constraint that the bands composing the feature-vector retain more than 95% of the maximum energy. For DCT, we use one-dimensional feature vectors to train k-NN, SVM, NN and a decision tree.

b) For DWT and WPD, the features are, one at a time, the mean, the entropy, or the energy of the sub-images. The additional novel feature extraction technique used here is the four-channel entropy aggregation. For DWT and WPD we use one-dimensional feature vectors to train the same ML classifiers as for DCT, but also the graph structuring approach of features in order to train a GNN. For the graph structured features, we use the classic mean, entropy and energy of the sub-images.

c) For TI-WPD and RI-WPD, we tested only the vertical, horizontal, N-directional, N-zonal and the four-channel aggregation techniques with the entropy as the means to characterize the sub-images of the decomposition. The number of directions in the N-directional framework is 8, and the number of zones in the N-zonal is 9. Both the one-dimensional feature vector and the graph approach are tested in these cases. The

'Coiflet3' wavelet family is used throughout the analysis of all wavelet-based approaches.

Regarding the ML classifiers, the following paragraph, emphasizes the most important aspects of their configuration. As the possibilities of the classifiers' configuration are indefinite, we choose to implement the most common architectures from the literature [122]: the k-NN algorithm uses $k = 3$ neighbors to classify the data [123]. The SVM uses a non-linear cubic kernel [124] to capture more complex relationships between data, thus to have a better classification capacity. The NN is a MLP with an input layer equal in size to the length of the feature vector, two layers of 128 neurons and a classification layer with a dimension that varies with the number of classes in each database. The decision tree has a maximum flexibility of 20 splits to help avoid overfitting. The GNN architecture, as it was described earlier, uses three GGC layers, each having 128 neurons per linear layers. The classification network, following the convolutional layers, is a MLP with two 128 neurons layers and an additional output layer with a varying number of neurons, based on the number of classes.

In terms of setup and training parameters, in all scenarios, the learning rate is 0.001. The training process is continued until the validation accuracy is not improving for five consecutive iterations with more than 1%. The learning algorithm is Adam [125]. The batch size used during training is 16.

The evaluation databases consist of images usually depicting patterns under multiple view-points variations including translations and rotations. Two of the databases exploit the THz frequency range to create the image of the pattern: THz database 1 and THz database 2. The other one exploits the optical range and is publicly available: Brodatz [126].

Analysis	Features	Analysis	Features
DCT	Full coefficients Zig-zag mask Regional mask Jing mask	TI-WPD	Vertical Entropy Horizontal Entropy N-directional Entropy N-zonal Entropy Channel agg. (entropy)
DWT	Mean Entropy Energy Channel agg. (entropy)	RI-WPD	Vertical Entropy Horizontal Entropy N-directional Entropy N-zonal Entropy Channel agg. (entropy)
WPD	Mean Entropy Energy Channel agg. (entropy)	**ML classifiers:** k-NN, SVM, NN, decision tree and GNN	

Evaluated databases	THz database 1
	THz database 2
	Brodatz

Table 2. Synthesis of evaluated approaches in classification task

The following sections will describe each database and the analysis of the classifiers' performance.

5.4.1 THz dataset 1

In this scenario we use improved THz images [80] of a sample comprising a paper substrate (95%) and randomly embedded copper fibers (5%). The enhanced images are the result of using the Phase Diagram Dynamic Characterization Image (PD-DCI) reconstruction method which exploits the phase diagram trajectory's dynamism using Frenet-Serret framework [85]. The algorithm is presented in Chapter 3. The database consists of 100 images per class, with a resolution of 64×64 pixels. Figure 5-7 illustrates the optical image of the sample alongside the corresponding THz image obtained using our THz imaging system and a raster-scanning approach. It is noteworthy to mention that by using THz frequency range to create the image of a sample, we emphasize characteristics that are not visible in the optical range.

To create this image database, we partition the original image into 9 patterns as in Figure 5-7. Within each class region, we apply a circular mask to highlight the contained information. This circular mask allows for the translation and rotation of an additional sampling mask without the possibility of mixing pixel information between classes. The aim is to ensure the independence of the images generated for each class.

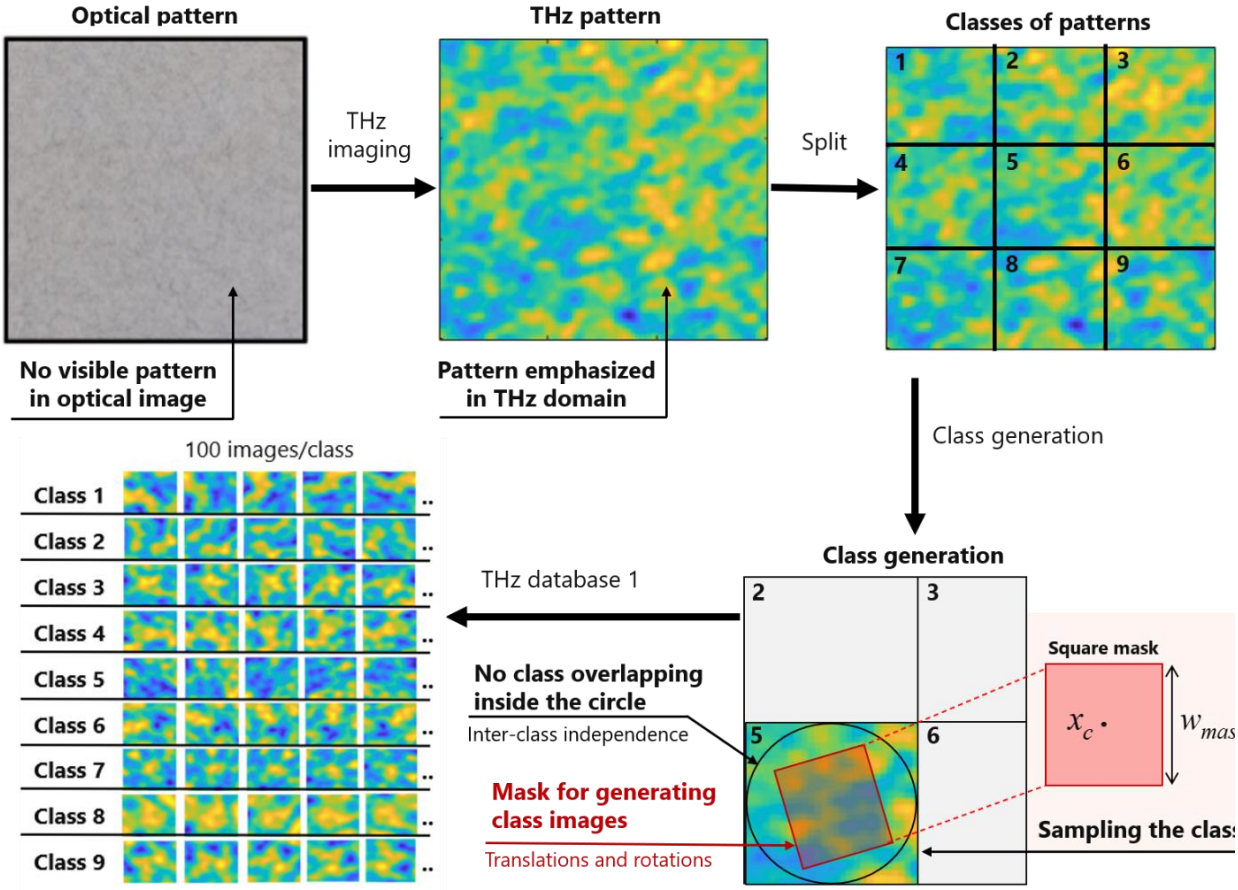


Figure 5-7. THz database 1 generation: 9 classes, 100 images per class

The sampling mask creates the images of a class by sampling at random and unique positions defined by the mask's center position (x_c, y_c), its width w_{mask}, and angle θ_{mask}, such that the square is inscribed in the circle. The arbitrary positions and orientations of the sampling mask assures different view-points on the pattern of interest. The number of images per class is 100 with a resolution of 64×64 pixels.

To enhance the learning and evaluation process, we divide the database with the ratio 60/20/20 corresponding for training, validation and testing sets.

The classes depict blobs of different dimensions and forms. In this database, the images of a class almost depict the same pattern, but with varying information due to the translations and rotations. Both of these geometric transformations imply changes of the information contained in the image, even if we refer about pixel loss due to translations, or pixels' intensities deformations due to digital rotations. Examples of images from each class are presented in Figure 5-7, where we observe the same pattern for a class but under multiple perspectives.

Regarding the performance analysis on the classification task, we provide individual results for each image analysis method.

a) In Figure 5-8 (a), we present the accuracy on the test THz dataset 1 of DCT-based features. We determine that the full DCT coefficients used with SVM yields the highest accuracy of 69.3%, followed by 62.7% achieved by the same feature vector with the NN.

   Overall, the accuracy of the DCT-based approaches is low as the transform cannot efficiently handle rotations and translations, resulting in high intra-class variability of the energy distribution among the transform coefficients. The consequence is an erroneous classification. Moreover, the mathematical model behind DCT is not optimal for emphasizing and localizing features.

b) In Figure 5-8 (b) the performance analysis of DWT is presented. It is showed that the four-channel entropy aggregation with a NN yields the highest accuracy of 78.65%. In the second place it is the decision tree with an accuracy of 72.6%.

c) Figure 5-8 (c) show that the WPD-based features extraction methods, with the graph-structured entropy features and a GNN achieves the highest accuracy of 83.05%. In the second place comes the four-channel entropy aggregation with NN with an accuracy of 81.95%.

   It is important to note that starting with DWT and its generalization, WPD, the accuracy scores increase as the wavelet-based transforms are suited for characterization of patterns due to the multiresolution concept and space-scale localization of features. However, as for DCT, the translation and rotation constraints are not handled by neither DWT or WPD in their classical implementation. This results in lower classification rates of datasets, such as THz database 1, that includes variations of the view-point.

   It is noteworthy to mention that through this analysis we observed that the four-channel aggregation features increased the accuracy of all classifiers based on DWT and WPD, demonstrating better discrimination abilities than the classic mean, energy and entropy features.

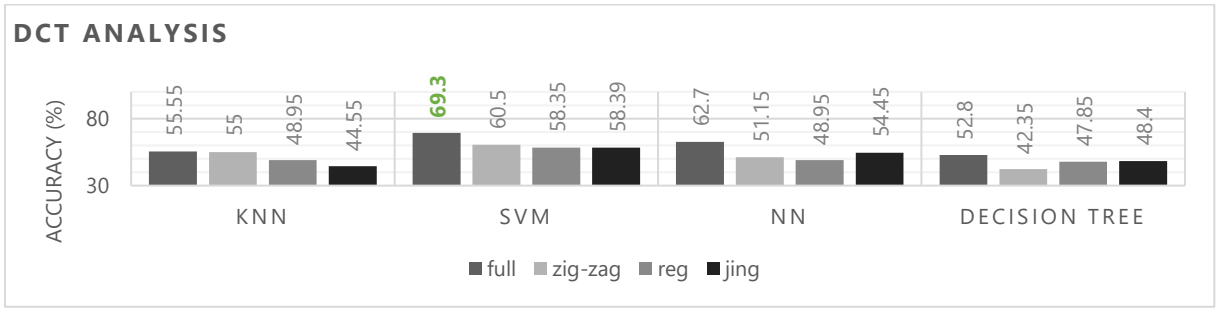
   The extended versatility of WPD throughout the frequency partitioning and selection is also reflected in the accuracy results, WPD achieving better scores than its counterpart, DWT. Additionally, we observe that the DCT-based approaches yield

some of the lowest accuracy scores compared to the wavelet-based features considering the disadvantages it has in the pattern classification task.

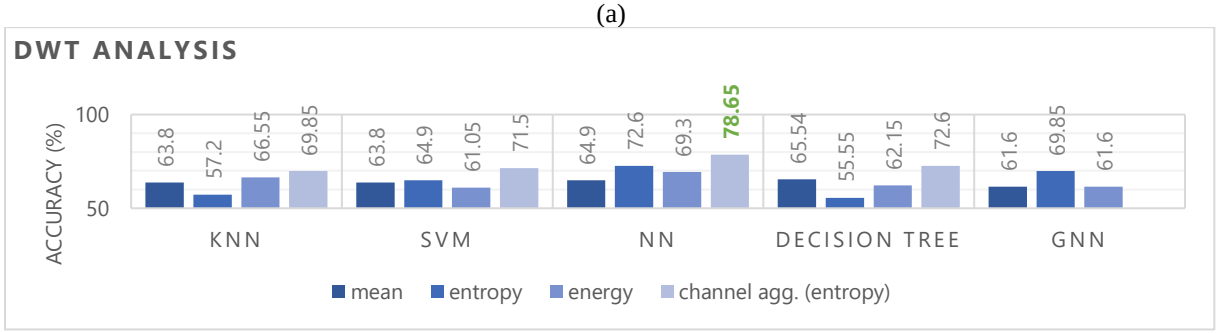
d) Figure 5-8 (d) and (e) presents the performance of TI-WPD and RI-WPD-based feature extraction techniques. It is worth noting than even the lowest accuracy of TI-WPD or RI-WPD scenarios, 84.7%, is greater than the highest achieved by the classic image decompositions such as DCT, DWT or WPD and the most common features, 83.05%.

This is expected, as the classic image decompositions vary from image to image, while for TI-WPD and RI-WPD, the "best"-bases converge to a quasi-unique representation. In other words, the TI-WPD and RI-WPD "best"-bases are similar, in their own context, and their variability is lowered. This suggests that addressing translation and rotation constraints at the decomposition level can significantly enhance the accuracy of the classifiers. The N-zonal features of TI-WPD achieves an accuracy score of 92.95%, while the N-directional features of RI-WPD goes up to 95.6%.

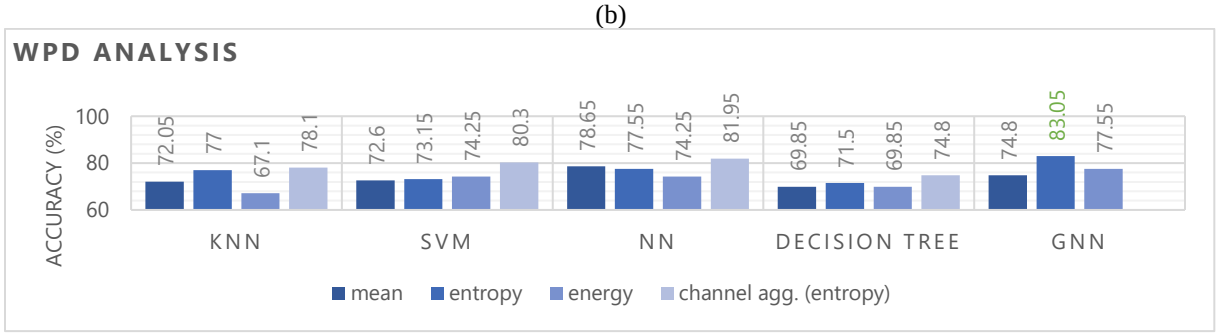
Among the introduced features, the vertical and horizontal frameworks have the lowest score. The reason behind this is that the features emphasize the sub-images' characteristics in only one direction, either vertically or horizontally. The accuracy scores are increased by using image areas such as the slices of the N-directional framework, or rectangles of the N-zonal framework. In this manner, the pattern is better analyzed and leads to more discriminatory features, thus, a more robust classifier.



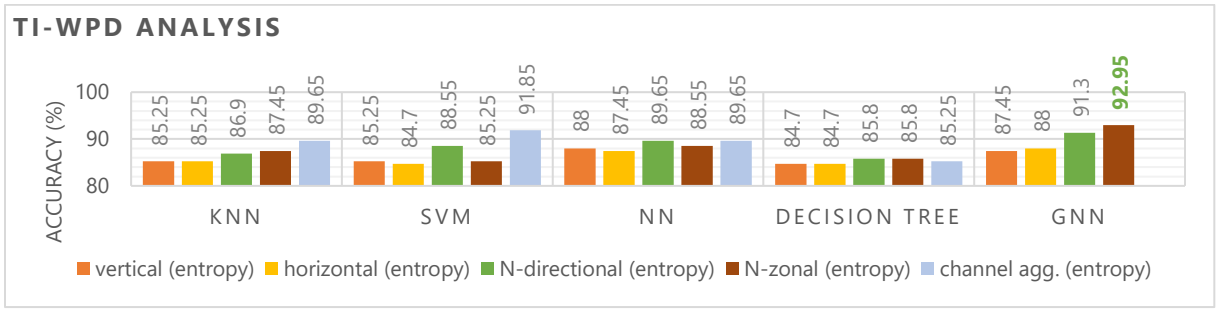
(a)



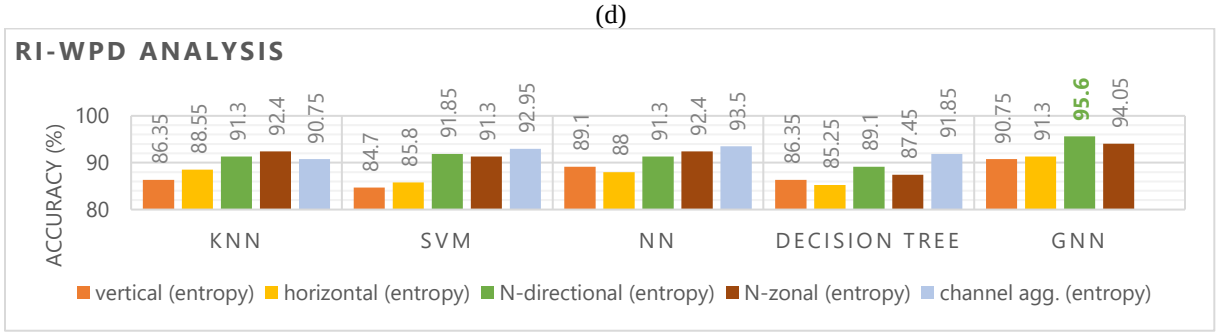
(b)



(c)



(d)



(e)

Figure 5-8. Overview of the (a) DCT features performance; (b) DWT feature performance; (c) WPD feature performance; (d) TI-WPD feature performance; and (e) RI-WPD feature performance

Finally, Figure 5-9 shows the best outcome score for every image decomposition approach and ranks them by the accuracy obtained on the test dataset. The RI-WPD with N-directional features and TI-WPD with N-zonal features achieve the highest and the second highest accuracy on this dataset. It is worth noting that for RI-WPD and TI-WPD, the ML classifier was the GNN. Its ability to exploit the frequency partitioning increases the robustness of the classifier. Considering that the "best" bases of images of the same class are quasi-unique, the discriminatory ability is increased as the different frequency structures of each class are included as features. This was not the case of DWT or WPD. The main reason the GNN does not performed well on DWT is due to the same frequency partitioning for all classes. For WPD, the issue stirs up from the high variability of the frequency partitioning. Finally, the third and fourth place are taken by WPD and DWT with four channel entropy aggregation. The lowest accuracy is obtained by using all DCT coefficients as features.

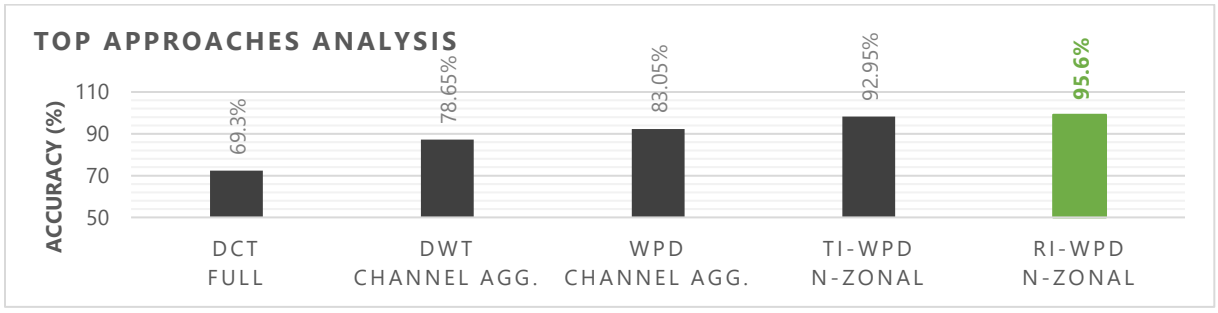


Figure 5-9. Overview of the approaches with the highest accuracy with every image decomposition.

The confusion matrix in Figure 5-10 continues the analysis of the GNN classifier by offering details about its performances. The confusions were between classes 2, 3, 4 and 9 which all exhibit similar patterns to an extent. Despite the visual similarity, the proposed approach was

able to achieve at least 90% accuracy on these classes, while the others were correctly assessed in their corresponding category.

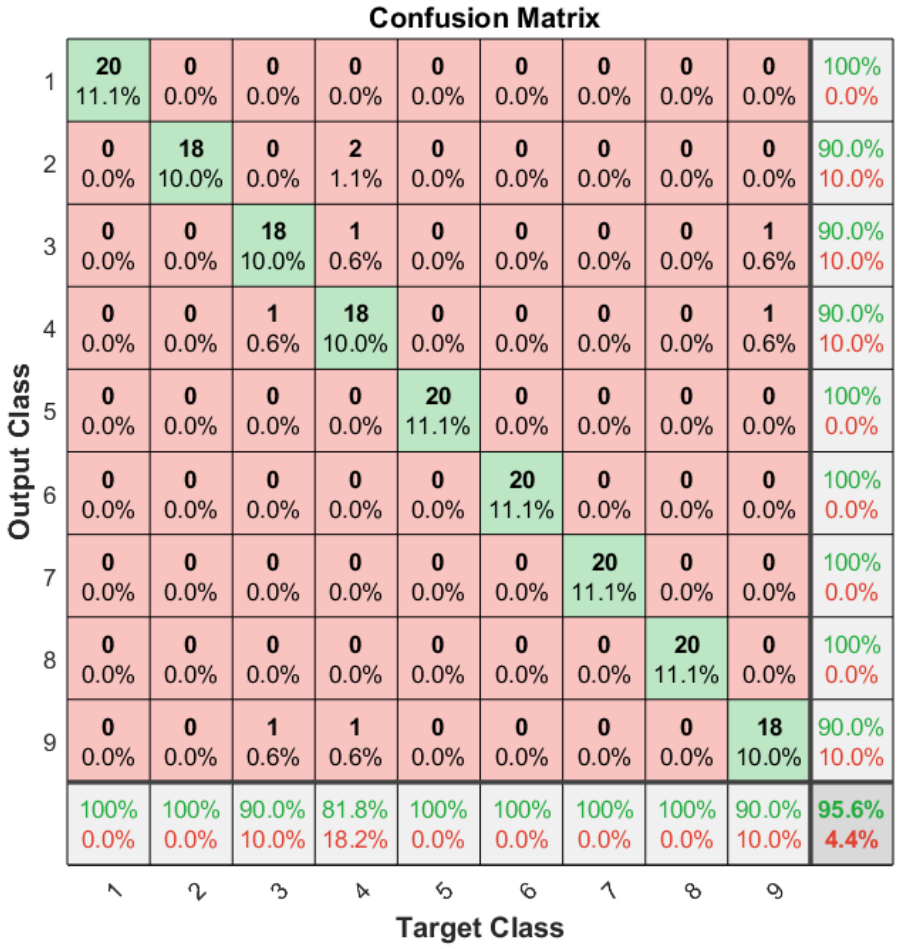
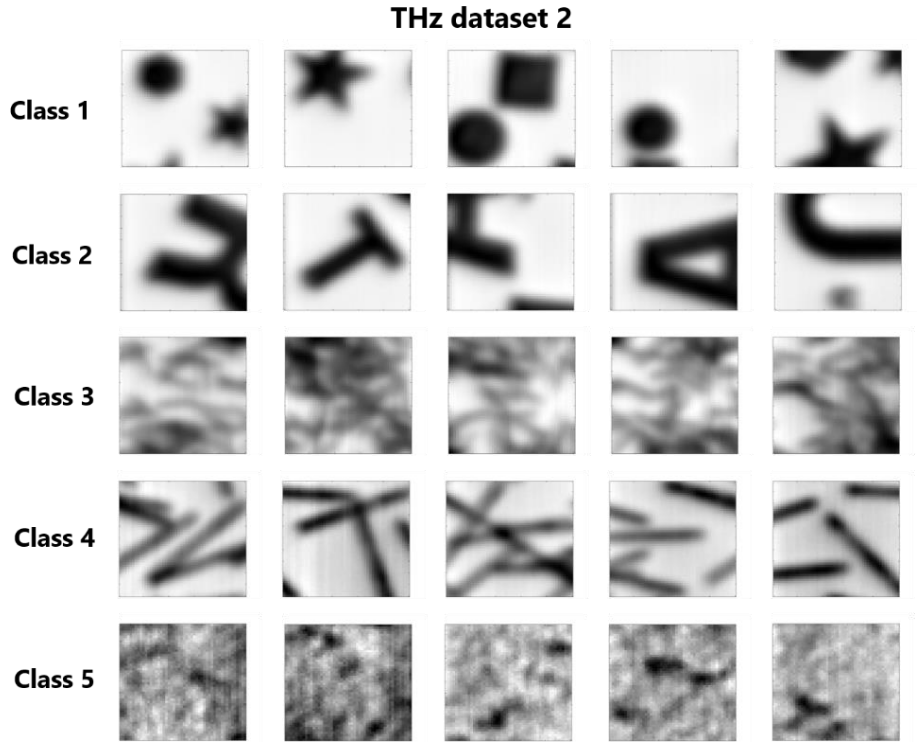


Figure 5-10. Confusion matrix of the classifier based on RI-WPD and N-directional features.

5.4.2 THz dataset 2

For our second performance review, we use another THz database containing 480 images per class. The images have a resolution of 128-by-128 pixels composing 5 classes that exhibit random natural translations and rotations, as opposed to the artificially generated translations and rotations from THz dataset 1. Each class consists of 480 images split into training, validation and test sets with 60/20/20 proportionality.

Figure 5-11 presents examples of images from each one of the five classes. The first class depicts shapes such as squares, circles, hexagons and stars. The second class contains letters “R”, “T”, “A” and “U”. The patterns of the first and second class can be partially or fully present in one image. Additionally, some of the images may have two or more shapes or letters. The third and fourth class contain small and large bars randomly distributed in the image. The sample of the fifth class is composed of copper fiber randomly distributed on paper substrate. All the other samples used to create these images are composed of polyethylene substrate with metallic-ink patterns. The images were acquired using the same THz imaging system.



	Class 1	Class 2	Class 3	Class 4	Class 5
Substrate	Polyethylene	Polyethylene	Polyethylene	Polyethylene	Paper
Patterns	Square, stars, circles, hexagons	Letters “R”, “T”, “A”, “U”	Small copper bars	Big copper bars	Copper fibers

Figure 5-11. Examples of images from the four classes of THz dataset 2

In the same manner as with THz database 1, the performance analysis is presented in order starting with DCT, DWT and WPD image decompositions, followed by TI-WPD and RI-WPD approaches.

a) In Figure 5-12 (a), we showcase the test THz dataset 2 accuracy for DCT-based features. Our analysis reveals that employing full DCT coefficients in conjunction with NN produces the highest accuracy at 74.62%, while the same feature vector with the SVM achieves a slightly lower accuracy of 71.22%. We observe that with this dataset, DCT-based methods also struggle with accuracy due to issues in handling rotations and translations.

b) The performance evaluation of DWT is depicted in Figure 5-12 (b). The results indicate that employing four-channel entropy aggregation with SVM yields the highest accuracy at 81.43%. Following closely is the NN, achieving an accuracy of 82.7%. In Figure 5-12 (c), the WPD-based feature extraction methods reveal that utilizing graph-structured entropy features with GNN attains the highest accuracy of 83.37%.

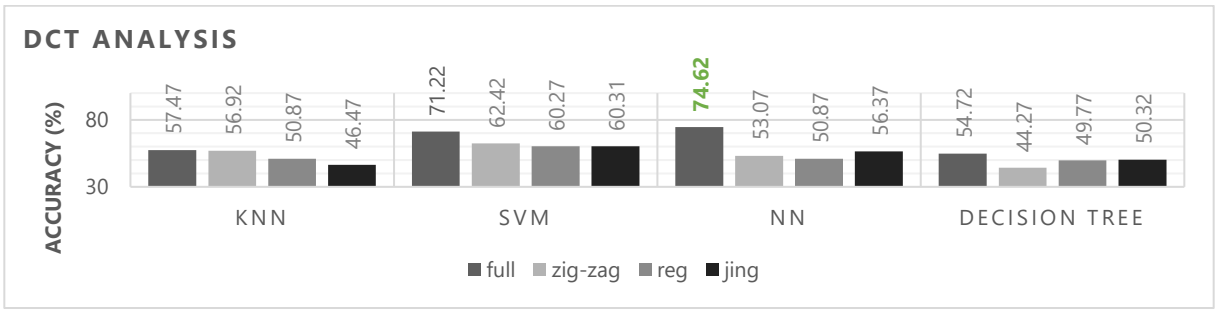
In the second position is the four-channel entropy aggregation with NN, achieving an accuracy of 82.27%. Starting with DWT and WPD, accuracy improves due to wavelet-based applicability to pattern characterization. However, like DCT, these

methods struggle with translation and rotation constraints, leading to lower classification rates.

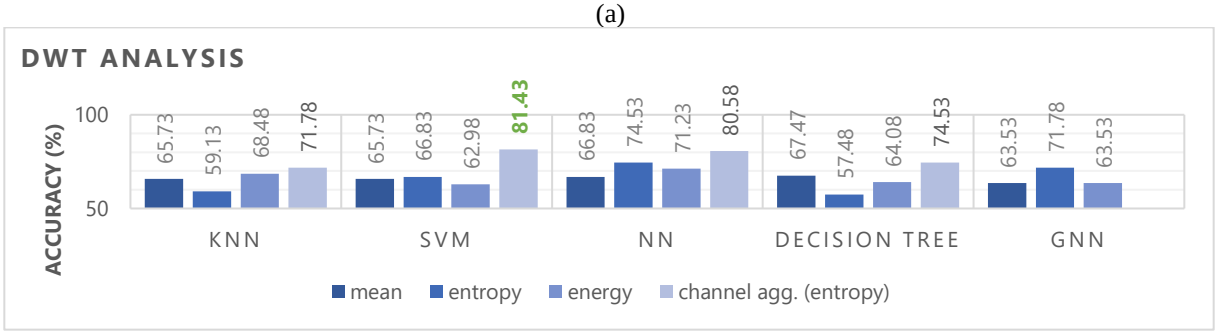
In our analysis, four-channel aggregation features notably improved accuracy for DWT and WPD classifiers, surpassing traditional mean, energy, and entropy features. WPD, with its enhanced versatility in frequency partitioning, outperformed DWT in accuracy. Conversely, DCT-based approaches showed lower accuracy, highlighting their limitations.

c) In Figure 5-12 (d) and (e), the efficiency of TI-WPD and RI-WPD-based feature extraction methods is presented. Notably, the lowest accuracy within TI-WPD or RI-WPD scenarios, at 86.1%, surpasses the highest achieved by traditional image decompositions like DCT, DWT, or WPD and classic features, which is 83.37%. Regarding their best performance, GNN classifier using the N-zonal features of TI-WPD achieves 97.89% accuracy, while the same classifier with the same features of RI-WPD reaches 99.8%.

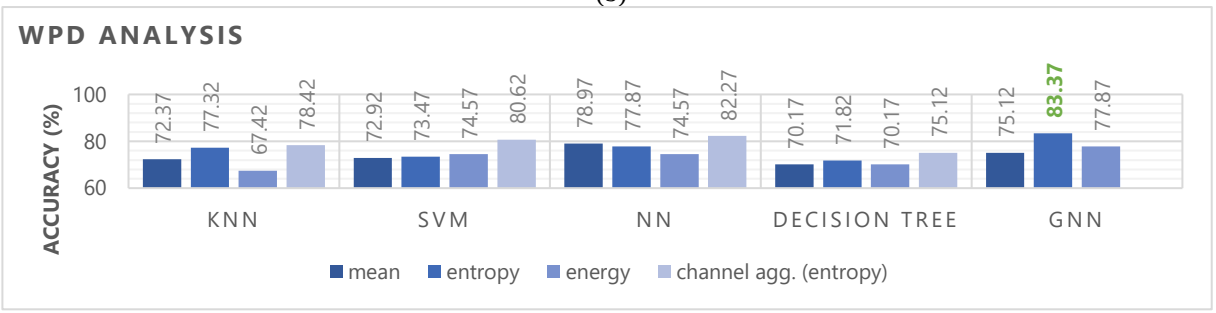
Vertical and horizontal frameworks, introduced in the features, score lower as they emphasize characteristics in one direction. Improved accuracy comes from using image areas like N-directional slices or N-zonal rectangles, enhancing pattern analysis for a robust classifier.



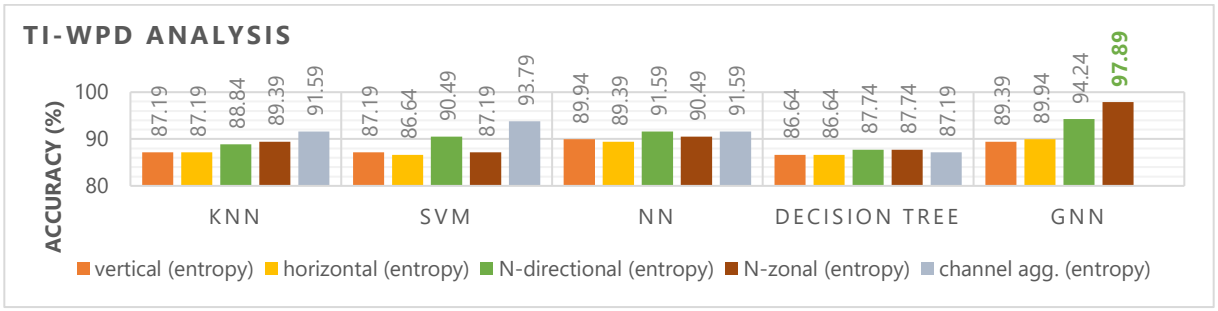
(a)



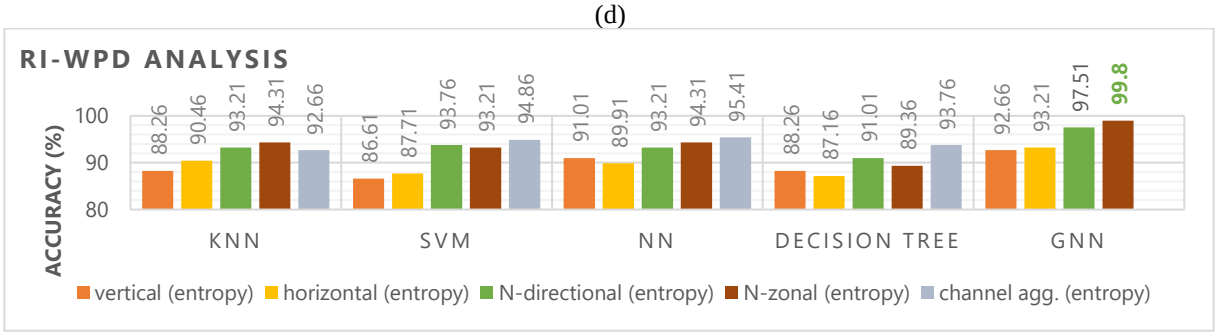
(b)



(c)



(d)



(e)

Figure 5-12. Overview of the (a) DCT features performance; (b) DWT feature performance; (c) WPD feature performance; (d) TI-WPD feature performance; and (e) RI-WPD feature performance

In Figure 5-13, the optimal outcome scores for various image decomposition approaches are displayed, ranked based on their accuracy on the test THz dataset 2. The TI-WPD and RI-WPD with N-zonal features secure the highest and second-highest accuracy on this dataset. Notably, GNN, employed as the ML classifier for RI-WPD and TI-WPD, enhances robustness by exploiting frequency partitioning. The quasi-uniqueness of "best" bases within the same class boosts discriminatory ability, an advantage not shared by DWT or WPD. Third and fourth places are secured by WPD with four-channel entropy aggregation and GNN, and DWT with the same features and SVM. The lowest accuracy results from using all DCT coefficients as features with NN.

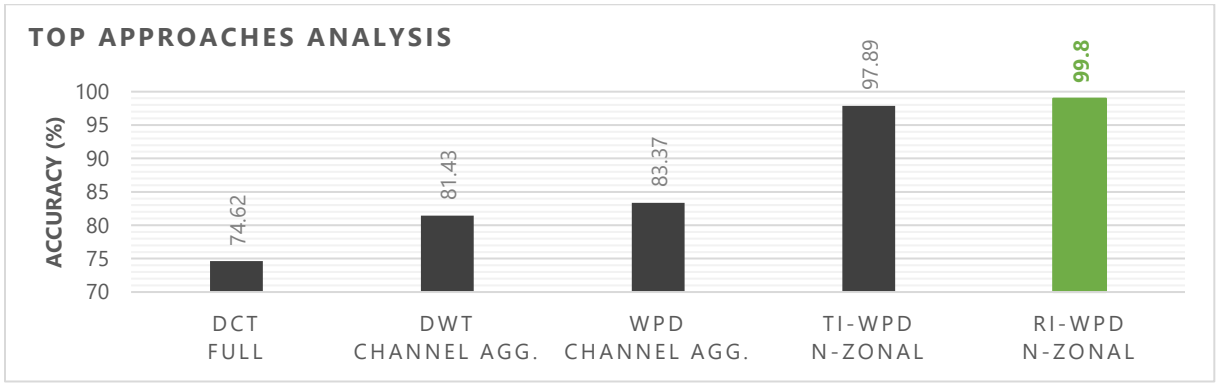


Figure 5-13. Overview of the approaches with the highest accuracy with every image decomposition

The confusion matrix in Figure 5-14 depicts the individual performance on each class of the classifier. As we observe, only one sample is mislabeled, while the others are successfully classified. The confused classes are the third and the fifth as they can share some visible similarities.

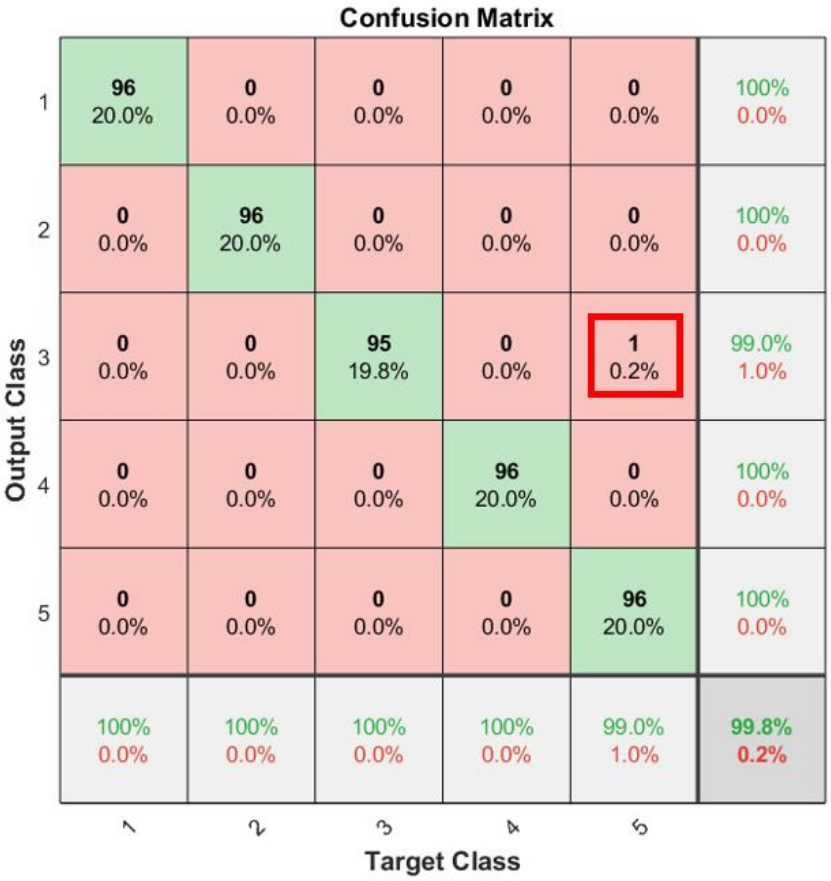


Figure 5-14. Confusion matrix of the classifier based on RI-WPD and N-zonal features.

Inferring from the two results obtained on THz dataset 1 and THz dataset 2, we assess that the most promising classification approach implies the invariant versions of the wavelet decompositions, TI-WPD and RI-WPD.

The introduced feature extraction methods have shown to increase the accuracy and robustness of ML classifiers, especially the four-channel aggregation approach. However, the GNN classifier exhibited the highest accuracy score only if it is used along with TI-WPD or RI-WPD, as they also naturally discriminate between patterns by having different "best" bases for each class.

In the following subsection we briefly present initial classification results on a public benchmark dataset, "Brodatz"[126].

5.5 Public benchmark dataset

The original "Brodatz" dataset contains 112 texture images of 640 by 640 pixels. The version of "Brodatz" dataset used in this study contains only 13 classes including surfaces of stone, bark, fabric, etc. For each class, from the initial texture image, we generated 48 images of 256 by 256 pixels. We do this, by splitting the image in four, and then rotate them with varying angles, $\theta \in \{0, 30, 60, ..., 330\}$. Finally, the images have their resolution reduced to 256 by 256 pixels to fit the dyadic nature of the wavelet decompositions. Figure 5-15 shows some example patches from each texture class. The class numbers start from the upper left with label "1", and continue left to right, up to down, to label "13".

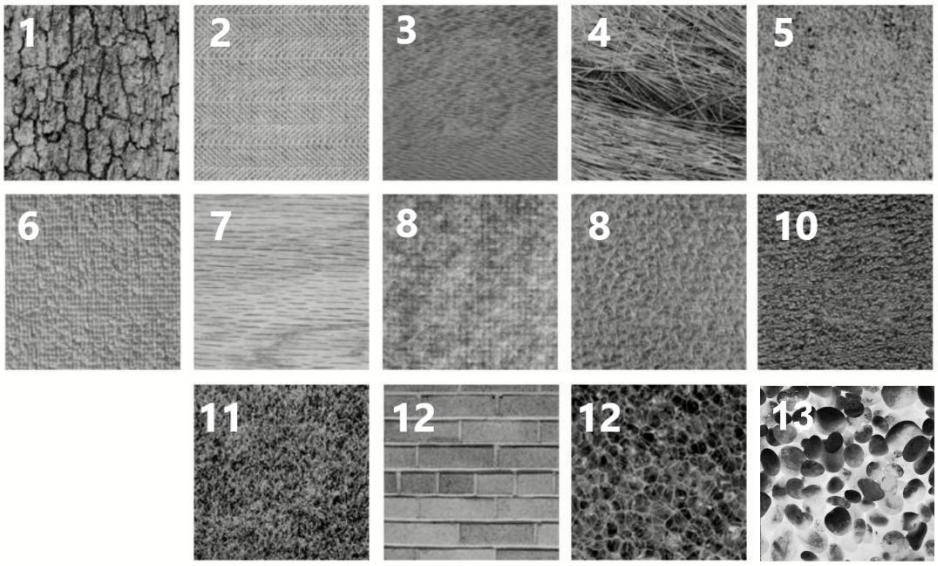


Figure 5-15. The thirteen classes of "Brodatz" dataset used in our research

Based on the results on THz, we computed only the best classifier for each image decomposition. For DCT, we trained a SVM on full coefficients. For DWT and WPD we used four channel entropy aggregation with NN and finally, for TI-WPD and RI-WPD we trained a GNN with N-zonal entropy features. The results are presented in Figure 5-16. As it can be seen, the highest accuracy was attained by the RI-WPD with N-zonal features and GNN, followed slightly behind by TI-WPD with the same features and classifier.

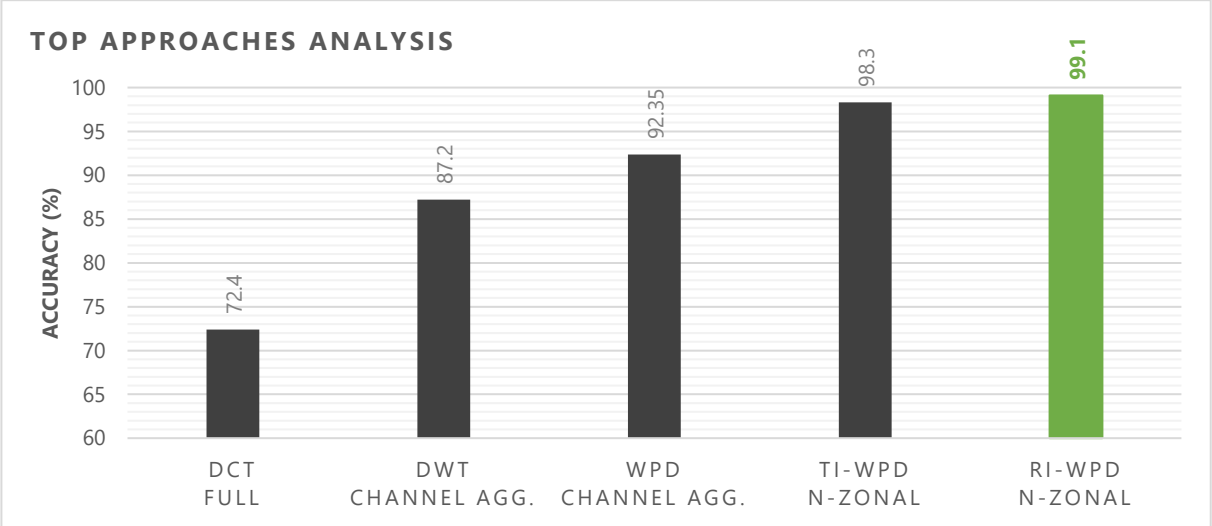


Figure 5-16. Overview of the approaches used for "Brodatz" dataset

The purpose of this analysis is to demonstrate that the introduced features and classification approaches based on GNN can be easily extended to any database of patterns, regardless of the imaging technology. Compared to some published state-of-the-art results [127], [128], we consider that despite having a different training approach, the results are comparable. The main difference between the training approaches is that they used the original training dataset, without artificially generating translations and rotations of the patterns as in our case. From this point of view, we

consider that the complexity of the dataset is increased by adding different perspectives of the patterns.

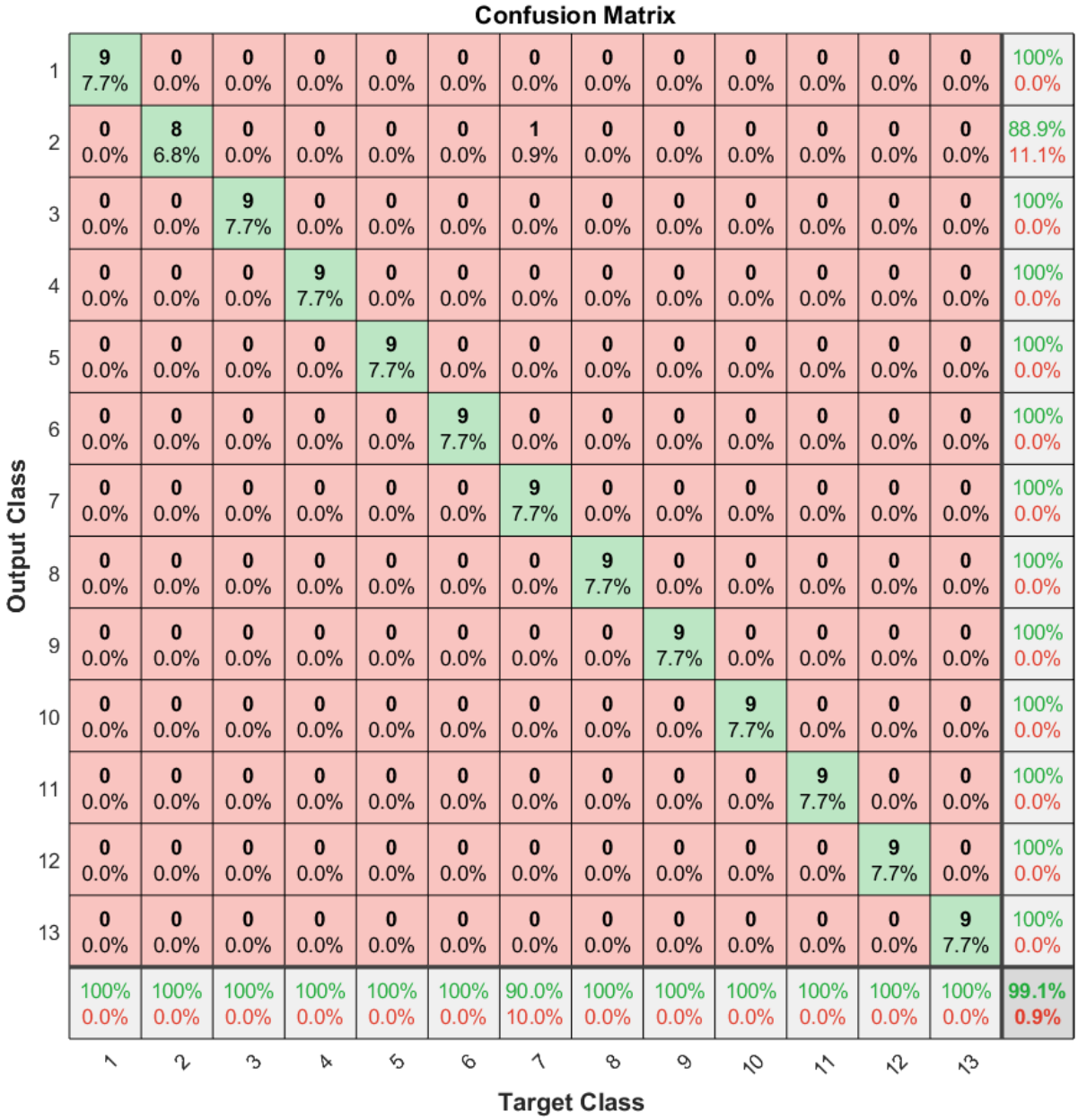


Figure 5-17. Confusion matrix of our approach that uses RI-WPD, N-zonal features and GNN as classifier.

5.6 Comments on DL approaches

DL approaches stand out for their hierarchical feature extraction capabilities. However, the application of DL in THz imaging faces early-stage challenges primarily due to the scarcity of datasets. The time-consuming nature of image acquisition in practical THz Time-Domain Spectroscopy (TDS) systems, taking minutes to hours, poses a significant obstacle in creating substantial THz image datasets.

However, given the scarcity of THz data and challenges in dataset creation, we evaluate a Deep Embedded K-Means (DEKM) clustering [129] approach to classify THz dataset 1. Briefly, DEKM involves three steps, starting with generating a lower dimensional space with an autoencoder network [130]. Next, it evaluates the embedded space and detects clusters using the K-Means algorithm [131]. Finally, it optimizes the embedded space to increase inter-class distance and lower the intra-class variation.

The results of DEKM on THz dataset 1 are presented in Figure 5-18. DEKM clusters, where we depict the clusters in a three-dimensional embedding space. We emphasize the 9 classes of THz dataset 1 with different colors and we observe that the DL approach cannot separate and identify all classes, as opposed to our approaches. Classes 1 and 3 are completely overlapped, while the others are intertwined and hardly separable.

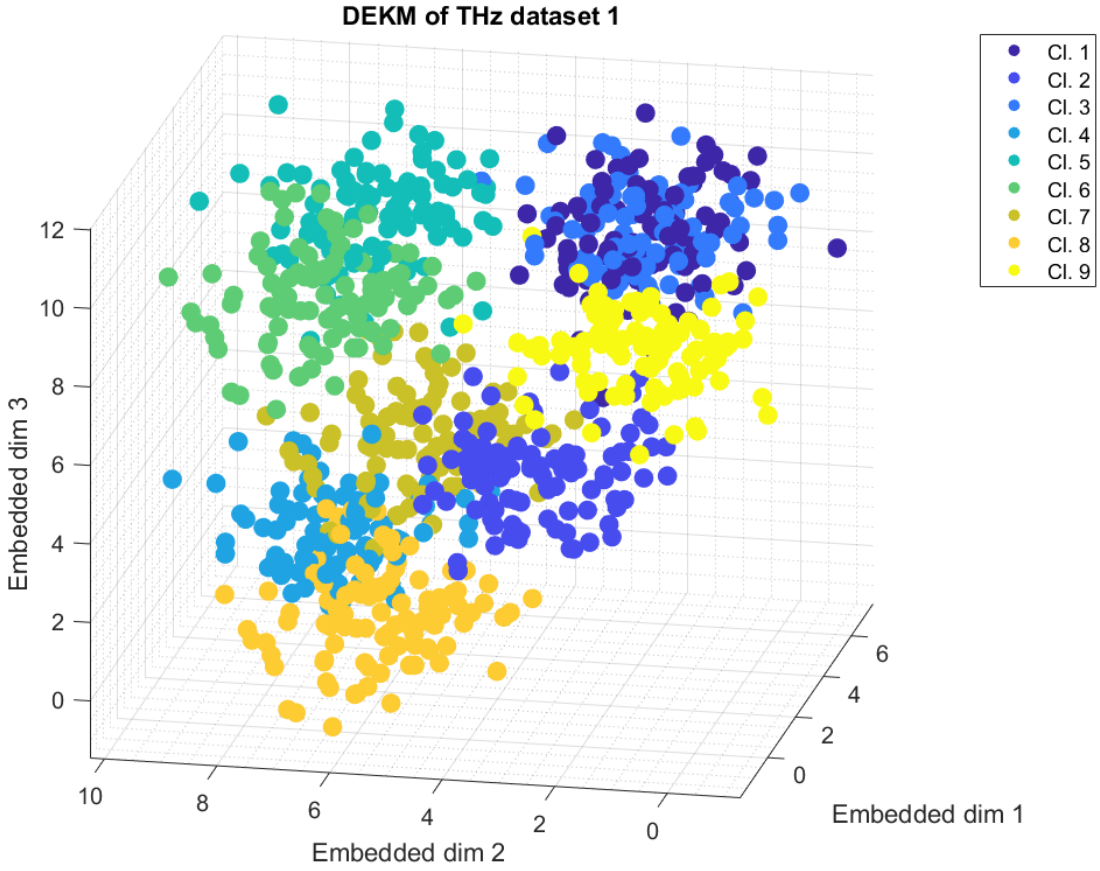


Figure 5-18. DEKM clusters

5.7 Chapter summary

In addressing the challenge of image classification, particularly in the context of limited THz data, the chapter opts for machine learning (ML) classifiers over a deep learning (DL) approach. The chapter introduces innovative feature assembling techniques, focusing on patterns and structures derived from graph representations of wavelet decomposition. These techniques leverage physical considerations such as wavelet frequency sub-bands, resulting in features adaptable for training graph neural networks (GNNs) for image classification. The second proposed idea involves aggregating features based on their depth and type which are the identifiers

of wavelet sub-images. These aggregated features are then fed into classical ML classifiers like k-NN, SVM, and decision trees.

The chapter concludes with a presentation of comparative results of classical features from the previous chapter with features generated by the proposed frameworks. Multiple ML classifiers were tested in order to have more reliable results about the improvement introduced by the proposed features. Extensive testing on proprietary THz datasets ensures a thorough evaluation of the approaches.

The discussion delves into the details of assembling techniques and their alignment with a "best" basis from the wavelet packet decomposition (WPD) family, specifically considering WPD, TI-WPD, and RI-WPD due to their multiresolution concept and invariant nature. Finally, most of the proposed approaches surpassed the classic feature extraction methods in image classification task with a large margin.

In all tests, the four-channel aggregation framework increased the robustness of the classifier regardless of its type (k-NN, SVM, NN, decision tree). Moreover, one of the most powerful classification approaches is to exploit the quasi-invariancy of the TI-WPD and RI-WPD “best”-bases in conjunction with GNN. The proposal was also successfully tested on the public benchmark dataset, “Brodatz”.

Chapter 6. Conclusions and perspectives

6.1 Conclusions

The research conducted during this thesis analyses and offers solutions to computer vision tasks in THz range. Generally, these tasks involve image reconstruction, analysis and classification. As THz is a relatively new domain, it lacks comprehensive analysis of the reconstruction methods. Moreover, due to the scarcity of THz images, their analysis and classification is not thoroughly developed and tested. However, considering the advantages of THz range in the domain of non-intrusive inspection, it is desirable to develop the field to a maturation stage to be efficiently exploited in a wide variety of medical, security and defense applications, etc.

Based on these facts, we start our research by tackling the image reconstruction task. First of all, we evaluate the state-of-the art methods used in nowadays THz-TDS imagery systems. We introduce a multiplicative perturbation that affects the phase information and which is factually more complex to remove than the common additive noise. The perturbation is multiplicative in its nature considering the exponential modeling of the THz signals. Thus, any component that changes or deforms the phase of the THz signal can be expressed as a multiplication between two exponential signals, the original THz signal and the noise component. To be more precise, we exemplify the following possible perturbations: the temperature effect on laser; delay line motion faults such as rolling and pitching which are in the translation stage; and finally, the noise resulted from the THz detection chain. These types of perturbations have low-frequency range compared to THz. In this scenario, when the multiplicative noise is present, THz images exhibit contour information loss, contrast degradation and additional artefacts unrelated to the original information.

We demonstrate that most of the common THz reconstruction methods fail to retrieve the initial information of the image when the multiplicative noise is considered. Starting from this challenge, we developed five novel methods that can be used to reconstruct THz images. Two of these methods exploit the MF concept in the phase diagram domain. They provide a good starting point as the MF quantifies the similarity of a THz reference template with other received THz templates. However, the complex deformations introduced by the multiplicative type of noise render MF unusable in this scenario. In order to increase the modeling power of the reconstruction method, we choose the implement DTW in phase diagram domain by introducing two other reconstruction methods. With DTW, we introduce an additional degree of freedom that helps better fit a THz reference template. This additional degree of freedom comes from the dilation and contraction of the template until a cost function is minimized. Despite the results slightly improving, the effects of the multiplicative type of noise are still present in the final image. The previous results conducted the research in the direction of intrinsically characterizing the phase diagram of received THz signals, while ignoring the characteristics of the reference THz template. The final method exploits the curvatures' evolution and their quantification. The higher the deformation of the phase space, the higher the values of the curvatures. By exploiting this concept, we are able to reconstruct THz images in both ideal and noisy context with high fidelity with respect to a reference image. We evaluate all the reconstruction methods with image quality metrics and conclude that in most of the cases, the curvature-based method successfully surpasses all other classic methods in terms of MSE, pSNR and SSIM. Thus, the improved amplitude estimation at reception offers a higher quality image with almost no effects coming from the multiplicative noise.

The next challenge tackled by our research regards the THz image analysis and classification methods. Considering that we optimally reconstructed a THz image such that the information is clearly emphasized, we want to minimize the effect of instabilities such as view-point variations. Generally, these instabilities conduct to different sets of features and thus, to an inefficient classification.

Based on the presented challenge, we present the state-of-the-art image decomposition that we demonstrate to be suboptimal when image translations and rotations are implied. The idea is to have a unique representation of an image regardless of translations and rotations. While the classic image decomposition methods perform well in straightforward situations without viewpoint variations, real-world cases often involve diverse images of patterns or subjects. Thus, we propose to use invariant versions of WPD such as TI-WPD and RI-WPD in order to eliminate the effect of the considered instabilities. Thus, by having an invariant decomposition, we introduce several feature extraction frameworks that can be further refined and organized to be compatible with various ML classifiers. Moreover, these frameworks are adaptable to multiple statistical measures, allowing the computation of complementary features and providing diverse perspectives on the analyzed image.

Innovative feature assembly techniques are introduced throughout the thesis, emphasizing patterns and structures derived from graph representations of wavelet decomposition. These techniques take into account physical considerations, such as wavelet frequency sub-bands, resulting in features that are suitable for training GNNs in the context of image classification. The second proposed idea involves aggregating features based on their depth and type, which serve as identifiers for wavelet sub-images.

We evaluate a large pool of image decomposition methods starting from DCT, DWT and WPD, and ending with TI-WPD and RI-WPD. For each decomposition method we consider multiple classic features such as the energy or the entropy of decomposition coefficients, but also the novel features. Finally, all these approaches are cross-evaluated on multiple ML classifiers in order to determine the efficiency of all classification frameworks. The final results shows that the novel features increase the classification accuracy of most ML models. The proposed methodology was successfully tested on two proprietary THz datasets and one public benchmark dataset, "Brodatz".

6.2 Perspectives

In terms of perspectives, we can structure a roadmap for further research in three main directions: the image reconstruction challenge, the analysis and classification tasks, and computer vision (edge detection).

6.2.1 Image Reconstruction Challenge

Given the unique characteristics and data-driven nature of the phase diagram, we can uncover novel insights about signal characteristics that are unobservable in other representation spaces. Despite its low complexity, the phase diagram enhances amplitude estimation in imaging systems. Leveraging these advantages, we can explore various characterizations and approaches to reconstruct images with increased contrast. For example, incorporating the trajectory’s torsion as complementary information to the curvature can potentially improve image quality. Using these two curve descriptors together, we can integrate them to generate THz images from raw data in both ideal and multiplicative contexts.

This improved amplitude estimation could also allow for an increased effective bandwidth by reducing the impact of noise at reception, resulting in higher quality image reconstructions.

Further research will involve reconstructing 2.5-dimensional images with more complex samples to fully exploit THz characteristics. These samples could include different density paper and multi-layered samples, which harbor hidden information that can be revealed through advanced imaging techniques and characteristics of the THz radiation.

6.2.2 Image Analysis and Classification Tasks

For image analysis and classification tasks, enhancing the robustness of classifiers can be achieved by integrating two invariant decompositions: Translation-Invariant Wavelet Packet Decomposition and Rotation-Invariant Wavelet Packet Decomposition. By utilizing features from both decompositions, we can strengthen the discriminative capabilities of classifiers.

Additionally, exploring another invariant version of wavelet packet decomposition that addresses shearing could further improve classification performance. The inherent scale-invariance in all wavelet packet decompositions, particularly in the approximation coefficients, should also be incorporated into the image processing pipeline to enhance analysis.

6.2.3 Computer Vision (Edge Detection)

In the context of computer vision, the phase diagram can be effectively utilized for edge detection. Edges represent small transitions between regions in an image, and the phase diagram can emphasize these transitions, as demonstrated with THz pulses. A straightforward edge detection procedure involves identifying transients in each column and row using the phase diagram, and then integrating the results into a single image. This method leverages the phase diagram's capability to highlight subtle changes, thereby improving edge detection accuracy.

By pursuing these research directions, we can advance both image reconstruction and analysis, leading to improved imaging systems, more robust classification methodologies, and enhanced edge detection techniques in computer vision.

List of publications

Journal

1. **Nastasiu D.**, Scripcaru R., Digulescu A., Ioana C., De Amorim R. Jr., Barbot N., Siragusa R., Perret E., Popescu F., A New Method of Secure Authentication Based on Electromagnetic Signatures of Chipless RFID Tags and Machine Learning Approaches. Sensors. 2020; 20(21):6385. https://doi.org/10.3390/s20216385

2. Digulescu A., Despina C., Popescu F., Stanescu D., **Nastasiu D.**, Sburlan D., UWB Sensing for UAV and Human Comparative Movement Characterization. *Sensors*. 2023; 23(4):1956. https://doi.org/10.3390/s23041956

3. Stanescu D., **Nastasiu D.**, Ioana C. et al. Characterization of digital modulations using the phase diagram analysis. *Eur. Phys. J. Spec. Top.* **232**, 187–199 (2023). https://doi.org/10.1140/epjs/s11734-022-00744-x

4. Digulescu A., Sarbu A., Stanescu D., **Nastasiu D.**, Despina C., Ioana C., Mansour A. Detection of OFDM modulations based on the characterization in the phase diagram domain. Frontiers in Signal Processing. 2023, 3. 10.3389/frsip.2023.1197590.

5. **Nastasiu D.**, Stănescu D., Despina C., Digulescu A., Ioana C., Șerbănescu A., Surface Roughness Classification Using a Translation Invariant Wavelet Packet Decomposition, Journal of Military Technology, Vol 5, nr 2, 2022.

6. Celmare A., Digulescu A., Leca C., Despina-Stoian C., Stanescu D., **Nastasiu D.**, Spectral Analysis in Mobile Communications, Journal of Military Technology, Vol 4, nr 2, 2021.

Conferences

1. **Nastasiu D**., Bernier M., Ioana C., Tréhoult C., Lyannaz L., Garet F., Phase diagram method for efficient THz images reconstructing, 2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz), Chengdu, China, 2021, pp. 1-2, doi: 10.1109/IRMMW-THz50926.2021.9566864.

2. Stanescu D., **Nastasiu D.**, Ioana C., Digulescu A., Serbanescu A, Low Complexity Acoustic Imaging System Based on Time of Arrivals Dynamic Estimation, 2021 10th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2021, pp. 1-4, doi: 10.1109/MECO52532.2021.9460241.

3. **Nastasiu D.**, Berniner M., Digulescu A., Garet F., Ioana C., Image pattern classification under invariant constraints. *GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images*, Sep 2022, Nancy, France.

4. **Nastasiu D.**, Digulescu A., Ioana C., Bernier M., Garet F., Serbanescu A., "A novel machine learning approach in Image Pattern Recognition under invariance constraints," *2022 30th European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia, 2022, pp. 558-562, doi: 10.23919/EUSIPCO55093.2022.9909907.

Résumé - L'imagerie THz est un domaine émergent depuis les avancées technologiques en termes d'émission de rayonnement THz et d'équipement de détection. L'objectif principal de la thèse est de contribuer et d'améliorer les systèmes d'imagerie THz, de la reconstruction et de l'analyse d'images aux tâches de classification d'images. Dans la première partie de la thèse, nous nous attaquons au défi de l'estimation de l'amplitude dans des conditions de bruit idéal et multiplicatif. Le bruit multiplicatif déforme la phase et introduit des artefacts complexes, tels que la perte d'information sur les contours et la dégradation du contraste, qui ne peuvent être éliminés à l'aide des techniques de reconstruction d'image les plus récentes. À cet égard, nous présentons cinq nouvelles méthodes de reconstruction qui exploitent la représentation du diagramme de phase des signaux. Deux de ces méthodes sont basées sur le filtrage du diagramme de phase pour estimer l'amplitude dans les deux conditions. Deux autres méthodes utilisent le concept de déformation temporelle dynamique (DTW) pour augmenter la capacité à modéliser le type de bruit multiplicatif. Enfin, nous exploitons la dynamique de la trajectoire de phase décrite par les courbures pour reconstruire l'image. Parmi le grand nombre de méthodes, nous évaluons tout au long de la thèse que la méthode basée sur les courbures reconstruit efficacement l'image dans des contextes idéaux et bruités. Après une reconstruction efficace de l'image, la deuxième partie de la thèse, nous étudions les méthodes d'analyse et de classification d'images en tenant compte des instabilités des systèmes d'imagerie du monde réel, telles que les translations et les rotations. Dans ce sens, nous proposons d'utiliser des décompositions de paquets d'ondelettes invariantes par rapport à la translation et à la rotation, qui fournissent une représentation unique et optimale d'une image, indépendamment de la translation ou de la rotation de l'image. Sur la base des représentations d'images invariantes, de nouvelles techniques d'extraction de caractéristiques sont introduites, telles que les cadres verticaux, horizontaux, N-directionnels et N-zonaux. En outre, deux structures de caractéristiques sont introduites, qui prennent en compte le partitionnement en fréquence de la décomposition en ondelettes et sont adaptées pour fonctionner avec des réseaux neuronaux graphiques (GNN) et des classificateurs ML classiques tels que les k-voisins les plus proches (k-NN), les machines à vecteurs de support (SVM), etc. Dans l'ensemble, les approches que nous proposons augmentent la précision de tous les classificateurs.

Mots clés: méthodes de reconstruction d'images, caractérisation du diagramme de phase, décomposition en paquets d'ondelettes invariants, apprentissage automatique, réseaux neuronaux graphiques, méthodes d'extraction de caractéristiques, classification d'images.

Abstract – THz imaging is an emerging field since the technological advances in terms of THz radiation emission and detection equipment. The main objective of the thesis is to contribute and to improve THz imaging systems, from image reconstruction and analysis to image classification tasks. In the first part of the thesis, we tackle the amplitude estimation challenge under ideal and multiplicative noise conditions. The multiplicative noise deforms the phase and introduces complex artefacts, such as contour information loss and contrast degradation, that cannot be eliminated using state-of-the-art image reconstruction techniques. In this regard, we introduce five novel reconstruction methods which exploit the phase diagram representation of signals. Two of the methods are based on phase-diagram match filtering to estimate the amplitude in both conditions. Another two methods use the concept of dynamic time warping (DTW) to increase the capability to model the multiplicative type of noise. Lastly, we exploit the dynamic of the phase trajectory described by the curvatures to reconstruct the image. From the large pool of methods, we evaluate throughout the thesis that the curvature-based method efficiently reconstructs the image in both ideal and noisy contexts. After an efficient image reconstruction, the second part of the thesis, we study image analysis and classification methods considering the instabilities of real-world imaging systems, such as translations and rotations. In this sense, we propose to use translation and rotation invariant wavelet packet decompositions, that provide a unique and optimal representation of an image, regardless if the image is translated or rotated. Based on the invariant image representations, novel feature extraction techniques are introduced such as vertical, horizontal, N-directional and N-zonal frameworks. Additionally, two feature structures are introduced and that consider the frequency partitioning of the wavelet decomposition and are adapted to work with Graph Neural Networks (GNNs) and classic ML classifiers such as k-nearest neighbors (k-NN), support vector machine (SVM), etc. Overall, our proposed approaches increase the accuracy of all classifiers.

Keywords: image reconstruction methods, phase diagram characterization, invariant wavelet packet decomposition, Machine Learning, Graph Neural Networks, feature extraction methods, image classification.

Bibliography

[1] M. Sagrillo, R. R. Guerra, F. M. Bayer, and R. MacHado, 'The LADistribution: An Approximation of the G0ADistribution for Amplitude SAR Image Modeling', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, 2023, doi: 10.1109/TGRS.2023.3262480.

[2] G. Valušis, A. Lisauskas, H. Yuan, W. Knap, and H. G. Roskos, 'Roadmap of Terahertz Imaging 2021', *Sensors 2021, Vol. 21, Page 4092*, vol. 21, no. 12, p. 4092, Jun. 2021, doi: 10.3390/S21124092.

[3] Y. S. Lee, 'Principles of terahertz science and technology', *Principles of Terahertz Science and Technology*, pp. 1–340, 2009, doi: 10.1007/978-0-387-09540-0/COVER.

[4] A. Dobroiu, C. Otani, and K. Kawase, 'Terahertz-wave sources and imaging applications', *Meas Sci Technol*, vol. 17, no. 11, p. R161, Sep. 2006, doi: 10.1088/0957-0233/17/11/R01.

[5] C. Wang *et al.*, 'Application of terahertz time-domain spectroscopy in intracellular metabolite detection', *J Biophotonics*, vol. 3, no. 10–11, pp. 641–645, Oct. 2010, doi: 10.1002/JBIO.201000043.

[6] V. Petrov, J. Kokkoniemi, D. Moltchanov, J. Lehtomäki, and Y. Koucheryavy, 'Enabling simultaneous cooling and data transmission in the terahertz band for board-to-board communications', *Physical Communication*, vol. 22, pp. 9–18, Mar. 2017, doi: 10.1016/J.PHYCOM.2016.11.002.

[7] I. Duling and D. Zimdars, 'Terahertz imaging: Revealing hidden defects', *Nat Photonics*, vol. 3, no. 11, pp. 630–632, Nov. 2009, doi: 10.1038/NPHOTON.2009.206.

[8] X. Yang *et al.*, 'Rapid and label-free detection and assessment of bacteria by terahertz time-domain spectroscopy', *J Biophotonics*, vol. 9, no. 10, pp. 1050–1058, Oct. 2016, doi: 10.1002/JBIO.201500270.

[9] S. Kar, 'Terahertz Technology—Emerging Trends and Application Viewpoints', *Terahertz Biomedical and Healthcare Technologies: Materials to Devices*, pp. 89–111, Jan. 2020, doi: 10.1016/B978-0-12-818556-8.00005-7.

[10] B. Recur *et al.*, 'Terahertz radiation for tomographic inspection', *https://doi.org/10.1117/1.OE.51.9.091609*, vol. 51, no. 9, p. 091609, May 2012, doi: 10.1117/1.OE.51.9.091609.

[11] Y. C. Shen, T. Lo, P. F. Taday, B. E. Cole, W. R. Tribe, and M. C. Kemp, 'Detection and identification of explosives using terahertz pulsed spectroscopic imaging', *Appl Phys Lett*, vol. 86, no. 24, pp. 1–3, Jun. 2005, doi: 10.1063/1.1946192.

[12] R. M. Woodward *et al.*, 'Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue', *Phys Med Biol*, vol. 47, no. 21, pp. 3853–3863, Nov. 2002, doi: 10.1088/0031-9155/47/21/325.

[13] G. Reese *et al.*, 'Using terahertz pulsed imaging (TPI) to identify colonic pathology', *33rd International Conference on Infrared and Millimeter Waves and the 16th International Conference on Terahertz Electronics, 2008, IRMMW-THz 2008*, 2008, doi: 10.1109/ICIMW.2008.4665690.

[14] F. Wahaia *et al.*, 'Detection of colon cancer by terahertz techniques', *J Mol Struct*, vol. 1006, no. 1–3, pp. 77–82, Dec. 2011, doi: 10.1016/J.MOLSTRUC.2011.05.049.

[15] A. Al-Ibadi *et al.*, 'Terahertz biomedical imaging: From multivariate analysis and detection to material parameter extraction', *Progress in Electromagnetics Research Symposium*, pp. 2756–2762, May 2017, doi: 10.1109/PIERS.2017.8262222.

[16] E. Leiss-Holzinger *et al.*, ‘Imaging of the inner structure of cave bear teeth by novel non-destructive techniques’, *Palaeontologia Electronica*, vol. 18, no. 1, 2015, doi: 10.26879/489.
[17] T. Yasui, T. Yasuda, K. I. Sawanaka, and T. Araki, ‘Terahertz paintmeter for noncontact monitoring of thickness and drying progress in paint film’, *Appl Opt*, vol. 44, no. 32, pp. 6849–6856, Nov. 2005, doi: 10.1364/AO.44.006849.
[18] T. Löffler *et al.*, ‘All-optoelectronic continuous-wave terahertz systems’, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 362, no. 1815, pp. 263–281, Feb. 2004, doi: 10.1098/RSTA.2003.1326.
[19] S. Wietzke, C. Jansen, F. Rutz, D. M. Mittleman, and M. Koch, ‘Determination of additive content in polymeric compounds with terahertz time-domain spectroscopy’, *Polym Test*, vol. 26, no. 5, pp. 614–618, Aug. 2007, doi: 10.1016/J.POLYMERTESTING.2007.03.002.
[20] H. Zhong *et al.*, ‘Inspection of space shuttle insulation foam defects using a 0.2 THz Gunn diode oscillator’, *Conference Digest of the 2004 Joint 29th International Conference on Infrared and Millimeter Waves and 12th International Conference on Terahertz Electronics*, pp. 753–754, 2004, doi: 10.1109/ICIMW.2004.1422311.
[21] R. F. Anastasi and E. I. Madaras, ‘Terahertz NDE for Metallic Surface Roughness Evaluation’.
[22] C. L. K. Dandolo and P. U. Jepsen, ‘Wall Painting Investigation by Means of Non-invasive Terahertz Time-Domain Imaging (THz-TDI): Inspection of Subsurface Structures Buried in Historical Plasters’, *J Infrared Millim Terahertz Waves*, vol. 37, no. 2, pp. 198–208, Feb. 2016, doi: 10.1007/S10762-015-0218-9.
[23] H. Cummins, ‘Challenges to terahertz counter-terrorism and security-related applications’, *NATO Science for Peace and Security Series B: Physics and Biophysics*, pp. 205–224, 2007, doi: 10.1007/978-1-4020-6503-3_14/COVER.
[24] Y. Watanabe, Y. Ogawa, K. Kawase, and H. Inoue, ‘Non-destructive terahertz imaging of illicit drugs using spectral fingerprints’, *Optics Express, Vol. 11, Issue 20, pp. 2549-2554*, vol. 11, no. 20, pp. 2549–2554, Oct. 2003, doi: 10.1364/OE.11.002549.
[25] A. Hirata *et al.*, ‘120-GHz-band millimeter-wave photonic wireless link for 10-Gb/s data transmission’, *IEEE Trans Microw Theory Tech*, vol. 54, no. 5, pp. 1937–1942, May 2006, doi: 10.1109/TMTT.2006.872798.
[26] T. Nagatsuma, G. Ducournau, and C. C. Renaud, ‘Advances in terahertz communications accelerated by photonics’, *Nature Photonics 2016 10:6*, vol. 10, no. 6, pp. 371–379, May 2016, doi: 10.1038/nphoton.2016.65.
[27] N. Karpowicz, H. Zhong, J. Xu, K. I. Lin, J. S. Hwang, and X. C. Zhang, ‘Comparison between pulsed terahertz time-domain imaging and continuous wave terahertz imaging’, *Semicond Sci Technol*, vol. 20, no. 7, Jul. 2005, doi: 10.1088/0268-1242/20/7/021.
[28] N. T. Yardimci, S. Cakmakyapan, S. Hemmati, and M. Jarrahi, ‘A High-Power Broadband Terahertz Source Enabled by Three-Dimensional Light Confinement in a Plasmonic Nanocavity’, *Scientific Reports 2017 7:1*, vol. 7, no. 1, pp. 1–8, Jun. 2017, doi: 10.1038/s41598-017-04553-4.
[29] E. Roussel *et al.*, ‘Phase Diversity Electro-optic Sampling: A new approach to single-shot terahertz waveform recording’, *Light: Science & Applications 2022 11:1*, vol. 11, no. 1, pp. 1–14, Jan. 2022, doi: 10.1038/s41377-021-00696-2.
[30] N. M. Burford and M. O. El-Shenawee, ‘Review of terahertz photoconductive antenna technology’, *Optical Engineering*, vol. 56, no. 1, p. 010901, Jan. 2017, doi: 10.1117/1.OE.56.1.010901.
[31] J. T. Good, D. B. Holland, I. A. Finneran, P. B. Carroll, M. J. Kelley, and G. A. Blake, ‘A decade-spanning high-resolution asynchronous optical sampling terahertz time-domain and

frequency comb spectrometer', *Review of Scientific Instruments*, vol. 86, no. 10, p. 103107, Oct. 2015, doi: 10.1063/1.4932567/15868598/103107_1_ACCEPTED_MANUSCRIPT.PDF.
[32] M. Schall, H. Helm, and S. R. Keiding, 'Far infrared properties of electro-optic crystals measured by THz time-domain spectroscopy', *Int J Infrared Millimeter Waves*, vol. 20, no. 4, pp. 595–604, 1999, doi: 10.1023/A:1022636421426/METRICS.
[33] J. R. Demers, J.-L. Coutaz, and F. Garet, 'A UAV-mounted THz spectrometer for real-time gas analysis', *https://doi.org/10.1117/12.2292783*, vol. 10531, pp. 83–89, Feb. 2018, doi: 10.1117/12.2292783.
[34] D. H. Auston, K. P. Cheung, and P. R. Smith, 'Picosecond photoconducting Hertzian dipoles', *Appl Phys Lett*, vol. 45, no. 3, pp. 284–286, Aug. 1984, doi: 10.1063/1.95174.
[35] D. H. Auston, X.-C. Zhang, J. T. Darrow, and B. B. Hu, 'Subpicosecond electromagnetic pulses from large-aperture photoconducting antennas', *Optics Letters, Vol. 15, Issue 6, pp. 323-325*, vol. 15, no. 6, pp. 323–325, Mar. 1990, doi: 10.1364/OL.15.000323.
[36] M. van Exter and D. R. Grischkowsky, 'Characterization of an Optoelectronic Terahertz Beam System', *IEEE Trans Microw Theory Tech*, vol. 38, no. 11, pp. 1684–1691, 1990, doi: 10.1109/22.60016.
[37] T. Bowman, M. El-Shenawee, and L. K. Campbell, 'Terahertz transmission vs reflection imaging and model-based characterization for excised breast carcinomas', *Biomedical Optics Express, Vol. 7, Issue 9, pp. 3756-3783*, vol. 7, no. 9, pp. 3756–3783, Sep. 2016, doi: 10.1364/BOE.7.003756.
[38] S. R. Tripathi, Y. Sugiyama, K. Murate, K. Imayama, and K. Kawase, 'Terahertz wave three-dimensional computed tomography based on injection-seeded terahertz wave parametric emitter and detector', *Opt Express*, vol. 24, no. 6, p. 6433, Mar. 2016, doi: 10.1364/OE.24.006433.
[39] T. Mohr *et al.*, '2D tomographic terahertz imaging using a single pixel detector', *Optics Express, Vol. 26, Issue 3, pp. 3353-3367*, vol. 26, no. 3, pp. 3353–3367, Feb. 2018, doi: 10.1364/OE.26.003353.
[40] E. Abraham *et al.*, 'Fast three-dimensional terahertz computed tomography using real-time line projection of intense terahertz pulse', *Optics Express, Vol. 21, Issue 2, pp. 2423-2433*, vol. 21, no. 2, pp. 2423–2433, Jan. 2013, doi: 10.1364/OE.21.002423.
[41] S. Zhang and Y. Xia, 'CT image reconstruction algorithms: A comprehensive survey', *Concurr Comput*, vol. 33, no. 8, p. e5506, Apr. 2021, doi: 10.1002/CPE.5506.
[42] J. M. Wandeto, H. Nyongesa, Y. Rémond, and B. Dresp-Langley, 'Detection of small changes in medical and random-dot images comparing self-organizing map performance to human detection', *Inform Med Unlocked*, vol. 7, pp. 39–45, Jan. 2017, doi: 10.1016/J.IMU.2017.03.001.
[43] K. Ong *et al.*, 'Detection of subtle white matter lesions in MRI through texture feature extraction and boundary delineation using an embedded clustering strategy', *Sci Rep*, vol. 12, no. 1, p. 4433, Dec. 2022, doi: 10.1038/S41598-022-07843-8.
[44] Y. Zhang, C. Ma, L. Zhuo, and J. Li, 'Arbitrary-Oriented Object Detection in Aerial Images with Dynamic Deformable Convolution and Self-Normalizing Channel Attention', *Electronics 2023, Vol. 12, Page 2132*, vol. 12, no. 9, p. 2132, May 2023, doi: 10.3390/ELECTRONICS12092132.
[45] S. H. Kim and D. Cho, 'Viewpoint-Aware Action Recognition Using Skeleton-Based Features from Still Images', *Electronics 2021, Vol. 10, Page 1118*, vol. 10, no. 9, p. 1118, May 2021, doi: 10.3390/ELECTRONICS10091118.

[46] M. Zaffar, A. Khaliq, S. Ehsan, M. Milford, K. Alexis, and K. McDonald-Maier, 'Are State-of-the-art Visual Place Recognition Techniques any Good for Aerial Robotics?', Apr. 2019, Accessed: Jun. 29, 2023. [Online]. Available: http://arxiv.org/abs/1904.07967
[47] C. B. Madsen, 'Viewpoint variation in the noise sensitivity of pose estimation', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 41–46, 1996, doi: 10.1109/CVPR.1996.517051.
[48] W. Jiang *et al.*, 'FC-ACGAN-based data augmentation for terahertz time-domain spectral concealed hazardous materials identification', *International Journal of Intelligent Systems*, vol. 37, no. 11, pp. 9562–9579, Nov. 2022, doi: 10.1002/INT.23013.
[49] H. Park and J. H. Son, 'Machine Learning Techniques for THz Imaging and Time-Domain Spectroscopy', *Sensors 2021, Vol. 21, Page 1186*, vol. 21, no. 4, p. 1186, Feb. 2021, doi: 10.3390/S21041186.
[50] M. Di Fabrizio, A. D'Arco, S. Mou, L. Palumbo, M. Petrarca, and S. Lupi, 'Performance Evaluation of a THz Pulsed Imaging System: Point Spread Function, Broadband THz Beam Visualization and Image Reconstruction', *Applied Sciences 2021, Vol. 11, Page 562*, vol. 11, no. 2, p. 562, Jan. 2021, doi: 10.3390/APP11020562.
[51] Z. Zhang, Y. Lu, C. Lv, Q. Mao, S. Wang, and S. Yan, 'Restoration of integrated circuit terahertz image based on wavelet denoising technique and the point spread function model', *Opt Lasers Eng*, vol. 138, p. 106413, Mar. 2021, doi: 10.1016/J.OPTLASENG.2020.106413.
[52] S. M. Zhu, B. W. H. Ng, B. M. Fischer, and D. Abbott, 'THz signal denoising via redundant representation', *34th International Conference on Infrared, Millimeter, and Terahertz Waves, IRMMW-THz 2009*, 2009, doi: 10.1109/ICIMW.2009.5325609.
[53] L. Duvillaret, J.-L. Coutaz, and F. Garet, 'Influence of noise on the characterization of materials by terahertz time-domain spectroscopy', *JOSA B, Vol. 17, Issue 3, pp. 452-461*, vol. 17, no. 3, pp. 452–461, Mar. 2000, doi: 10.1364/JOSAB.17.000452.
[54] T. Song and L. S. Uhm, 'Multiplicative noise model and composite signal detection', *IEE Proceedings, Part F: Radar and Signal Processing*, vol. 138, no. 6, pp. 531–538, 1991, doi: 10.1049/IP-F-2.1991.0071.
[55] M. G. Shama, 'Multiplicative noise removal using complementary total generalized variation and nonlocal low rank regularization', *Optik (Stuttg)*, vol. 291, p. 171374, Nov. 2023, doi: 10.1016/J.IJLEO.2023.171374.
[56] M. A. Ammar, H. A. Hassan, M. S. Abdel-Latif, and S. A. Elgamel, 'Performance evaluation of SAR in presence of multiplicative noise jamming', *National Radio Science Conference, NRSC, Proceedings*, pp. 213–220, Apr. 2017, doi: 10.1109/NRSC.2017.7893506.
[57] S. Batool, F. Frezza, F. Mangini, and P. Simeoni, 'Introduction to Radar Scattering Application in Remote Sensing and Diagnostics: Review', *Atmosphere 2020, Vol. 11, Page 517*, vol. 11, no. 5, p. 517, May 2020, doi: 10.3390/ATMOS11050517.
[58] R. Heremans, M. Vandewal, and M. Acheroy, 'Space-time versus frequency domain signal processing for 3D THz imaging', *Proceedings of IEEE Sensors*, pp. 739–744, 2009, doi: 10.1109/ICSENS.2009.5398545.
[59] L. Zanotto *et al.*, 'Time-domain terahertz compressive imaging', *Optics Express, Vol. 28, Issue 3, pp. 3795-3802*, vol. 28, no. 3, pp. 3795–3802, Feb. 2020, doi: 10.1364/OE.384134.
[60] Q. Wang, L. Xie, and Y. Ying, 'Overview of imaging methods based on terahertz time-domain spectroscopy', *Appl Spectrosc Rev*, vol. 57, no. 3, pp. 249–264, 2022, doi: 10.1080/05704928.2021.1875480.

[61] L. Hui, W. Jingzhu, L. Cuiling, S. Xiaorong, and Y. le, ‘Study on Pretreatment Methods of Terahertz Time Domain Spectral Image for Maize Seeds’, *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 206–210, Jan. 2018, doi: 10.1016/J.IFACOL.2018.08.142.
[62] H. ; Yoo *et al.*, ‘High-Speed THz Time-of-Flight Imaging with Reflective Optics’, *Sensors 2023, Vol. 23, Page 873*, vol. 23, no. 2, p. 873, Jan. 2023, doi: 10.3390/S23020873.
[63] C. Zhang, J. Li, B. Li, and W. Ma, ‘Blind Matching Filtering Algorithm for Spectrum Sensing under Multi-Path Channel Environment’, *Electronics 2023, Vol. 12, Page 2499*, vol. 12, no. 11, p. 2499, Jun. 2023, doi: 10.3390/ELECTRONICS12112499.
[64] X. Ren and Y. Jiang, ‘Spatial Domain Terahertz Image Reconstruction Based on Dual Sparsity Constraints’, *Sensors 2021, Vol. 21, Page 4116*, vol. 21, no. 12, p. 4116, Jun. 2021, doi: 10.3390/S21124116.
[65] A. R. Barket, A. R. Barket, W. Hu, B. Wang, W. Shahzad, and J. S. Malik, ‘Selection criteria of image reconstruction algorithms for terahertz short-range imaging applications’, *Optics Express, Vol. 30, Issue 13, pp. 23398-23416*, vol. 30, no. 13, pp. 23398–23416, Jun. 2022, doi: 10.1364/OE.457840.
[66] I. Catapano and F. Soldovieri, ‘THz imaging and data processing: State of the art and perspective’, *Innovation in Near-Surface Geophysics: Instrumentation, Application, and Data Processing Methods*, pp. 399–417, Jan. 2019, doi: 10.1016/B978-0-12-812429-1.00011-8.
[67] J. Guo, X. Wang, Y. Cheng, W. A. Mustafa, M. Mydin, and M. A. Kader, ‘A Review of Histogram Equalization Techniques in Image Enhancement Application’, *J Phys Conf Ser*, vol. 1019, no. 1, p. 012026, Jun. 2018, doi: 10.1088/1742-6596/1019/1/012026.
[68] K. Nagamani, K. Divya, K. Sujatha, K. R. Bonagiri, G. B. Kande, and P. S. S. Kumar, ‘Adaptive histogram equalization of wavelet sub bands for the enhancement of contrast in aerial images’, *Mater Today Proc*, vol. 52, pp. 898–901, Jan. 2022, doi: 10.1016/J.MATPR.2021.10.297.
[69] P. Maragos, ‘Morphological Filtering for Image Enhancement and Feature Detection’, *Handbook of Image and Video Processing*, pp. 135–156, 2005, doi: 10.1016/B978-012119792-6/50072-3.
[70] K. A. M. Said and A. B. Jambek, ‘Analysis of Image Processing Using Morphological Erosion and Dilation’, *J Phys Conf Ser*, vol. 2071, no. 1, p. 012033, Oct. 2021, doi: 10.1088/1742-6596/2071/1/012033.
[71] J. Chen, J. Benesty, Y. Huang, and S. Doclo, ‘New insights into the noise reduction Wiener filter’, *IEEE Trans Audio Speech Lang Process*, vol. 14, no. 4, pp. 1218–1233, 2006, doi: 10.1109/TSA.2005.860851.
[72] A. S. Abdul-Nasir, M. Y. Mashor, and Z. Mohamed, ‘Modified global and modified linear contrast stretching algorithms: New colour contrast enhancement techniques for microscopic analysis of malaria slide images’, *Comput Math Methods Med*, vol. 2012, 2012, doi: 10.1155/2012/637360.
[73] Y. Zhu and C. Huang, ‘An Improved Median Filtering Algorithm for Image Noise Reduction’, *Phys Procedia*, vol. 25, pp. 609–616, 2012, doi: 10.1016/J.PHPRO.2012.03.133.
[74] Z. Al-Ameen, A. Muttar, and G. Al-Badrani, ‘Improving the Sharpness of Digital Image Using an Amended Unsharp Mask Filter’, *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 3, pp. 1–9, Mar. 2019, doi: 10.5815/IJIGSP.2019.03.01.
[75] S. Takagi, S. Takahashi, K. Takeya, and S. R. Tripathi, ‘Influence of delay stage positioning error on signal-to-noise ratio, dynamic range, and bandwidth of terahertz time-domain spectroscopy’, *Appl Opt*, vol. 59, no. 3, p. 841, Jan. 2020, doi: 10.1364/AO.378756.

[76] A. Digulescu, C. Ioana, and A. Serbanescu, ‘Phase Diagram-Based Sensing with Adaptive Waveform Design and Recurrent States Quantification for the Instantaneous Frequency Law Tracking’, *Sensors (Basel)*, vol. 19, no. 11, Jun. 2019, doi: 10.3390/S19112434.
[77] M. Zhang, W. D. Kalies, J. A. S. Kelso, and E. Tognoli, ‘Topological portraits of multiscale coordination dynamics’, *J Neurosci Methods*, vol. 339, Jun. 2020, doi: 10.1016/J.JNEUMETH.2020.108672.
[78] R. Scripcaru, D. Nastasiu, A. Digulescu, D. Stanescu, C. Ioana, and A. Serbanescu, ‘On the potential of phase diagram analysis to identify the wide band modulations’, *2020 13th International Conference on Communications, COMM 2020 - Proceedings*, pp. 55–58, Jun. 2020, doi: 10.1109/COMM48946.2020.9141963.
[79] A. Digulescu, C. Despina-Stoian, F. Popescu, D. Stanescu, D. Nastasiu, and D. Sburlan, ‘UWB Sensing for UAV and Human Comparative Movement Characterization’, *Sensors 2023, Vol. 23, Page 1956*, vol. 23, no. 4, p. 1956, Feb. 2023, doi: 10.3390/S23041956.
[80] D. Nastasiu, M. Bernier, C. Ioana, C. Tréhoult, L. Lyannaz, and F. Garet, ‘Phase diagram method for efficient THz images reconstructing’, *International Conference on Infrared, Millimeter, and Terahertz Waves, IRMMW-THz*, vol. 2021-August, 2021, doi: 10.1109/IRMMW-THZ50926.2021.9566864.
[81] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, ‘Recurrence plots for the analysis of complex systems’, *Phys Rep*, vol. 438, no. 5–6, pp. 237–329, Jan. 2007, doi: 10.1016/J.PHYSREP.2006.11.001.
[82] G. L. Turin, ‘An Introduction to Matched Filters’, *IRE Transactions on Information Theory*, vol. 6, no. 3, pp. 311–329, 1960, doi: 10.1109/TIT.1960.1057571.
[83] M. Müller, ‘Dynamic Time Warping’, *Information Retrieval for Music and Motion*, pp. 69–84, 2007, doi: 10.1007/978-3-540-74048-3_4.
[84] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, ‘Generalizing DTW to the multi-dimensional case requires an adaptive approach’, *Data Min Knowl Discov*, vol. 31, no. 1, p. 1, Jan. 2017, doi: 10.1007/S10618-016-0455-0.
[85] M. Pilte, S. Bonnabel, and F. Barbaresco, ‘Tracking the Frenet-Serret frame associated to a highly maneuvering target in 3D’, *2017 IEEE 56th Annual Conference on Decision and Control, CDC 2017*, vol. 2018-January, pp. 1969–1974, Jan. 2018, doi: 10.1109/CDC.2017.8263937.
[86] S. L. Bangare, A. Dubal, P. S. Bangare, and S. T. Patil, ‘Reviewing otsu’s method for image thresholding’, *International Journal of Applied Engineering Research*, vol. 10, no. 9, pp. 21777–21783, 2015, doi: 10.37622/IJAER/10.9.2015.21777-21783.
[87] L. Alzubaidi *et al.*, ‘Review of deep learning: concepts, CNN architectures, challenges, applications, future directions’, *Journal of Big Data 2021 8:1*, vol. 8, no. 1, pp. 1–74, Mar. 2021, doi: 10.1186/S40537-021-00444-8.
[88] Y. Zhang, Z. Su, F. Qi, J. Zhou, and X. P. Zhang, ‘TERAHERTZ IMAGE RESTORATION BENCHMARKING DATASET’, *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2022-May, pp. 1700–1704, 2022, doi: 10.1109/ICASSP43922.2022.9747368.
[89] D. Liang, F. Xue, and L. Li, ‘Active Terahertz Imaging Dataset for Concealed Object Detection’, May 2021, Accessed: Aug. 10, 2023. [Online]. Available: https://arxiv.org/abs/2105.03677v1
[90] K. Arai, ‘Discrete Cosine Transformation based Image Data Compression Considering Image Restoration’, *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, pp. 125–131, 2020, doi: 10.14569/IJACSA.2020.0110616.
[91] L. Leng, J. Zhang, J. Xu, M. K. Khan, and K. Alghathbar, ‘Dynamic weighted discrimination power analysis in DCT domain for face and palmprint recognition’, *2010*

International Conference on Information and Communication Technology Convergence, ICTC 2010, pp. 467–471, 2010, doi: 10.1109/ICTC.2010.5674791.
[92] S. G. Mallat, 'A Theory for Multiresolution Signal Decomposition: The Wavelet Representation', *IEEE Trans Pattern Anal Mach Intell*, vol. 11, no. 7, pp. 674–693, 1989, doi: 10.1109/34.192463.
[93] S. Mallat, 'A Wavelet Tour of Signal Processing', *A Wavelet Tour of Signal Processing*, 2009, doi: 10.1016/B978-0-12-374370-1.X0001-8.
[94] G. G. Yen, 'Wavelet packet feature extraction for vibration monitoring', *IEEE Transactions on Industrial Electronics*, vol. 47, no. 3, pp. 650–667, 2000, doi: 10.1109/41.847906.
[95] W. Ting, Y. Guo-zheng, Y. Bang-hua, and S. Hong, 'EEG feature extraction based on wavelet packet decomposition for brain computer interface', *Measurement*, vol. 41, no. 6, pp. 618–625, Jul. 2008, doi: 10.1016/J.MEASUREMENT.2007.07.007.
[96] D. Wang, D. Miao, and C. Xie, 'Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection', *Expert Syst Appl*, vol. 38, no. 11, pp. 14314–14320, Oct. 2011, doi: 10.1016/J.ESWA.2011.05.096.
[97] P. Saraiva, 'On Shannon entropy and its applications', *Kuwait Journal of Science*, vol. 50, no. 3, pp. 194–199, Jul. 2023, doi: 10.1016/J.KJS.2023.05.004.
[98] K. Huang and S. Aviyente, 'Choosing best basis in wavelet packets for fingerprint matching', *Proceedings - International Conference on Image Processing, ICIP*, vol. 5, pp. 1249–1252, 2004, doi: 10.1109/ICIP.2004.1419724.
[99] R. A.M, K. W.M, E. M. A, and W. Ahmed, 'Jpeg Image Compression Using Discrete Cosine Transform - A Survey', *International Journal of Computer Science & Engineering Survey*, vol. 5, no. 2, pp. 39–47, Apr. 2014, doi: 10.5121/IJCSES.2014.5204.
[100] N. Dragoș, D. Angela, I. Cornel, B. Maxime, G. Fréderic, and Șerbănescu Alexandru, 'A novel machine learning approach in Image Pattern Recognition under invariance constraints', *European Signal Processing Conference*, vol. 2022-August, pp. 558–562, 2022, doi: 10.23919/EUSIPCO55093.2022.9909907.
[101] T. Li, M. Zhou, C. M. Travieso-González, and J. B. Alonso-Hernández, 'ECG Classification Using Wavelet Packet Entropy and Random Forests', *Entropy 2016, Vol. 18, Page 285*, vol. 18, no. 8, p. 285, Aug. 2016, doi: 10.3390/E18080285.
[102] I. Cohen, S. Raz, and D. Malah, 'Orthonormal shift-invariant wavelet packet decomposition and representation', *Signal Processing*, vol. 57, no. 3, pp. 251–270, Mar. 1997, doi: 10.1016/S0165-1684(97)00007-8.
[103] C.-M. Pun, 'Efficient and Adaptive Rotation Invariant Wavelet Transform', *https://doi.org/10.1142/S0219691303000219*, vol. 01, no. 03, pp. 353–372, Jan. 2012, doi: 10.1142/S0219691303000219.
[104] T. N. Kipf and M. Welling, 'Semi-Supervised Classification with Graph Convolutional Networks', *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, Sep. 2016, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/1609.02907v4
[105] B. Yu, H. Yin, and Z. Zhu, 'Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting', *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2018-July, pp. 3634–3640, Sep. 2017, doi: 10.24963/ijcai.2018/505.
[106] J. You, R. Ying, and J. Leskovec, 'Design Space for Graph Neural Networks', *Adv Neural Inf Process Syst*, vol. 2020-December, Nov. 2020, Accessed: Nov. 28, 2023. [Online]. Available: https://arxiv.org/abs/2011.08843v2
[107] D. Zhu, P. Cui, Z. Zhang, and W. Zhu, 'Robust graph convolutional networks against adversarial attacks', *Proceedings of the ACM SIGKDD International Conference on*

Knowledge Discovery and Data Mining, pp. 1399–1407, Jul. 2019, doi: 10.1145/3292500.3330851.

[108] W. Hu *et al.*, 'Strategies for Pre-training Graph Neural Networks', *8th International Conference on Learning Representations, ICLR 2020*, May 2019, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/1905.12265v3

[109] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, 'Neural Message Passing for Quantum Chemistry', *34th International Conference on Machine Learning, ICML 2017*, vol. 3, pp. 2053–2070, Apr. 2017, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/1704.01212v2

[110] W. L. Hamilton, R. Ying, and J. Leskovec, 'Inductive Representation Learning on Large Graphs', *Adv Neural Inf Process Syst*, vol. 2017-December, pp. 1025–1035, Jun. 2017, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/1706.02216v4

[111] K. Xu, S. Jegelka, W. Hu, and J. Leskovec, 'How Powerful are Graph Neural Networks?', *7th International Conference on Learning Representations, ICLR 2019*, Oct. 2018, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/1810.00826v3

[112] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. Prasanna, 'GraphSAINT: Graph Sampling Based Inductive Learning Method', *8th International Conference on Learning Representations, ICLR 2020*, Jul. 2019, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/1907.04931v4

[113] P. Veličković, A. Casanova, P. Liò, G. Cucurull, A. Romero, and Y. Bengio, 'Graph Attention Networks', *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, Oct. 2017, doi: 10.1007/978-3-031-01587-8_7.

[114] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, 'Simple and Deep Graph Convolutional Networks', *37th International Conference on Machine Learning, ICML 2020*, vol. PartF168147-3, pp. 1703–1713, Jul. 2020, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/2007.02133v1

[115] L. Zhao and L. Akoglu, 'PairNorm: Tackling Oversmoothing in GNNs', *8th International Conference on Learning Representations, ICLR 2020*, Sep. 2019, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/1909.12223v2

[116] Y. Chen, X. Tang, X. Qi, C. G. Li, and R. Xiao, 'Learning Graph Normalization for Graph Neural Networks', *Neurocomputing*, vol. 493, pp. 613–625, Sep. 2020, doi: 10.1016/j.neucom.2022.01.003.

[117] B. Xu, H. Shen, Q. Cao, Y. Qiu, and X. Cheng, 'Graph Wavelet Neural Network', *7th International Conference on Learning Representations, ICLR 2019*, Apr. 2019, Accessed: Nov. 27, 2023. [Online]. Available: https://arxiv.org/abs/1904.07785v1

[118] G. Li, M. Muller, A. Thabet, and B. Ghanem, 'DeepGCNs: Can GCNs Go as Deep as CNNs?', *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 9266–9275, Apr. 2019, doi: 10.1109/ICCV.2019.00936.

[119] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, 'Understanding deep learning requires rethinking generalization', *Commun ACM*, vol. 64, no. 3, pp. 107–115, Nov. 2016, doi: 10.1145/3446776.

[120] T. Pearce, A. Brintrup, and J. Zhu, 'Understanding Softmax Confidence and Uncertainty', Jun. 2021, Accessed: Dec. 04, 2023. [Online]. Available: https://arxiv.org/abs/2106.04972v1

[121] E. Kee, J. J. Chong, Z. J. Choong, and M. Lau, 'A Comparative Analysis of Cross-Validation Techniques for a Smart and Lean Pick-and-Place Solution with Deep Learning', *Electronics 2023, Vol. 12, Page 2371*, vol. 12, no. 11, p. 2371, May 2023, doi: 10.3390/ELECTRONICS12112371.

[122] I. H. Sarker, ‘Machine Learning: Algorithms, Real-World Applications and Research Directions’, *SN Comput Sci*, vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/S42979-021-00592-X/FIGURES/11.
[123] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, ‘Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction’, *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 1–11, Apr. 2022, doi: 10.1038/s41598-022-10358-x.
[124] Muhathir and Al-Khowarizmi, ‘Measuring the Accuracy of SVM with Varying Kernel Function for Classification of Indonesian Wayang on Images’, *2020 International Conference on Decision Aid Sciences and Application, DASA 2020*, pp. 1190–1196, Nov. 2020, doi: 10.1109/DASA51403.2020.9317197.
[125] D. P. Kingma and J. L. Ba, ‘Adam: A Method for Stochastic Optimization’, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Dec. 2014, Accessed: Feb. 18, 2022. [Online]. Available: https://arxiv.org/abs/1412.6980v9
[126] Phil. Brodatz, ‘Textures; a photographic album for artists and designers.’, p. 112, 1966, Accessed: Dec. 29, 2023. [Online]. Available: https://books.google.com/books/about/Textures.html?id=4rBk_SemB2QC
[127] H. G. Nguyen, R. Fablet, and J. M. Boucher, ‘Visual textures as realizations of multivariate log-Gaussian Cox processes’, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2945–2952, 2011, doi: 10.1109/CVPR.2011.5995340.
[128] X. Qi, R. Xiao, C. G. Li, Y. Qiao, J. Guo, and X. Tang, ‘Pairwise rotation invariant co-occurrence local binary pattern’, *IEEE Trans Pattern Anal Mach Intell*, vol. 36, no. 11, pp. 2199–2213, Nov. 2014, doi: 10.1109/TPAMI.2014.2316826.
[129] W. Guo, K. Lin, and W. Ye, ‘Deep Embedded K-Means Clustering’, *2021 International Conference on Data Mining Workshops (ICDMW)*, vol. 2021-December, pp. 686–694, Dec. 2021, doi: 10.1109/ICDMW53433.2021.00090.
[130] W. H. Lopez Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, ‘Autoencoders’, *Machine Learning: Methods and Applications to Brain Disorders*, pp. 193–208, Mar. 2020, doi: 10.1016/B978-0-12-815739-8.00011-0.
[131] M. Ahmed, R. Seraj, and S. M. S. Islam, ‘The k-means Algorithm: A Comprehensive Survey and Performance Evaluation’, *Electronics 2020, Vol. 9, Page 1295*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/ELECTRONICS9081295.