Integration of a priori data to optimise industrial X-ray tomographic reconstruction

Victor Bussy

To cite this version:

Victor Bussy. Integration of a priori data to optimise industrial X-ray tomographic reconstruction. Signal and Image processing. INSA de Lyon, 2024. English. NNT : 2024ISAL0116 . tel-04961293

HAL Id: tel-04961293
https://theses.hal.science/tel-04961293v1

Submitted on 21 Feb 2025

N° d'ordre NNT : 2024ISAL0116

THESE de DOCTORAT DE L'INSA LYON, membre de l'Université de Lyon

Ecole Doctorale 160
Électronique, Électrotechnique, Automatique

Spécialité de doctorat :

Traitement du Signal et de l'Image

Soutenue publiquement à Paris le 04/12/2024, par :

Victor Bussy

Integration of a priori data to optimise industrial X-ray tomographic reconstruction

---

Devant le jury composé de :

Desbat	Laurent	Professeur des Universités	Université de Grenoble	Rapporteur
Sijbers	Jan	Professeur des Universités	University of Antwerp	Rapporteur
Herl	Gabriel	Professeur des Universités	Deggendorf Institute of Technology	Examinateur
Quinsat	Yann	Professeur Agrégé	ENS Paris-Saclay	Examinateur
Kaftandjian	Valérie	Professeur des Universités	INSA-Lyon	Directrice de thèse
Vienne	Caroline	Ingénieure Chercheuse	CEA-Saclay	Examinatrice
Escoda	Julie	Ingénieure Chercheuse	CEA-Saclay	Examinatrice

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
ED 206 CHIMIE	**CHIMIE DE LYON** **https://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage **secretariat@edchimie-lyon.fr**	**M. Stéphane DANIELE** C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne **directeur@edchimie-lyon.fr**
ED 341 E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 **secretariat.e2m2@univ-lyon1.fr**	**Mme Sandrine CHARLES** Université Claude Bernard Lyon 1 UFR Biosciences Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69622 Villeurbanne CEDEX **e2m2.codir@listes.univ-lyon1.fr**
ED 205 EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://ediss.universite-lyon.fr** Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 **secretariat.ediss@univ-lyon1.fr**	**Mme Sylvie RICARD-BLUM** Laboratoire ICBMS - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 **sylvie.ricard-blum@univ-lyon1.fr**
ED 34 EDML	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 **yann.de-ordenana@ec-lyon.fr**	**M. Stéphane BENAYOUN** Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 **stephane.benayoun@ec-lyon.fr**
ED 160 EEA	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **https://edeea.universite-lyon.fr** Sec. : Philomène TRECOURT Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 **secretariat.edeea@insa-lyon.fr**	**M. Philippe DELACHARTRE** INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 **philippe.delachartre@insa-lyon.fr**
ED 512 INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 **infomaths@univ-lyon1.fr**	**M. Hamamache KHEDDOUCI** Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 **direction.infomaths@listes.univ-lyon1.fr**
ED 162 MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Philomène TRECOURT Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon **mega@insa-lyon.fr**	**M. Etienne PARIZET** INSA Lyon Laboratoire LVA Bâtiment St. Exupéry 25 bis av. Jean Capelle 69621 Villeurbanne CEDEX **etienne.parizet@insa-lyon.fr**
ED 483 ScSo	**ScSo**[1] **https://edsciencessociales.universite-lyon.fr** Sec. : Mélina FAVETON Tél : 04.78.69.77.79 **melina.faveton@univ-lyon2.fr**	**M. Bruno MILLY** (INSA : J.Y. TOUSSAINT) Univ. Lyon 2 Campus Berges du Rhône 18, quai Claude Bernard 69365 LYON CEDEX 07 Bureau BEL 319 **bruno.milly@univ-lyon2.fr**

---

[1] ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Contents

List of Figures

List of Tables

Acronyms

AABB Axis-Aligned-Bounding-Boxes 123

ADMM Alternating Directions Method of Multipliers 52, 58, 139, 141, 144, 148, 153, 156, 208

ADSIR Adaptive Dictionary SIR 151, 152

ART Algebraic Reconstruction Technique 49, 50

AUC Area Under Curve 57

BM3D Block-Matching and 3D Filtering 134, 135

BP Basis Pursuit 136

BVH Bounding Volume Hierarchy 123–125, 130, 204, 205

CAD Computer-Aided Design 21–23, 61, 67, 73, 75, 81, 82, 91–94, 98, 101, 104, 109, 113, 115, 121–124, 126, 130, 159, 160, 196

CGLS Conjuguate Gradient Least Squares 50, 51, 58

CNR Contrast Noise Ratio 57

CSC Convolutional Sparse Coding 143–145, 148, 152, 156, 160, 207, 208

CT Computed Tomography 21, 29, 31, 33, 35, 38, 50, 51, 54, 61, 64, 70, 133, 145, 208

DCT Discrete Cosine Transform 139, 145, 208

DEIM Discrete Empirical Interpolation Method 67–71, 73, 75, 76, 159

DLT Direct Linear Transform 84

DOE Design of Experiments 71, 72

DoG Difference of Gaussians 86

EIM Empirical Interpolation Method 67, 68, 70, 73

FBP Filtered Backprojection 41–43, 58, 64, 133

FDDL Fischer Discrimination Dictionary Learning 156

FDK Feldkamp Davis Kress 43, 76, 122

FISTA Fast Iterative Thresholding Algorithm 138, 139

FOV Field Of View 35

GDSIR Global Dictionary SIR 151, 152

GPU Graphics Processing Unit 113, 123, 125

IBVS Image Based Visual Servoing 87, 88, 91, 92, 101

ICA Independent Component Analysis 74

ICP Iterative Closest Point 23, 92, 102, 104, 107, 189, 197, 198

ISTA Iterative Thresholding Algorithm 138, 139

KL Kullback-Leibler 142

LARS Least-Angle Regression 141

LPBF Laser Powder Bed Fusion 157

LRSDL Low-Rank Share Dictionary Learning 155, 156

MFD Measurement Field Diameter 35, 62

ML-CSC Multi-Layer Convolutional Sparse Coding 144

MLEM Maximum Likelihood Expectation Maximization 51, 58

MOD Method of Optimal Directions 141

MSE Mean Square Error 55, 58

NDT Non Destructive Testing 21, 22

NLM Non-Local Means 134, 135

OMP Orthogonal Matching Pursuit 137, 138

P^3 Plug-and-Play 134

PBVS Position Based Visual Servoing 87, 91

PCA Principal Component Analysis 67, 69, 145

PDHG Primal-Dual Hybrid Gradient 52, 53, 58

PnP Perspective-n-Points 83, 85, 86, 92

POCS Projection Onto Convex Sets 119, 120

POD Proper Orthogonal Decomposition 67–69, 74

PSNR Peak Signal Noise Ratio 55, 58, 75, 76, 121, 145, 150, 159

RANSAC Random Sample Consensus 85, 86

RED Regularisation by Denoising 134

ROC Receiver Operating Characteristic 57

ROI Region Of Interest 61, 79

ROM Reduced Order Models 67

SART Simultaneous Algebraic Reconstruction 49, 50, 58, 117, 128, 129, 153, 205

SIFT Scale-Invariant Feature Transform 86

SIR Statistical Iterative Reconstruction 133

SIRT Simultaneous Iterative Reconstruction 49, 50, 58, 117–121, 202

SNR Signal Noise Ratio 57, 74

SSIM Structural Similarity Index 56, 58, 75, 76, 121, 145, 150

STL STereoLithography 94, 96, 123

SVD Singular Value Decomposition 69, 85, 93

Introduction

Context

This thesis explores research topics within the realm of X-ray industrial non-destructive testing (NDT). NDT is an invaluable technique as it enables the evaluation of a material or system's properties without causing any damage. Among the various NDT methodologies (such as ultrasonic testing, thermography, eddy-current testing, radiography, ...), X-ray computed tomography (CT) stands out as a powerful tool for characterising or locating internal flaws and verifying the geometric conformity of an object. The application of CT has broadened significantly and its utilisation has surged in numerous industrial sectors for the inspection of components for flaw detection, geometric measurement, and reverse engineering applications. CT can be employed for the inspection of components during manufacturing, thereby ensuring product integrity.

Although its primary use is medical imaging, CT inspection is now well-established in the industrial domain but must continually evolve due to the increasing demands and constraints on inspection processes. Whether in terms of reconstruction quality or inspection time, CT technology is continuously advancing, particularly in what is known as the *sparse-view strategy*. This strategy involves reconstructing an object using CT with the minimum possible number of radiological projections while maintaining reasonable reconstruction quality. This approach reduces acquisition times and associated costs. Sparse-view reconstruction is a real challenge as the tomographic problem is ill-posed. Numerous techniques have been developed to overcome this issue, many of which are based on incorporating prior information during the reconstruction process. By utilising data and knowledge available prior to the experiment, it is possible to enhance the reconstruction result despite the reduced number of projections and hence the information acquired about the studied system.

Typically, the integration of *a priori* information is accomplished via a regularisation term that models knowledge about the object or functions learned through machine learning from samples similar to the one under study. However, this approach can be extended to many other stages of the tomographic process beyond the reconstruction itself. Whether it is the expertise of the experimenter selecting the optimal machine parameters, the precise positioning of objects, the ideal trajectory, or even specific knowledge about the type of inspected part: prior information can take various forms. All these insights can be leveraged to improve sparse-view CT reconstruction.

In our industrial context, for example, the computer-aided design (CAD) model of the object is often available, representing valuable information about the geometry of the object under study. Nevertheless, it is important to note that CAD provides only an approximate object representation. In NDT or metrology, it is precisely the differences between an object and its CAD model that are of interest. Consequently, instilling *a priori* information is challenging because the information is often approximate and cannot be used as-is. Therefore, in this thesis, we do not aim to incorporate *a priori* information obtained from vast databases to create a regularisation *a priori* term. Instead, we propose to judiciously utilise the geometric information available from the CAD model at each stage of the process. We thus propose a methodology tailored integrating the geometric *a priori* information into the tomographic reconstruction.

The CEA LIST at Paris-Saclay, where this thesis was conducted, is a laboratory specialising in NDT, and it is equipped with an advanced robotic X-ray platform. This robotic cell consists of two multi-axis robotic arms, shown in Figure 1. One arm holds an X-ray source and is referred to as the *sniper*, while the second arm carries a detector and is known as the *target*. This platform offers significant flexibility in both the experimental process and the reconstruction, allowing us to address a wide range of topics throughout this thesis.

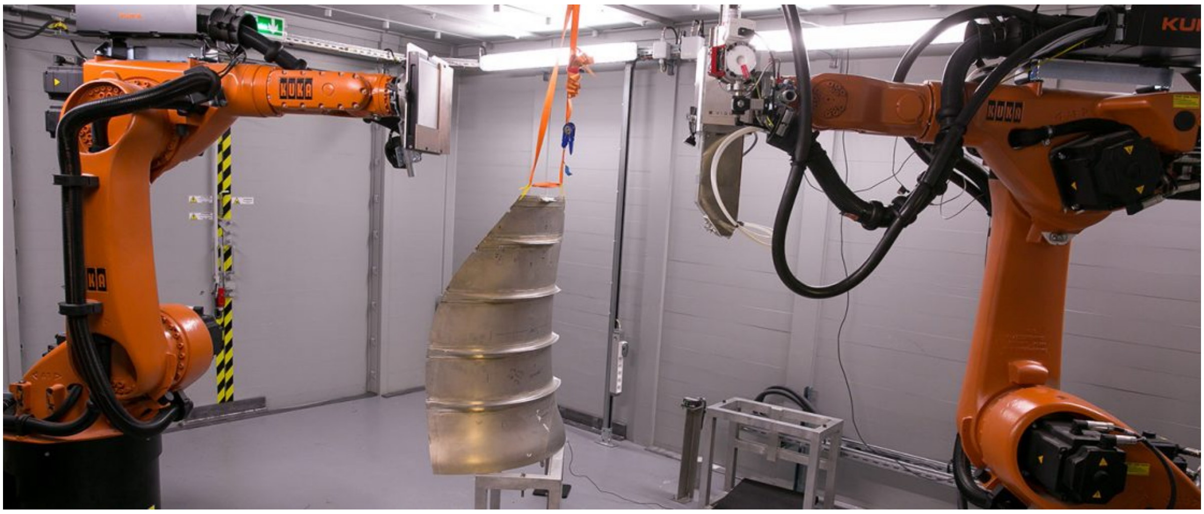


Figure 1: Photograph of the robotic cell at CEA LIST. On the left is the target, and on the right is the sniper.

In [11], we present a whole optimised methodology using *a priori* information to improve reconstruction quality on an additive manufactured metallic part. The methodology is summed up in Figure 2 below. This thesis revisits most elements of this methodology. In each chapter, we will examine one component of this methodology.

Chapter 1 provides a general introduction to tomography, covering both the physical components for X-ray generation and detection, as well as the mathematical components for solving the tomographic problem.

Chapter 2 presents a state-of-the-art review of projection acquisition trajectory optimisation methods and the methods we have developed. Our algorithms, based on the *Empirical Interpolation Method*, use CAD as *a priori* geometric information to define the most relevant projections for reconstruction. As our research progressed, several variants were proposed depending on the type of implementation and the incorporation of constraint terms. The methods have been tested on many objects and results have already been published in [4, 9]. The presented methods will be evaluated on a new experimental setup and a plastic additive manufactured part.

Chapter 3 introduces 2D/3D registration techniques. Registration is a preliminary step essential for incorporating *a priori* information, which needs to be positioned within a reference frame. We introduce fundamental concepts for registering CAD models with real objects. Two registration techniques were developed and are presented: *Robust 3D/2D hybrid registration* and *Convex Hull Inverse Iterative Perspective Matching*. Each has advantages depending on the required quality or speed. The first one uses visual servoing and image moments, while the second uses an iterative scheme close to the ICP to register a CAD on experimental projections. Simulated results from these methods have already been published in [12, 13]. In this thesis, the presented methods will be tested and evaluated on an experimental piece.

Chapter 4 proposes using our *a priori* geometric information to provide a mathematical description suited to the studied object. By using CAD, it is possible to adapt the discretisation of the object to reduce the number of variables or improve resolution. We will present results on so-called "masked" reconstructions as well as hierarchical structures. Our work on mask reconstruction was published in [11, 13] where the reconstruction was coupled with an upstream CAD registration technique.

Chapter 5 does not use geometric *a priori* information but rather focuses on the texture and appearance of the reconstructed objects. This chapter explains the fundamental concepts of sparse coding and dictionary learning to denoise sparse-view reconstructions. We compare existing algorithms and present an extension of these using convolution. Results on denoising, regularisation, and even classification will be presented. Those results have also been showed in [14].

Finally, Chapter 6 concludes our research and presents future perspectives.

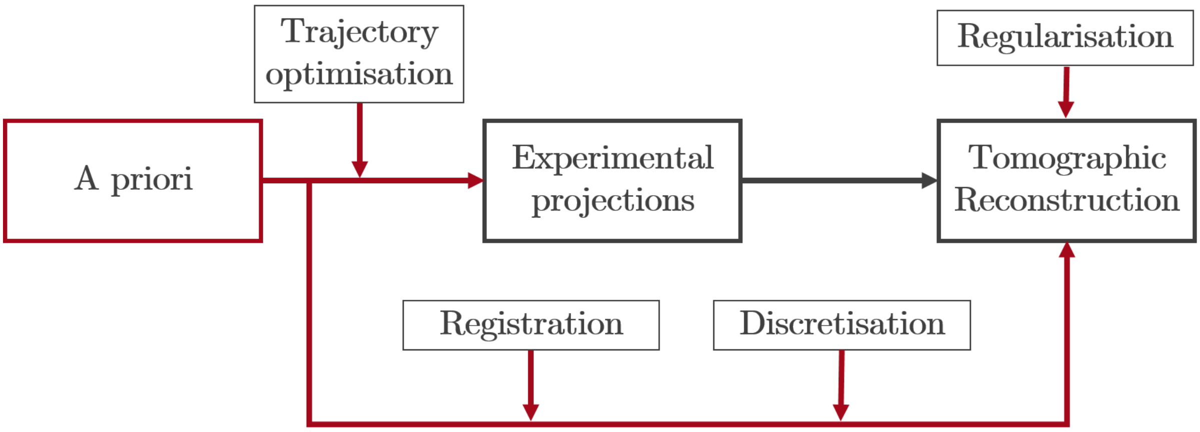


Figure 2: Diagram of the tomographic process. *A priori* information are used at different steps to optimise reconstruction quality in sparse-view.

Contributions

Throughout my research, I have had the opportunity to publish several contributions.

- In our first paper, we presented trajectory optimisation algorithms based on empirical interpolation methods, exploiting CAD models to define the most relevant

projections, thereby enhancing the quality of reconstructions for complex objects.
Victor Bussy, Caroline Vienne, and Valérie Kaftandjian. Fast algorithms based on empirical interpolation methods for selecting best projections in sparse-view x-ray computed tomography using a priori information. *NDT&E International*, 2022.

- We further extended our work on trajectories by introducing a physical constraint into the previous method.
  Victor Bussy, Caroline Vienne, and Valérie Kaftandjian. Best projections selection algorithm based on constrained qdeim for sparse-views x-ray computed tomography. In *12th Conference on Industrial Computed Tomography (iCT) 2023*, 27 February - 2 March 2023 in Fürth, Germany. *e-Journal of Nondestructive Testing* Vol. 28(3), 2023.

- Subsequently, for the registration, we published a paper on visual servoing based on points and image moments.
  Victor Bussy, Caroline Vienne. Robust 3D/2D hybrid registration for integrating a priori CAD model into X-ray Computed Tomography. In *ORASIS 2021*, Saint Ferréol, France, September 2021. Centre National de la Recherche Scientifique [CNRS].

- This registration method was further complemented by another paper on fast registration based on inverse perspective matching of convex hull points. In this paper, we applied this registration to introduce the notion of a mask in reconstructions, which reduces the number of variables and significantly decreases artefacts.
  Victor Bussy, Caroline Vienne, Julie Escoda, and Valérie Kaftandjian. Sparse-View X-Ray CT Reconstruction using CAD Model Registration. *49th Annual Review of Progress in Quantitative Nondestructive Evaluation*, 2022.

- The idea of reducing the number of variables was further pursued, where we explored hierarchical tree structures combined with meshes for reconstruction, which resulted in a patent submission.

- Our work on convolutional sparse coding and its applications in denoising reconstructed volumes from sparse views has been validated on manufactured parts, demonstrating promising results, paving the way for future improvements in nondestructive testing and industrial metrology.
  Chuan Huang, Paul Vaska, Yongfeng Gao, Shaojie Chang, Thomas Wesley Holmes, Amir Pourmorteza, and Jerome Liang. *Proceedings of the 17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2023.

- We have also published and presented a french version of this last article.
  Victor Bussy, Caroline Vienne, Julie Escoda & Valérie Kaftandjian. Convolutional Sparse Coding et Dictionary Learning pour la reconstruction tomographique par rayons X pour le contrôle de pièces de fabrication additive. LES JOURNÉES COFREND/ COFREND DAYS, France, Marseille, 06 - 08 June 2023. e-Journal of Nondestructive Testing Vol. 28(9), 2023.

- We applied our entire methodology to the reconstruction of a metallic object in additive manufacturing. This approach significantly improved the precision of the reconstructions while reducing acquisition times.
  Victor Bussy, Caroline Vienne, Julie Escoda, and Valérie Kaftandjian. Méthodologie

optimisée pour la reconstruction tomographique avec ajout d'informations a priori pour l'inspection par rayons x. *e-Journal of Nondestructive Testing*, 28, 09 2023.

Chapter 1

Introduction

1.1 Fundamentals of X-ray tomography

Given that we are proposing an optimised methodology at various stages, it is important first of all to be familiar with the entire tomographic process. In this section, we endeavour to delineate all the stages of tomographic reconstruction. We shall first address the physical concepts necessary for understanding X-rays and the tomographic apparatus, before progressing to the mathematical concepts essential for reconstruction and quality metrics. The concepts discussed in this section are elaborated upon in the references [1, 3, 15].

1.1.1 X-Ray Generation

Wilhelm Conrad Röntgen is celebrated for his discovery of X-rays, an achievement that earned him the Nobel Prize in 1901. The significance of X-rays lies in their ability to penetrate organic matter, allowing for the clear differentiation of tissues and bones in radiographic images. The wavelength of X-rays ranges from 0.01 nm (hard X-rays) to 10 nm (soft X-rays), but it is generally more practical to discuss them in terms of energy. As X-rays are electromagnetic waves, their energy is given by:

$$E = h\nu = h\frac{c}{\lambda}, \tag{1.1}$$

with $h = 6.63 \times 10^{-34}$ J $\cdot$ s being Planck's constant, and $c = 3 \times 10^{8}$ m $\cdot$ s^{-1}, which spans approximately from 100 eV to 1 MeV. Several devices can generate X-rays, including synchrotrons [15], but the most common in medical and industrial fields are X-ray tubes [1]. Among these, the *Coolidge* tube is the most common. Figure 1.1 shows a schematic diagram of how a Coolidge tube works.

A Coolidge tube comprises a cathode and an anode. Essentially, the cathode is heated, liberating electrons from its outer layers, a process known as thermionic emission. The density of emitted electrons is described by the Richardson-Dushman equation. The voltage between the cathode and anode accelerates the liberated electrons towards the anode. Upon collision with the anode, several interactions can occur:

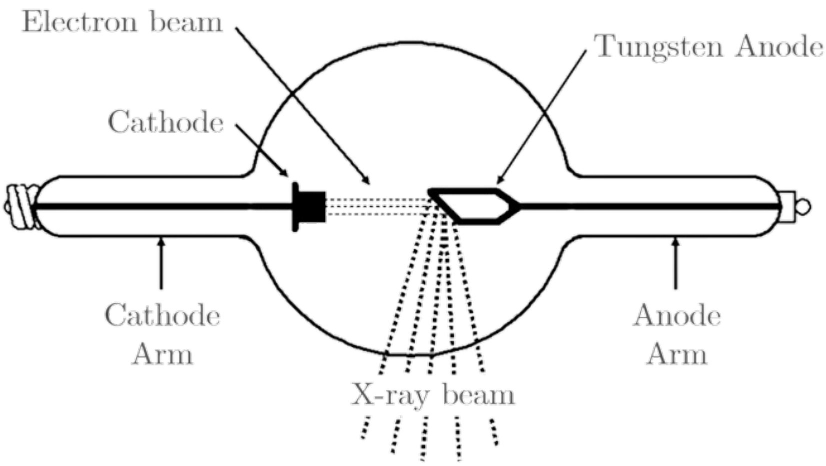


Figure 1.1: Coolidge X-Ray Tube. Image adapted from Oak Ridge Associated Universities.

- The electrons may be deflected, resulting in the *Bremsstrahlung* radiation. As the electron approaches an atom in the anode, it is attracted to the positively charged nucleus due to Coulomb's law. This causes the electron to deviate from its path and decelerate. The loss of kinetic energy results in the emission of radiation (see Figure 1.2a). The greater the change in the electron's kinetic energy, the higher the energy of the emitted photon. Bremsstrahlung produces a continuous spectrum, also known as white radiation or continuous braking radiation.

- The incident electron may collide with an electron in a deep shell of an atom, ejecting it and ionizing the atom. An electron from a higher shell then fills the vacancy (see Figure 1.2b). This transition releases energy in the form of characteristic radiation, the energy of which depends on the specific electron shells involved. The energy differences between the various shells of an atom are fixed and characteristic of that atom, creating distinct lines in the emission spectrum.

- In a less common interaction, the electron can directly collide with the nucleus of an atom. In this scenario, the electron's entire energy is converted into a photon according to the Bremsstrahlung effect. This interaction produces X-rays of maximum energy, though the probability of such a collision is very low (see Figure 1.2c).

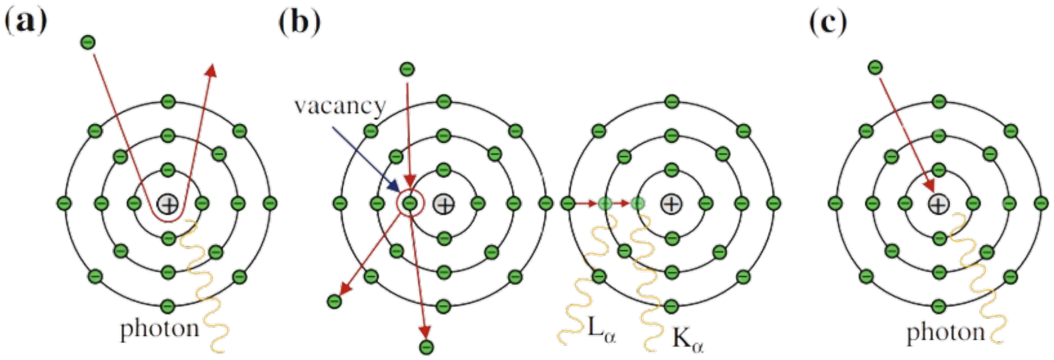


Figure 1.2: X-Ray generation principle. Image adapted from [1]. (a) Bremsstrahlung effect. (b) Characteristic radiations. (c) Direct electron/nucleus interaction.

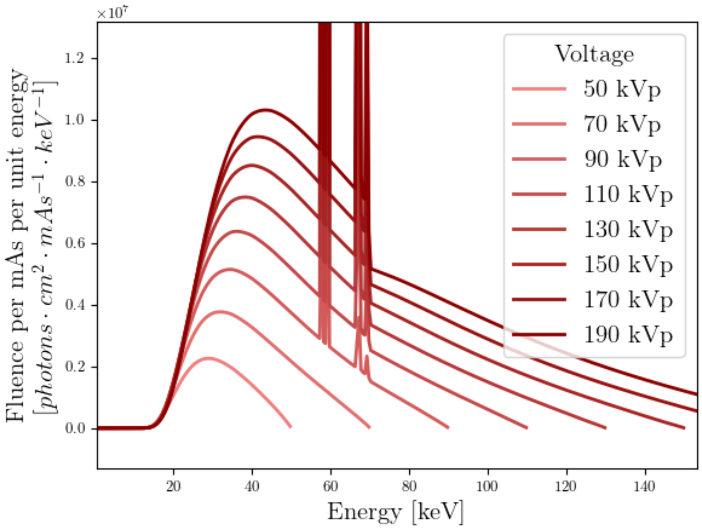


Figure 1.3: Example of a filtered X-ray radiation spectrum with a tungsten anode. The anode angle is 12°, and the tube potential varies from 50 to 190 kV. Spectrum range increases with voltage.

To better illustrate the effects of these interactions, Figures 1.3, 1.4 and 1.5 show typical spectra from an X-ray tube with a tungsten anode and filtration[1]. They all show the continuous spectrum of Bremsstrahlung radiation as well as some characteristic emission lines. Among all the parameters that can influence the spectrum, some are more significant. The tube current is one of the most important parameters for example, but it is relatively simple to account for as it only affects the amplitude of the spectra. Another important parameter is the acceleration voltage which typically increases the amplitude and broadens the spectrum. Figure 1.3 shows different spectra for different voltages from 50kV to 190kV. As acceleration increases, the spectrum is shifted to higher energies, which increases the penetration of the X-ray beam.

It is common to remove low-energy X-rays, and for this purpose, the beam is filtered. Figure 1.4 shows the impact of filtration with 1mm Be and 2mm Al. The amplitudes of the spectrum are attenuated at all energies, but primarily at lower energies.

The material of the anode will endure intense heat due to the electron flux, so to prevent the target from melting, one must expand the focal spot by increasing the angle between the flux and the anode. However, this leads to a blur, known as the *penumbra effect*. In practice, one prioritises image quality over photon quantity. In high-resolution industrial CT, spot sizes are about 0.1 mm. The anode angle non-linearly increases the flux, especially at low energy. Figure 1.5 shows the impact of the anode angle on the spectrum.

1.1.2 Absorption

Once the X-rays are generated, they will penetrate the object under study and undergo one or several interactions with its material. Generally, four interactions are identified:

---

[1] Spectra are made with SpekPy `https://bitbucket.org/spekpy/spekpy_release/wiki/Home`

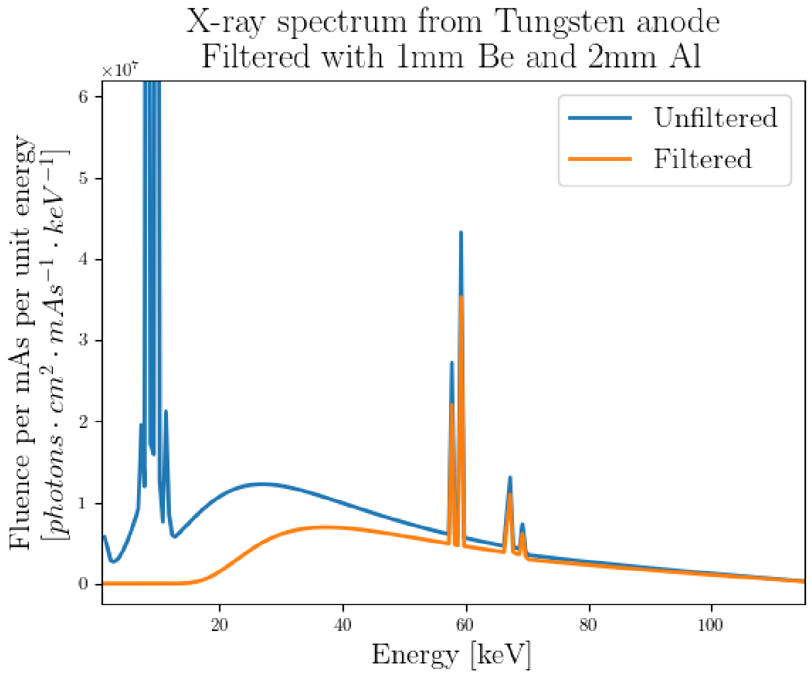


Figure 1.4: Example of an X-ray radiation spectrum with a tungsten anode. The blue curve represents the unfiltered spectrum, while the orange spectrum is filtered by 1 mm of Be and 2 mm of Al. The anode angle is 12°, and the tube potential is 120 kV. Filtering removes low energy.

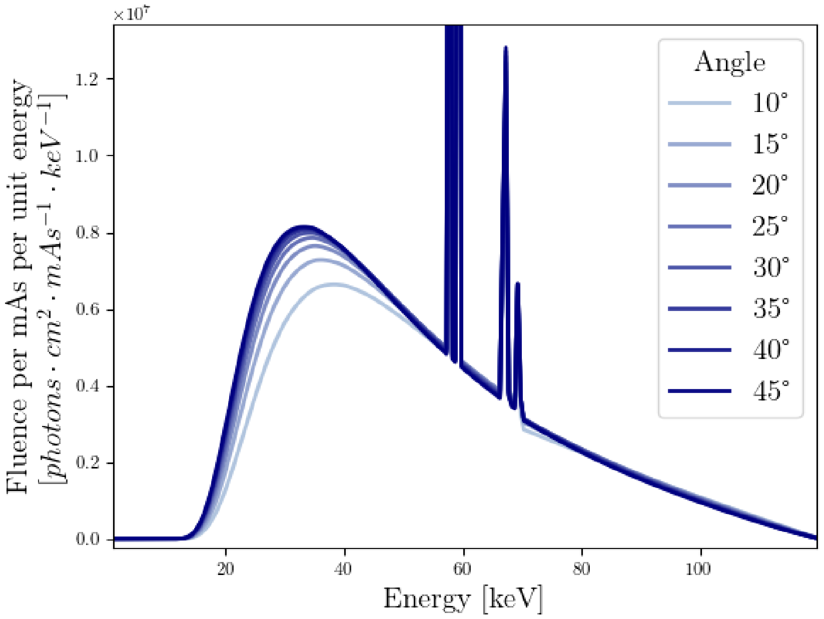


Figure 1.5: Example of a filtered X-ray radiation spectrum with a tungsten anode. The anode angle varies from 10 to 45°. Increasing the angle slightly increases the amplitude of the spectrum and shifts it towards low energies.

Photoelectric effect, Incoherent Compton effect, Rayleigh / Thompson scattering and Pair production.

The photoelectric effect, whose explanation is due to Albert Einstein, involves the absorption of an incident photon and the release of an electron (called a photoelectron) if its binding energy is lower than that of the incident photon. A vacancy is created in the inner shell, and an electron from a higher shell replaces it, resulting in a characteristic line, known as fluorescence. Occasionally, another electron may be ejected instead of X-ray fluorescence; this is known as an Auger electron. The absorption coefficient depends on Z (atomic number) and the energy of the photon, but various models with different choices for the coefficients exist, sometimes scaling with Z^4, sometimes with Z^5. However, it is understood that higher-Z materials absorb more, allowing differentiation of materials in radiography. This is the key idea behind radiology and tomography.

During the incoherent Compton effect, a high-energy incident photon collides with an electron of the atom but is not absorbed. The electron is ejected, and the incident photon, having lost some of its energy, is scattered in a new direction, the angle of which depends on the initial energy and the energy lost (the direction is given by the Klein-Nishina formula, which allows for the estimation of the differential cross section in barns per steradian). Compton scattering is also known as incoherent scattering and occurs in the same energy range as the photoelectric effect. Its probability depends only on the electron density of the material (the number of electrons in the valence shell) and not the atomic number. The lack of correlation with the atomic number does not provide good contrast between different materials of similar density, and the random angle of the scattered photon creates noise. Therefore, tomography devices attempt to minimise the impact of the Compton effect.

Rayleigh coherent scattering occurs at low energy when the incident photon does not interact with the electron cloud but the atom emits another photon of equivalent energy in a different direction. This is equivalent to a deviation of the incident photon. It is rare in CT because it only occurs at low energy.

Pair production is very rare. For this phenomenon to occur, the energy of the photon must be greater than twice the rest mass of the electron, which is 1.02 MeV, generally beyond what X-ray tubes can provide for industrial tomography. In this case, an electron and its antiparticle are created.

The nature of the interaction depends mainly on the energy of the incident ray and the nature of the traversed material. For each material and each interaction, one can define a linear attenuation coefficient $\boldsymbol{\mu}$ in cm^{-1} (or in *Hounsfield Units* (HU) directly related to the attenuation of water), which depends on the photon energy. Figure 1.6 shows the contribution of each interaction for the iron at different energies. The photoelectric interaction is predominant at low energies, while incoherent Compton scattering becomes more significant at higher energies. The data used for the figure comes from XrayDB[2]. Figure 1.7 shows the evolution of the total mass attenuation coefficient for aluminium, copper, and iron between 1 and 100 keV. Since the various linear attenuation coefficients are proportional to the mass density ρ, it is more common to use the mass attenuation coefficient μ/ρ in cm^2/g. Being independent of the mass density, it depends only on the chemical composition of the material and not on its physical state.

---

[2] XrayDB is X-ray data SQLite library `https://xraypy.github.io/XrayDB/index.html`

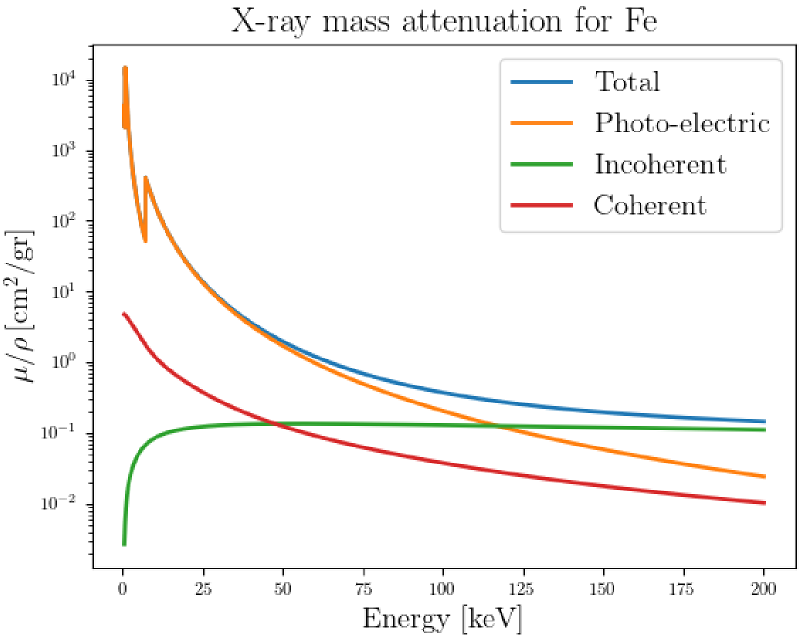


Figure 1.6: Attenuation values resulting from each interaction for iron (Fe). The photoelectric effect predominates at low energies, while the Compton effect is more significant at high energies.

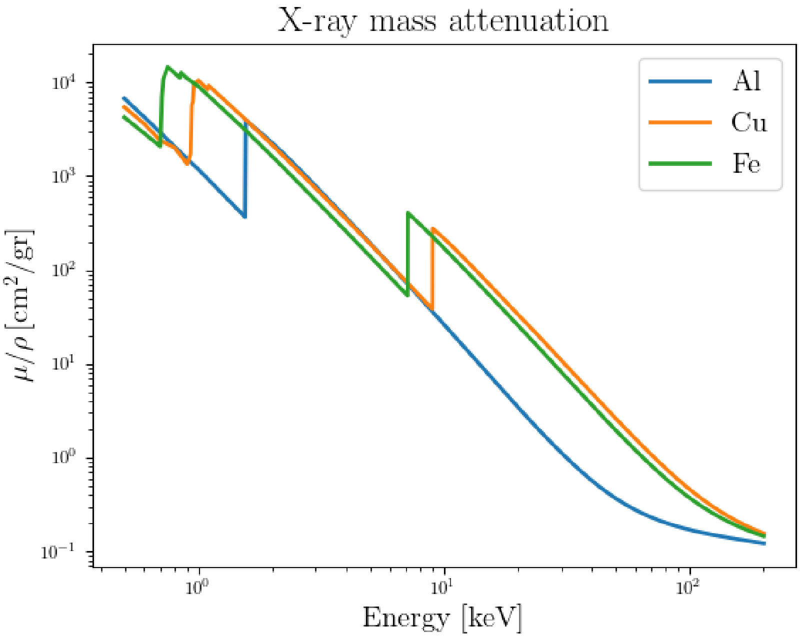


Figure 1.7: Total X-ray mass attenuation for aluminium, copper and iron. In our industrial applications, the energy will always be above 10 keV, right side of the figure.

As the number of generated rays is too high to maintain a ray-by-ray approach and focus on the nature of each interaction, one will instead consider the total attenuation due to all interactions and adopt a beam vision. Ultimately, for radiological and tomographic applications, what is interesting is not so much the interactions or the probability of interaction of a ray, but rather the ratio of transmitted rays to emitted ones in the whole beam. In other words, one will only consider the ratio between the transmitted I and initial I_0 X-ray beam intensity. By definition of the linear attenuation coefficient, the evolution of the beam intensity can be described using Beer-Lambert's law:

$$\mathrm{d}I = -\mu I \mathrm{d}L, \tag{1.2}$$

where $\mathrm{d}L$ represents an infinitesimal distance travelled by the beam of rays. To obtain the relationship between the initial beam and the one received after passing through the object, this equation must be integrated along the entire path L of the beam. Since μ depends on the location within the object and on the energy, we must also integrate over energy, leading to the following equation:

$$I = \int I_0(E) e^{-\int_{\mathbf{x}\in L} \mu(E,\mathbf{x})\mathrm{d}L} \mathrm{d}E. \tag{1.3}$$

Equation 1.3 takes into account what is called the *polychromaticity* of the spectrum. As mentioned earlier, the attenuation coefficient varies not only depending on the path (which part of the object we integrate over) but also on energy. Since attenuation is stronger at lower energies, we notice that the first layers of material will filter the beam and only let through the higher energies. The layers behind will therefore only see high energies, resulting in less attenuation there. This phenomenon is called *beam hardening*. In practice, this dependence is neglected, but it is a significant source of artefacts. Figure 1.8a shows the ratio I/I_0 along the penetrated length in air, water and iron. In air and even water, beam hardening effects are negligible, but not in metals.

To evaluate the impact of beam hardening, we illustrated in Figure 1.8b the evolution of effective attenuation through iron. The curves represent μL i.e. $\ln(^{I_0}/_{I})$, with L the traversed length. In the monochromatic case, the value of μ is constant, and we indeed find a line with slope μ. In the polychromatic case, the effective attenuation coefficient decreases with the length traversed. The average path of an X-ray is on the order of $1/\mu$; for iron at 120 keV, it is approximately 1 mm (values from NIST[3]). To reduce the impact of beam hardening, it is essential to remove the low-energy photons from the spectrum, as they are predominantly responsible for this effect. This explains why the incident beam is often filtered initially.

Since it is difficult to compute Equation 1.3 rapidly, one must make the simplifying assumption that attenuation is constant across energy. This is the *monochromatic* assumption. Equation 1.3 becomes:

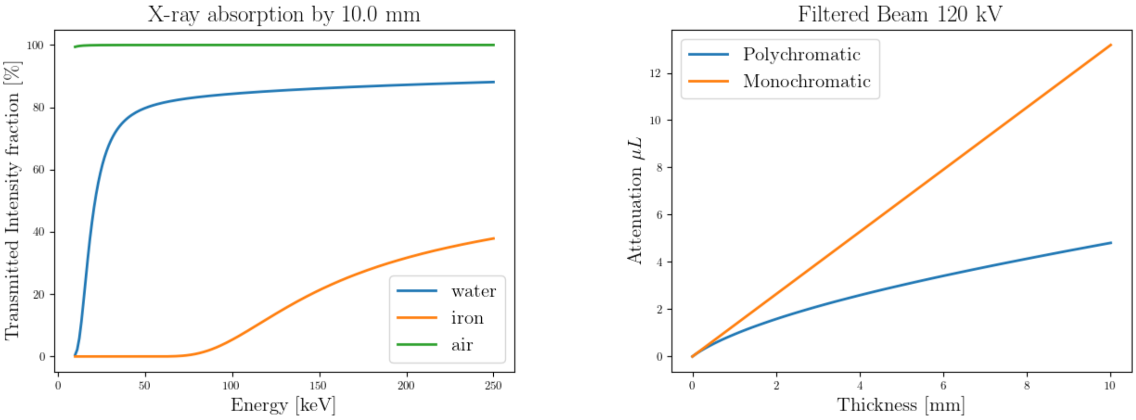
$$I = I_0 e^{-\int_{\mathbf{x}\in L} \mu(\mathbf{x})\mathrm{d}L}. \tag{1.4}$$

Since CT deals with images (in pixels or voxels), the reconstruction space must be discretised. This step is necessitated by the nature of our numerical computation methods. In its discrete form, Beer-Lambert's law is often written as follows:

$$\ln\left(\frac{I}{I_0}\right) = -\sum_i \boldsymbol{\mu}_i \ell_i, \tag{1.5}$$

---

[3] `https://physics.nist.gov/PhysRefData/XrayMassCoef/tab3.html`



(a) X-Ray transmission by 10.0mm of three different materials (water, iron and air). Energy dependency is more pronounced for metals, and so is beam hardening.

(b) Evolution of the effective attenuation coefficient of the iron times length given an unfiltered polychromatic one of maximum energy 120keV or an equivalent monochromatic X-ray beam.

Figure 1.8

where $\boldsymbol{\mu}_i$ represents the attenuation of the i^{th} pixel/voxel and ℓ_i represents the contribution of this pixel/voxel. There are numerous methods for taking into account the contribution of a particular zone. One can consider traversed length, surface area or even just a boolean coefficient. More details on the implementation of ray tracing are given in Section 1.3.2.

1.1.3 Signal detection and pre-processing

Once the X-ray beam has interacted with the object, the attenuated radiation is detected by the imaging system. Historically, X-ray detectors were based on film technology; however, these have increasingly been supplanted by digital detectors. X-ray detectors can be classified into three principal categories: ionization gas detectors, semiconductor detectors, and scintillator detectors [16]. Among these, scintillator detectors are predominant in contemporary industrial computed tomography systems. These detectors comprise a scintillation crystal coupled with a photon detector. When X-rays strike the scintillator layer, they are converted into visible light photons. This light is then absorbed by the photon detector, which generates an electric charge proportional to the X-ray radiation energy. Common scintillating materials include sodium iodide doped with thallium, caesium iodide, and cadmium tungstate ($CdWO_4$) [1, 17]. For more detailed information on the physics of detectors, their performance, and their characteristics, the reader is referred to [18].

Detectors simply produce an image that represents the number of photons that have struck them, so before proceeding with reconstruction, a pre-processing step is essential to account for the detector's response. Initially, images without the object are acquired to determine the blank factor or *flat-field* $\mathbf{F}$ (which differs from $\mathbf{I}_0$). Additionally, images of the detector's response without irradiation, known as the *dark frame* $\mathbf{D}$, are taken to detect and correct issues such as dead pixels. This operation is crucial for ensuring the accuracy of the detector. Subsequently, the object is positioned, and measurements are taken at different angles to obtain the necessary projections $\mathbf{I}$ for reconstruction. The experimental projections are corrected according to:

$$\mathbf{I}_{corr} = \frac{\mathbf{I} - \mathbf{D}}{\mathbf{F} - \mathbf{D}}. \tag{1.6}$$

Afterwards, to define $\mathbf{I}_0$, the intensity without attenuation, one generally takes an average over the edges of the image or any non-attenuated part of the image. The corrected projections are then given by:

$$\mathbf{p} = -\ln\left(\frac{\mathbf{I}_{corr}}{\mathbf{I}_0}\right). \tag{1.7}$$

At this point, projections $\mathbf{p}$ are ready for the reconstruction algorithm.

1.1.4 Practical aspect of tomographic reconstruction

The primary objective of CT tomography is to reconstruct the attenuation coefficient function $\mu(\mathbf{x})$ for every point in a 3D volume from 2D projections that contain information about the lines traversed by X-rays. Typically, we aim to reconstruct a parallelepiped of voxels encompassing the studied object. Since a single projection only contains information along lines rather than specific points in space, multiple projections from various angles are necessary to solve the tomographic problem.

In the vast majority of cases, CT scans employ a "circular" trajectory. The detector and the source are stationary while the object rotates on a turntable. Alternatively, this trajectory can be achieved with a stationary object and the source-detector pair moving along a circle centred on the object. Although numerous other trajectories exist, they are less popular due to their implementation challenges. Tomographic inspection trajectories will be discussed in greater detail in Section 2.1.

The field of view (FOV) of a projection is defined as the set of half-lines extending from the source point to the detector screen (considered finite, see Figure 1.9). This definition can be extended to parallel projections by considering the source point at infinity (see Figure 1.10). When discussing the field of view of a reconstruction, we generally refer to the intersection of the fields of view of each projection. For instance, in a circular trajectory around the object in 2D, the FOV of each projection forms a triangle with the detector and the source point. The resulting field of view for the reconstruction is a disc whose diameter is called the Measurement Field Diameter (MFD). If a point belongs to this disc, then for each projection, it belongs to a measured line, and it can be reconstructed stably [19]. Therefore, even though 2D reconstruction is generally performed on a square grid of pixels, efforts will always be made to ensure the object is included within this disc and to focus on its reconstruction. The field of view is also related to the *magnification factor*. The magnification factor is defined as the ratio between the size of the object's projection and the size of the object itself, i.e., the distance from the source to the detector over the distance from the source to the object. These concepts are crucial in the design of a CT scan and play a pivotal role in achieving high resolution. It is essential to optimise these parameters to strike a balance between the FOV, resolution, and the signal-to-noise ratio of the projections.

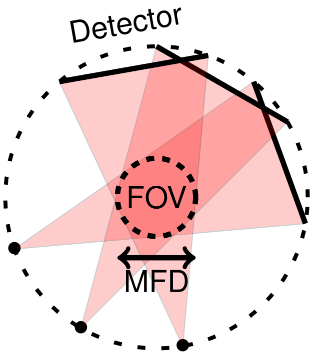
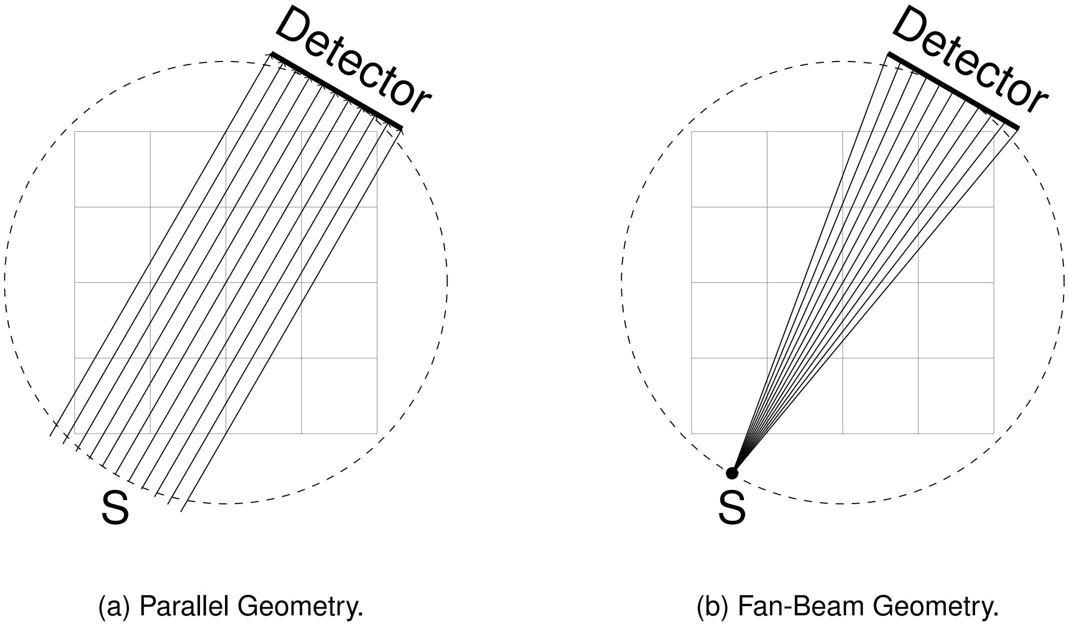


Figure 1.9: Illustration of the FOV and MFD for a fan-beam geometry on a circular trajectory. The intersection of the beams creates a circle with a diameter MFD, called the FOV.

1.2 Analytical reconstruction methods

Reconstruction methods are classified into two categories: analytical and iterative. Historically, analytical methods emerged first. They are known for their speed and high quality when the number of projections is sufficient. We will illustrate the concepts in two dimensions before extending them to three-dimensional spaces. First, we will detail the modelling of ray casting in both parallel and fan-beam cases (see Figure 1.10), followed by a description of the reconstruction process itself.



(a) Parallel Geometry.

(b) Fan-Beam Geometry.

Figure 1.10: Illustration of 2D parallel and fan-beam geometries. The image to be reconstructed is a 4×4 pixel square. In the general case of circular acquisition, the source S and detector move together along a circle around the object. In this illustration, the projections are truncated, i.e. the image to be reconstructed is not in the FOV, which is not generally the case.

1.2.1 Parallel geometry

To model ray tracing in a parallel geometry, we will use the Radon transform. The Radon transform is a fundamental concept in X-ray tomography. It represents the line integral of the attenuation function. The Radon transform of a function μ for a line L in the plane is defined as:

$$\mathcal{R}\mu(L) = \int_{\mathbf{x}\in L} \mu(\mathbf{x})\,\mathrm{d}L. \tag{1.8}$$

This transform naturally corresponds to the monochromatic Beer-Lambert law explained earlier (see Eq. 1.4) when rays are parallel. Typically, the data acquisition is performed along a circular trajectory, allowing us to adopt a more suitable coordinate system. For a line $L_{\boldsymbol{\alpha},r}$ with a normal unit vector $\boldsymbol{\alpha}$ passing at a distance r from the origin, Equation 1.8 becomes:

$$\mathcal{R}\mu(L_{\boldsymbol{\alpha},r}) = \int_{\mathbf{y}\in\boldsymbol{\alpha}^{\perp}} \mu(r\boldsymbol{\alpha} + \mathbf{y})\mathrm{d}\mathbf{y}, \tag{1.9}$$

where $\boldsymbol{\alpha}^{\perp} = \{\mathbf{y} \in \mathbb{R}^2 | \mathbf{y}\cdot\boldsymbol{\alpha} = 0\}$. In 2D, there is another common notation which uses two orthogonal unit vectors set with $\varphi \in [0, \pi[$:

$$\boldsymbol{\alpha}_{\varphi} = (\cos\varphi, \sin\varphi), \tag{1.10}$$

$$\boldsymbol{\beta}_{\varphi} = (-\sin\varphi, \cos\varphi). \tag{1.11}$$

Then, the 2D Radon Transform can be rewritten in this more convenient expression:

$$\mathcal{R}\mu(\varphi, r) = \int_{\mathbb{R}} \mu(r\boldsymbol{\alpha_{\varphi}} + s\boldsymbol{\beta_{\varphi}})\mathrm{d}s. \tag{1.12}$$

This allows to simply define the projection p_{φ} at the angle φ by all the Radon Transforms, such that:

$$p_{\varphi}(r) = \mathcal{R}\mu(\varphi, r). \tag{1.13}$$

Figure 1.11 shows the described coordinates. The entirety of these lines yields what is termed the *sinogram*. It is called so because the Radon transform of a point will result in a sinusoidal. The Radon transform can also be expressed for the 3D case. In this scenario, we no longer consider a line integral but rather a plane integral with unit vector $\boldsymbol{\alpha}$ and a parameter r for the distance from the origin. To extend the Radon Transform, we will define:

$$\boldsymbol{\alpha_{\varphi,\theta}} = (\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta), \tag{1.14}$$

$$\boldsymbol{\beta_{\varphi,\theta}} = (-\sin\varphi, \cos\varphi, 0), \tag{1.15}$$

$$\boldsymbol{\gamma_{\varphi,\theta}} = (-\cos\varphi\cos\theta, -\sin\varphi\cos\theta, \sin\theta),\ \ \varphi \in [0, 2\pi[, \theta \in [0, \frac{\pi}{2}[. \tag{1.16}$$

The 3D Radon Transform can then be expressed as:

$$\mathcal{R}\mu(\varphi, \theta, r) = \iint_{\mathbb{R}^2} \mu(\boldsymbol{\alpha_{\varphi,\theta}} + s\boldsymbol{\beta_{\varphi,\theta}} + t\boldsymbol{\gamma_{\varphi,\theta}})\mathrm{d}s\mathrm{d}t. \tag{1.17}$$

Another fundamental concept that goes hand in hand with the Radon transform is the *backprojection*. Backprojection is the dual operator of the projection. It involves

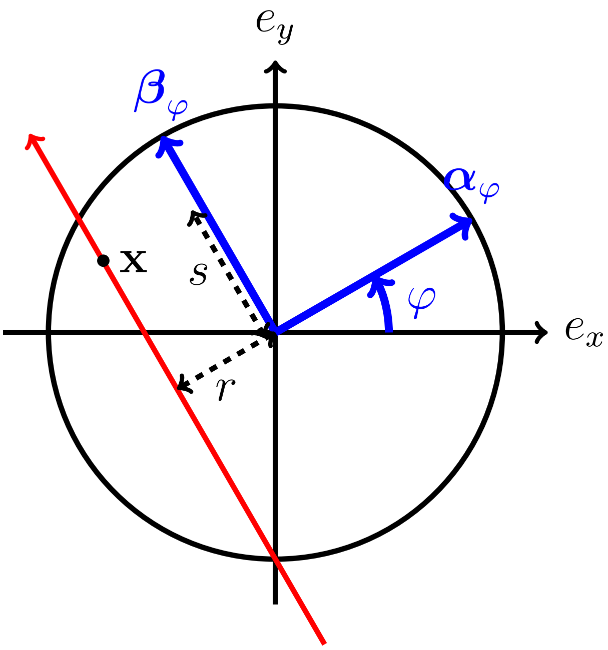


Figure 1.11: 2D coordinates representation for the Radon transform. A point $\mathbf{x}$ is described by its coordinates r and s. The angle of the projection is defined by φ.

projecting, along the lines, the Radon transforms' result. In 2D, the backprojection $\mathcal{B}p_\varphi$ of the projection at angle φ is:

$$\mathcal{B}p_\varphi(\mathbf{x}) = p_\varphi(\mathbf{x} \cdot \boldsymbol{\alpha}_\varphi). \tag{1.18}$$

Generally, the backprojection is defined by considering all projections around a circular trajectory. Thus, the backprojection is often written as:

$$\mathcal{B}p(\mathbf{x}) = \int_0^\pi p_\varphi(r)\Big|_{r=\boldsymbol{\alpha}_\varphi \cdot \mathbf{x}} \mathrm{d}\varphi. \tag{1.19}$$

This operation is essential because it is a core component in all reconstruction algorithms.

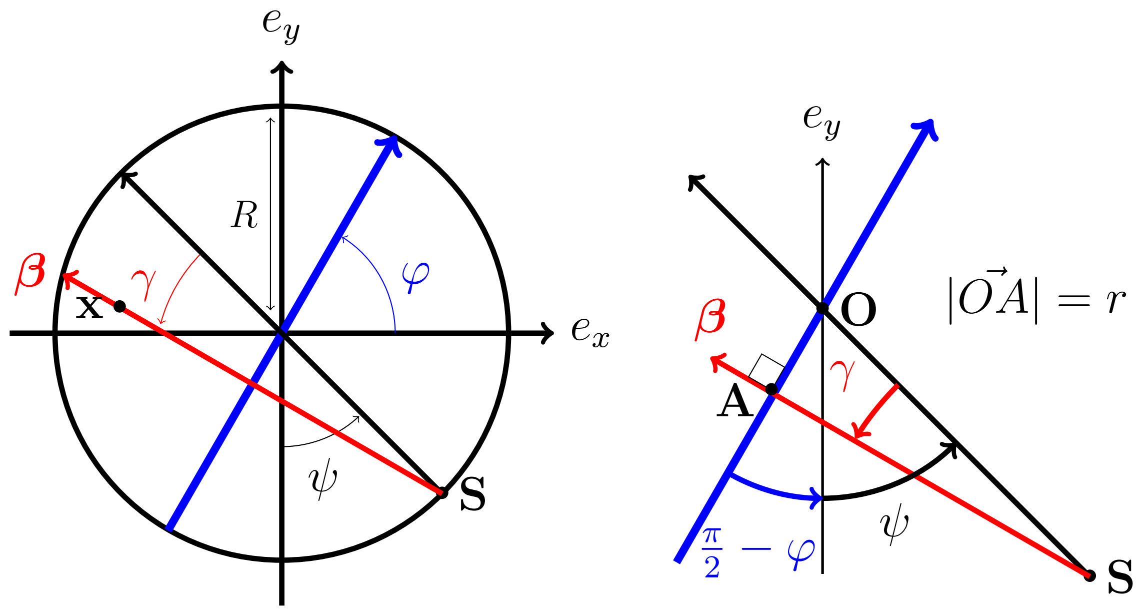
1.2.2 Fan-beam and Cone-beam geometry

The formalism of the previous section is adapted for parallel projections, where the emitted rays are parallel to each other. These projections are defined by an angle φ and a distance to the origin. However, in our industrial CT applications (unlike synchrotrons), the X-ray source is considered a point source (see Figure 1.10b). Therefore, it is appropriate to use different tools to model the projector. The Radon transform integrates over hyperplanes rather than lines, meaning that it is not appropriate for 3D applications. The X-ray transform $\mathcal{X}$, on the other hand, continues to sum over lines for higher dimensions, making it more suitable than Radon for fan-beam projections (this model is called *fan-beam* in 2D and *cone-beam* in 3D). In practice, there are two possible designs for fan-beam geometries: the equiangular fan-beam and the equal-spaced fan-beam. Here, we consider the equiangular fan-beam. In 2D, for a source position $\mathbf{S} \in \mathbb{R}^2$, the X-ray transform for a fan-beam geometry is defined as:

$$\mathcal{X}\mu(\mathbf{S}, \boldsymbol{\beta}) = \int_0^{+\infty} \mu(\mathbf{S} + s\boldsymbol{\beta})\,\mathrm{d}s, \quad \boldsymbol{\beta} \in \mathbb{U} \text{ (the unit circle)}. \tag{1.20}$$

Thus, the fan-beam projection is defined as:

$$p(\mathbf{S}, \boldsymbol{\beta}) = \mathcal{X}\mu(\mathbf{S}, \boldsymbol{\beta}). \tag{1.21}$$



(a) 2D coordinates for the fan-beam geometry and its rebinning.

(b) Relation between parallel and fan-beam geometry.

Figure 1.12: Equivalence between the 2D fan-beam and parallel geometry. The angle φ is the same as the one defined for the parallel beam. A relationship can be established between φ, γ, and ψ, which enables rebinning.

It is notable that, for a circular trajectory of radius R, the same line can be parameterised in both parallel and fan-beam geometries, leading to the equivalences shown in Figure 1.12. Since each straight line intersects the circle at two points, this change of coordinates is described by the following two relations:

$$\begin{cases} \varphi = \psi + \gamma \\ r = -R\sin\gamma \end{cases} \quad \text{or} \quad \begin{array}{l} \varphi = \psi + \gamma + \pi \\ r = R\sin\gamma \end{array} \tag{1.22}$$

If, in parallel geometry, a line is defined by its coordinates (φ, r), we can find its equivalent in fan-beam geometry (ψ, γ). This process is known as *rebinning*. Figure 1.12b illustrates these relations.

The principle of reconstruction algorithms by rearrangement into parallel data involves calculating the equivalent parallel sinogram by applying these variable transformation formulas and then reconstructing the associated image using the filtered back-projection algorithm described in the following section. This geometric change necessitates performing interpolation for each pair of parallel coordinates between the measured fan-beam coordinate pairs surrounding the desired values. In practice, this method is rarely used, instead the reconstruction formula is modified directly.

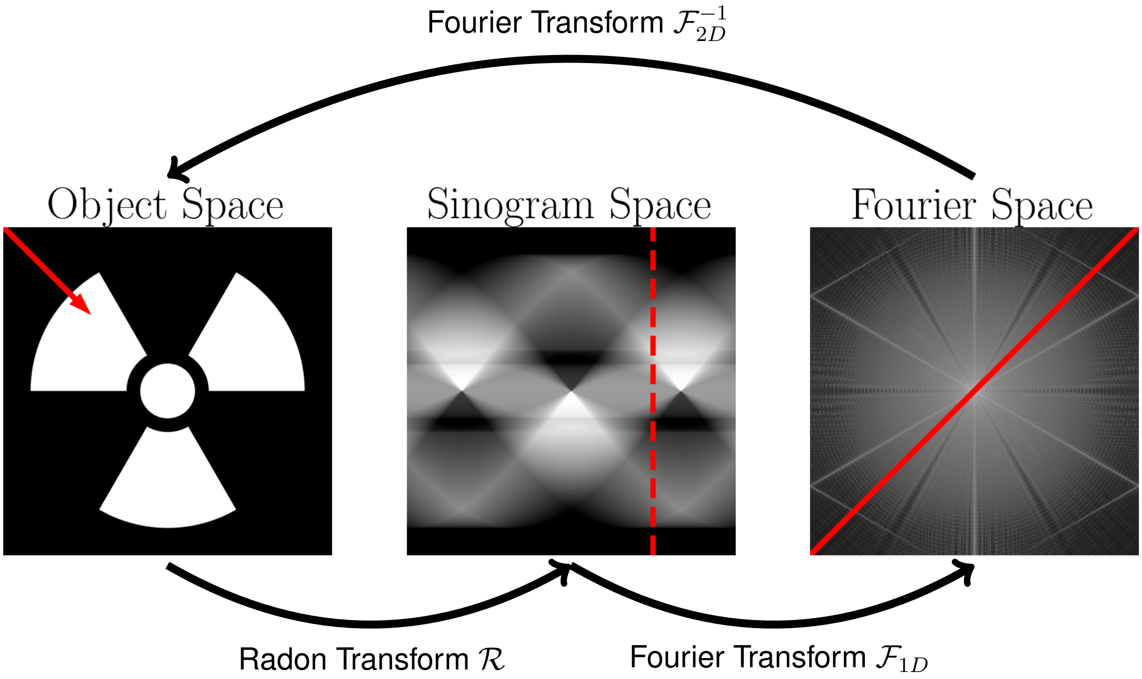


Figure 1.13: Fourier Slice Theorem. The emitted rays are parallel to the red arrow in the image within the object space. The Radon transform at this angle forms a column of the sinogram, shown here as a red dashed line. This transform also corresponds to a line passing through the centre in Fourier space. The central slice theorem connects these spaces and allows for the reconstruction of the object.

1.2.3 Analytical Reconstruction

We have described the projection and backprojection operators for parallel and fan-beam geometries. We will now demonstrate how these operators are utilised in the reconstruction algorithm itself.

Fourier Slice Theorem

The most famous theorem in tomography is undoubtedly the Fourier slice theorem. This theorem connects the Radon transform of an object to its Fourier transform. It states that, $\forall \rho \in \mathbb{R}, \forall \varphi \in [0, 2\pi[$:

$$\mathcal{F}_{1D}(\mathcal{R}\mu(\varphi, \rho)) = \mathcal{F}_{2D}(\mu(\rho \boldsymbol{\alpha}_{\varphi})), \tag{1.23}$$

where $\mathcal{F}_{1D}$ and $\mathcal{F}_{2D}$ are respectively the 1D and 2D Fourier transform, and rho the Fourier variable associated with r. In essence, this theorem provides a reconstruction method. The data processing diagram is illustrated in Figure 1.13. However, this method is rarely used because it suffers from numerous issues due to interpolations in the Fourier domain, making it often ineffective. The projections are arranged radially in Fourier space, but the inverse Fourier transform is generally implemented for a Cartesian grid. It will therefore be necessary to interpolate, which is difficult as depicted in Figure 1.14.

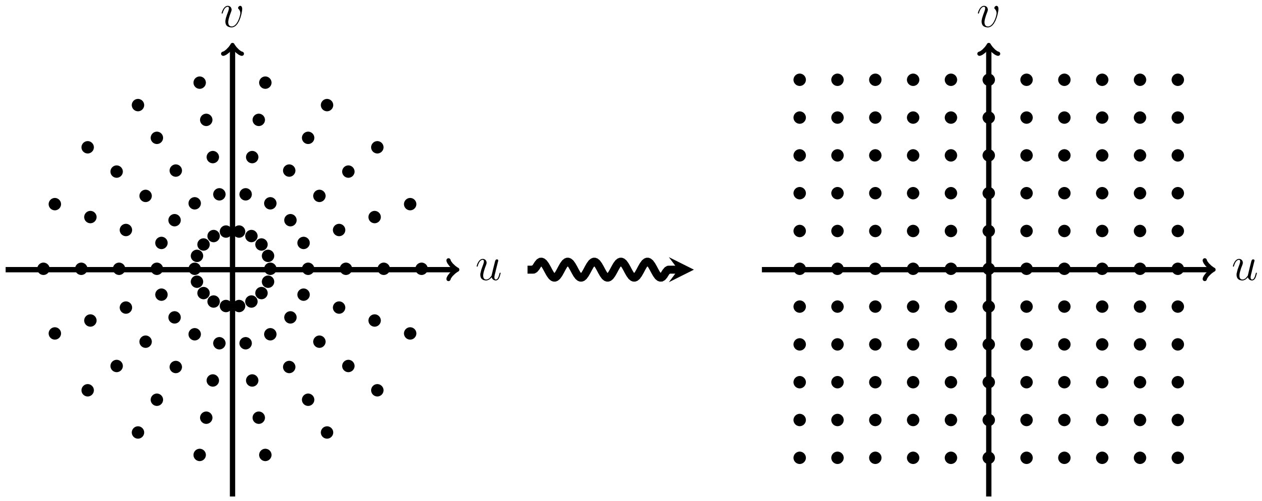


Figure 1.14: Polar to Cartesian interpolation in the Fourier Domain. During data acquisition, low frequencies are sampled more frequently than high frequencies. Additionally, the polar grid must be interpolated to convert it to a Cartesian grid, a step that introduces errors.

Filtered Back-Projection on Parallel Geometries

The Fourier Slice Theorem is a theoretical basis for reconstruction but, as we have pointed out, is never used directly. A commonly used algorithm for tomography is the Filtered Back-Projection (FBP). We will explain this method in the context of a circular trajectory with 2D parallel geometry.

Firstly, we recall that an image, represented by a function $\mu(x, y)$, can be recovered from its Fourier transform $\mathcal{F}_\mu(u, v)$ by the inverse Fourier transform:

$$\mu(x,y) = \int\int_{-\infty}^{+\infty} \mathcal{F}_\mu(u,v) e^{2\pi j(ux+vy)} \mathrm{d}u\mathrm{d}v. \tag{1.24}$$

To demonstrate how FBP works, we will switch from Cartesian coordinates (u, v) to polar coordinates. The purpose of this switch is to express the quantity $\mathcal{F}_\mu(u, v)$ in the form in which the data are naturally collected (the Fourier transform of each parallel projection falls on a polar grid). The coordinate transformation is given by:

$$\begin{cases} u = & \omega\cos(\theta), \\ v = & \omega\sin(\theta). \end{cases} \tag{1.25}$$

Which induces the following relation:

$$\mathrm{d}u\mathrm{d}v = \mathbf{J}\mathrm{d}\omega\mathrm{d}\theta = \begin{vmatrix} \frac{\partial u}{\partial\omega} & \frac{\partial u}{\partial\theta} \\ \frac{\partial v}{\partial\omega} & \frac{\partial v}{\partial\theta} \end{vmatrix} \mathrm{d}\omega\mathrm{d}\theta = \omega\mathrm{d}\omega\mathrm{d}\theta, \tag{1.26}$$

where $\mathbf{J}$ is the jacobian determinant. Now Equation 1.24 can be rewritten in the new coordinates system, which leads to:

$$\mu(x,y) = \int_{\theta=0}^{2\pi}\int_{\omega=0}^{+\infty} \mathcal{F}_\mu(\omega\cos\theta, \omega\sin\theta) e^{2\pi j\omega(x\cos\theta + y\sin\theta)} \mathrm{d}\omega\mathrm{d}\theta. \tag{1.27}$$

Then, using the Fourier slice theorem, we can write:

$$\mu(x,y) = \int_{\theta=0}^{2\pi}\int_{\omega=0}^{+\infty} \tilde{P}(\omega,\theta)e^{2\pi j\omega(x\cos\theta+y\sin\theta)}\omega \mathrm{d}\omega\mathrm{d}\theta \tag{1.28}$$

$$\begin{aligned} &= \int_{\theta=0}^{\pi}\int_{\omega=0}^{+\infty} \tilde{P}(\omega,\theta)e^{2\pi j\omega(x\cos\theta+y\sin\theta)}\omega \mathrm{d}\omega\mathrm{d}\theta \\ &+ \int_{\theta=0}^{\pi}\int_{\omega=0}^{+\infty} \tilde{P}(\omega,\theta+\pi)e^{-2\pi j\omega(x\cos\theta+y\sin\theta)}\omega \mathrm{d}\omega\mathrm{d}\theta. \end{aligned} \tag{1.29}$$

Here, $\tilde{P}(\omega,\theta)$ is the Fourier transform of the projection at angle θ. For the last equality, we used a convenient symmetry property among the projection samples for a parallel sampling geometry:

$$p(\epsilon,\theta+\pi) = p(-\epsilon,\theta), \tag{1.30}$$

$$\tilde{P}(\omega,\theta+\pi) = \tilde{P}(-\omega,\theta). \tag{1.31}$$

The entire expression can be rewritten as:

$$\mu(x,y) = \int_{\theta=0}^{\pi}\int_{\omega=-\infty}^{+\infty} \tilde{P}(\omega,\theta)e^{2\pi j\omega(x\cos\theta+y\sin\theta)}|\omega| \mathrm{d}\omega\mathrm{d}\theta. \tag{1.32}$$

The inner integral is the inverse Fourier transform of the quantity $\tilde{P}(\omega,\theta)|\omega|$. In the spatial domain, this represents a projection filtered by a function whose frequency domain response is $|\omega|$, and is therefore called a *filtered projection*. If we denote the filtered projection at angle θ by $g(r,\theta)$, represented by the inner integral of Equation 1.32, then:

$$g(r,\theta) = g(x\cos\theta+y\sin\theta) = \int_{-\infty}^{+\infty} \tilde{P}(\omega,\theta)e^{2\pi j\omega(x\cos\theta+y\sin\theta)}|\omega| \mathrm{d}\omega, \tag{1.33}$$

$$\mu(x,y) = \int_{0}^{\pi} g(x\cos\theta+y\sin\theta)\mathrm{d}\theta. \tag{1.34}$$

The variable $x\cos\theta + y\sin\theta$ is simply the signed distance of the point (x,y) to a line that goes through the origin of the coordinate system and forms an angle θ with respect to the x axis. The previous equation states that the reconstructed image $\mu(x,y)$ at location (x,y) is the summation of all filtered projection samples that pass through that point. Alternatively, we can focus on a particular filtered projection sample and examine its contribution to the reconstructed image. Since $(\theta, x\cos\theta + y\sin\theta)$ represents a straight line that overlaps the ray path that produces the projection sample, the intensity of $g(x\cos\theta + y\sin\theta)$ is added uniformly to the reconstructed image along the straight line. Consequently, the value of the filtered projection sample is painted or superimposed along the entire straight-line path [17].

The ramp filter can also be considered as the compositin of the Hilbert transform and the derivative (with a factor of $\pi/2$). This observation forms the basis of other reconstruction methods, such as Differentiated Backprojection [20]. Implementing FBP involves addressing several challenges. Since the backprojection implementation is a pixel-driven algorithm, it necessitates interpolating the projections (More details on the pixel-driven concept are given in Section 1.3.2). Moreover, transitioning into Fourier space to apply the filter and subsequently reverting to the spatial domain is inefficient. Ideally, these steps would be avoided; however, there is no convolution kernel that spatially corresponds to the ramp filter in the frequency domain, as the ramp filter diverges at infinity. Therefore, bounds are introduced where the filter will be zero below and beyond certain limits (resulting in an 'M'-shaped filter). Various filters can be utilised, created by adding what is known as a window function (for the 'M' shape, it is rectangular). The most well-known filters include Hanning, cosine, and sinc filters. The choice of filter and the cut-off frequency significantly impact the reconstruction [1, 17, 21].

Filtered Back-Projection on Fan-Beam Geometries

Filtered Back-Projection can also be effectively utilised in fan-beam geometries. This can be achieved either by rebinning the data to apply parallel FBP (Eq. 1.34) or by employing the variable transformation previously described within the FBP formula (Eq. 1.22). The direct algorithm processes each acquisition sequentially as soon as it is performed, thus circumventing the need for interpolation and the intermediate storage of rearranged data. Despite its advantages, the computational complexity increases due to the weighting factors introduced by the fan-beam geometry, which samples the volume differently from the parallel beam configuration. The numerical implementation of backprojection in fan-beam geometry can also introduce artefacts, such as aliasing at the edges of objects, resulting from undersampling [22].

FDK Algorithm

The FDK algorithm, named after its authors Feldkamp, Davis, and Kress, is specifically designed for three-dimensional cone-beam circular trajectories and utilises a filtered backprojection method, making it straightforward to implement. The FDK algorithm is a modified version of the fan-beam FBP algorithm and comprises three primary steps. Initially, the projections are pre-scaled by the cosine of the angle between the cone-beam ray and the central ray of the projection. Subsequently, a row-by-row ramp filter is applied to the data. Finally, a cone-beam backprojection of the filtered data is performed, incorporating a weighting function based on the distance from the reconstruction point to the focal point [3].

However, due to the circular trajectory's non-compliance with Tuy's condition (see Section 1.2.3), the algorithm yields only approximate reconstructions, which may introduce artefacts. These artefacts are particularly noticeable in regions distant from the orbit plane. Despite these limitations, the FDK algorithm is renowned for its practicality and robustness. The cone angle plays a crucial role in cone-beam imaging, with smaller angles (less than 10°) producing relatively accurate images. At the orbit plane, the algorithm achieves exact reconstructions [22].

Grangeat Method

For the three-dimensional case, P. Grangeat established the exact relationship between the 3D X-ray transform in cone-beam geometry and the first derivative of the 3D Radon transform [15]. Consequently, it is possible to compute the derivative of the 3D Radon transform and then apply filtered backprojection on the derivative. Grangeat's algorithm differs from traditional filtered backprojection algorithms and necessitates data rebinning, a process that can result in significant interpolation errors.

However, a more direct algorithm, as detailed in [23, 24], utilises Grangeat's approach without rebinning but with appropriate weighting. This algorithm offers a more streamlined procedure with reduced data storage requirements. Unlike the two-step method, there is no need to calculate and store Radon transforms (in the form of a 3D matrix) before performing the backprojection. The data processing becomes sequential, partly due to the weighting function that manages redundancies in the cone-beam projections.

Data-sufficiency conditions

To achieve an exact reconstruction, it is necessary to obtain all the sufficient information. There are sampling conditions for exact reconstruction in parallel geometry (Orlov) and cone-beam geometry (Tuy) [15]. The Orlov condition is a criterion in parallel 3D geometry that ensures an exact reconstruction. If we assume that parallel projections are measured for a set of directions $\{\boldsymbol{\alpha}_i\}_i$, which is a subset of the unit sphere $\mathcal{S}^2$, and that all lines parallel to $\boldsymbol{\alpha}_i$ are measured for each direction $\boldsymbol{\alpha}_i$ (meaning the 2D projection is untruncated), then the Fourier slice theorem enables the calculation of the 3D Fourier transform of $\mu(\mathbf{x})$ for a frequency $\boldsymbol{\nu} \in \mathbb{R}^3$ from a 2D projection in any direction $\boldsymbol{\alpha}_i$ that satisfies $\boldsymbol{\alpha}_i \cdot \boldsymbol{\nu} = 0$. The sufficiency condition for stable reconstruction of a unique image μ is met if and only if all great circles on the unit sphere have a non-empty intersection with the set of measured directions $\{\boldsymbol{\alpha}_i\}_i$.

In fan-beam geometry, the sufficiency condition for data to fill the Fourier space is that every line intersecting the object of interest must contain a fan-beam focal point (i.e., the X-ray source). The Tuy condition is quite similar but applies to cone-beam geometries: every plane intersecting the object of interest must contain a cone-beam focal point. The Tuy condition is not satisfied for planar trajectories in 3D objects, such as the circular trajectory (even though this trajectory is still widely used for approximate reconstructions of satisfactory quality) [3, 15].

1.3 Iterative reconstruction methods

1.3.1 Discrete Projectors and Back-Projectors

In this section, we will adopt a different formalism to describe and solve the tomographic problem. This approach will involve a discrete description, akin to the one used for the Beer-Lambert equation (Eq.1.5). For each ray, if the contribution of each pixel/voxel to be reconstructed in the attenuation can be defined, one can formulate the tomographic problem as a linear system:

$$\mathbf{p} = \mathbf{A}\boldsymbol{\mu} + \mathbf{n}, \tag{1.35}$$

where $\mathbf{p}$ represents the experimental projections processed as explained in Section 1.1.3, $\boldsymbol{\mu}$ is the vector representing the flattened voxel grid to be reconstructed, $\mathbf{n}$ is the vector accounting for noises, and $\mathbf{A}$ is the projection operator. Its adjoint, $\mathbf{A}^\mathsf{T}$, represents the backprojection operator. These operators model the interaction between rays and matter, from a geometrical perspective. The projection operator is a discretised form of

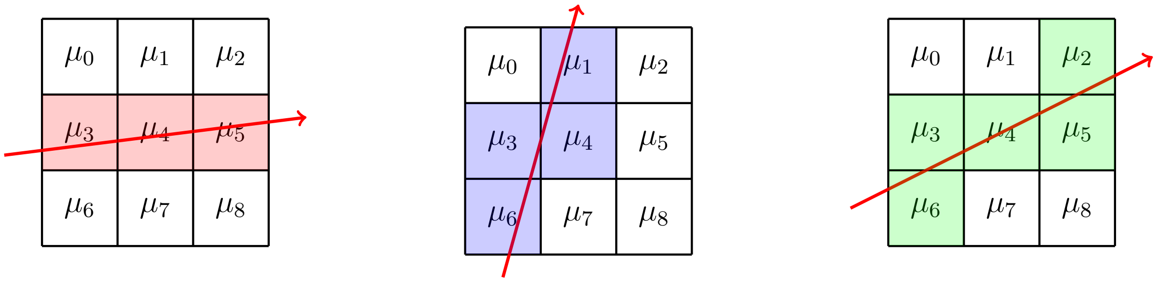


Figure 1.15: Illustration of three rays on a 2D image composed of 9 pixels. The system matrix corresponding to the interactions of these rays on this image is given in Eq. 1.36.

the X-ray transform. Mathematically, the projection operator is described as a matrix where the number of rows corresponds to the number of rays, and the number of columns corresponds to the number of voxels to be reconstructed. Each row contains coefficients indicating the contribution of each voxel to the value of the projection for a specific ray, while each column can be analysed through backprojection and indicates how rays go through a specific voxel. Figure 1.15 illustrates the passage of three rays through an image composed of nine pixels. Each pixel represents an unknown in the linear system to be solved. The system matrix depends solely on the acquisition geometry and the parameters of the pixel/voxel grid to be reconstructed, but not on the object itself. Iterative methods can therefore be used in parallel and cone-beam configurations, as well as on arbitrary trajectories. The system matrix associated with these rays is given by Equation 1.36:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & \ell_{\mu_3} & \ell_{\mu_4} & \ell_{\mu_5} & 0 & 0 & 0 \\ 0 & \ell_{\mu_1} & 0 & \ell_{\mu_3} & \ell_{\mu_4} & 0 & \ell_{\mu_6} & 0 & 0 \\ 0 & 0 & \ell_{\mu_2} & \ell_{\mu_3} & \ell_{\mu_4} & \ell_{\mu_5} & \ell_{\mu_6} & 0 & 0 \end{pmatrix}. \tag{1.36}$$

The coefficients ℓ_{μ_i} reflect the contribution of each pixel to the attenuation. In the classical case, the matrix is highly sparse and ill-conditioned. The inversion of $\mathbf{A}$ is generally non-trivial or even infeasible, given that it is nearly singular and of substantial size. Additionally, Equation 1.35 can only be solved approximately due to the presence of noise in the data, precluding an exact solution. Generally, the matrix $\mathbf{A}$ is designed to be over-determined for a full dense trajectory acquisition. However, in the sparse-view strategy, the number of projections is relatively low. In all cases, the optimisation problem is:

$$\boldsymbol{\mu}^* \in \underset{\boldsymbol{\mu}}{\operatorname{argmin}} \|\mathbf{A}\boldsymbol{\mu} - \mathbf{p}\|_2^2, \tag{1.37}$$

that adheres to the least-squares criterion. This corresponds to the Moore-Penrose pseudo-inverse, theoretically computable via singular value decomposition. Practically, however, this method is only viable for very small dimensions of $\mathbf{A}$. The reconstruction process can also be formulated in a Bayesian framework of maximising the *a posteriori* probability:

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\mu}}{\operatorname{argmax}} P(\boldsymbol{\mu}|\mathbf{p}). \tag{1.38}$$

Based on the Bayes rule, this probability can be expressed in the following form:

$$P(\boldsymbol{\mu} \mid \mathbf{p}) = \frac{P(\mathbf{p} \mid \boldsymbol{\mu})P(\boldsymbol{\mu})}{P(\mathbf{p})}. \tag{1.39}$$

Usually, no *a priori* information on the projections is available, and the optimisation problem is reformulated:

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\mu}}{\operatorname{argmin}}\big(-\ln P(\mathbf{p} \mid \boldsymbol{\mu}) - \ln P(\boldsymbol{\mu})\big). \tag{1.40}$$

The first term is called the *log-likelihood* or the *data fidelity term*. The second term reflects our *a priori* knowledge about the reconstruction; it is the *regularisation* term. In practice, due to miscounts by the detector cells and other uncertainties, the photon count is modelled by a Poisson process with a mean given by the Beer-Lambert law:

$$I_i = \mathcal{P}(I_0 e^{-[\mathbf{A}\boldsymbol{\mu}]_i + \mathbf{n}_i}), \tag{1.41}$$

where $\mathbf{n}_i$ represents a Gaussian noise. By incorporating this distribution into the log-likelihood, we can reformulate the tomographic problem (Eq 1.35) in its classical least-squares form. Ultimately, the formulation:

$$\boldsymbol{\mu}^* = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} \|\mathbf{A}\boldsymbol{\mu} - \mathbf{p}\|_W^2 + \lambda \mathbf{R}(\boldsymbol{\mu}) \tag{1.42}$$

will be the most frequently employed. The factor λ manages the trade-off between data fidelity and regularisation $\mathbf{R}$. The norm $\|\cdot\|_W$ represents the Euclidean norm weighted according to the variance in the measurements. Depending on the nature of $\mathbf{R}$, the complexity of solving the problem can vary significantly.

1.3.2 Operator Implementation Details

There are several methods to model and implement the projection and its adjoint. The first family of methods is called *pixel-driven* or *voxel-driven*. Each pixel or voxel is simply projected onto the detector, with values being summed and interpolated along the detector (see Figure 1.16). This approach is straightforward to understand but can induce artefacts. In the fan-beam geometry, some detector elements might be updated more frequently than their neighbours because the pixel ray path is not always perpendicular to the detector, leading to non-uniform interaction of ray paths with detector elements (see Figure 1.18b). This can result in the fan-beam geometry being less accurate and more likely to introduce artefacts compared to parallel-beam geometry [25, 26]. The pixel-driven method is more appropriate for backprojection because it considers all voxels, but it can lead to artefacts in projection if the voxels are too large compared to the detector elements.

Another approach is *ray-driven*, where each detector element determines a ray. The length traversed in each pixel is calculated and contributions are summed (see Figure 1.17a). There are several variants of ray-driven methods, including grid sampling methods that sample the voxel grid (see Figure 1.17b). In this version, the sampling step is crucial. There are also area-integrated methods that consider areas or volumes traversed (see Figure 1.17c). These methods are more "physical" but computationally intensive. In general, ray-driven techniques have limited cache utilisation in parallel systems, impacting computational efficiency. The ray-driven version is better suited for direct projection but can induce artefacts in backprojection if the detector resolution is too fine compared to the voxel grid (see Figure 1.18a) [27, 28, 29, 30].

There is also the *distance-driven* method, where ray-voxel and detector intersections are projected onto a mid-plane and values are accumulated there. The distance-driven

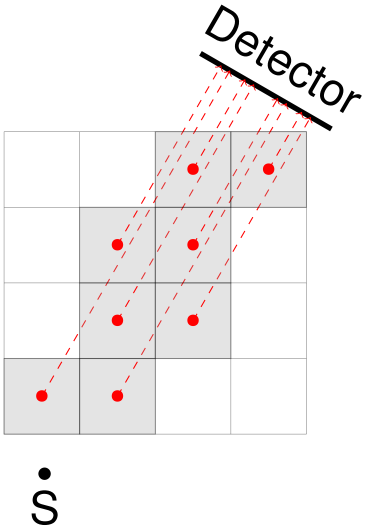


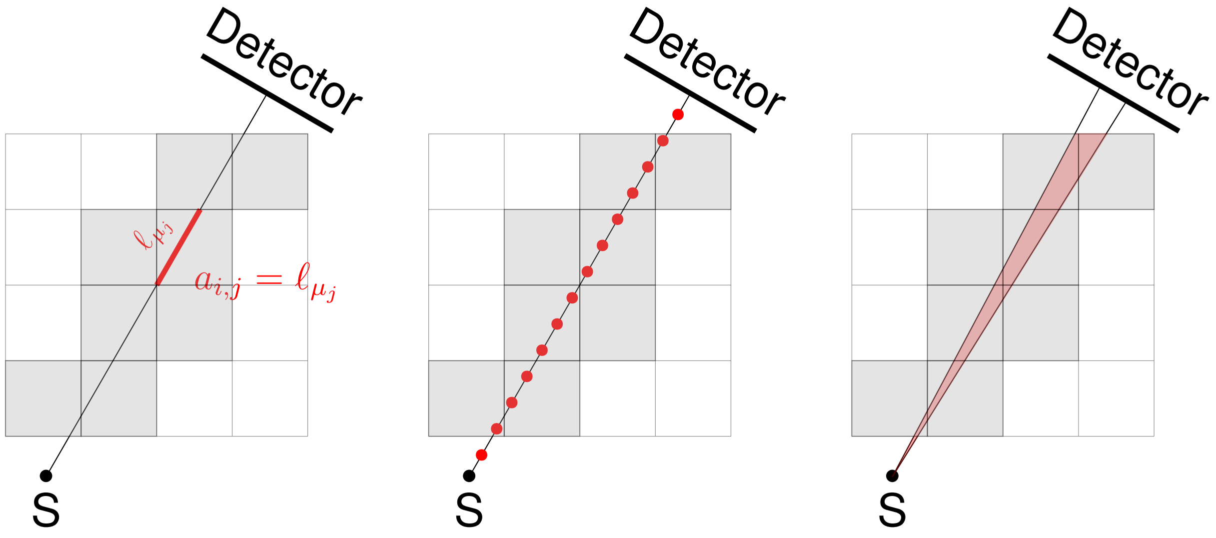
Figure 1.16: Pixel-driven strategy for projection

method provides a remedy for artefacts and retains simplicity in implementation, making it highly suitable for hardware integration [31]. Another approach is the separable footprints technique, which approximates the voxel footprint on the detector for speed-up. According to the authors, it is more accurate than the distance-driven projection and faster than both voxel-driven and distance-driven methods.

Following our discussion on the various aspects of implementations, it may be desirable to use different implementations for the projection and backprojection. There is no general rule regarding the results of reconstructions using an unmatched projector $\mathbf{A}$/backprojector $\mathbf{B}^\mathsf{T}$ pair. Sometimes, it can even yield better results in terms of artefact reduction. In an industrial context, the determining criterion for the choice of operators is often the computational time. When an unmatched projector/backprojector pair is employed in an iterative algorithm, the result depends on both the projector and backprojector. The unmatched pair solves the system $\mathbf{B}^\mathsf{T}\mathbf{A}\boldsymbol{\mu} = \mathbf{B}^\mathsf{T}\mathbf{p}$. Note that $\mathbf{B}^\mathsf{T}\mathbf{A}$ is not necessarily symmetric. The extent to which these differences affect the reconstructed image is problem-specific [22].

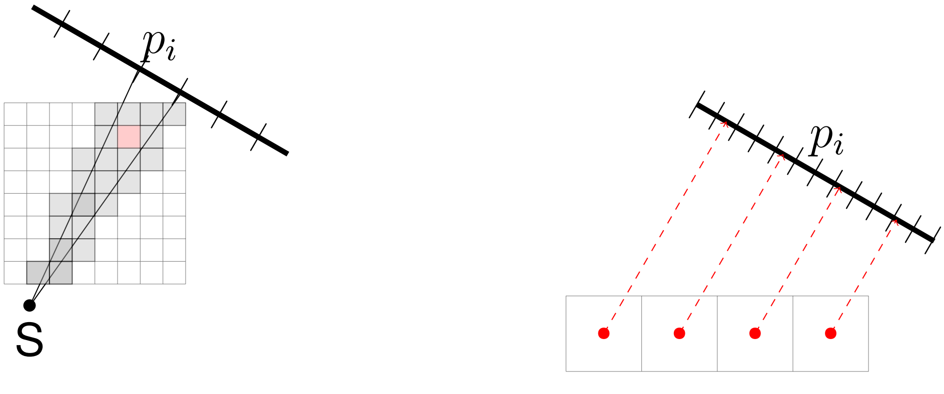
1.3.3 Reconstruction Algorithm

There are numerous algorithms available for solving the tomographic problem. The most well-known are collectively referred to as *Landweber* algorithms. Conjugate gradient methods are also commonly employed. When the regularisation term is not smooth, proximal algorithms can be utilised; these are becoming increasingly popular. The tomography problem is a classical linear optimisation problem, but due to its large size, the system matrix is never assembled. Optimisation methods must therefore be adapted to this constraint.



(a) Siddon (b) Grid sampling (c) Area integrated

Figure 1.17: Different strategies for ray-driven implementation



(a) Ray-driven limitations (b) Pixel-driven limitations

Figure 1.18: Illustration of limitations according to strategies. The size of voxels and detector pixels must be relevant. In the first case, the contribution of the red voxel is neglected. In the second case, some detector pixels are not illuminated.

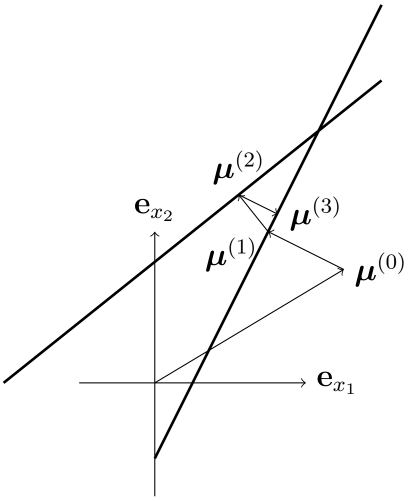


Figure 1.19: Kacmarz's method. The vector $\boldsymbol{\mu}$ is projected onto the hyperplane generated by each row of the system matrix (here two lines). This operation is repeated until convergence. A complete iteration occurs when all the hyperplanes have been used. Several iterations are required.

Landweber methods: ART, SART & SIRT

The *Landweber* algorithms are the most popular for iterative X-ray reconstruction. They are based on the Landweber iteration, which consists of applying the adjoint operator $\mathbf{A}^{\mathsf{T}}$ (or conjugate in complex numbers) to the residual $(\mathbf{A}\boldsymbol{\mu} - \mathbf{p})$, which can be interpreted as a gradient descent. The differences between the methods in this family arise from variations in the definition of the operator and its implementation. The most popular methods (ART, SART, SIRT) are described below.

The ART method (Algebraic Reconstruction Technique) was the first iterative approach proposed for X-ray tomographic reconstruction [21, 32]. Stemming from the work of Gordon et al. [33], it is essentially an adaptation of Kaczmarz's minimisation scheme to the tomographic problem. As a starting point, an initial image, $\boldsymbol{\mu}^{(0)}$, is required to initiate the iteration. This image could be derived from an analytical reconstruction, for example. However, any random vector serves the purpose equally well. This vector is projected perpendicularly onto the hyperplane generated by the first equation in the system matrix $\mathbf{a_1}$, representing the first X-ray, $\mathbf{p}_1$, to obtain a new and improved image, $\boldsymbol{\mu}^{(1)}$. This image is then projected perpendicularly onto the second equation, yielding an image that is improved with respect to $\boldsymbol{\mu}^{(1)}$ because $\boldsymbol{\mu}^{(2)}$ lies closer to the intersection point of the straight lines than its two predecessors. A relaxation factor λ is used for controlling the updates. The iteration formula for the projection on the i^{th} hyperplane is given by:

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \lambda \frac{\mathbf{a_i}\boldsymbol{\mu}^{(k)} - \mathbf{p}_i}{\|\mathbf{a_i}\|^2} \mathbf{a_i}^{\mathsf{T}}. \tag{1.43}$$

In practice, for each complete iteration of the algorithm, the data must be projected at least once onto all hyperplanes. Figure 1.19 shows the procedure on a 2D space with two equations in the system matrix. In practical cases, due to noise, the intersection of hyperplanes formed by each equation never occurs precisely at a single point. Unfortunately, in each iteration of ART, only a single ray is considered, which updates very few elements of $\boldsymbol{\mu}$ at each step, rendering the procedure extremely slow. To expedite convergence, the SART (Simultaneous ART) method proposes using all rays of a projection in each iteration. In this approach, most elements of $\boldsymbol{\mu}$ are updated. SART is significantly faster; however, like ART, the convergence speed depends on the order in which the data are processed, as two adjacent views or rays typically provide similar information.

One can advance even further by updating the reconstruction with multiple projections. This approach, known as OS-SART (Ordered Subset - SART), uses a set of projections at each iteration. While convergence is indeed faster, it demands more memory resources since the experimental projections need to fit into the RAM. If all projections are used, the technique is referred to as SIRT (Simultaneous Iterative Reconstruction Technique) [34].

We consider the system matrix $\mathbf{A}$, which accounts for all projections. A single iteration of SIRT can thus be expressed as:

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \lambda \frac{\mathbf{A}\boldsymbol{\mu}^{(k)} - \mathbf{p}}{\|\mathbf{A}\|^2} \mathbf{A}^\mathsf{T}. \tag{1.44}$$

To align closer with practical implementation, the SIRT algorithm can also be formulated as follows:

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \lambda \mathbf{N_c} \mathbf{A}^\mathsf{T} \mathbf{N_r} (\mathbf{p} - \mathbf{A}\boldsymbol{\mu}^{(k)}), \tag{1.45}$$

which further illustrates the normalisation step implementation. Here, $\mathbf{N_r}$ and $\mathbf{N_c}$ are diagonal matrices defined as follows:

$$\mathbf{N_r}_{ii} = \frac{1}{\sum_j \mathbf{A}_{ij}}, \tag{1.46}$$

$$\mathbf{N_c}_{jj} = \frac{1}{\sum_i \mathbf{A}_{ij}}. \tag{1.47}$$

In practice, $\mathbf{N_r}$ and $\mathbf{N_c}$ matrices are not assembled. Unitary coefficients are used for projection and backprojection to compute them.

Conjugate Gradient Least Squares

The conjugate gradient algorithm is one of the most commonly used gradient descent methods. Using the steepest descent approach, it appears that successive descent directions follow more or less the same general orientation. Consequently, it would be more advantageous to obtain, for each iteration, a more favourable direction to find a more efficient path to the desired minimiser. The conjugate gradient has been created with this idea, it makes a Krylov space with the searching directions at each iteration of the procedure to fasten the convergence.

The conjugate gradient algorithm is often derived for a positive symmetric matrix. To apply it to a CT system, we can simply apply the method to the normal equations:

$$\mathbf{A}^\mathsf{T}\mathbf{A}\boldsymbol{\mu} = \mathbf{A}^\mathsf{T}\mathbf{p}. \tag{1.48}$$

The algorithm resulting from the application of conjugate gradient to normal equations is called conjugate gradient least squares (CGLS).

Algorithm 1: Conjugate Gradient Least Squares (CGLS)

```
Data: Initial guess $\boldsymbol{\mu}^{(0)}$
Result: Reconstruction $\tilde{\boldsymbol{\mu}}$
$k = 0$
$\mathbf{r}^{(0)} = \mathbf{p} - \mathbf{A}\boldsymbol{\mu}^{(0)}$
$\mathbf{d}^{(0)} = \mathbf{s}^{(0)} = \mathbf{A}^{\mathsf{T}}\mathbf{r}^{(0)}$
$\gamma^{(0)} = \|\mathbf{s}^{(0)}\|^2$
while not convergence and not max iterations reached do
    $k = k + 1$
    $\mathbf{q}^{(k)} = \mathbf{A}\mathbf{d}^{(k)}$
    $\alpha^{(k)} = \gamma^{(k)}/\|\mathbf{q}^{(k)}\|^2$
    $\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)}$
    $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} + \alpha^{(k)}\mathbf{q}^{(k)}$
    $\mathbf{s}^{(k+1)} = \mathbf{A}^{\mathsf{T}}\mathbf{r}^{(k+1)}$
    $\gamma^{(k+1)} = \|\mathbf{s}^{(k+1)}\|^2$
    $\beta^{(k)} = \gamma^{(k+1)}/\gamma^{(k)}$
    $\mathbf{d}^{(k+1)} = \mathbf{s}^{(k)} + \beta^{(k)}\mathbf{d}^{(k)}$
end
```

Maximum Likelihood Expectation Maximization

The Maximum Likelihood Expectation Maximization algorithm (MLEM) [35] was originally developed for reconstruction in emission tomography but has also been adapted for attenuation tomography. In our context, the likelihood L of obtaining $\mathbf{p}$ given $\boldsymbol{\mu}$ is defined as:

$$L(\boldsymbol{\mu}) = -\ln P(\mathbf{p}|\boldsymbol{\mu}) = \sum_i \left(\mathbf{p}_i[\mathbf{A}\boldsymbol{\mu}]_i + I_0 e^{-[\mathbf{A}\boldsymbol{\mu}]_i}\right). \tag{1.49}$$

Since $\boldsymbol{\mu}$ is unknown, the EM algorithm introduces a latent variable Z representing the contribution of voxel j in the volume to the value of pixel i in the detector. Thus, $\mathbf{p}_i = \sum_j z_{ij}$. $\boldsymbol{\mu}$ becomes a parameter of the random variable Z, and the goal is to find the parameter value that maximises the probability of obtaining the data $\mathbf{p}$. The EM algorithm maximises the likelihood Q of a random variable depending on a parameter and unobserved data in two steps. First, it estimates the log-likelihood for a given parameter $\boldsymbol{\mu}$, known as the expectation step:

$$Q(\boldsymbol{\mu}|\boldsymbol{\mu}^{(k)}) = \mathbb{E}_Z\left[L(\boldsymbol{\mu})|\boldsymbol{\mu}^{(k)}\right], \tag{1.50}$$

where $\boldsymbol{\mu}^{(k)}$ denotes the current reconstruction. This step compensates for the incomplete knowledge of the variable Z by estimating the expectation of its likelihood with an arbitrary parameter $\boldsymbol{\mu}$ initially. Once Q is estimated, the algorithm maximises its value with respect to $\boldsymbol{\mu}$, known as the maximization step:

$$\boldsymbol{\mu}^{(k+1)} = \underset{\boldsymbol{\mu}}{\operatorname{argmax}}\, Q(\boldsymbol{\mu}|\boldsymbol{\mu}^{(k)}). \tag{1.51}$$

The MLEM algorithm is not widely used in CT due to its sensitivity to noise. Artefacts accumulate throughout the iterations, and determining the optimal number of iterations, which is often high, is challenging. Moreover, convergence rates vary across different areas of the image.

Alternating Directions Methods of Multipliers

The Alternating Directions Methods of Multipliers (ADMM), also known as *Douglas-Rachford splitting*, is a convenient algorithm for solving constrained optimisation problems. ADMM is a variant of the augmented Lagrangian scheme that uses partial updates for the dual variables. ADMM is made to solve the typical problem:

$$\min_{\boldsymbol{\mu}} f(\boldsymbol{\mu}) + \lambda g(\boldsymbol{\mu}). \tag{1.52}$$

where f, $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex functions, λ is a positive scalar to control the trade-off between data fidelity and regularisation. The functions f and g can both not be smooth, but usually, f in tomography application represents the fidelity term and g a non-smooth regularisation term. However, by choosing the following equivalent formulation:

$$\min_{\mathbf{x},\mathbf{y}} f(\mathbf{x}) + \lambda g(\mathbf{y}), \quad \text{subject to} \quad \mathbf{x} = \mathbf{y}, \tag{1.53}$$

the objective function is separable in $\mathbf{x}$ and $\mathbf{y}$. This property is particularly interesting for large dimensions problems and complex regularisation terms. The algorithm is summed up in Algorithm 2.

Algorithm 2: ADMM

```
Data:   Initial guess $\boldsymbol{\mu}^0$, $\lambda > 0$
Result: Reconstruction $\tilde{\boldsymbol{\mu}}$
Initialize $\mathbf{z}^0 = \boldsymbol{\mu}^0, \mathbf{u}^k = \mathbf{0}$
while the convergence criteria have not been met do
    $\boldsymbol{\mu}^{k+1} = \mathbf{prox}_{\lambda f}(\mathbf{z}^k - \mathbf{u}^k)$
    $\mathbf{z}^{k+1} = \mathbf{prox}_{\lambda g}(\boldsymbol{\mu}^{k+1} - \mathbf{u}^k)$
    $\mathbf{u}^k = \mathbf{u}^k + \boldsymbol{\mu}^{k+1} - \mathbf{z}^{k+1}$
end
```

The proximal operator is an operator associated with a proper, lower semi-continuous convex function f from a Hilbert space $\mathcal{X}$ to $[-\infty, +\infty]$, and is defined by:

$$\mathbf{prox}_f(v) = \operatorname*{argmin}_{x \in \mathcal{X}} \left(f(x) + \frac{1}{2} \|x - v\|^2_{\mathcal{X}} \right). \tag{1.54}$$

Primal-Dual Hybrid Gradient Algorithm

The Chambolle-Pock method, also known as the Primal-Dual Hybrid Gradient (PDHG) method, stands out as a robust and efficient approach for solving a diverse range of constrained and non-differentiable optimisation problems. In contrast to methods like the ADMM, PDHG typically avoids costly minimisation sub-steps. The essence of the PDHG algorithm lies in its alternating strategy: it iteratively updates a dual variable $\mathbf{y}$ using a gradient-like ascent, while concurrently updating a primal variable $\boldsymbol{\mu}$ through a gradient-like descent. This dual-primal interplay is further enhanced by an over-relaxation step applied to the primal variable, which can accelerate convergence in practice. The method's versatility and computational efficiency make it particularly well-suited for problems involving large-scale data, sparse solutions, and regularisation constraints across various fields of computational science and engineering [36, 37]. PDHG was formulated to solve the minimisation problem:

$$\min_{\boldsymbol{\mu}} f(\mathbf{K}\boldsymbol{\mu}) + g(\boldsymbol{\mu}), \tag{1.55}$$

where $\mathbf{K}$ is a linear operator, and the functions f and g are as defined previously. The procedure is outlined in Algorithm 3. Here, $\mathbf{K}^*$ denotes the adjoint of $\mathbf{K}$, and f^* represents the Fenchel conjugate of f.

Algorithm 3: Primal Dual Hybrid Gradient

```
Data:   Initial guess $\boldsymbol{\mu}^0$, $\tau > 0$, $\sigma > 0$, $\theta \in [0, 1]$
Result: Reconstruction $\tilde{\boldsymbol{\mu}}$
Initialize $\bar{\boldsymbol{\mu}}^0 = \mathbf{y}^0 = \boldsymbol{\mu}^0$
while the convergence criteria have not been met do
    $\mathbf{y}^{k+1} = \mathbf{prox}_{\sigma f^*}(\mathbf{y}^k + \sigma \mathbf{K} \bar{\boldsymbol{\mu}}^k)$
    $\boldsymbol{\mu}^{k+1} = \mathbf{prox}_{\tau g}(\boldsymbol{\mu}^k - \tau \mathbf{K}^* \mathbf{y}^{k+1})$
    $\bar{\boldsymbol{\mu}}^{k+1} = \boldsymbol{\mu}^k + \theta(\boldsymbol{\mu}^{k+1} - \boldsymbol{\mu}^k)$
end
```

1.4 Reconstruction quality

The assumptions made to enable the formulation of the tomographic problem and the measurement noise lead to the presence of artefacts in the reconstructed image, of which the most well-known are presented here. Additionally, this section introduces metrics for evaluating the quality of the images. There are numerous sources of noise and artefacts. We will list the most significant ones and illustrate them in Figure 1.21. Then, we will list the most popular metrics and use them to evaluate different reconstruction algorithms.

1.4.1 Noises and artefacts

As stated before, beam hardening occurs because X-ray beams are polychromatic. When X-rays pass through an object, the lower-energy photons are more likely to be absorbed by the material, while the higher-energy photons tend to pass through with less attenuation. As the X-rays traverse the object, they lose these low-energy photons, causing the X-ray beam to become *harder* i.e. more penetrating. Beam hardening artefacts create an overestimation of the attenuation at the edges and an underestimation at the centre of the object. This phenomenon, referred to as the *cupping artefact*, is represented in Figure 1.20. It can be negligible when the object's absorption rate does not vary much within the X-ray energy distribution or when monochromatic X-rays are used.

Compton scattering is a fundamental interaction between X-ray photons and electrons within the atoms of an object being imaged. When high-energy X-ray photons interact with electrons, they can transfer some of their energy to these electrons, causing them to scatter in different directions. This phenomenon is called Compton scattering. Artefacts occur because scattered X-ray photons do not follow the expected straight-line path from the X-ray source to the detector. Instead, they deviate from this path due to the scattering process. When these scattered photons are detected, they can be mistakenly interpreted

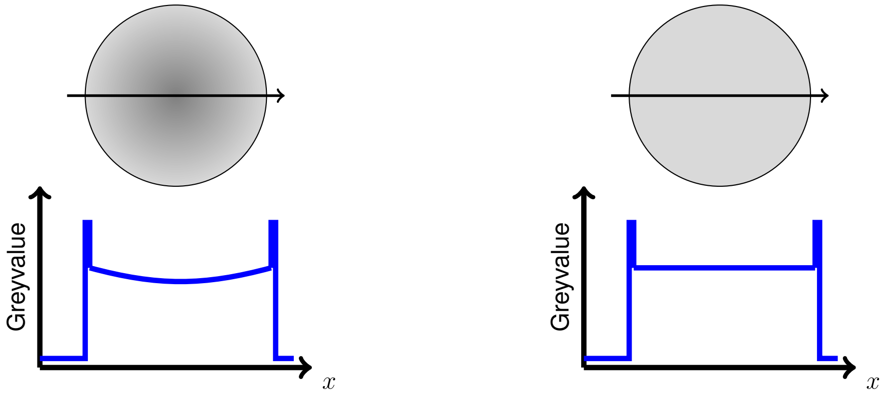


Figure 1.20: Cupping artefact due to beam hardening. On the left, the centre of the object is underestimated due to the artefact. On the right, the object's value is well determined.

as originating from a different location in the object being imaged. This misinterpretation can lead to inaccuracies in the reconstructed image and the appearance of artefacts.

Undersampling, i.e. acquiring an insufficient number of projections, is the primary cause of artefacts in sparse-view strategies. This often results in pronounced streaks radiating from the centre of the image. Acquiring additional projections helps to smooth out these artefacts and enhances the overall reconstruction quality.

Cone beam artefacts occur when using a cone beam geometry, particularly in scenarios where the object extends beyond the field of view of the cone beam. These artefacts manifest as distortions at the extremities and streaks in the reconstructed images, often due to incomplete or inconsistent data sampling. Cone beam artefacts can degrade image quality and obscure important details. Accurate calibration and sophisticated reconstruction algorithms, such as iterative reconstruction techniques, are essential to mitigate cone beam artefacts.

A motion artefact occurs when the structure of the object being inspected changes during acquisition. If the attenuation coefficient varies over time, data measurements will be inconsistent. This type of artefact can appear in medical imaging due to patient movement, but correction techniques are available. In most industrial systems, the inspected objects are typically immobile and stable on a suitable support. However, in robotic cells where robotic arms are used to move the source and detector, errors can occur due to incorrect positioning of the arms. Proper position calibration of the source orientation is essential; otherwise, projection realignment will be necessary. This issue will be discussed in detail in subsequent chapters.

Ring artefacts occur in CT when there are inconsistencies or defects in the detector elements. These artefacts appear as concentric rings centred around the rotational axis of the scanner. Ring artefacts are particularly problematic because they can obscure critical details in the reconstructed images. Ring artefacts may result from faulty or miscalibrated detector elements. Regular maintenance and calibration of the CT system are essential to minimise these artefacts.

Noise can also occur on projections prior to reconstruction, with serious consequences for the latter. The noise of an X-ray detector is a combination of several types of noises: quantum noise from the counting of X-ray photons and electronic noise from photodiodes during the conversion of visible light intensity into electronic signal. Quantum noise is often approximated by a Poisson law, electronic noise is often described with a Gaussian distribution. The combination of a Poisson and a Gaussian distribution is not frequently studied in the literature. This is because the discrete Gaussian distribution is already a good approximation of the Poisson distribution when the number of detected X-ray photons is sufficiently large. The random variable resulting from the two effects is thereby directly modelled as a Gaussian, allowing to use of simpler models for X-ray detection.

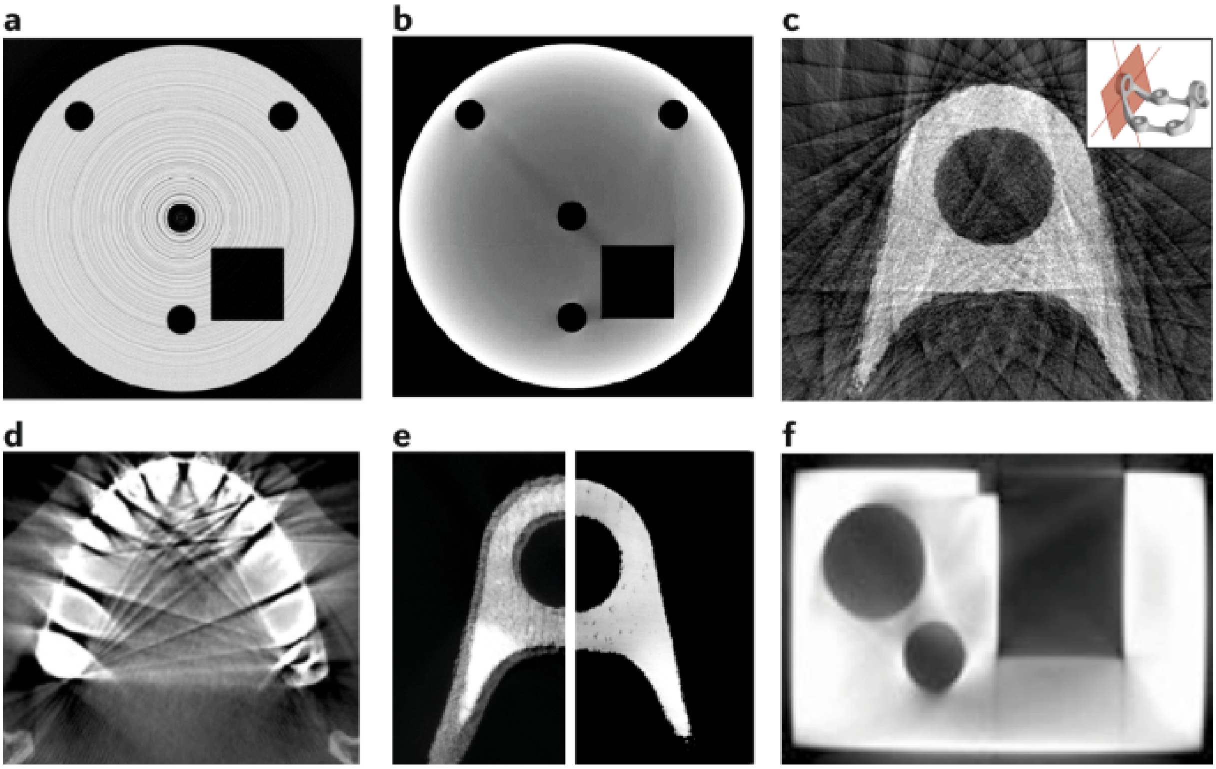


Figure 1.21: Computed tomography slices illustrating common reconstruction artefacts. Image from [2]. a) Ring artefacts. b) Beam hardening artefacts. c) Limited (here, 30 projections) projection artefacts for the reconstruction of a titanium additively manufactured bracket. d) Streak artefacts in a section of the upper jaw due to silver amalgam crowns. e) Titanium additively manufactured bracket with (left) and without (right) motion artefact. f) Cone beam distortion.

1.4.2 Quality Metrics for Image Evaluation

To evaluate, or even train methods using machine learning algorithms, quality metrics are indispensable. Defining quality objectively is inherently challenging, as it is a context-sensitive concept. Most traditional metrics rely on a pixel-by-pixel comparison between a test image I_{test} and a reference image I_{ref}. The most well-known metrics are the *Mean Square Error* (MSE or RMSE for Root Mean Square Error, RMSE $= \sqrt{\text{MSE}}$) and the *Peak Signal-to-Noise Ratio* (PSNR).

$$\text{MSE} = \frac{1}{m\,n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{\text{test}}(i,j) - I_{\text{ref}}(i,j)]^2. \tag{1.56}$$

$$\begin{aligned}\mathrm{PSNR} &= 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \\ &= 20 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}}}\right) \\ &= 20 \cdot \log_{10}(\mathrm{MAX}_I) - 10 \cdot \log_{10}(\mathrm{MSE}).\end{aligned} \tag{1.57}$$

Here, n and m represent the height and width of the images, and MAX_I denotes the maximum possible pixel value of the reference image. These metrics are particularly useful for examining reconstruction quality or the theoretical differences between two images. They are particularly well-suited for comparing images where noise is simulated [38].

However, these metrics have the limitation of being less suitable for comparing experimental images. By comparing only pixel values, they do not adequately account for the information present in the images and are highly sensitive to misalignment and distortion. The Structural Similarity Index (SSIM) considers luminance l, contrast c, and structure s, providing a metric between zero and one (with one indicating strong similarity between images) that better models human visual perception than previous error metrics. Luminance measures the average value of an image, while contrast indicates the variability within the image. Once two images are normalised to have the same luminance and contrast levels, their structures can be compared. The SSIM is computed over multiple sliding of an image. The index between two windows X and Y of common size is defined as:

$$\mathrm{SSIM}(X,Y) = l(X,Y) \cdot c(X,Y) \cdot s(X,Y) = \frac{(2m_X m_Y + c_1)(2\sigma_X\sigma_Y + c_2)(\mathrm{cov}_{XY} + c_3)}{(m_X^2 + m_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)(\sigma_X\sigma_Y + c_3)}. \tag{1.58}$$

In this formula, m_X and m_Y represent the pixels means of X and Y, σ_X and σ_Y denote the pixels standard deviations, cov_{XY} is the covariance of X and Y, and c_1, c_2, and c_3 are constants to stabilise the division.

Metrics derived from information theory can also be employed. Images are treated as random variables, allowing the use of metrics based on entropy, which is one of the most popular concepts. The entropy H of a discrete random variable X with values in $\mathscr{X}$ and probability p is defined as:

$$\mathrm{H}(X) = -\sum_{x\in\mathscr{X}} p(x)\log_2 p(x). \tag{1.59}$$

Entropy measures the expected amount of information conveyed by identifying the outcome of a random trial; the term $-\log_2 p(x)$ is known as *surprisal* or *self-information.* To measure the dependence between two variables from an information-theoretic perspective, we compute how much information is obtained about one variable upon knowing the value of another. This leads to concepts such as conditional entropy and mutual information. The conditional entropy quantifies the amount of information needed to describe the outcome of a random variable Y given that the value of another random variable X is known. It is defined as:

$$\mathrm{H}(Y|X) = -\sum_{x\in\mathscr{X},y\in\mathscr{Y}} p(x,y)\log\frac{p(x,y)}{p(x)}. \tag{1.60}$$

This notion is closely related to mutual information, which measures the dependence between two variables:

$$MI(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{1.61}$$

$$MI(X,Y) = \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)}. \tag{1.62}$$

Moreover, mutual information can also be defined using the Kullback-Leibler divergence D_{KL}:

$$MI(X,Y) = D_{\mathrm{KL}}(P_{(X,Y)}\|P_X \otimes P_Y), \tag{1.63}$$

where $P_{(X,Y)}$ is the joint distribution, and $P_X \otimes P_Y$ is the outer product of the marginal distributions.

Other metrics do not require a reference image but necessitate the ability to isolate an image from a particular feature of interest. These metrics are particularly suited for studying a defect or a region of interest. For example, the Signal-to-Noise Ratio (SNR) is simply defined as the ratio of the signal amplitude to the standard deviation of the noise:

$$\mathrm{SNR} = \left|\frac{\mu_{\text{signal}}}{\sigma_{\text{noise}}}\right|. \tag{1.64}$$

The contrast-to-noise ratio (CNR) is also frequently utilised:

$$\mathrm{CNR} = \frac{\mu_{\text{signal}} - \mu_{\text{noise}}}{\sigma_{\text{noise}}}. \tag{1.65}$$

CNR is not considered an optimal image quality metric because it does not account for the size of the object and the pixel size of the image [38].

As many imaging systems are designed for very specific purposes, task-based metrics can be created, which quantify the ability of the imaging system to perform a good diagnosis. For example, the method of Receiver Operating Characteristics (ROC) quantifies the performance of radiological technology based on a statistical analysis. ROC curves are common tools to evaluate the quality of an imaging system. The ROC curve displays the fraction of true positives versus the fraction of false positives for a test. The area under the curve (AUC) itself serves as a metric of the reliability of the statistical test. Unfortunately, ROC analyses are time-consuming and complex, prone to bias if proper precautions are not taken, and can only be conducted on available equipment [39]. Ideally, an easier-to-measure merit factor, which could be estimated *a priori*, would be preferable. This requires modelling the ROC curves. In a statistical test where we observe the variable $\mathbf{g}$ and the possible outcomes are: $\mathcal{H}_1$ for the presence of a certain feature, and $\mathcal{H}_0$ for its absence, we can define the detectability index analogously to the CNR as the ratio of the expectation value of the signal (the difference between the expectation values of both hypotheses) to the average of the variances of the conditional test statistics [10, 39]:

$$d = \mathrm{SNR}_{\mathbf{g}} = \frac{\langle \mathbf{g}|\mathcal{H}_1\rangle - \langle \mathbf{g}|\mathcal{H}_0\rangle}{\sqrt{\frac{1}{2}(\sigma^2_{\mathcal{H}_0} + \sigma^2_{\mathcal{H}_1})}}. \quad (1.66)$$

The detectability index is central to the concept of model observers. These notions are particularly useful for optimising and modelling a human observer who wishes to quantify the visibility of a certain feature in a signal [10, 40].

1.4.3 Summary of Reconstruction Methods and their Evaluation

To practically illustrate the previously discussed concepts, we propose to reconstruct the 1024×1024 pixels Shepp-Logan phantom in sparse-view using several reconstruction methods and evaluate them with different metrics. In this example, Gaussian noise is added to the sinogram. The sinogram consists of 50 parallel-beam projections acquired along a circular trajectory around the phantom. The detector consists of 1024 pixels. All iterative algorithms performed 100 iterations. This choice of number of iterations is arbitrary, not all methods converge at the same speed. Proximal methods, ADMM and PDHG are regularised with a Total Variation term (i.e. a ℓ_1 norm constraint on the image gradient). Figure 1.22 shows the phantom as well as the reconstructions.

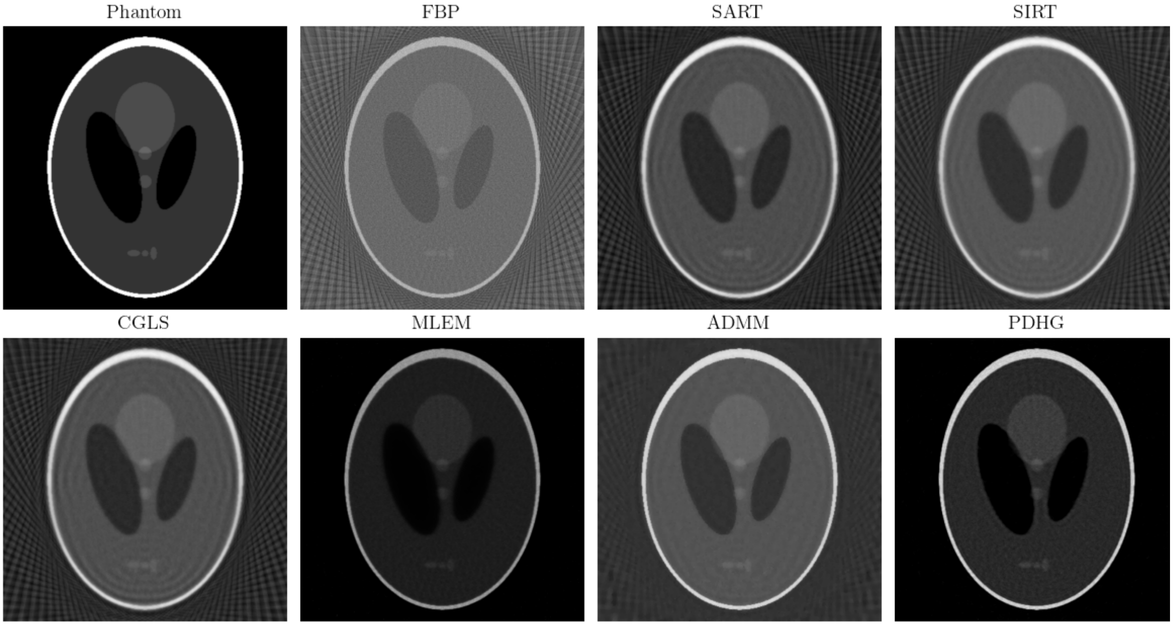


Figure 1.22: Comparison of different reconstruction algorithms on the Shepp-Logan phantom in a sparse and noisy context.

In general, it is observed that the lack of projections induces streak artefacts in the background of the image, which are particularly pronounced in FBP, SART, SIRT, and CGLS. It is noteworthy that MLEM and the regularised methods do not exhibit as many streak artefacts. The added Gaussian noise on the sinogram creates a blurring effect in most reconstructions. This blurring is less pronounced in FBP, but it shows a stronger granularity in the texture of the reconstruction. The reconstructions are evaluated according to the MSE, PSNR, SSIM and Normalised Mutual Information (NMI) metrics. The results are summarized in Table 1.1. We observe that depending on the metric used, the conclusion about the best method varies; however, overall, the more sophisticated iterative methods provide better reconstructions.

Table 1.1: Evaluation metrics for different reconstruction methods.

Method	MSE	PSNR [dB]	SSIM	NMI
FBP	0.8294	13.92	0.0738	1.0583
SART	0.2572	24.09	0.4512	1.2408
SIRT	0.2797	23.36	0.4754	1.2428
CGLS	0.2581	24.06	0.4316	1.2386
MLEM	0.1286	30.12	**0.8466**	**1.4097**
ADMM	**0.1215**	**30.61**	0.6848	1.3227
PDHG	0.1235	30.47	0.8012	1.3974

1.5 Experimental setup

To test our algorithms, we used both simulated objects and a real object throughout all parts of this thesis. For each method presented in this thesis, we validated our results on a plastic additive manufacturing part. The object is a star-shaped trophy. The projections are acquired following a circular trajectory centred around the object. In this setup, the object is placed on a rotating stage, while the detector and source are fixed. The source-to-object distance is 390 mm, and the source-to-detector distance is 668 mm. Figure 1.23 shows the experimental setup.

A total of 928 projections are acquired, spaced regularly every 0.388°. This number of projections was chosen based on the RX Solutions machine's criteria, which aims to acquire enough views to achieve the best quality without making the scan too long. The projections have a resolution of 1880×1212 pixels, with a pixel size of 127 μm. The exposure time is set to 100 ms. The X-ray source is set to a voltage of 100 kV, the current is 500 mA and the beam is filtered by 0.6 mm of copper. Most of our reconstructions are done at a resolution of 1204×1772×1204 voxels, with each voxel measuring 74.1 μm^3. We would like to thank RX Solutions, and especially Morgane Gelin, for the assistance provided during this acquisition.

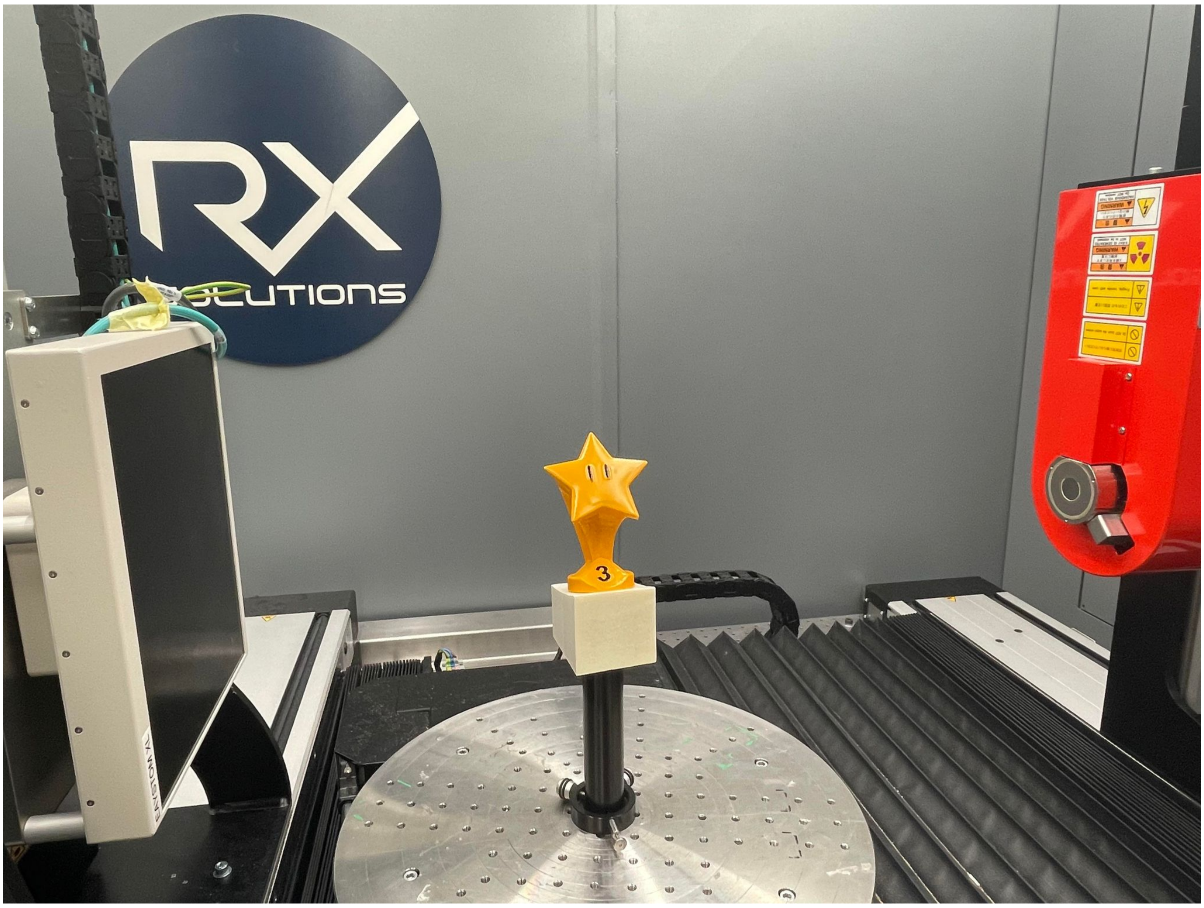


Figure 1.23: Photo of the experimental set-up with the studied piece at the centre of the turntable. The detector is on the left of the object, and the source is on the right.

Chapter 2

Trajectory optimisation

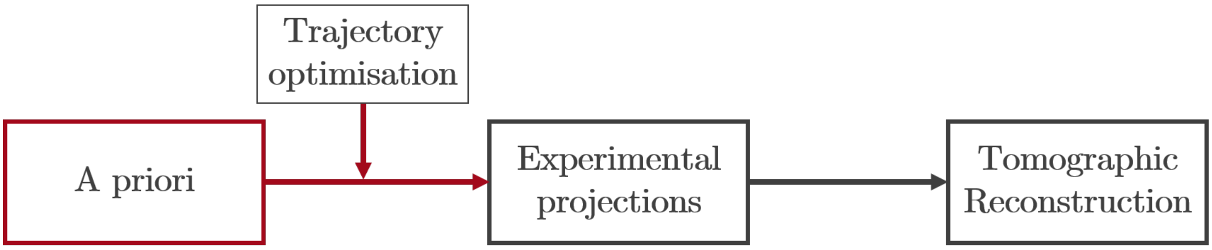


Figure 2.1: Diagram of the tomographic process. In this section, we study trajectory optimisation. The aim is to select the best projections so that an object can be reconstructed with sufficient quality using the minimum number of projections possible.

As depicted in the Diagram 2.1, the first optimisation for sparse-view CT involves selecting the most informative views. Projections often contain a substantial amount of redundant information, making it feasible to reduce their number without compromising the reconstruction quality. It is crucial, however, to avoid eliminating projections that provide indispensable information. Therefore, the goal is to optimise the selection of projections, which can also be referred to as optimising the acquisition trajectory. This optimisation can be theoretically approached without any prior knowledge of the object. However, if a CAD model is available, it provides valuable geometric information that can be utilised. In this section, we introduce the *Empirical Interpolation Method* and show how it can bring significant results in the choice of best projections and we demonstrate the performance of our method on an experimental case.

2.1 Projections selection

2.1.1 Sampling conditions and standard trajectories

Most often, data acquisition is carried out along a circular trajectory around the object. This trajectory is the easiest to implement since it involves placing the object on a rotating platform while keeping the X-ray source and detector stationary. It offers a single degree of freedom and does not require a difficult calibration. However, the circular trajectory does not meet the Tuy criterion for data completeness, which states that each plane containing a point within the Region Of Interest (ROI) must intersect the source trajectory

Table 2.1: Overview of X-Ray CT trajectories, adapted from [10].

Trajectory	Tuy-Smith criterion	Reference
Inclined circles	No	[42]
Two orthogonal circles	Yes	[43]
Partial circle & Arcs	Yes	[44]
Ellipse-Line-Ellipse	Yes	[45]
Dual Ellipse Cross Vertex	Yes	[46]
Two Concentric Arcs	Yes	[47]
N-sin closed sinusoidal	Yes	[48]
Low discrepancy sphere	Yes	[49]

non-tangentially. As a result, a multitude of alternative trajectories have been proposed. Helical trajectories, in particular, have gained significant prominence in medical imaging and have also found applications in industrial settings [41]. However many other methods have been proposed and led to the development of numerous trajectory designs and corresponding reconstruction methodologies. Table 2.1 presents a brief, non-exhaustive summary of the trajectories proposed in the literature and their compliance with the Tuy criterion.

Most of these novel trajectories have demonstrated the ability to satisfy the Tuy criterion. However, it is worth noting that this criterion is typically derived for noise-free untruncated projections on continuous trajectories and continuous perfect detectors. A real practical imaging setup does not conform to these ideal conditions. Nevertheless, there have been efforts to establish discrete equivalents of the Tuy criterion for cone-beam imaging, as exemplified by [50]. This discrete criterion is exclusively founded upon the properties of the Radon transform and Tuy conditions. However, there exist other criteria, such as described in [51], which incorporate notions from the Nyquist-Shannon sampling theorem. Coarsely, the Nyquist-Shannon theorem states that the maximum available frequency in the data spectrum ν_{max}, must be less than half the sampling rate $\Delta\xi$:

$$\nu_{max} < \frac{1}{2\Delta\xi}. \tag{2.1}$$

In our case, $\Delta\xi$ represents the detector pixel pitch. Each acquisition point is separated by $\Delta\gamma$. Figure 2.2 represents the acquisition parameters in red. In Radon space, the projections are not placed in columns as for the sinogram, but in a polar manner according to the angle of acquisition. In Fourier space, ν_{max} corresponds to the radius of the circle in which all information lies, hence:

$$2\nu_{max} = n_p \Delta\nu, \tag{2.2}$$

where n_p is the number of sampling elements in the detector, $\Delta\nu$ is the size of each of this n_p elements in Fourier. Moreover, the diameter of the Radon space is equal to the MFD in the spatial domain, $\mathrm{MFD} = n_p\Delta\xi$. Thereby, the previous equation can also be written as:

$$\nu_{max} < \frac{n_p}{2\mathrm{MFD}}, \tag{2.3}$$

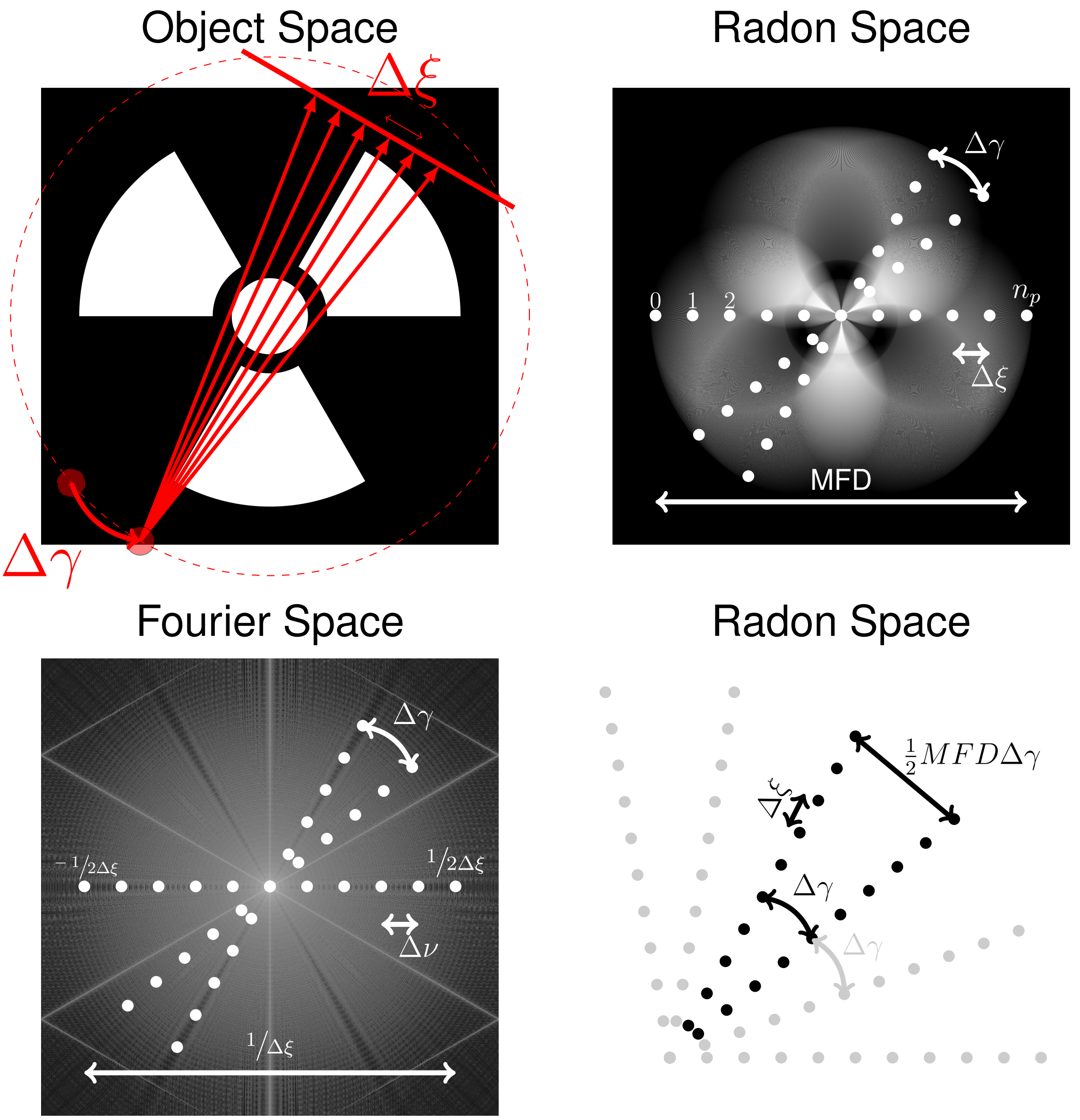


Figure 2.2: Illustration of the sampling points in the object, Radon and Fourier space [3].

which is really useful to estimate the minimum number of elements in the detector: $n_{pmin} = 2\text{MFD}\nu_{max}$. The Fourier analysis can even be taken further to estimate the number of necessary projections. Each acquisition samples a radial line in Fourier, whose distance of neighbouring in the Radon domain is $\frac{1}{2}\text{MFD}\Delta\gamma$. For a better interpolation grid, we want this distance close to $\Delta\xi$. To sample all over a half-circle of radius ν_{max}, the number of projections must satisfy:

$$N > \frac{\pi}{\Delta\gamma}. \tag{2.4}$$

Injecting the previous equations, the minimum number of projections N_{min} can be expressed as:

$$N_{min} = \frac{\pi}{2} n_{pmin}. \tag{2.5}$$

This relation can also be expressed with the diameter of the object and the resolution and is known as the *Crowther* criterion [52]. Most of the time, in sparse-view CT, this criterion is not respected and too few projections are acquired. Fortunately, developments in compressed sensing now make it possible to reconstruct a signal convincingly even if the Nyquist-Shannon criterion is not respected. Figure 2.2 shows how information about the inspected object is presented in the different spaces (Real, Radon and Fourier). This shows that the information is not necessarily distributed homogeneously and that intelligent sampling can avoid unnecessary projections [53].

2.1.2 Bibliography on the design of optimised trajectories

Several methods and criteria have been proposed through the years to optimise the reconstruction quality with limited projections. Obviously, each method depends on the application and the experimental devices, but we will mainly focus on the concepts that drive those methods rather than technical details. Because, while most methods rely on similar concepts, they do not share the same criteria to select the best projections; it is important to understand the big concepts. Moreover, some proposed methods are driven by multiple concepts.

The most recurrent idea in selecting the best projections is that tangential rays are imperative to reconstruct an edge properly. Those rays are more likely to capture the discontinuities and delimitations between two zones. To illustrate this idea, Figure 2.3 shows the FBP reconstruction of a homogeneous square with two parallel beam projections. On the left, the projections are salient, i.e. horizontal and vertical in this case, while on the right, the projections are diagonal. The edges of the square are better reconstructed with the salient views. This phenomenon is explained by the sampling conditions mentioned above [54]. The edges of the image will create characteristic lines in Fourier space, which must be sampled correctly in order to reconstruct them. In Figure 2.4, the edges of the radioactive symbol create sharp lines at all 60 °in Fourier. Whereas the rounded edges are responsible for the concentric circles that fill the Fourier image. So, to reconstruct the edges of the radioactive symbol correctly, we just need to send tangential rays to the edges, whereas to reconstruct the rounded edges we need to sample a wide range of angles [54].

The importance of tangential views has been clearly understood and exploited. Zheng and Mueller [55] use the Hough transform of the gradient image to detect salient object features and use an ant colony optimiser to solve the projection selection as a set-cover problem. In a related vein, Venere et al. [56] employed a geometric criterion as a basis for reconstructing cracks with elliptical shapes. They stipulate that better reconstructions can be achieved by accumulating more views about the main axis of the ellipse and using a genetic algorithm to select to next best view.

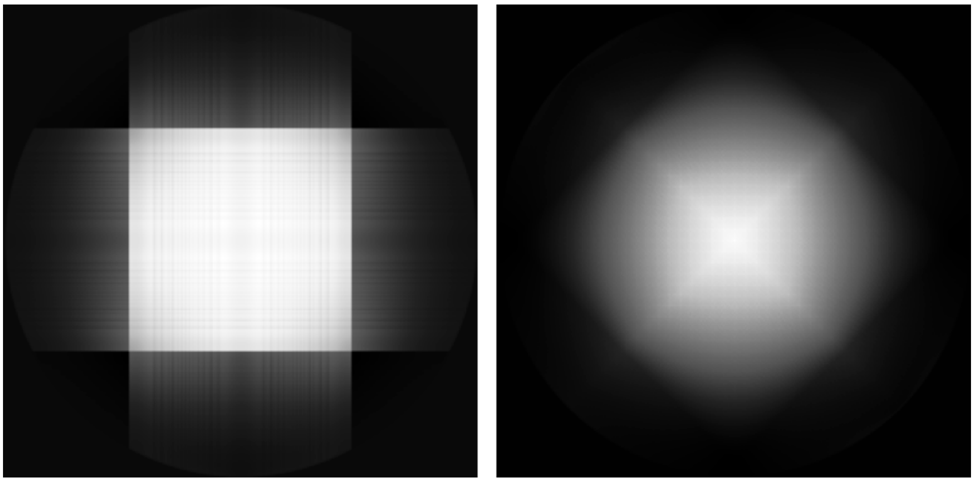


Figure 2.3: Reconstruction of a square phantom with only two projections. Left: projections are horizontal and vertical. Right: Projections are diagonal. Salient views allow to reconstruct the edges better

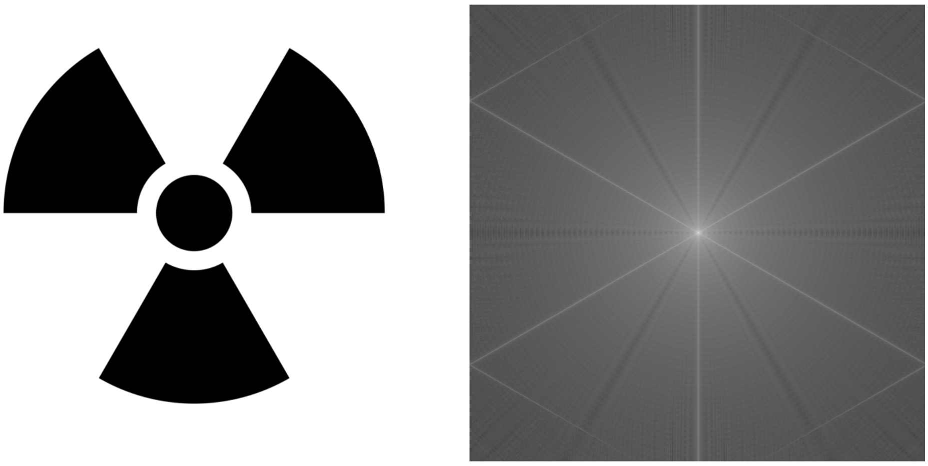


Figure 2.4: Fourier Transform of the radioactive symbol. The main streaks correspond to the salient edges while the concentric circles forming the background are due to the rounded parts of the object. Log scaling is applied for visualisation.

However, the idea of tangential rays can also be declined in other forms. Capturing strong variations can mean many different things and many criteria can be created from this ambition. For example, Haque et al. [57] use the variation of the spectral richness in the projections as a criterion to adapt the angle step size during an acquisition. Strong variations in the projections also imply strong variations in the object and therefore must be sampled more. Placidi et al. [58] have formulated an algorithm wherein the optimal views are selected based on an entropy measure. The acquisition process commences with an initial set of uniformly distributed projections, followed by subsequent acquisitions occurring between the two projections exhibiting the highest discrepancy in entropy.

This approach finds application within the domain of nuclear magnetic resonance imaging, demonstrating a reduction in the overall requisite number of projections when the specimen exhibits certain regularities or symmetries. Matz et al. [59] use the wavelet decomposition to detect edges in the projections. They select the best projections by optimising a compromise between a high amount of edge information and diversity to prevent redundancies.

The need for tangential rays to stably detect an edge can be mathematically explained with properties of the X-ray transform. Besides, the global version of this result is the Tuy condition [54], which is also one of the main criteria for selecting the best projections. Based on a local and quantitative description of the completeness of the projection data [60], Herl et al. formulate a new algorithm to both optimise projections completeness and avoid highly attenuated rays [61]. Similar ideas are found in medical applications to avoid metal-artefacts [62]. Moreover, some studies also consider the stability of the part orientation [63].

Many studies take noise and artefacts into account when choosing views [64]. This is an important parameter with a strong influence on reconstruction quality [65, 66]. Although absorption reduction can be a criterion for view selection, it is often combined with other criteria as shown in the previous examples.

Other studies select projections by directly optimising reconstruction quality. Often based on a greedy algorithm, projections are gradually added to the trajectory to maximise reconstruction quality at each step [67]. Brierley et al. [68] go further by finding the best projections set while including simulated defects in the studied part. This approach uses *a priori* information on defects to create a dataset including them, in order to select projections that maximise their visibility and detection.

A recent criterion for selecting the best trajectories has become increasingly popular. Introduced by Stayman et al. [69], the detectability index, described in Eq. 1.66, uses model observers to quantify the performance of a set of projections for a given task. Modulation transfer function and noise power spectrum are used to quantify the information brought by the projections. Other developments followed, including Fischer et al. [70] who suggested a local non-pre-whitening observer model, and Bauer et al. [71] who included this criterion in a complete optimisation framework with a geometrical weighting and experimentally tested the results. Subsequently, several elements of machine learning were incorporated into the initial methods. To avoid metal artefacts, Thies et al. [72] propose an online trajectory adjustment pipeline based on a neural network for predicting detectability. Further developments even include both index detectability and data completeness in a machine learning framework [73].

2.2 Our methods

2.2.1 Empirical Interpolation Methods

Although many methods and criteria have been proposed, none prevails yet. Each method has its own specificities of use which can be prohibitive in certain situations. Due to the lack of training data, machine-learning methods are impossible in our case. Many of the methods presented are less effective for the 3D case, and calculating the detectability index takes too long for our applications. In the industrial context in which we work, the selection of projections must be carried out quickly. We will assume that the CAD model is available and will be used as *a priori* information. However, we are assuming some discrepancy between the 3D model and the experimental part. The method we propose is inspired by Reduced Order Models (ROM). Since the aim is to reduce the number of projections while maintaining good reconstruction quality, ROM and data compression methods seem particularly relevant. The *Empirical Interpolation Method* (EIM) and its discrete version (DEIM) belong precisely to this category [74]. Initially developed for sparse sampling and approximating computationally expensive non-linear operators in differential equations, it has been used in many other fields such as finite elements. DEIM is particularly used for high-dimensional and complex systems because the computational cost of evaluating the non-linearity is scaled with the rank of the reduced Proper Orthogonal Decomposition (POD) basis [75].

The POD is certainly one of the most used techniques for dimensionality reduction,it leverages the low-rank patterns of a complex system to model it accurately. Usually, POD is used for spatial-temporal system, it decomposes a field $\mathbf{b}(\mathbf{x}, t)$ into a set of spatial vectors $\{\boldsymbol{\psi}_i\}_{i=1}^m$ and time coefficients $\{a_i\}_{i=1}^m$, such that:

$$\mathbf{b}(\mathbf{x}, t) = \sum_{i=1}^{m} a_i(t)\boldsymbol{\psi}_i(\mathbf{x}). \tag{2.6}$$

In order to find the POD basis, i.e the set of vectors $\{\boldsymbol{\psi}_i\}_{i=1}^m$, the field $\mathbf{b}(\mathbf{x}, t)$ is first sampled at all locations and times to make the so-called *snapshot matrix*, where each row corresponds to a measurement of the system at a different state:

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}(\mathbf{x}_1, t_1) & \cdots & \mathbf{b}(\mathbf{x}_m, t_1) \\ \vdots & & \vdots \\ \mathbf{b}(\mathbf{x}_1, t_p) & \cdots & \mathbf{b}(\mathbf{x}_m, t_p) \end{pmatrix}, \tag{2.7}$$

where m is the number of spatial sampling locations and p is the number of time samples. Next, the covariance $\mathbf{B}^\mathsf{T}\mathbf{B}$ of the snapshot matrix is diagonalised. Its eigenvectors, sorted by decreasing eigenvalues order, form the POD basis $\{\boldsymbol{\psi}_i\}$. Obviously, for the purposes of dimension reduction, the index m will have to be relatively low to reduce the calculation costs, and then the sum of the approximation of $\mathbf{b}$ shall be truncated at a lower dimension. POD is similar to the *Principal Component Analysis* (PCA) and the *Karhunen–Loève theorem*, the difference is that POD is applied to a physical field.

To return to the DEIM algorithm, it precisely takes as an input argument the POD basis $\{\boldsymbol{\psi}\}_{i=1}^m$ of a non-linear *snapshots* matrix. The first step of the DEIM is to find the maximum of $|\boldsymbol{\psi}_1|$, the first and most meaningful vector of the POD basis, and its index $\wp_1$. This index represents the first selected sampling location. The measurements $\mathbf{P}$ matrix, which contains the information about the sampling locations, is initialised with the canonical unitary vector $\mathbf{e}_{\wp_1}$. Then, for each iteration, an approximation of the projection of $\boldsymbol{\psi}_i$ over $\boldsymbol{\Psi}$, the set of all considered eigenvectors during the algorithm, is done. Unlike the initialisation, the next sampling location is calculated by taking the index of the maximum of the residual $\mathbf{r}$, i.e., the point where the approximation of $\boldsymbol{\psi}_i$ is the worst. A new singular vector is added to $\boldsymbol{\Psi}$, and a new vector to $\mathbf{P}$. Thereby, $\mathbf{P}$ contains the sampling locations in descending order of importance. During the whole procedure, it is important to check that no indexes are selected twice. The fact that the

Algorithm 4: Discrete Empirical Interpolation Method (DEIM) [75]

```
**Data:** ${\psi}_{i=1}^{m} \subset \mathbb{R}^n$ linearly independent
**Result:** $\boldsymbol{\wp} = [\wp_1, \wp_2, \dots, \wp_m]^\mathsf{T} \in \mathbb{R}^m$
$\wp_1 = \text{argmax}\{|\boldsymbol{\psi}_1|\}$
$\boldsymbol{\Psi} = [\boldsymbol{\psi}_1], \mathbf{P} = [\mathbf{e}_{\wp_1}], \boldsymbol{\wp} = [\wp_1]$
for $i = 2$ to m do
    Solve $(\mathbf{P}^\mathsf{T}\boldsymbol{\Psi})\mathbf{c} = \mathbf{P}^\mathsf{T}\boldsymbol{\psi}_\mathbf{i}$ for $\mathbf{c}$
    $\mathbf{r} = \boldsymbol{\psi}_i - \boldsymbol{\Psi}\mathbf{c}$
    $\wp_i = \text{argmax}\{|\mathbf{r}|\}$
    $\boldsymbol{\Psi} \leftarrow \begin{bmatrix}\boldsymbol{\Psi} & \boldsymbol{\psi}_i\end{bmatrix}, \mathbf{P} \leftarrow \begin{bmatrix}\mathbf{P} & \mathbf{e}_{\wp_i}\end{bmatrix}, \boldsymbol{\wp} \leftarrow \begin{bmatrix}\boldsymbol{\wp} \\ \wp_i\end{bmatrix}$
end
```

basis $\boldsymbol{\Psi}$ originates from a POD ensures that the DEIM algorithm will always converge and that the error growth will be limited. The procedure is detailed in Algorithm 4.

DEIM has proved to be effective and easy to implement in many areas, including interpolation. The EIM, sometimes even called *magic points* for interpolation, has been compared with classical polynomial interpolation techniques such as Gauss-Lobatto-Legendre or Chebyschev nodes [76]. Studies show that the nodes selected with EIM give good interpolation points with Lebesgue constants [1] close to the optimal values (see Table 2.2). In Figure 2.5a, we have plotted different sampling strategies for interpolating a Vandermonde basis. DEIM shares huge similarities with classical polynomial interpolation techniques. Figure 2.5b shows the mean square error of the interpolation according to the sampling strategy. DEIM shows better results than equidistant and Chebyshev nodes. Moreover, EIM has been generalised in [77], where snapshots are replaced by more general linear forms.

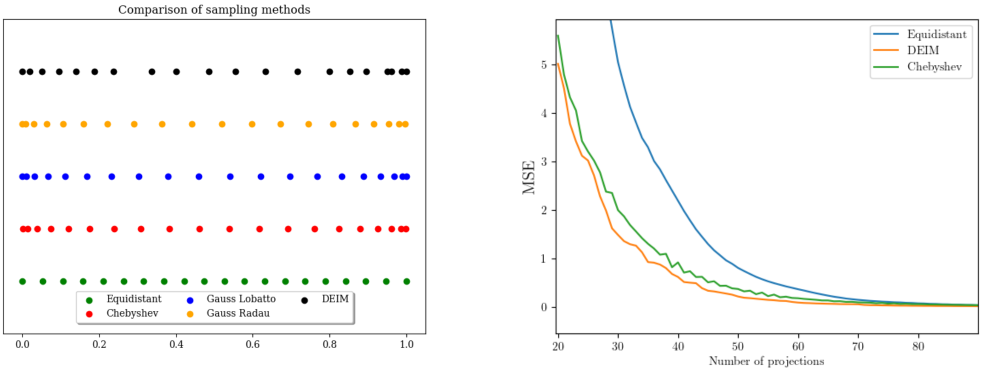
Table 2.2: Lebesgue constant of different nodes selection methods

Method	Lebesgue Constant
Equidistant	5886
Random	1.061e+11
Chebyshev	2.869
Gauss-Lobatto	2.575
Gauss-Radau	9.626
DEIM	3.722

The Empirical Interpolation Methods have aroused a lot of interest and have been extensively studied. Many variants came out including the Q-DEIM algorithm which uses the pivoted QR decomposition to select iteratively the best sampling locations [74].

---

[1] Lebesgue constants Λ gives an estimator of how good the interpolant of a function is in comparison with the best polynomial approximation of the function of degree n on the interval $[a, b]$ with the given nodes x_i : $\Lambda_n(\{x_i\}_{i\leq n}) = \max_{x\in[a,b]} \sum_{j=0}^{n} |\prod_{\substack{i=0 \\ j\neq i}}^{n} \frac{x - x_i}{x_j - x_i}|$.



(a) Comparison of different sampling strategies on an ill-conditioned Vandermonde basis for 20 nodes.

(b) Mean square interpolation error on polynomial interpolation according to the number of interpolation nodes.

Figure 2.5: Comparison of different node selection techniques on a Vandermonde basis and their interpolation error

The QR decomposition is a really simple mathematical tool to factorise a matrix in an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$. The QR decomposition was introduced for the solution of least squares problems in 1965 by Businger and Golub [78].

QR decomposition is generally done with one of the three following methods: Gram-Schmidt projections, Householder reflections and Givens rotations [79]. Gram-Schmidt projections are easy to implement but numerically unstable, Householder reflections are simple and stable but are bandwidth-heavy, Givens rotations are more difficult to implement but are parallelisable and bandwidth efficient. To illustrate our algorithms and simplify understanding, we will nevertheless use the Gram-Schmidt method. It is easy to understand that at each iteration of the Algorithm 5, the columns which have not yet been selected are orthogonally projected onto those which have been.

In the classic method, the columns are selected one by one, but in the pivoted version, the column with the greatest norm is greedily chosen. This trick was specially made for rank-deficient matrices, it greatly improves the numerical stability of the factorization. With the column pivoting, the QR decomposition becomes $\mathbf{XP} = \mathbf{QR}$, where $\mathbf{P}$ is a permutation matrix so that the diagonal elements of $\mathbf{R}$ are decreasing: $|\mathbf{R}_{11}| \geq |\mathbf{R}_{22}| \geq \cdots \geq |\mathbf{R}_{nn}|$. This property can slightly recall the structure of the diagonal matrix of the SVD, because there is a very close link between the QR decomposition and the PCA. QR decomposition is even often used for computing the PCA [80]. This is also why even if Q-DEIM does not use a POD base properly speaking, the method remains similar to DEIM.

Moreover, a variant of Q-DEIM uses the POD basis of the snapshot matrix, then does a QR decomposition and selects the sampling points using the pivot [81]. In this particular work, Manohar et al. first calculate the POD modes of a snapshot matrix with a SVD decomposition: $\mathbf{X} = \mathbf{\Psi\Sigma V}^\mathsf{T}$. The POD basis is truncated to a lower rank r, then the κ most important sensor locations are determined by the permutation matrix $\mathbf{P}$:

$$\text{if } \kappa\text{=r, } \mathbf{\Psi}_r\mathbf{P} = \mathbf{QR}, \tag{2.8}$$

$$\text{if } \kappa\text{>r, } (\mathbf{\Psi}_r\mathbf{\Psi}_r)^\mathsf{T}\mathbf{P} = \mathbf{QR}. \tag{2.9}$$

This technique has shown good results and is quite interesting, because, by using another basis than that of snapshots, we find concepts similar to the generalised EIM mentioned above. In the remainder of this manuscript, this method will be referred to as POD-QDEIM.

Algorithm 5: Pivoted QR algorithm

```
Data: $\mathbf{X} = [\,\mathbf{X}_1 \;\dots\; \mathbf{X}_m\,] \in \mathbb{R}^{n\times m}$
Result: P Permutation matrix such that XP = QR
$\mathbf{P} = [\;]$
for $k = 1$ to m do
    $i = \underset{j}{argmax}\; ||\mathbf{X}_j||^2$
    $\mathbf{X} \leftarrow \mathbf{X}(\mathbf{I} - \mathbf{X}_i^{\top}\mathbf{X}_i/||\mathbf{X}_i||^2)$ #Gram-Schmidt projection
    $\mathbf{P} \leftarrow [\mathbf{P}\;\mathbf{e}_i]$
end
```

2.2.2 Connections between EIM and other methods

The field of application of the *Empirical Interpolation Methods* might seem quite far from the X-ray tomography but many connections can be established between them and the other works for choosing the best projections. The EIMs are used to optimise sensor locations and wisely sample a physical field. The sinogram - or the set of projections for non-circular trajectory - can be seen as such a field. The spatiotemporal evolution in a physical field can be compared to the spatio-angular coordinates of the sinogram, which is interpreted as a dynamical system. From this standpoint, EIMs can be directly applied to the sinogram to determine the best angles to acquire. It could be argued that the method does not take into account the reconstruction stage and only considers the projections, which could be a limitation. However, it is also an advantage because the methods we propose are independent of the reconstruction algorithm, they apply to all types of non-standard trajectories, parallel and cone-beam, and as they only consider projections and not volumes, operators are generally of small size which makes the methods fast.

To clearly understand the choice of EIMs approaches for sparse view CT, we will link their behaviour to other works. To illustrate our points, we will use the 2D phantom in Fig. 2.6. The first criterion to select the best projections was to favour rays tangential to edges. Even if EIMs do not directly optimise that, it can be empirically verified. For example, in DEIM each sampling location is selected as the worst interpolated point for each mode. Strong variations and discontinuities are often difficult to predict if the interpolation nodes are too far apart. For the phantom, the first selected projections are at angles 0°and 90°(horizontal and vertical directions, respectively). These directions offer rays particularly tangent to the protruding edges of the object. Figure 2.6 shows the coordinates and the first selected angles.

As previously outlined, extant research endeavours have employed a spectral richness metric as a mean to modulate the step size applied in circular trajectories, thereby effecting a reduction in step size specifically proximate to image edges. Employing the same spectral richness metric employed in [57], it is discernible that the phantom spectral richness in Figure 2.7a exhibits two localised maxima characterised by markedly elevated spectral

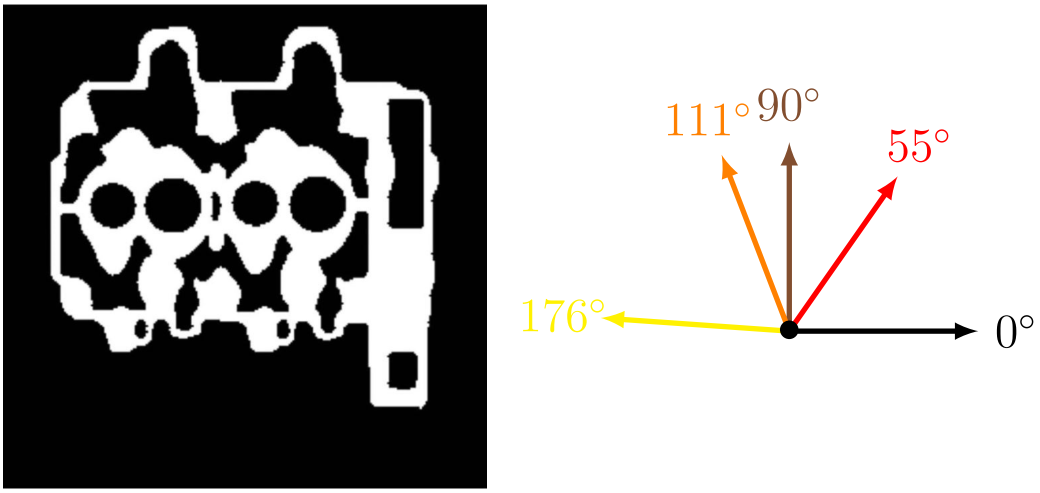
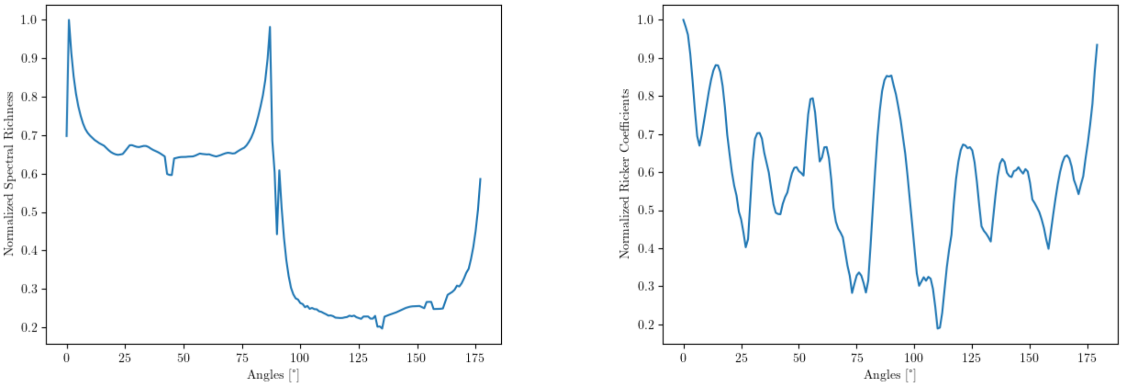


Figure 2.6: Phantom 2D. We will use this phantom to test the various state-of-the-art methods in this section. The five first angles selected by DEIM are 0,90,55,111 and 176 °.

richness values at orientations of 1°and 87°, a phenomenon that aligns with the initial angles chosen by the DEIM.



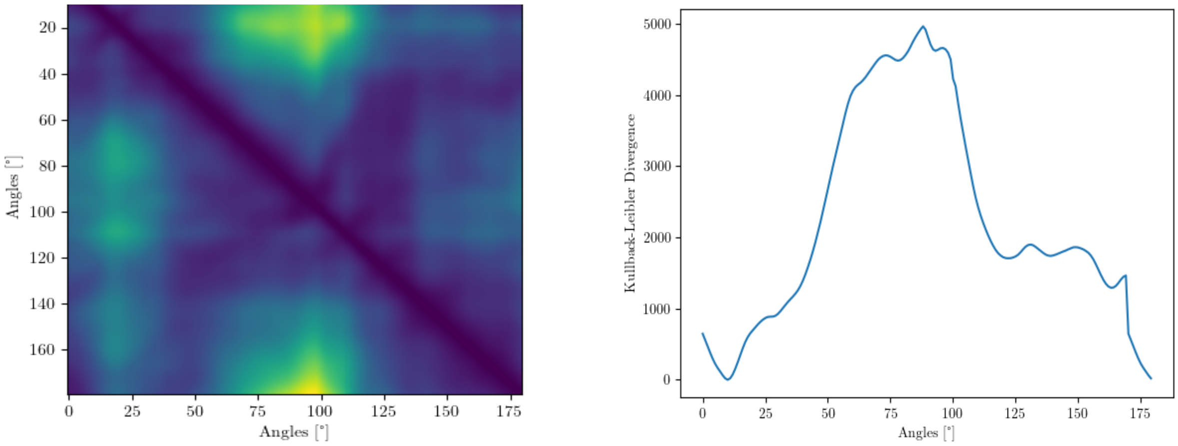
(a) Spectral richness of the phantom. Two maxima at orientations of 1°and 87°are visible. They correspond to the two most important views in terms of information.

(b) Normalised wavelets coefficients. Two maxima are seen at 0°and 89°. The third projection selected by DEIM is 55°which corresponds to another local maxima.

Figure 2.7: Spectral richness and normalised wavelets coefficients of the phantom

In another study by Matz et al. [59], rather than relying on spectral richness, the selection of projections is predicated upon the identification of angles associated with the highest absolute wavelet coefficients. Employing Ricker wavelets, it is observed that the maximum coefficients manifest at angular orientations of 0°and 89°, see Figure 2.7b. It is noteworthy that the concept of salient edges emerges as a recurring motif across various methodologies employed for projection selection. DEIM, in a more abstract sense, also manifests an inherent inclination toward projections with salient edges.

An alternative perspective on the selection of views can be linked to the principles of a *Design of Experiments* (DOE). In this analogy, the sinogram is regarded as a comprehensive DOE, while the optimised strategies aim to choose a specific set of projections



(a) Relative entropy between all projections

(b) Relative entropy with the 0°projection

Figure 2.8: Illustration of the relative entropy between projections for the phantom sinogram. The maximum relative entropy is between 0 and 90°.

for reconstructing an object using minimal data. Within the field of optimal design, the *D-optimal* design criterion seeks to maximise the determinant of the matrix $\mathbf{X}^\mathsf{T}\mathbf{X}$, where $\mathbf{X}$ represents the DOE matrix. Such optimisation leads to the maximisation of the differential entropy associated with parameter estimates [82]. Q-DEIM is directly linked to the D-optimal criterion as the pivoted QR factorisation selects at each iteration the column that maximises the determinant of the $\mathbf{X}^\mathsf{T}\mathbf{X}$ [83]. In fact, at each iteration the algorithm imposes a diagonal dominance structure in $\mathbf{R}$, leading to:

$$|\mathbf{R}_{11}| \geq |\mathbf{R}_{22}| \geq \ldots \geq |\mathbf{R}_{nn}|, \tag{2.10}$$

which maximises the determinant of the left-hand submatrix. This finding can also be associated with numerous methodologies employing entropy as a criterion for the selection of measurements [58, 84]. In various domains, such as Image-Based Modeling/Rendering and Virtual Reality, entropy-based metrics have been introduced [85], and they continue to be utilised in modern viewpoint selection algorithms [86, 87]. The underlying concept behind employing entropy as a criterion is to prioritise measurements that exhibit the greatest dissimilarity from one another, with the aim of imparting fresh information to the computed tomography system. For the sinogram of the phantom, it is noteworthy that the maximum relative entropy between two columns is observed at 0°and 90°(see Figure 2.8a), corresponding to the initial two measurements selected by the Discrete Empirical Interpolation Method. Figure 2.8b shows the relative entropy between all projections and the horizontal projection.

Furthermore, QR decomposition is sometimes used in the process of selecting columns within subsets to create smaller subsets with advantageous algebraic properties [88]. Notably, certain studies have employed the Gram determinant for quantifying the quality of a projection set [58, 89]. This approach proves to be valuable as it facilitates the quantification of the information contributed by a new projection, irrespective of the chosen reconstruction methodology. An intriguing observation pertains to the fact that the Gram determinant serves as a measure of orthogonality between a vector and a subspace—a concept central to the pivot selection in Q-DEIM and POD-QDEIM. By iteratively selecting the view that exhibits the highest degree of orthogonality in relation to the previously selected views, Q-DEIM effectively maximises the Gram determinant.

2.2.3 Constrained Q-DEIM

EIMs are connected to many other methods and appear promising for projection selection. However, these techniques come from purely algebraic reasoning. EIMs are only based on the approximate geometric information of the studied part. Space coverage, noise, absorption and artefacts of all kinds are not considered. For certain objects, this can be a real obstacle, because the view providing the most information can also be the most noisy, for example, the most salient view is also that of maximum traversed thickness. To improve the results of the Q-DEIM, we propose to consider the physical aspects of the X-rays. A constraint is imposed on the QR decomposition pivot to integrate information on the physics of the acquisition. Each potential projection of the complete trajectory is assigned a weight. These weights are completely task-dependent. They can correspond to noises, stability of the robotic arm, etc. In their studies, Clark et al. [90] use the cost of the sensors as a constraint. The QR decomposition algorithm including the cost vector $\boldsymbol{\eta}$, called *Constrained Q-DEIM* (CQDEIM) is summarised in the Algorithm 6. As before, the Gram-Schmidt projection is used here to illustrate, but in practice, the Householder method is used in the implementation for greater robustness.

Algorithm 6: Constrained Q-DEIM algorithm (CQDEIM)

```
Data: $\gamma \in \mathbb{R}$, cost vector $\boldsymbol{\eta} \in \mathbb{R}^m$, $\mathbf{X} = [\ \mathbf{x}_1 \ \dots \ \mathbf{x}_n\ ] \in \mathbb{R}^{n \times m}$
Result: P Permutation matrix such that XP = QR
$\mathbf{P} = [\ ]$
for $k = 1$ to m do
    $i = \underset{j, j \notin P}{argmax} \ (||\mathbf{x}_j||^2 - \gamma \boldsymbol{\eta}_j)$
    $\mathbf{X} \leftarrow \mathbf{X}(\mathbf{I} - \mathbf{x}_i^T \mathbf{x}_i / ||\mathbf{x}_i||^2)$ #Gram-Schmidt projection
    $\mathbf{P} \leftarrow [\mathbf{P}\ \mathbf{e}_i]$
end
```

This method can prove particularly advantageous for certain applications, as it genuinely enables the integration of any constraint expressible within the cost vector. The additional computation time is contingent solely upon the complexity of computing the vector. An illustrative application of this method in the field of X-ray 3D imaging is provided in [9], where the absorption is used as a constraint on a dense object.

2.3 Results

Simulated results concerning these methodologies have been previously published in [4]. In this previous work, we have applied our methods on the CAD of a metallic additive manufactured grate. The outcomes showed tangible contributions to the method. The details of the parts were better reconstructed by optimising the views. The structure of the object is better defined, especially when subsampling a spherical trajectory. Results are shown in Figure 2.9. Direct comparison among DEIM, Q-DEIM and POD-QDEIM methodologies presents challenges due to their tendency to yield comparable results. Efforts were made to substitute the POD-QDEIM method, which relies on POD bases, with alternative bases such as Independent Component Analysis (ICA) or tailored sparse dictionaries, yet the resultant outcomes remained commensurate. The constrained variant was also implemented and results were shown in [9]. In this context, attenuation, noise and beam hardening were considered in the projections and leveraged as a regularisation technique. The addition of a physical constraint brings very interesting results by seeking to find the best compromise between the geometric aspect described above, which includes protruding edges, and the notion of SNR with angles that reduce scattering. For both publications, circular and spherical trajectories were utilised. However, the improvements and potential of the methods are more pronounced with spherical trajectories. Throughout our studies, we confined ourselves to trajectories with constant magnification, which makes it easier to compare the projections. This limitation stems from the method's algebraic comparison of projections, and it appears more logical to compare vectors representing similar images.

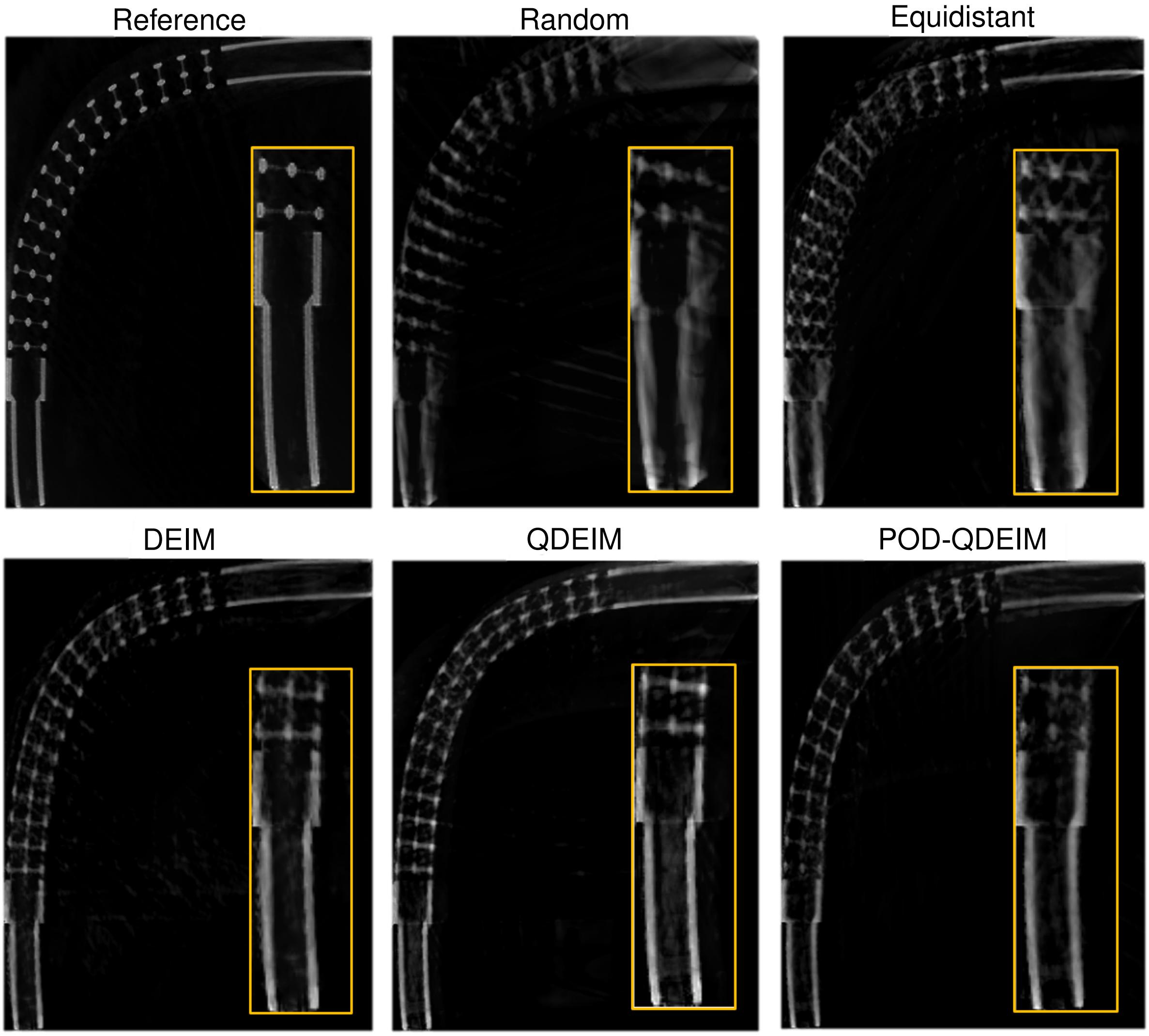


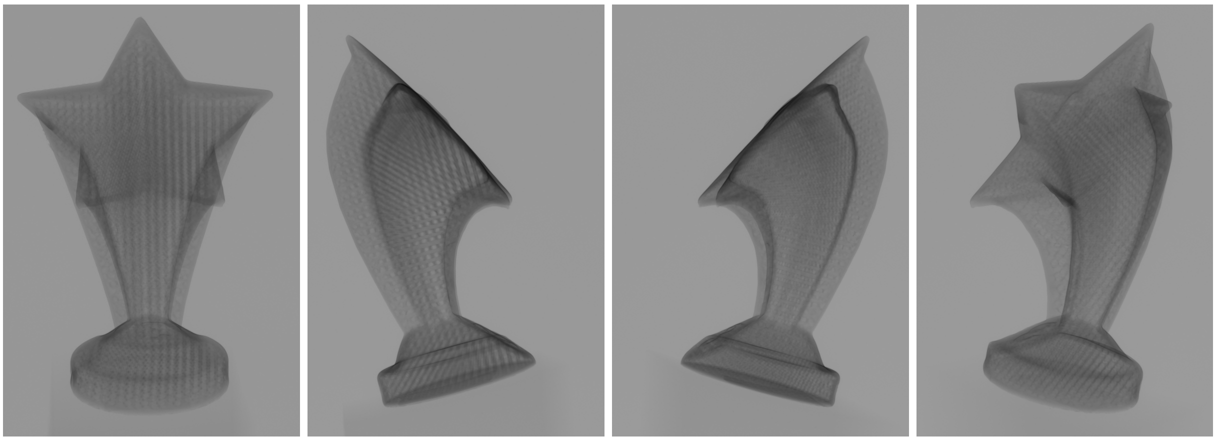
Figure 2.9: Cross-section of the reconstructed grate with 10 projections for the spherical trajectory using different acquisition strategies. Image from [4]. Trajectory optimisation significantly improves reconstruction quality, especially in the case of non-planar trajectories.

In this section, we have tested our methods on experimental projections obtained along a circular trajectory. The used object is the star-shaped trophy described in Section 1.5. The CAD model of the object is shown in Figure 2.10. Unlike our previous studies, which were simulated, here we have the experimental projections, there was no need to simulate the projections from CAD. The methods were directly applied to experimental projections.

Figure 2.10: CAD model of the studied experimental part

The trajectory consists of 928 views evenly distributed on a circle. The projections are cropped to consider only the region of interest at the centre (1880×1212 pixels) and then resized to 235×151 pixels for view selection. This step is necessary to increase robustness to small details and irregularities of the object and to accelerate the procedure. Projections should be selected according to the overall geometry of the object, details that may vary between the model and the real object should not be considered. We tested the DEIM, Q-DEIM, CQDEIM with an absorption constraint and CQDEIM with a coverage constraint. In the latter, whenever an angle is selected, adjacent positions receive an additional cost, proportionally to the distance with the selected position, to promote spatial spacing between projections. In this example, the implementation slightly deviates from that outlined in Algorithm 6, as the cost vector updates throughout the iterations. This capability also paves the way for adaptive sampling, where the next view depends on the preceding ones. Figure 2.11 shows the first four projections selected by the DEIM algorithm. The selected projections are well in line with the predictions. Projections show strong edges. The first three views are about 90°apart. The fourth is diagonal to the previous ones.

Figure 2.12 shows the reconstruction results for various numbers of projections on a circular trajectory. The PSNR is computed over the entire volume, while the SSIM results are provided for the central horizontal slice of the reconstructions. In terms of PSNR, our



(a) Trophy front projection 0° (b) Left side projection 88.4° (c) Right side projection 257.6° (d) Projection at an angle 305.7°

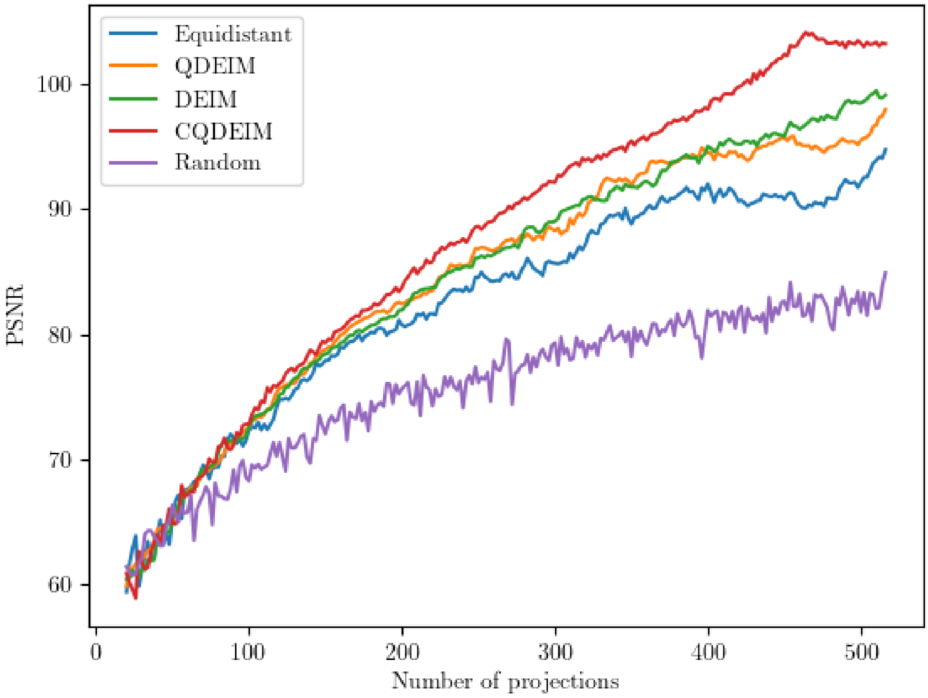
Figure 2.11: First four projections selected by the DEIM algorithm

methods offer an advantage compared to equiangular sampling. The PSNR of optimised methods is generally one to two dB higher. Comparable quality can be achieved with fewer projections. The SSIM results are less conclusive. Reconstruction qualities appear similar regardless of the sampling method chosen, except for random sampling. To delve deeper, we present the FDK reconstruction of another slice using the studied trajectories. Figure 2.13 illustrates the reconstructions for 300 projections, and Table 2.3 presents the reconstruction quality in terms of SSIM. The table demonstrates a slight improvement in quality due to our methods. However, visually, the quantity of artefacts is reduced. The streak artefact appears diminished, and values outside the object are lower. In this study, we did not impose a noise constraint as in [9], but rather applied a regularisation term on the spread of source points. The object we examined had low attenuation, and the noise in the acquired projections was minimal, so a noise constraint would have had little effect. It is observed that the CQDEIM method only shows improvement when the number of projections is high, as the acquired views are already naturally well-distributed along the trajectory.

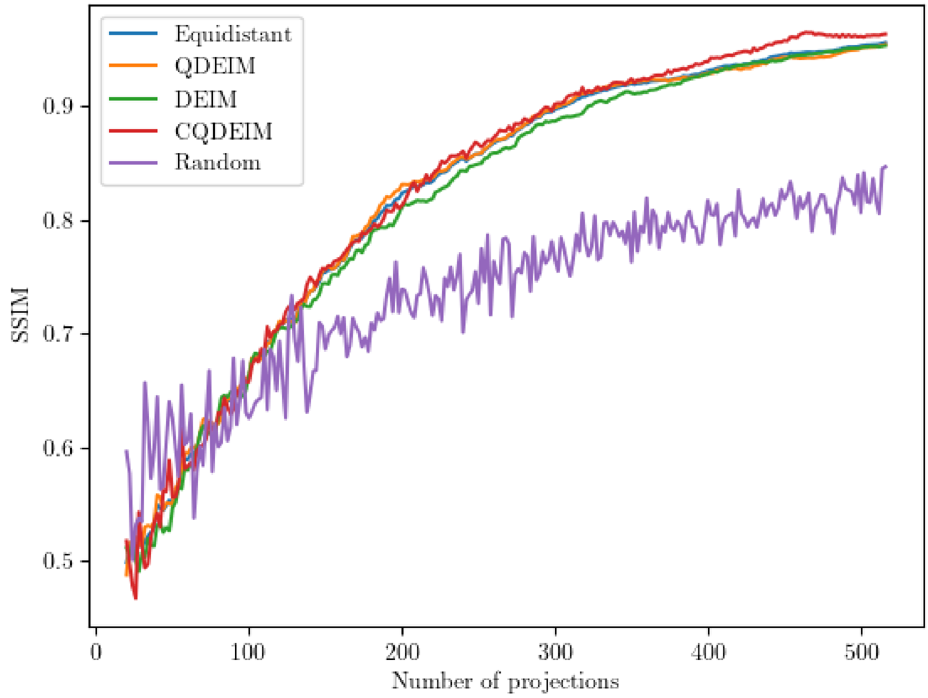
Table 2.3: Comparison of the sampling strategies for the shown slice

Method	SSIM
Equidistant	0.902
Random	0.795
QDEIM	0.923
DEIM	0.913
CQDEIM	0.923

In this section, we have presented trajectory selection algorithms and show results from our own algorithms. The methods we have proposed are quick and easy to implement, as they do not require any parameters to be set or in-depth knowledge of the object under study. They enable us to improve the quality of reconstruction using a sparse-view strategy. There is no great difference in quality between DEIM and Q-DEIM, the two

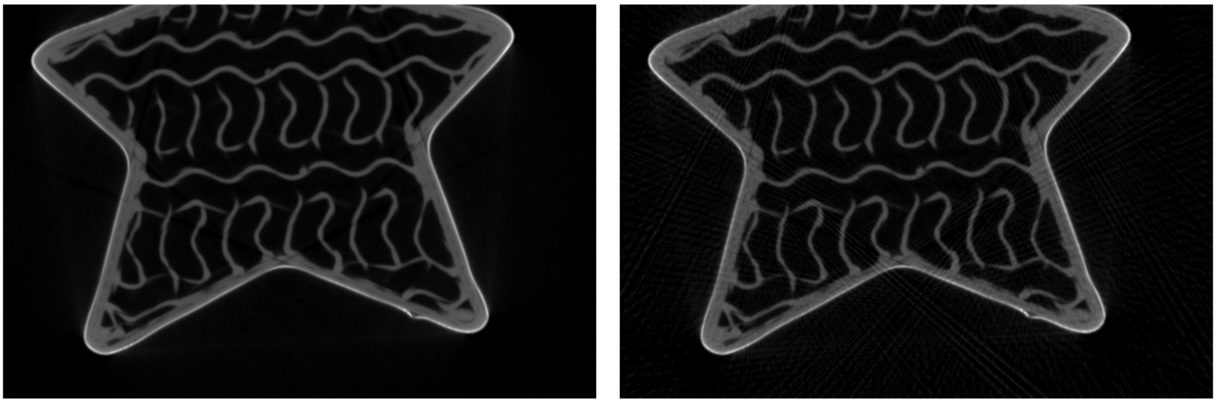


(a) PSNR

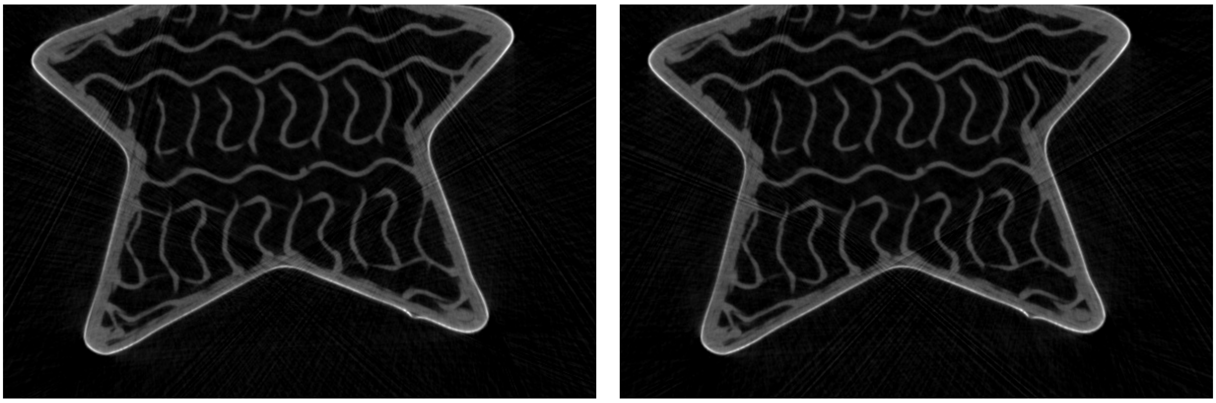


(b) SSIM

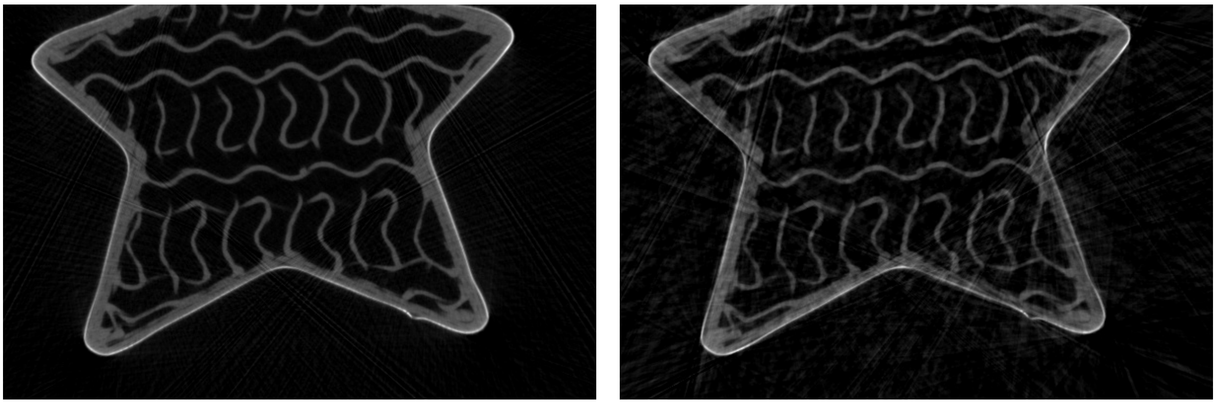
Figure 2.12: Reconstruction quality according to the number of projections used during reconstruction



(a) Dense reconstruction (b) Equidistant sampling



(c) QDEIM (d) DEIM



(e) CQDEIM (f) Random sampling

Figure 2.13: Cross-sections of the reconstructions performed with 300 views according to the various trajectories.

methods can be used interchangeably except when one wants to include a specific task. In this case, CQDEIM should be used. By comparing our previous articles and this study, we can see that the gains made by trajectory optimisation methods can be achieved for different numbers of projections. Parts with very prominent faces, little symmetry and few curves are more amenable to trajectory optimisation, and from the first few tens of views, our methods show a clear difference in reconstruction quality. In the case of symmetrical parts, such as the trophy, with few preferred directions, the trajectory has less of an impact on the reconstruction and the improvement can be seen from a higher number of views. In order to apply the methods presented correctly, it is necessary to know the context in which the experimenter is carrying out its study, particularly in order to determine the most appropriate type of constraint to add. For future developments, we aim to develop new functionalities in trajectory optimisation. Incorporating the notion of ROI in view selection would be particularly beneficial. If a specific area is prone to defects or features requiring diagnosis, enhancing the reconstruction quality in that region becomes crucial. Additional developments are planned to explore trajectories beyond those with constant magnification. Moreover, it would be intriguing to optimise the detector orientation.

Chapter 3

The preliminary step of Registration

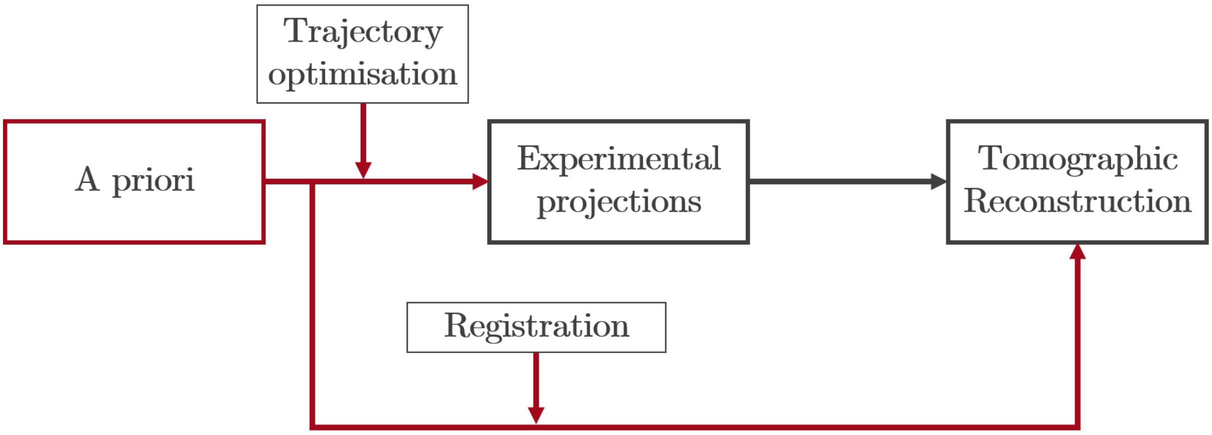


Figure 3.1: Diagram of the tomographic process. In this section, we study the Registration. As we go along in this thesis, we will use *a priori* information in all stages of the process.

As this thesis aims to inject *a priori* information, it is necessary that our prior knowledge corresponds as much as possible to the real object. Having a CAD model aligned with the real object allows not only to compare the object to its model but also to be able to simulate the experimental environment, simulate projections, correct the position of the X-ray source and directly use *a priori* knowledge during the reconstruction, as depicted in Diagram 3.1. For the purpose of aligning the CAD model and the real object, many strategies are possible. The most commonly used option is the 3D/3D registration: the object is reconstructed, and then the volume is registered to the CAD model. This procedure is the most encountered in non-destructive testing as it is expected to directly compare the reconstruction and the CAD model to estimate the surface condition. A second option, called 3D/2D registration, is to register the CAD model from a few experimental projections. 3D/2D registration can be defined as establishing a projection mapping, from a 3D to a 2D coordinate system such that points in each space which correspond to the same physical point are mapped to each other [91, 92]. This strategy can be done without a reconstruction step and can be used to correct experimental projections and the trajectory. For these reasons, our study will focus more on 3D/2D methods[1].

---

[1] The term 3D/2D is sometimes also used for volume-slice registration, we will not talk about these methods, which can be considered an extreme case of 3D/3D registration [92].

3.1 3D/2D Registration with the CAD model

The objective of registration is to align the CAD model with the physical object within a common coordinate system. This necessitates determining a transformation between the CAD coordinate system and the *world* coordinate system. Given that most parts examined in this study are non-deformable, we will focus on rigid transformations. Rigid transformations involve only six parameters, representing a 3D translation and rotations around each axis. Conversely, non-rigid transformations are more complex, requiring many additional parameters to account for deformations. Such deformations are crucial in applications like medical procedures, where pre-interventional data can guide operations. However, in industrial contexts, most samples are non-deformable, making rigid transformations sufficient for part registration. It is initially assumed that the physical part and the CAD model have congruent shapes. Any discrepancies between the two will be addressed subsequently.

The 3D/2D registration can be done in many ways, but we propose classifying them according to their metric [92]. All methods share approximately the same procedure, there is always an iterative optimisation scheme whose metric is updated at each step. The main differences are due to the nature of the strategy to achieve spatial correspondences. There are three main categories:

- If the strategy relies on a *coarse reconstruction* at each iteration and the matching between this reconstruction and the 3D model, the procedure is *reconstruction-based* and is very similar to the 3D/3D registration. A fidelity criterion compares the selected features on the two 3D images (reconstruction and CAD) and updates the transformation [93].
- The registration may rely on *backprojections*, usually, these methods seek to minimise the distance between virtual rays and occluding contours [94].
- The strategy can also rely on *projections*, in this case, the CAD model is used to simulate projections. Then, features on simulated and experimental projections are compared to update the CAD position [95]. Figure 3.2 illustrates a 3D/2D registration based on projections.

For each case, the registration can rely on some fiducial markers or stereotactic frames. Sometimes called *extrinsic* registration, marker-based registrations are easier. Indeed, some correspondences have already been made between CAD 3D points and their 2D experimental projections. In that case, it becomes a *Pose Estimation Problem* whose resolution is relatively easy thanks to the numerous existing methods.

A contrario, for *intrinsic* registration, natural features are used. They can be of any sort, and the simplest are pixels [96] and points. Points-to-points registration methods often serve as a rough initial registration. They require finding particular landmarks in both projections and the 3D object and matching them together. *Curve-to-curve* equivalents also exist and often offer more robust matching. Some methods use non-local features that allow for getting rid of a matching step. Image moments, for example, use the information on the whole image. The features are not only based on the intensity of the image, they can be of any nature. In this thesis, we opted not to add markers to our images. All techniques employed will utilise natural features exclusively.

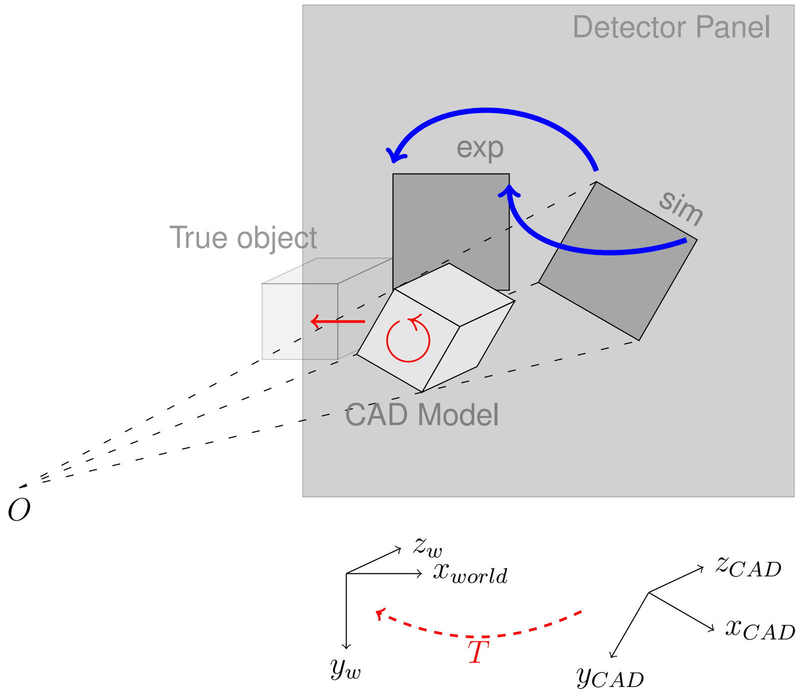


Figure 3.2: Geometric setup of the registration of the 3D CAD model to the experimental projections. In this example, the CAD model of the cube will be registered to the real object by comparing experimental projections to those simulated via the CAD. Here, the registration relies on a point-to-point comparison between the projections. The blue arrows denote correspondences between points in the real and simulated projections. The red arrows symbolise the translation and rotation to be determined in order to establish the mapping '*T*' between the CAD and real-world coordinates.

3.1.1 Pose Estimation Problem

The Pose Estimation problem is one of the most crucial challenges in registration and serves as the foundation for numerous other techniques. It involves determining the pose of a camera relative to a known 3D object based on correspondences between 2D image points and their corresponding 3D points in a scene. In our context, the problem is analogous but with a slight variation. In tomography, the source functions as the camera pinhole and the detector as the focal plane. Figure 3.3 illustrates the similarity between the pinhole camera model and the X-ray imaging setup. Unlike optical imaging problems, where the camera position is unknown, in tomography, the position of the camera, i.e., the X-ray source, is known, while the position of the physical object is unknown. However, it is sufficient to apply the inverse transformation found in the Pose Estimation Problem to the object rather than to the camera.

The Pose Estimation Problem, particularly the *Perspective-n-Points* (PnP) [97], is inherently nonlinear. The PnP problem consists of determining the six degrees of freedom of the camera pose by minimising an objective function that finds the best match between the 3D points and their corresponding 2D projections. PnP is nonlinear due to the relationship involved in the perspective projection of 3D points onto a 2D image plane. While the problem itself is nonlinear, specific methods and approximations may aim to linearise or simplify aspects of the problem for efficient computation. For instance, the

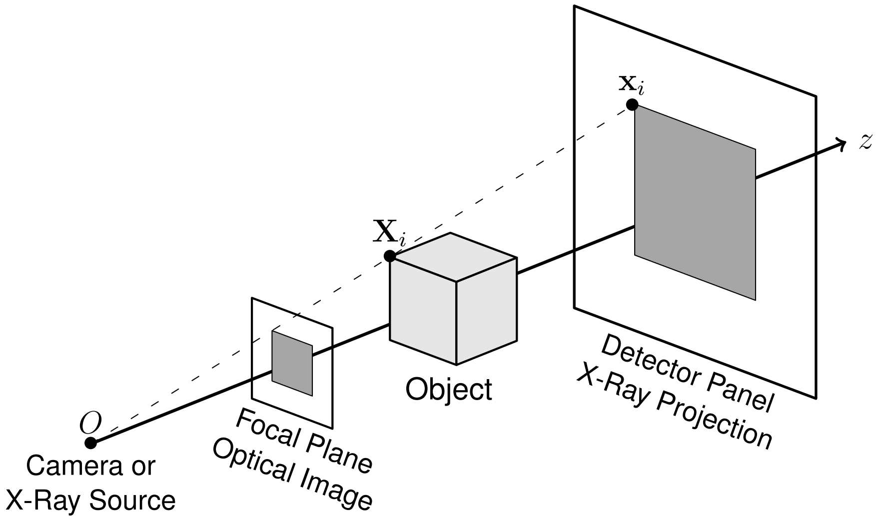


Figure 3.3: Comparison between the pinhole camera model for optical and X-ray imaging. In X-ray tomography, the detector acts as the focal plane, as if the image was formed behind the object. $\mathbf{X}_i$ represents a 3D point of the object and $\mathbf{x}_i$ its projection onto the detector.

Direct Linear Transform (DLT), one of the oldest methods for pose estimation [98, 99], cast the problem as a system and assumes the sought transformation has twelve degrees of freedom. DLT considers the following linear system to find the homography between two coordinates systems:

$$\mathbf{x}_i = \mathbf{\Pi}{}^{\mathbf{c}}\mathbf{T}_{\mathbf{w}}{}^{\mathbf{w}}\mathbf{X}_i, \tag{3.1}$$

where $\mathbf{x}_i$ represents a point onto the detector in homogeneous coordinates, $\mathbf{\Pi}$ is the perspective projection matrix, ${}^{\mathbf{c}}\mathbf{T}_{\mathbf{w}}$ the transformation between the **w**orld and the **c**amera coordinates and ${}^{\mathbf{w}}\mathbf{X}_i$ is an object point in the **w**orld coordinates in homogeneous coordinates. Generally, with the camera model, the third dimension of $\mathbf{x}_i$ is set to one and there are camera intrinsic parameters in the perspective matrix which normalise the vectors. The camera intrinsic parameters also take into account the focal lengths and an eventual skew parameter but, in an X-ray framework, those parameters do not appear. So in our case, instead of normalising the vectors, we will fill the third dimension of $\mathbf{x}_i$ with the source-detector distance d. The formulation of the points on the detector becomes: $\mathbf{x}_i = (x_i, y_i, d)$. The operators in Equation 3.1 can be rewritten as:

$$\mathbf{x}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} {}^{\mathbf{c}}\mathbf{R}_{\mathbf{w}} & {}^{\mathbf{c}}\mathbf{t}_{\mathbf{w}} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix} {}^{\mathbf{w}}\mathbf{X}_i, \tag{3.2}$$

where the transformation ${}^{\mathbf{c}}\mathbf{T}_{\mathbf{w}}$ is represented by the rotation matrix ${}^{\mathbf{c}}\mathbf{R}_{\mathbf{w}}$ and the translation ${}^{\mathbf{c}}\mathbf{t}_{\mathbf{w}}$. To show the homography ${}^{\mathbf{c}}\mathbf{H}_{\mathbf{w}}$ associated with the sought mapping, the equation can also be rewritten as:

$$\mathbf{x}_i = {}^{\mathbf{c}}\mathbf{H}_{\mathbf{w}}{}^{\mathbf{w}}\mathbf{X}_i. \tag{3.3}$$

The system is then transformed using the cross-product:

$$\mathbf{x}_i \times ({}^{\mathbf{c}}\mathbf{H}_{\mathbf{w}}{}^{\mathbf{w}}\mathbf{X}_i) = \mathbf{0}. \tag{3.4}$$

This equation is linear with the unknown $^{\mathbf{c}}\mathbf{H}_{\mathbf{w}}$. To facilitate analysis and understanding, we will rewrite this equation as the following linear system:

$$\mathbf{A}\mathbf{h} = \mathbf{0}, \tag{3.5}$$

$$\begin{pmatrix} \mathbf{0}^{\mathsf{T}} & -d\mathbf{X}_i^{\mathsf{T}} & y_i\mathbf{X}_i^{\mathsf{T}} \\ d\mathbf{X}_i^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}} & -x_i\mathbf{X}_i^{\mathsf{T}} \\ -y_i\mathbf{X}_i^{\mathsf{T}} & x_i\mathbf{X}_i^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}} \end{pmatrix} \begin{pmatrix} \mathbf{h_1} \\ \mathbf{h_2} \\ \mathbf{h_3} \end{pmatrix} = \mathbf{0}, \tag{3.6}$$

where $\mathbf{h_j}$ is the jth row of $^{\mathbf{c}}\mathbf{H}_{\mathbf{w}}$. It is important to acknowledge that although every set of coordinate matches results in a set of three equations, only two of these equations are linearly independent. Usually, the third one is disregarded.

Theoretically, since six independent parameters can represent the pose, three points should be sufficient to solve this problem. Many methods, called P3P, use only three points to deduce the pose. In practice, they are not robust because of the orthogonal constraint on the rotation matrix. Therefore, as we explained above, twelve unknowns are used instead of six for the homography. By consequence, as $\mathbf{A}$ has rank 8 [98], four points correspondences are needed for determining a unique solution. PnP can use as many points as available, but the system is over-determined. To determine the best homography, SVD is used along with the Kabsch-Umeyama algorithm [100] (more details in Section 3.2.1). Although there are many well-known solutions for the pose estimation problem, PnP algorithms are the most common as the pose accuracy increases with the number of points. Still, for more accuracy, PnP often needs to be combined with an outlier rejection process like the Random Sample Consensus (RANSAC) [101]. We will note, however, that P3P is sometimes preferred over the other methods for its rapid computational speed and requirement of only three correspondences [97, 102].

The POSIT algorithm, introduced by Dementhon [103], is another method to address the PnP problem. As with other methods, it assumes known 3D-2D point correspondences. POSIT combines two algorithms. The first one, POS (Pose from Orthography and Scaling), fastens the procedure and makes the problem linear by approximating the perspective projection with a scaled orthogonal projection. The second POSIT (POS with ITerations), an iterative technique, employs the previously approximated pose to compute improved scaled orthographic projections of the feature points. The algorithm's advantages lie in its computational efficiency and iterative refinement approach, obviating the necessity for an initial guess. However, POSIT assumes a perspective camera model, which may not be universally applicable, particularly in coplanar points scenarios. Additionally, it might converge to local minima if the initial estimate significantly deviates from the correct pose. A notable improvement of POSIT is SoftPOSIT, which simultaneously determines the pose and the correspondences between points [104]. SoftPOSIT merges the iterative softassign algorithm for computing correspondences [105] and the POSIT algorithm.

In the previous examples, homography was defined as the solution of a system. However, it is also possible to define pose estimation as a minimisation problem with regard to translation and rotation parameters. These so-called iterative methods are generally more complex, require a good initialisation to avoid local minima and require a stopping criterion. However, they allow much greater flexibility in terms of minimising errors. Iterative algorithms have the advantage that they can be used as long as the function to be minimised is well-defined. For example, in visual servoing, the distance between the real and simulated projected points is often chosen, but any other feature can also be used [106, 107]. Denoting $\mathbf{s} = ({}^{\mathbf{c}^*}\mathbf{t}_{\mathbf{w}}, \theta\mathbf{u})$, the representation of the translations and rotations parameters defining the rigid transformation ${}^{\mathbf{c}}\mathbf{T}_{\mathbf{w}}$, the pose estimation problem solves for:

$$\mathbf{s}^* = \underset{\mathbf{s}}{\operatorname{argmin}} \sum_i \|\mathbf{x}_i - \mathbf{\Pi}\,{}^{\mathbf{c}}\mathbf{T}_{\mathbf{w}}(\mathbf{s})\,{}^{\mathbf{w}}\mathbf{X}_i\|_2^2. \tag{3.7}$$

A major advantage of iterative methods is that the Jacobian of the function to be minimised can be approximated, which greatly improves the algorithms' convergence speed (more details in Section 3.1.2).

So far, we have assumed that correspondences between the image points and the 3D object are already known, but the pose estimation can also be solved with natural feature-based methods. Those techniques involve the following steps: key points extraction, feature description, matching and solving for the transformation. Considered a breakthrough in key points extraction, the scale-invariant feature transform (SIFT), proposed by David Lowe [108], allows to detect, describe, and match local features in images. First, SIFT uses the *Difference of Gaussians* (DoG) technique to identify potential key points across different levels of image blurring. The algorithm convolves the image with Gaussian filters at different scales and subtracts adjacent blurred images to generate a DoG pyramid. Local extrema in the DoG pyramid (points where a pixel is higher or lower than its neighbours across scale and space) are considered potential key points. Once potential key points are detected, SIFT applies a detailed localisation step to refine them. It examines the DoG pyramid to eliminate low-contrast key points and points on edges and discard points that are not sufficiently defined or stable across scales. Key points are further refined by fitting a 3D quadratic function to nearby sample points, determining sub-pixel accuracy in localisation. Then, SIFT assigns an orientation feature to each key point, which consists of assigning one or more orientations determined locally on the image from the direction of the gradients in a neighbourhood around the point. At this stage, SIFT constructs a descriptor for each key point, representing its local image region. A window around the key point is divided into subregions, and gradient histograms for each subregion are generated. These histograms are concatenated to form a high-dimensional vector that serves as the descriptor for the key point. Once descriptors are computed for key points in multiple images, SIFT enables the matching of corresponding key points across images. Figure 3.4 shows the results of the key points matching between two pictures taken at different angles [5]. Despite a few exceptions, most matches are correct. Finally, a PnP method is used to compute the transformation between the images. Given that the method generates many correspondences, the homography calculation must be robust and requires RANSAC.

The methods cited here are merely examples, yet they illustrate the general framework of registration techniques. Typically, correspondences between features are established, followed by a resolution method that provides the desired transformation. Visual servoing is no exception to this approach but also introduces numerous other aspects that we will utilise in our methods.

3.1.2 Visual servoing

Visual servoing is initially a technique for controlling the motion of a robot using visual information feedback from cameras. This technique can be applied in various fields such

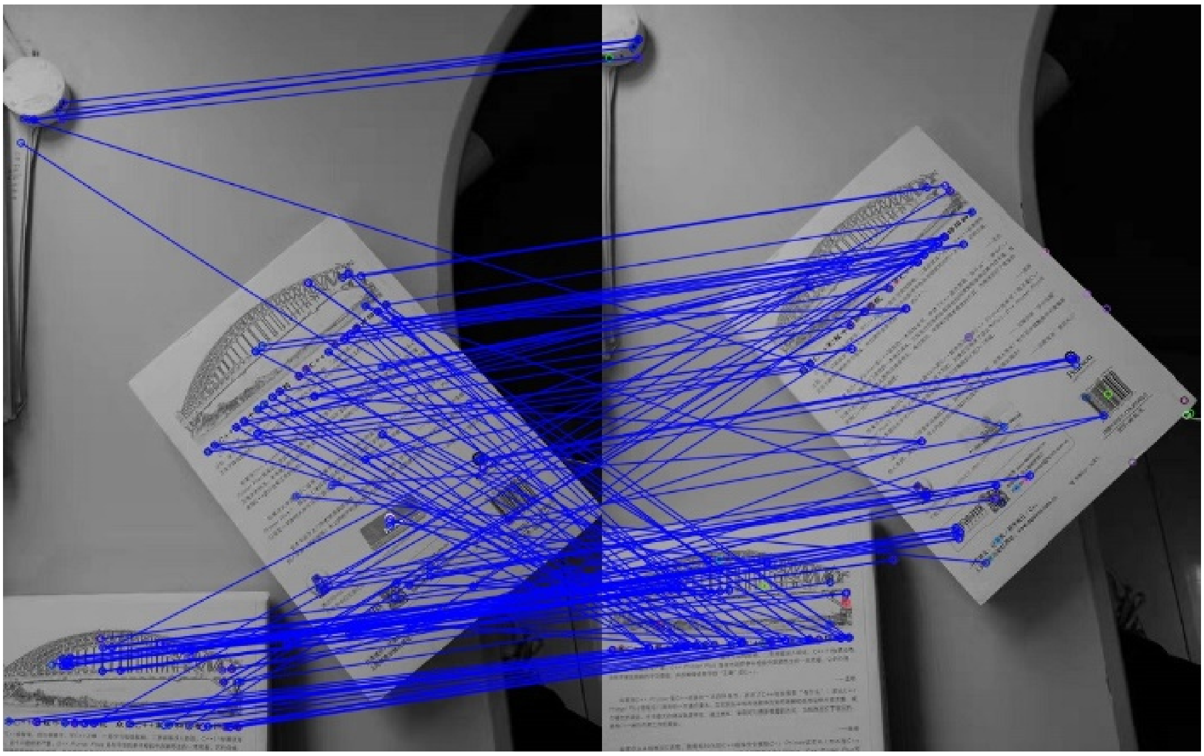


Figure 3.4: Results of the correspondences found by the SIFT algorithm between two images [5].

as industrial automation, robotics, object tracking, and navigation. Thanks to the similarities between camera systems and X-ray setup, this technique can be used for X-ray pose estimation too. Its formulation is general and encompasses many situations, the visual control scheme aims to minimise the error $\mathbf{e}$ defined as:

$$\mathbf{e} = (\mathbf{s}(\mathbf{x}) - \mathbf{s}^*), \tag{3.8}$$

where $\mathbf{x} = (x, y, z, \omega_x, \omega_y, \omega_z)$ represents the pose of the camera (i.e. X-ray source), the target $\mathbf{s}^*$ and $\mathbf{s}$ are either a set of features in the images or a pose, which must be estimated from image measurements. In the first case, the approach is called Image Based Visual Servoing (IBVS) and in the latter, it is called Pose Based Visual Servoing (sometimes Position) (PBVS) [106]. Within the framework of visual servoing, our focus shifts towards the position and orientation of the camera/source. Consequently, in this section, $\mathbf{x}$ denotes the source's position rather than the points of projection, as previously discussed.

Once the features on the projections have been selected, the control scheme can be straightforwardly deduced. Initially, the goal is to define $\mathbf{L_s} = \frac{\partial \mathbf{s}}{\partial \mathbf{x}}$, which represents the *interaction matrix*, also referred to as the *feature Jacobian*. This matrix delineates the alterations in the features $\mathbf{s}$ corresponding to the camera/source movement and, in conjunction with the chosen features, plays a crucial role in effectively guiding the system. It facilitates the determination of the camera's kinematic screw $\mathbf{v}$ required to minimise the visual error. The features time derivative is:

$$\begin{aligned}\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} &= \frac{\partial \mathbf{s}}{\partial \mathbf{x}}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} + \frac{\partial \mathbf{s}}{\partial t} \\ \frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} &= \mathbf{L}_{\mathbf{s}}\mathbf{v} + \frac{\partial \mathbf{s}}{\partial t}.\end{aligned} \tag{3.9}$$

As the object is motionless, the last term is zero. Using Eq. 3.8, the relationship between the error and the displacement $\mathbf{v} = (\mathbf{v}_c, \boldsymbol{\omega}_c)$ appears:

$$\frac{\mathrm{d}\mathbf{e}}{\mathrm{d}t} = \frac{\partial \mathbf{e}}{\partial \mathbf{s}}\frac{\partial \mathbf{s}}{\partial \mathbf{x}}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{L}_{\mathbf{e}}\mathbf{v}, \tag{3.10}$$

where $\mathbf{L}_{\mathbf{e}} = \mathbf{L}_{\mathbf{s}}$ [2]. If an exponential decrease in error is forced on the system, it gives the equation:

$$\frac{\mathrm{d}\mathbf{e}}{\mathrm{d}t} = -\lambda \mathbf{e}, \tag{3.11}$$

where λ is a positive constant. Combining Equations 3.10 and 3.11, we obtain a formulation of the displacement to be applied to the system:

$$\mathbf{v} = -\lambda {\mathbf{L}_{\mathbf{e}}}^{+}(\mathbf{s} - \mathbf{s}^*), \tag{3.12}$$

$$\frac{\mathrm{d}\mathbf{e}}{\mathrm{d}t} = -\lambda \mathbf{L}_{\mathbf{e}}{\mathbf{L}_{\mathbf{e}}}^{+}(\mathbf{s} - \mathbf{s}^*). \tag{3.13}$$

In practical implementations of visual servoing control laws, achieving perfect knowledge of either $\mathbf{L}_e$ or its pseudoinverse $\mathbf{L}_e^+$ is unattainable. Consequently, an approximation or estimation of one of these matrices becomes necessary. Henceforth, we denote both the pseudoinverse of the estimated interaction matrix and the estimated pseudoinverse of the interaction matrix using the symbol $\widehat{\mathbf{L}_{\mathbf{e}}}$. Employing this notation, the control law is effectively represented as:

$$\mathbf{v} = -\lambda \widehat{\mathbf{L}_{\mathbf{e}}}^{+}\mathbf{e}. \tag{3.14}$$

This formulation encapsulates the utilisation of an estimated or approximated pseudoinverse of the interaction matrix $\mathbf{L}_e$ within the control law. When $\mathbf{L}_{\mathbf{e}}\widehat{\mathbf{L}_{\mathbf{e}}}^{+}$ is positive, it indicates a decrease in the error. Conversely, if the result is negative, it suggests an increase in the error. Further explanation of this concept is provided in [106]. Figure 3.5 illustrates the visual servoing loop.

Image-Based Visual Servoing (IBVS)

To illustrate how visual servoing works and how the interaction matrix is derived, we will explain the example of servoing where the features are points. This is an example of IBVS. As before, we will call d the source-detector distance, $\mathbf{x}$ the coordinates of a point on the detector and $\mathbf{X}$ the corresponding point on the 3D object. For conventional cameras, the source-detector distance must be replaced by the lens's focal length. The perspective projection gives the equation:

$$\mathbf{x} = \frac{d}{Z}\mathbf{X}. \tag{3.15}$$

---

[2] Although this point is not elaborated in detail here, it is crucial for comprehending the method. For further information, the reader is encouraged to consult the following reviews. [106, 107].

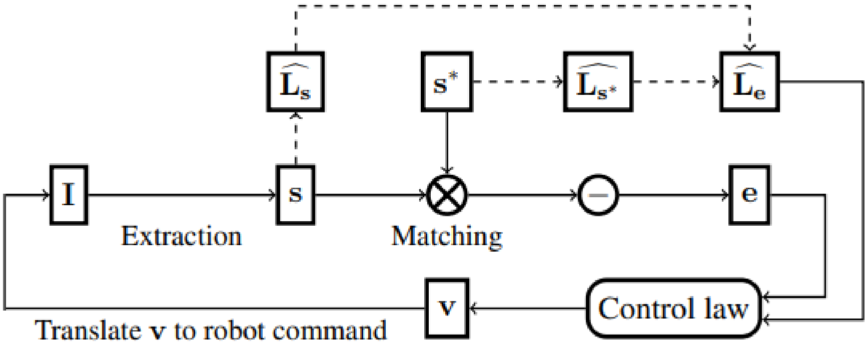


Figure 3.5: Visual Servoing Loop. From the image $\mathbf{I}$, a set of features $\mathbf{s}$ is extracted. The error $\mathbf{e} = (\mathbf{s}^* - \mathbf{s})$ is minimised thanks to the interaction matrix $\widehat{\mathbf{L_e}}$ which allows to compute $\mathbf{v}$ to move the camera position.

Note that the value of Z is approximate; it is not assumed to be known at the beginning of the algorithm. Furthermore, the velocity of the 3D point can be expressed as a function of the camera spatial velocity as follows:

$$\dot{\mathbf{X}} = -\mathbf{v}_c - \boldsymbol{\omega}_c \times \mathbf{X}. \tag{3.16}$$

Using both Equations 3.15 and 3.16, we obtain:

$$\dot{x} = -\frac{d}{Z}v_x + \frac{x}{Z}v_z + \frac{xy}{d}\omega_x - d(1+\frac{x^2}{d^2})\omega_y + y\omega_z, \tag{3.17}$$

$$\dot{y} = -\frac{d}{Z}v_y + \frac{y}{Z}v_z + d(1+\frac{y^2}{d^2})\omega_x - \frac{xy}{d}\omega_y - x\omega_z. \tag{3.18}$$

Rearranging the terms gives the desired result $\dot{\mathbf{x}} = \mathbf{L_e v}$, with:

$$\mathbf{L_e} = \begin{pmatrix} -\frac{d}{Z} & 0 & \frac{x}{Z} & \frac{xy}{d} & -(1+\frac{x^2}{d^2})d & y \\ 0 & -\frac{d}{Z} & \frac{y}{Z} & (1+\frac{y^2}{d^2})d & -\frac{xy}{d} & -x \end{pmatrix}. \tag{3.19}$$

The interaction matrix is contingent upon the object's position, a factor that remains unknown. Hence, an approximation, denoted as $\widehat{\mathbf{L_e}}$, addresses this limitation as explained before. The interaction matrix can also be approximated by $\mathbf{L_{e^*}}$, but this choice impacts the trajectory. To comprehend how to derive the interaction matrix for any generic feature, more details are provided in Appendix A.1.

The primary principle underlying this control methodology aims to minimise the error norm, enhancing the precision of both tracking and control:

$$\mathcal{L} = \frac{1}{2}\|\mathbf{e}\|_2^2. \tag{3.20}$$

When dealing with six features, assuming both $\mathbf{L_e}$ and $\widehat{\mathbf{L_e^+}}$ possess full rank, global asymptotic stability is achieved, indicated by:

$$\dot{\mathcal{L}} = \mathbf{e}^\mathsf{T}\dot{\mathbf{e}} = -\lambda\mathbf{e}^\mathsf{T}\mathbf{L_e}\widehat{\mathbf{L_e^+}}\mathbf{e}. \tag{3.21}$$

However, if the number of features exceeds six, the same criterion provides local asymptotic stability around $\mathbf{e}^*$. The global asymptotic stability of the system is guaranteed if

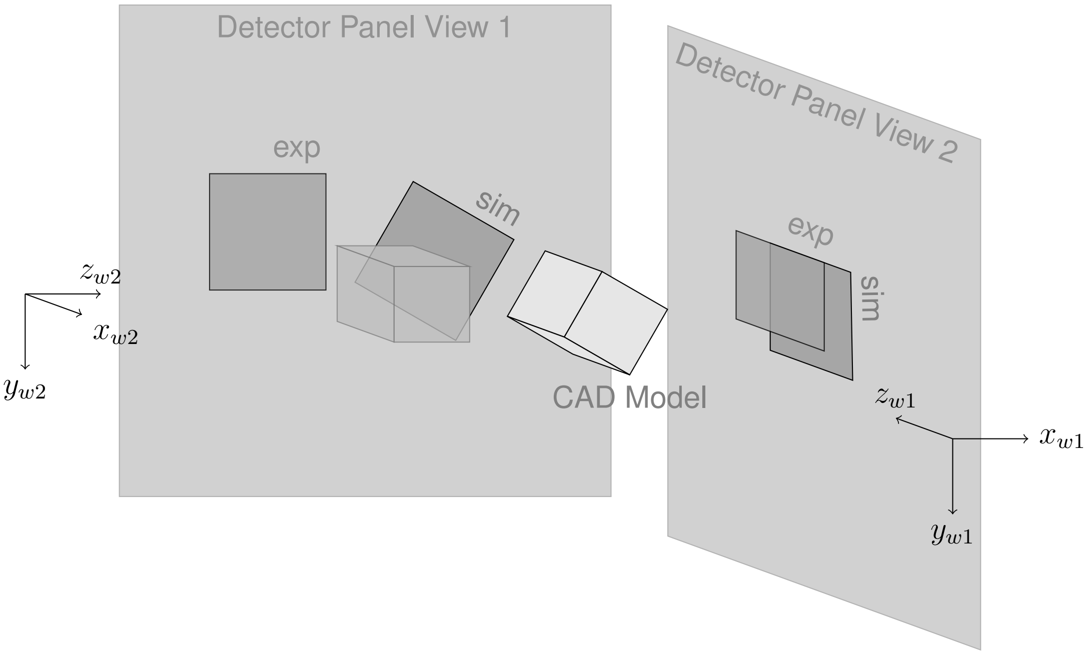


Figure 3.6: Stereovision system. By leveraging multiple views, the visual servoing approach mitigates ambiguities and avoids local minima. The goal is to converge the current features (from simulated views) with the desired ones (from experimental views) across both perspectives.

$\mathbf{L_e}\widehat{\mathbf{L_e^+}} > 0$ in the sense that a matrix $\mathbf{A}$ is considered positive if $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} > 0 \; \forall \; \mathbf{x}$. This condition aligns with Lyapunov stability principles.

If multiple projections are available for registration, we can concurrently utilise the features. This scenario closely resembles stereovision or multi-camera setups in visual servoing. Figure 3.6 shows an experimental setup with two perpendicular projections available for registration. The interaction matrix of the second view is expressed in the frame of the first view with:

$$^1\mathbf{L_{e_2}} = {}^2\mathbf{L_{e_2}} \cdot {}^2\mathbf{W}_1, \tag{3.22}$$

$$^2\mathbf{W}_1 = \begin{pmatrix} ^2\mathbf{R}_1 & [^2\mathbf{t}_1]_\times \, ^2\mathbf{R}_1 \\ \mathbf{0} & ^2\mathbf{R}_1 \end{pmatrix}. \tag{3.23}$$

Where $[\mathbf{t}]_\times$ is the skew-symmetric matrix associated with the vector $\mathbf{t}$ and $(^2\mathbf{R}_1, ^2\mathbf{t}_1)$ is the rigid-body transformation between the coordinates frames, $^1\mathbf{L_{e_2}}$ and $^2\mathbf{L_{e_2}}$ are the interaction matrices of the second view in the first and second frame respectively [109].

$$\dot{\mathbf{e_s}} = \begin{pmatrix} \dot{\mathbf{e_1}} \\ \dot{\mathbf{e_2}} \end{pmatrix} = \begin{pmatrix} \mathbf{L_{e_1}} \\ \mathbf{L_{e_2}} {}^2\mathbf{W}_1 \end{pmatrix} \mathbf{v} = \mathbf{L_{e_s}} \mathbf{v}. \tag{3.24}$$

Note that the epipolar constraints might link some equations.

Pose Based Visual Servoing

In image-based methods, the geometric model of the studied object is not explicitly used, only its projection and associated features. In X-rays, the positions of the source and the detector are well-known, equivalent to knowing the intrinsic camera parameters, allowing the use of the geometric model. PBVS involves a 3D reconstruction step.[3]

PBVS aims to directly estimate the source's (or camera's in conventional servo control) position and orientation. Essentially, PBVS mirrors the problem of repositioning the camera based on visual data at hand. Accurate estimation of the camera's pose demands a 3D model of the object and the camera's intrinsic parameters. This 3D model serves as a reference, facilitating the estimation of spatial relations between the camera and the object. Additionally, formulating the servoing error and computing the interaction matrix is crucial for effective servoing based on the estimated camera pose.

With ${}^{\mathbf{c}^*}\mathbf{t_c}$ the translation between the target and current frames of reference, and ${}^{\mathbf{c}^*}\mathbf{R_c}$ the rotation matrix between these frames, we define $\mathbf{s} = ({}^{\mathbf{c}^*}\mathbf{t_c}, \theta\mathbf{u})$ is feasible. In this formulation, $\mathbf{s}^* = \mathbf{0}$ and $\mathbf{s} = \mathbf{e}$. The servoing can be modelled with:

$$\mathbf{L_e} = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{L}_{\theta\mathbf{u}} \end{pmatrix}, \tag{3.25}$$

$$\mathbf{L}_{\theta\mathbf{u}} = \mathbf{I_3} + \frac{\theta}{2}[\mathbf{u}]_\times + \left(1 - \frac{\text{sinc } \theta}{\text{sinc}^2 \frac{\theta}{2}}\right)[\mathbf{u}]^{\mathbf{2}}_\times. \tag{3.26}$$

Thereby, the formula for the displacement gives:

$$\mathbf{v} = -\lambda \widehat{\mathbf{L}^+_{\mathbf{e}}} \mathbf{e}. \tag{3.27}$$

$$\begin{cases} \mathbf{v_c} = & -\lambda \mathbf{R}^{\top}\,{}^{\mathbf{c}^*}\mathbf{t_c} \\ \boldsymbol{\omega}_c = & -\lambda\theta\mathbf{u} \end{cases}. \tag{3.28}$$

Note the decoupling between translational and rotational motions. Pose-based visual servoing is particularly intriguing since, under perfect parameter estimation and given $\theta \neq 2n\pi$ with $n \in \mathbb{N}$, then $\mathbf{L_e}\widehat{\mathbf{L}^+_{\mathbf{e}}} = \mathbf{I_6}$, ensuring global asymptotic stability [106, 107].

3.2 Our methods

To address the registration, we propose two distinct methods. The first, termed *Convex Hulls Iterative Inverse Perspective Matching*, constitutes a PBVS approach. We prioritise this method because it is very well suited to components typical of the industry with polygonal parts. Given that PBVS methods do not necessitate projections computation, we have opted for this rapid method that circumvents this requirement. Notably, CAD models can be cumbersome, often necessitating preliminary steps such as decimation [112] or voxelization [6].

Subsequently, we present our IBVS method known as *Robust 3D/2D hybrid registration*. This method amalgamates point-based and image moment-based servoing techniques to yield improved convergence precision. The outcomes and evaluations of these two methods are already documented in our previous publications [12, 13] and we present here new results on the experimental data already considered in the previous chapters.

---

[3] Note that combining PVBS and IBVS is possible [107], such as in the $2^{1/2}$D visual servoing [110], or Degrees Of Freedom partitioned [111]. A distinction exists between visual servo methods (image and pose) and look-and-move methods. Look-and-move methods incorporate joint feedback. That is, a position is given to the robot to reach autonomously following the governing control laws, without visual feedback, and then the robot looks again and moves again. In contrast, visual servoing lacks joint feedback. The control is purely visual.

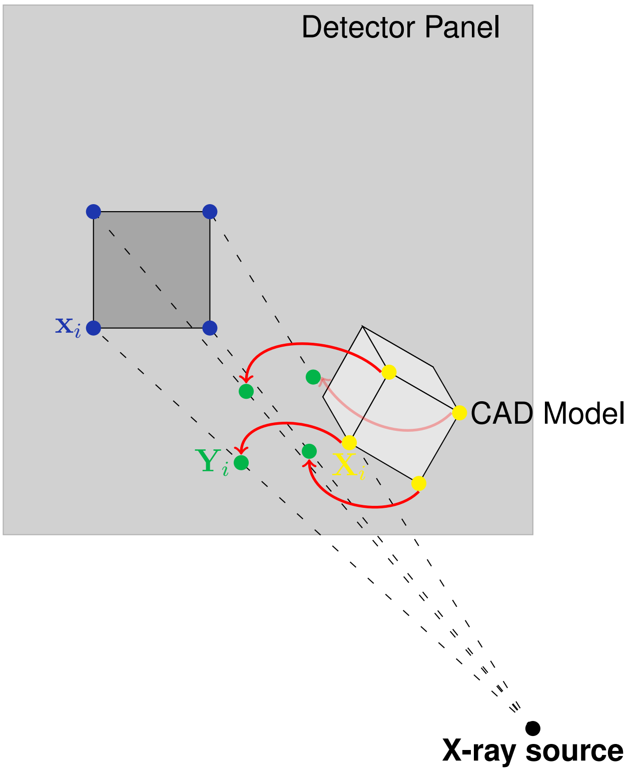


Figure 3.7: Illustration of the Registration setup. $\mathbf{M} = \{\mathbf{X}_i\}_i, \mathbf{F} = \{\mathbf{x}_j\}_j$.

3.2.1 Convex Hulls Iterative Inverse Perspective Matching

Inverse perspective matching is initially introduced by Wunsch et al. [113]. This method resembles the *Iterative Closest Point* (ICP) algorithm, albeit with a distinct approach involving the establishment of correspondences between lines and points. Inverse perspective matching fundamental concept is to establish correspondences between 2D image features, denoted as **F** on the imaging detector and the 3D points representing the CAD model, denoted as **M**. However, these correspondences are established within the 3D spatial domain, rather than in the 2D image space, as was the case with IBVS or PnP.

For each image feature point, $\mathbf{x} = (x_u, x_v) \in \mathbf{F}$, identified in the experimental projection, a ray $\mathcal{P}^{-1}(\mathbf{x})$ is back-projected from the X-ray source point using the following equation:

$$\mathcal{P}^{-1}(\mathbf{x}) = \{\mathbf{y} \mid \mathbf{y} = \lambda(x_u, x_v, d)^\mathsf{T}, \lambda \in \mathbb{R}\}. \tag{3.29}$$

Here, d denotes the source-detector distance, x_u and x_v represent the image coordinates of the point $\mathbf{x}$. It is assumed that the X-ray source is situated at the origin, and the detector is oriented perpendicularly to the source-detector axis.

Figure 3.7 illustrates the procedure. In this figure, the blue points on the detector panel represent image features **F**, each associated with a ray. Simultaneously, in Figure 3.7, yellow dots symbolise 3D points from the CAD model. The matching process involves associating each 3D CAD point with its closest ray and, consequently, with an image feature. For each correspondence established, the 3D CAD point is orthogonally projected onto its corresponding ray, resulting in the formation of the points cloud denoted as **Y** (depicted as green dots). This process ensures that for every $\mathbf{X}_i$ within the set $\mathbf{M}$, a 3D counterpart $\mathbf{Y}_i \in \mathcal{P}^{-1}(\mathbf{x})$ is defined with:

$$||\mathbf{X}_i - \mathbf{Y}_i|| = \min_{\mathbf{x} \in \mathbf{F}} ||\mathbf{X}_i - \mathcal{P}^{-1}(\mathbf{x})|| \ . \tag{3.30}$$

The point clouds $\mathbf{M}$ and $\mathbf{Y}$ are represented in the same coordinate system and have an identical number of points. Consequently, the problem has transitioned into a 3D/3D registration, generally more straightforward than the initial 3D/2D registration task. The primary objective function to be minimised, denoted as $\mathcal{L}$, can now be expressed as follows:

$$\mathcal{L}(\mathbf{R}, \mathbf{t}) = \sum_i \operatorname{dist}(\mathbf{Y}_i - (\mathbf{R}\mathbf{X}_i + \mathbf{t})) \ . \tag{3.31}$$

Here, $\mathbf{R}$ represents a rotation matrix, and $\mathbf{t}$ denotes a translation vector. Furthermore, in order to enhance the method's robustness, reduce the impact of outliers, and ensure numerical stability, the employed distance $\operatorname{dist}(\cdot, \mathbf{0})$ is the Huber M-estimator function, whose parameter k depends on the experimental setup and the 3D model to be aligned [114]:

$$\operatorname{dist}_k(x) = \begin{cases} \frac{x^2}{2} & \text{if } |x| < k, k \in \mathbb{R}^+ \\ k(|x| - \frac{k}{2}) & \text{if } |x| \geqslant k \end{cases} \ . \tag{3.32}$$

Equation 3.31 can be reformulated as a weighted least squares problem [115] and solved using the Kabsch-Umeyama algorithm [100], described below. This method yields the rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$ that minimises the loss function $\mathcal{L}$. The optimisation scheme divides the resolution into two stages: first, a translation is performed on the set of points $\mathbf{M}$ by $\mathbf{t}$, aligning its centroid with that of $\mathbf{Y}$. Subsequently, the covariance matrix $\mathbf{G}$ is computed using the equation:

$$\mathbf{G} = \mathbf{Y}^\intercal(\mathbf{M} - \mathbf{t}). \tag{3.33}$$

The optimal rotation between the two sets is theoretically given by:

$$\mathbf{R} = \left(\mathbf{G}^\intercal\mathbf{G}\right)^{\frac{1}{2}} \mathbf{G}^{-1}. \tag{3.34}$$

However, this computation is challenging. Instead, an SVD decomposition is applied to the covariance matrix:

$$\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal. \tag{3.35}$$

$\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Sigma}$ is diagonal. The desired rotation matrix is obtained as:

$$\mathbf{R} = \mathbf{V} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix} \mathbf{U}^\intercal. \tag{3.36}$$

The sign of the last coefficient is determined by $\operatorname{sign}\left(\det\left(\mathbf{V}\mathbf{U}^\intercal\right)\right)$. Each point $\mathbf{X}_i \in \mathbf{M}$, the CAD model, is updated using the transformation equation:

$$\mathbf{X}_i' = \mathbf{R}\mathbf{X}_i + \mathbf{t}. \tag{3.37}$$

The control scheme and position update process are relatively straightforward and computationally efficient. However, practical convergence becomes uncertain when the

Iterative Perspective Matching is applied as is. The crucial step of establishing point correspondences greatly influences the accuracy of displacement calculations. Incorrect associations can lead to erroneous movements. We decided to employ two projections to enhance the method's robustness, a strategy noted for its efficacy in improving convergence [113].

Moreover, the great particularity of our implementation is that we limit our focus solely to the convex hull of the 3D CAD model and the convex hull of the experimental projection. These points hold particular significance in our context. Firstly, they are relatively sparse for industrial parts. Challenges may arise with curves, as in CAD models defined by meshes such as STL files, where all points along a curve can potentially contribute to the convex hull. Furthermore, it can be demonstrated that the points constituting the convex hull in the CAD model will also form the convex hull in the 2D projection (see Appendix A.2).

To further reduce the number of points, the 3D hull points are themselves clipped. Typically, numerous points in the 3D hull are not responsible for the convex hull of the projection. Hence, the 3D convex hull is projected onto the detector in each iteration. Only the points contributing to the projections convex hulls are retained. This process eliminates points incapable of generating matches, enhancing method stability by curtailing erroneous matches, which can prolong the minimisation step due to significant errors. A proof of convergence is provided in Appendix A.3. Algorithm 7 recaps the whole procedure.

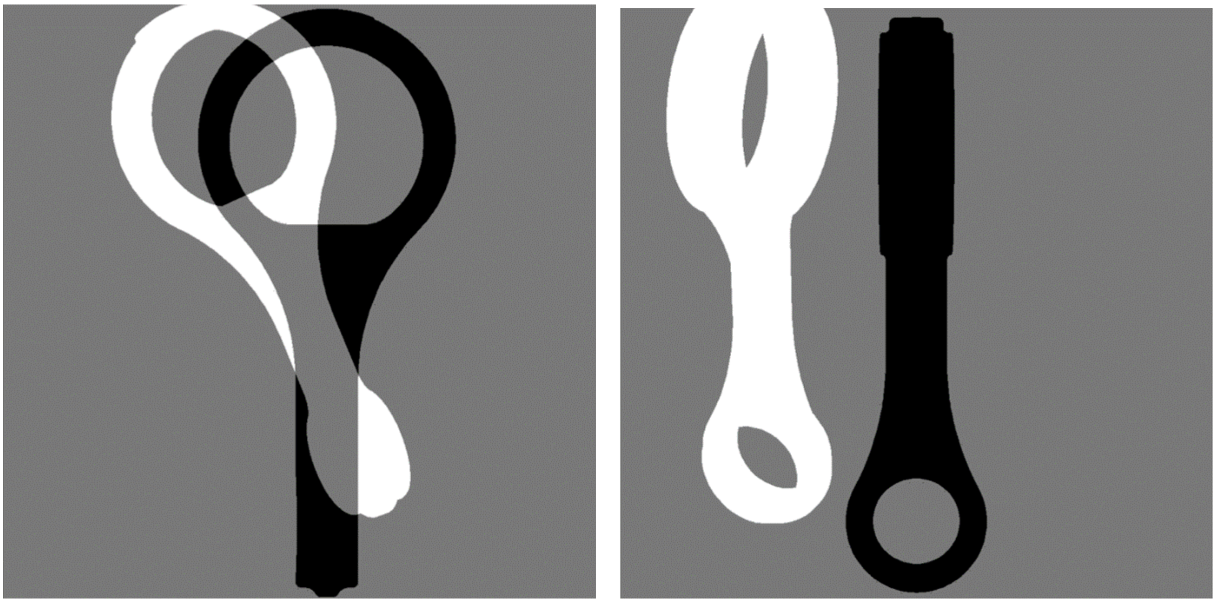
A crucial detail that has not been addressed is the initialisation of the object's position. Effective initialisation is essential for facilitating convergence and ensuring the success of the method. In this approach, poor initialisation can be particularly problematic, potentially causing the CAD object to move out of the X-ray beam's field of view. In our method, the initial orientation of the object is set randomly, and the CAD object is translated so that its bounding box fits within the field of view of each detector position. Given the method's speed, if the object moves out of the field of view, the process can be restarted with a different random orientation.

Algorithm 7: Convex Hulls Iterative Inverse Perspective Matching

```
Data: Image features $\mathbf{F}$, model shape $\mathbf{M}$, experimental setup parameters
Result: Registered CAD model $\mathbf{M}'$
Compute CAD 3D convex hull
Compute target projections convex hulls
while the convergence criteria have not been met do
    Curtail the 3D convex hull to keep only points of interest
    Compute $\mathbf{Y}$ by establishing matches between $\mathbf{M}$ and $\mathbf{F}$
    Solve $\underset{\mathbf{R},\mathbf{t}}{\mathrm{argmin}} \sum_{\mathbf{X}_i \in \mathbf{M}} \|\mathbf{Y}_i - \mathbf{R}\mathbf{X}_i - \mathbf{t}\|^2$ for $\mathbf{R}$ and $\mathbf{t}$
    Update $\mathbf{M} \leftarrow \mathbf{R}\mathbf{M} + \mathbf{t}$
end
```

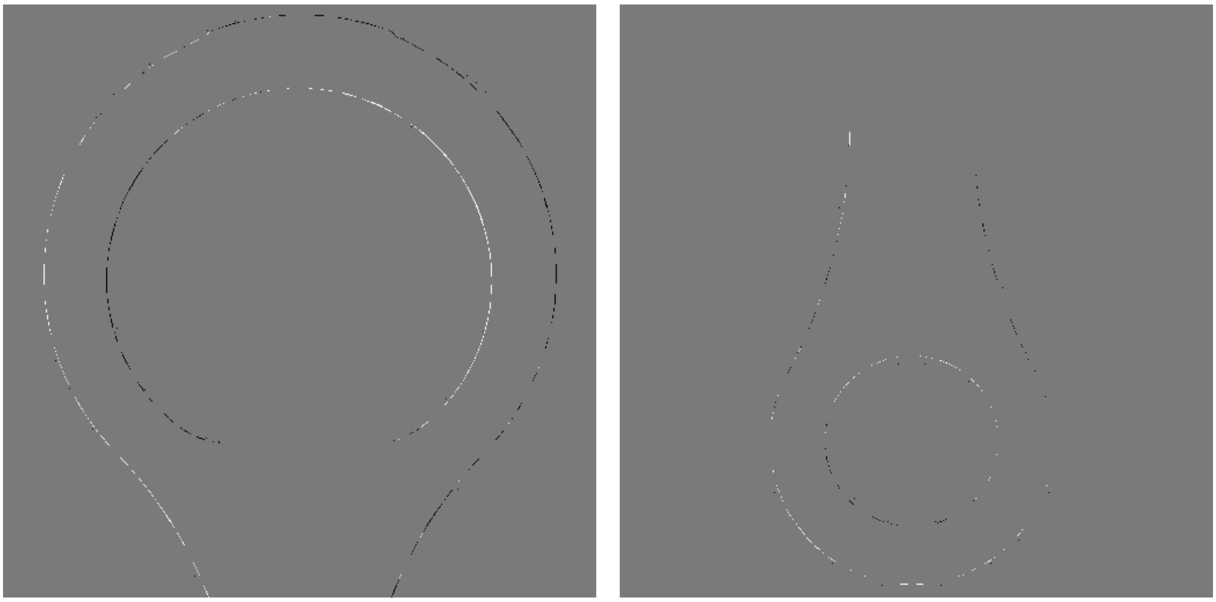
Results on an experimental part

The method has been firstly validated on simulated data and the corresponding results have been presented in [13] and are reported in Fig. 3.8. We show the registration of a connecting rod, the special feature of this object is that the two rings at the ends are perpendicular to each other. This means that at least one ring is visible on each front and transverse projections at the same time, which is convenient for assessing the quality of the reconstruction. The results were extremely encouraging. Using two perpendicular projections, we were able to align the connecting rod with an error of less than two pixels. This alignment was even used to apply a mask during reconstruction (see Chapter 4) [13].



(a) Initial Front view

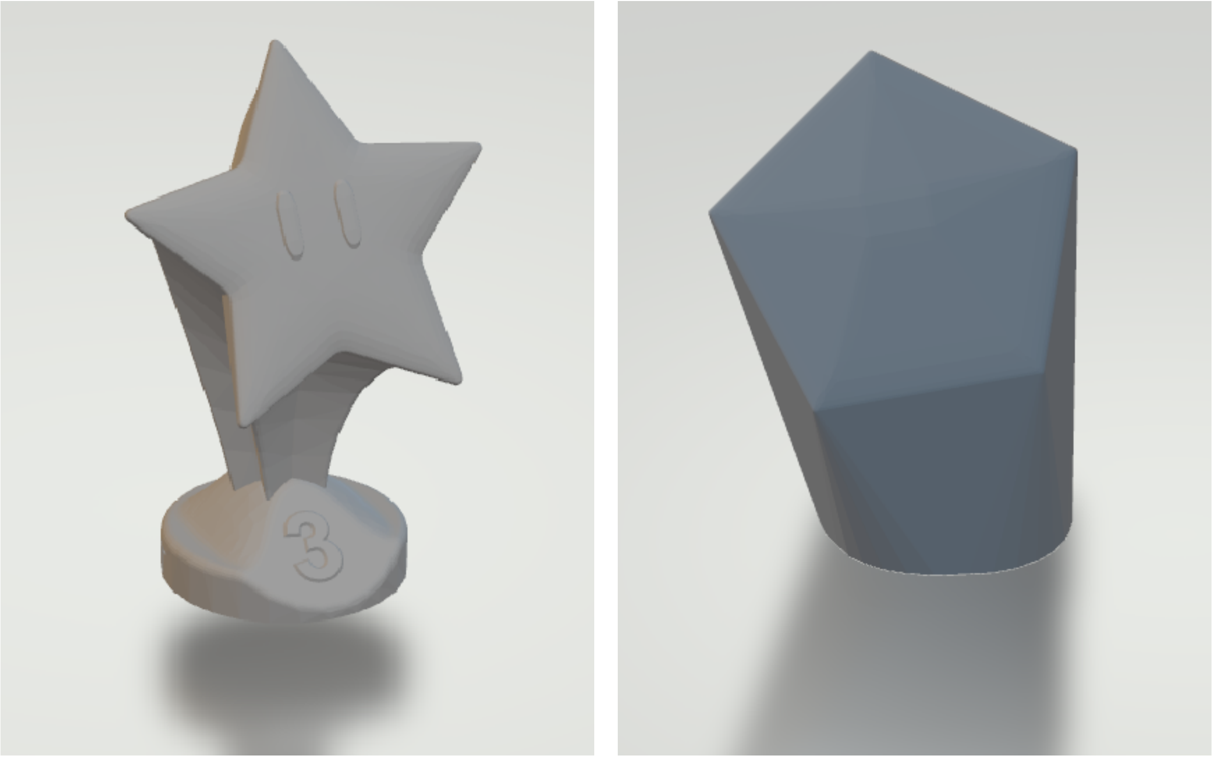
(b) Initial Transverse view



(c) Final Front view (zoom)

(d) Final Transverse view (zoom)

Figure 3.8: Registration of a connecting rod. Well-registered zones are grey, target projections are black and CAD projections are white.



(a) CAD Model. (b) CAD Convex Hull.

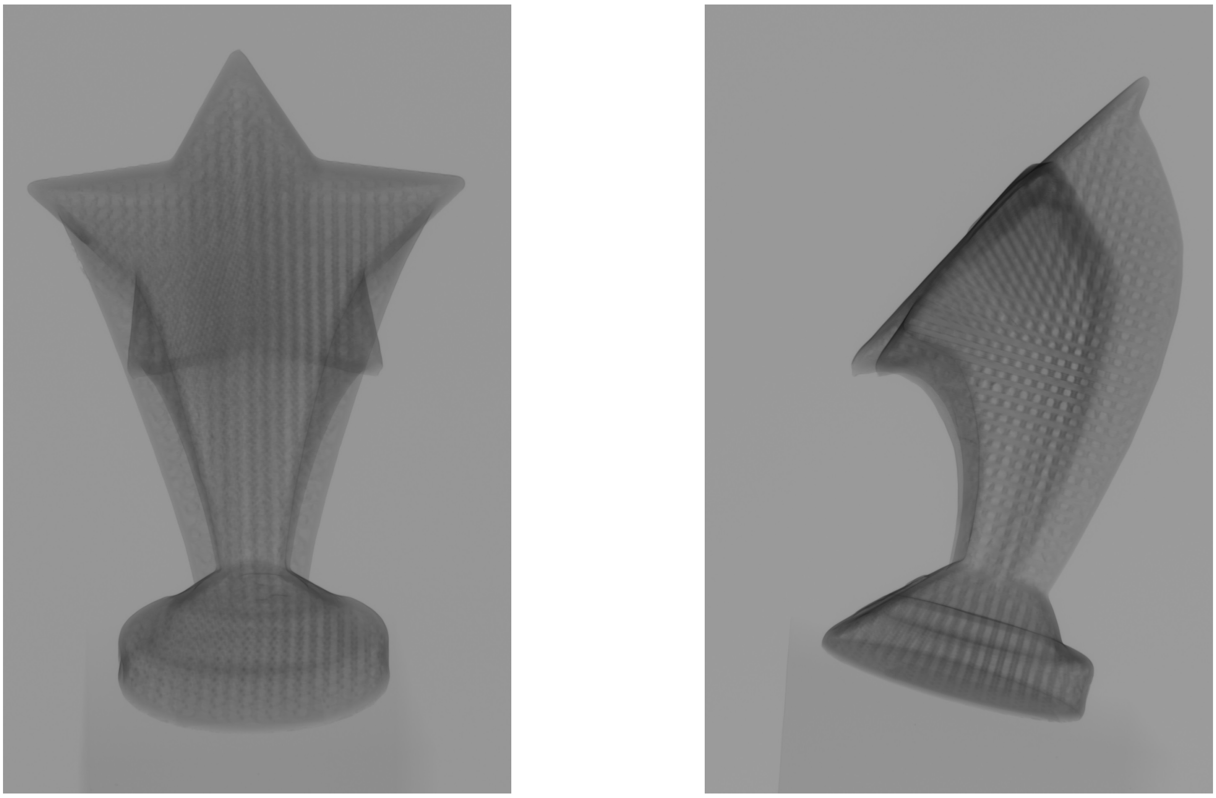
Figure 3.9: Convex hull of the 3D CAD Model.

The method has also been tested on a real part with experimental projections in [11], where the part studied had the particularity of being very different from its 3D model. Here, we will test the method on the experimental projections of the star-shaped trophy. In this scenario, there are variations between the 3D model and the real object, as well as potential noise in the projections, and the base of the object needs to be considered. We will therefore assess whether these additional challenges affect performance.

The first step is to compute the 3D convex hull. This step simplifies a lot the STL model. Allowing to consider 406 points for the convex hull instead of 1.767.648, that is to say, 0.023% of the original number of points (see Fig. 3.9). The number of points can even decrease further for pieces that do not have curves, here the circular base generates many points. The trophy remains quite interesting for this method because the points at the end of the star's branches are quite protruding and make registration easier.

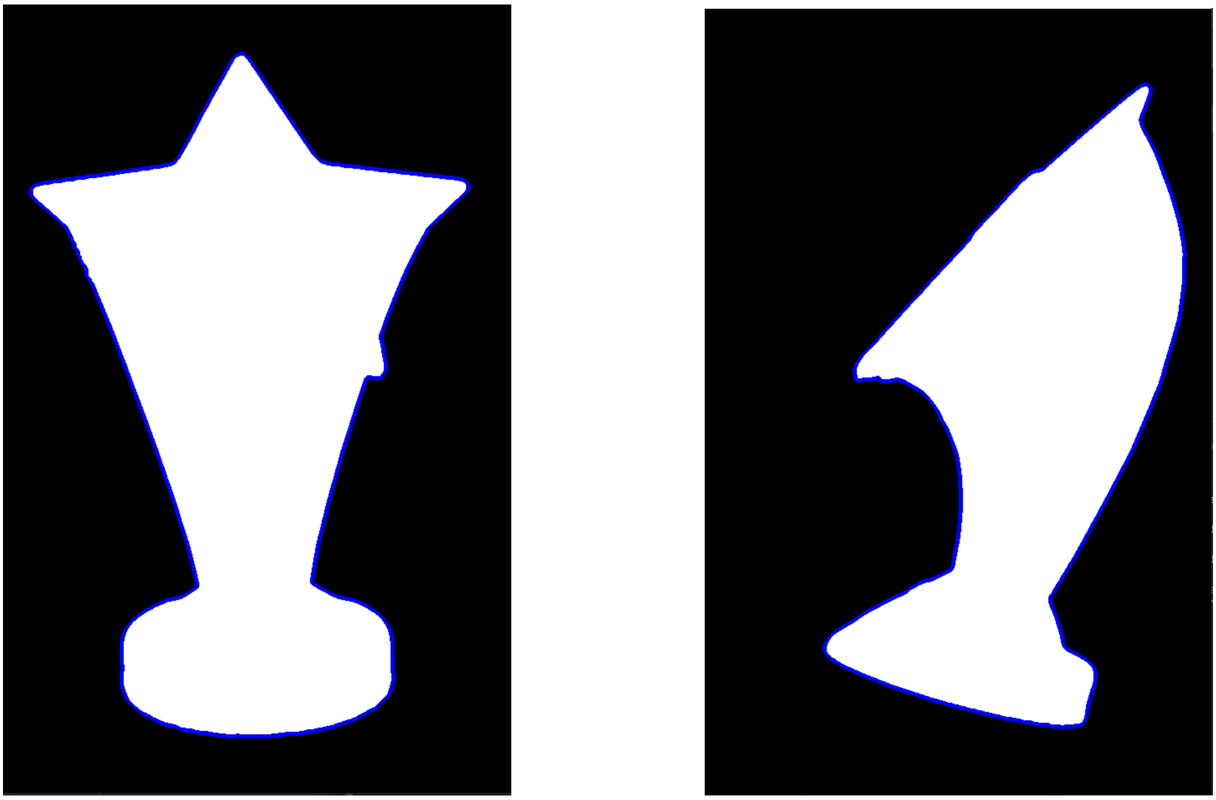
The next step is to find the 2D convex hulls of the projections. For this example, we consider two projections perpendicular to each other. The projections are processed to binarise the image and outline the object. On the experimental projections, we can see the polystyrene trophy base. Fortunately, it does not pose a significant issue; a simple threshold segmentation enables a satisfactory identification of the piece's contours. Figure 3.10 shows, for the front and transverse views, the results of the binarisation and the detected contours.

Afterwards, finding the projections' convex hulls is straightforward. Figure 3.11 shows the result. The front view convex hull is made of 37 points and the transverse of 53 points. The rationale behind binarisation and focusing solely on the convex hull becomes evident. Projections contain a vast amount of information, yet not all are essential for registration



(a) Front view

(b) Transverse view

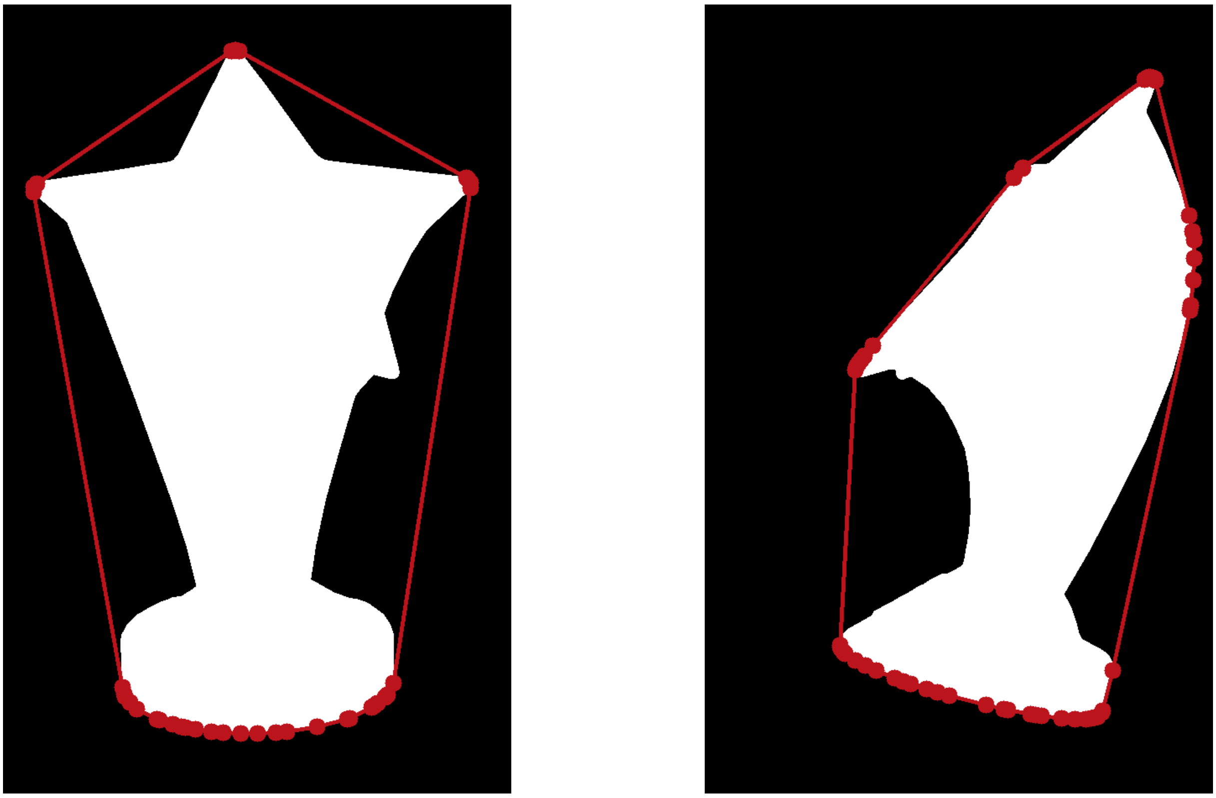


(c) Front view's contours

(d) Transverse view's contours

Figure 3.10: Two orthogonal views of the object before processing (top), after binarisation and contour extraction (bottom).

purposes. Restricting to a few points does not compromise the registration capabilities; instead, it streamlines and accelerates the algorithm.



(a) Front View Convex Hull

(b) Transverse View Convex Hull

Figure 3.11: Construction of the convex envelope of the experimental front and transverse projections.

The preliminary steps are done, and the iterations start now. To enhance the algorithm's robustness and speed, the points from the convex hull undergo further sorting. Only the points projecting onto the detector, forming the convex hull of the new temporary simulated projection, are taken into account. For the first iteration and the front view, only 24 points are kept. It now involves matching 37 keypoints $\in \mathbf{F}$ from the 2D convex hull to 24 points from the 3D convex hull $\in \mathbf{M}$ for the first view. In the second view, the 31 points of the 3D model that project onto the convex hull of the object are considered and matched to the 24 2D keypoints. The matching step becomes even easier. The orthogonal projections on the backprojected key points can be computed rapidly and each $\mathbf{X}_i$ can be associated to a $\mathcal{P}^{-1}(\mathbf{x}_j)$, solving for Eq. 3.30. Figure 3.12 illustrates the correspondences step as in Figure 3.7. Green points are the orthogonal projections of $\mathbf{M}$ on $\mathcal{P}^{-1}(\mathbf{x})$.

From the two sets $\mathbf{M}$ and $\mathbf{Y}$, the Kabsch-Umeyama algorithm is applied, and the parameter k is set to 2 in the Huber M-estimator in Equation 3.32. The iterations are repeated until convergence. Figure 3.13 shows the initial and final overlaps after 50 iterations.The whole procedure takes approximately 3 seconds. The target is represented in black, while the projection of the CAD model is in white. The overlap of the two is shown in grey. The goal is to maximise the grey area. The first row represents the initial overlap, and the second row shows the overlap after registration. The transverse

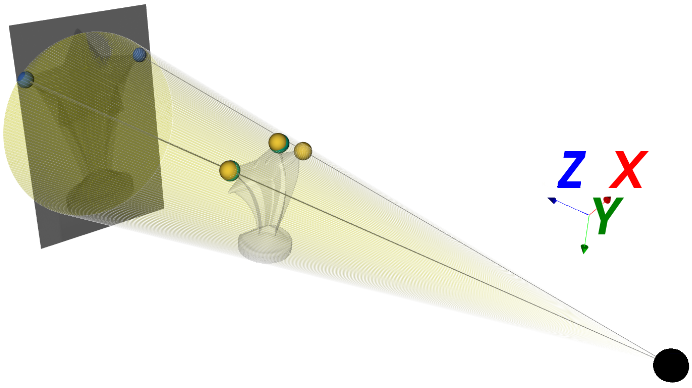


Figure 3.12: Visualisation of the registration setup. The matching is illustrated for three points on the CAD model $\mathbb{M}$, located at the tips of the star's branches. Each point on the 3D model (in yellow) is associated with a point on the 2D envelope (in blue). Subsequently, a point cloud (in green) is constructed.

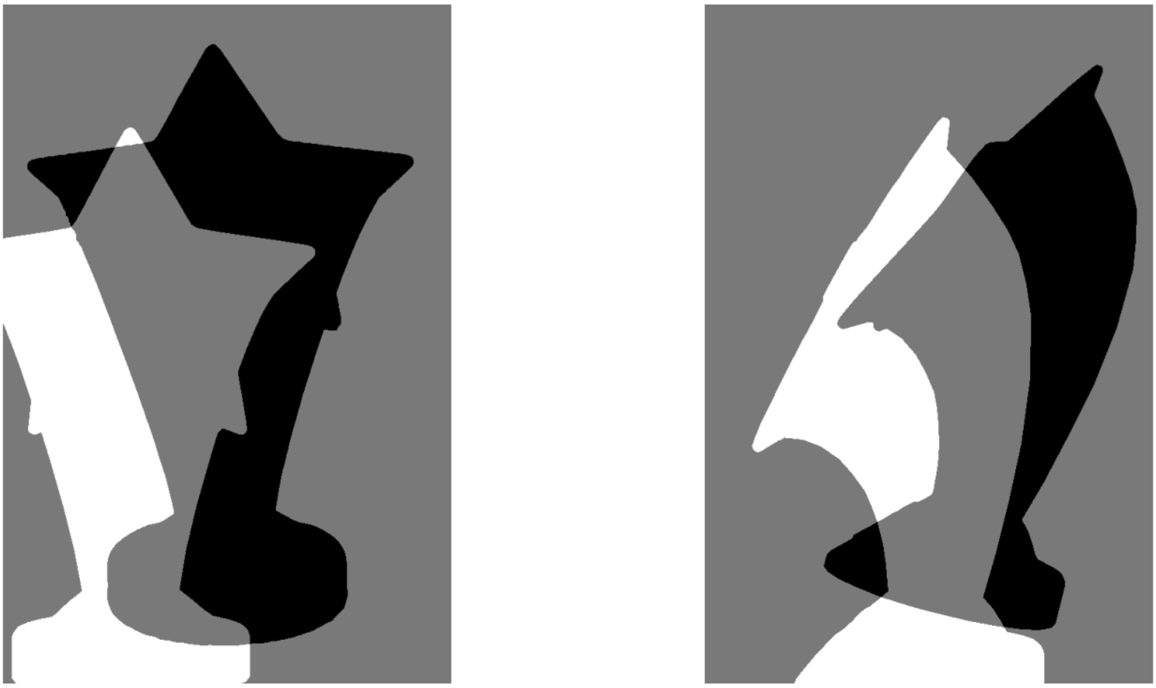
Table 3.1: Quantitative registration results for the star trophy.

Metric	Front View	Transversal View
Hausdorff distance [pixel]		
Before registration	181.1	218.8
After registration	13.15	4.000
Accuracy [%]		
Before registration	63.32	69.55
After registration	97.77	99.29

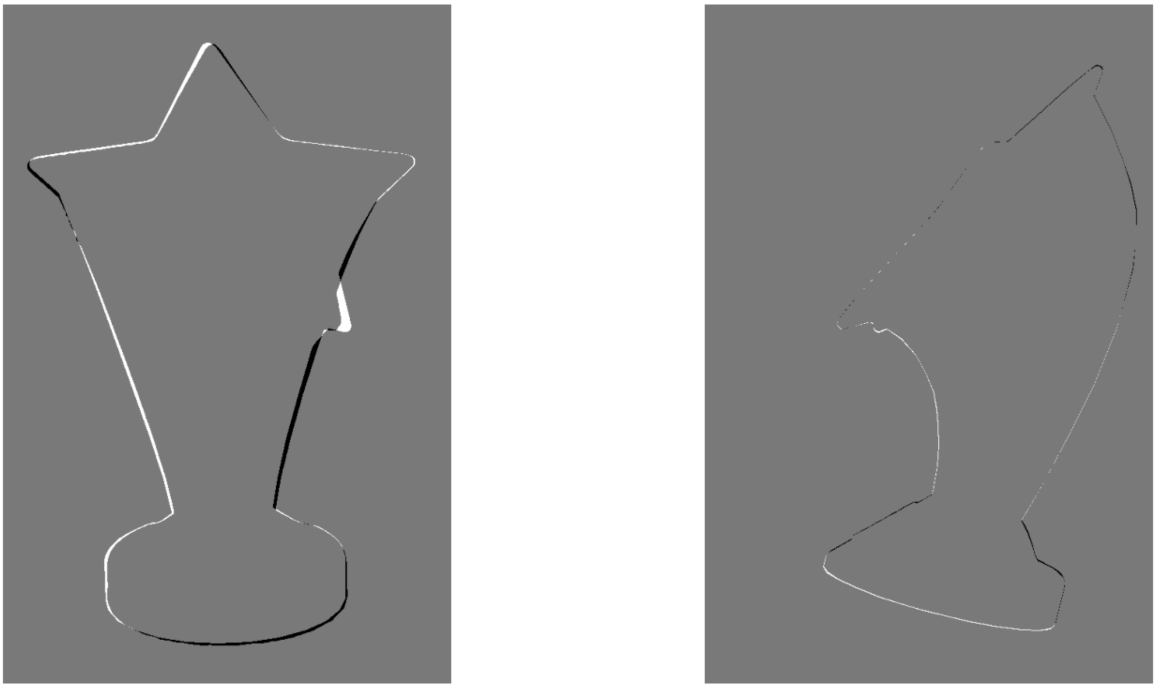
view demonstrates a very good alignment, with only a thin portion of the contours not overlapping. Table 3.1 shows the Haussdorff distance in pixels and the accuracy for the registration. The pixels on the detector are $127\mu\text{m}^2$. The Haussdorff distance d_H between two non-empty bounded closed sets X and Y in a metric space (E, δ) is defined as:

$$d_H(X,Y) = \max\left\{\sup_{y\in Y}\delta(X,y),\ \sup_{x\in X}\delta(x,Y)\right\} = \max\left\{\sup_{y\in Y}\inf_{x\in X}\delta(x,y),\ \sup_{x\in X}\inf_{y\in Y}\delta(x,y)\right\} \tag{3.38}$$

In a few words, it is the maximum distance in pixels between the contours of the experimental and registered simulated image. This distance is useful because when we will apply the mask in the next section, it will be dilated to compensate for registration errors. This expansion will be a factor of d_H divided by the magnification factor. The accuracy is defined as the number of pixels where the image overlay is correct (grey area) divided by the total number of pixels. Figure 3.14 illustrates the convergence of the six registration parameters (t_x, t_y, t_z, ω_x, ω_y, ω_z) over the iterations. The top row shows the rotation parameters while the bottom row shows the translations. The parameters are defined in relation to the point of convergence, i.e. the point obtained when the algorithm is run until the result is stationary. It can be observed that ω_y exhibits a slower convergence compared to the others. This outcome was anticipated as it represents the vertical O_y axis that is not directly observed. Translations compensate for errors in rotations and are



(a) Front view initial overlap (b) Transverse view initial overlap



(c) Front view final overlap (d) Transverse view final overlap

Figure 3.13: Initial and final overlaps

relatively easy to stabilise. Slight oscillations around the final value can be noticed. The graph is centred around the final value to ease interpretation. The ordinate represents rotations in degrees and translations in millimetres.

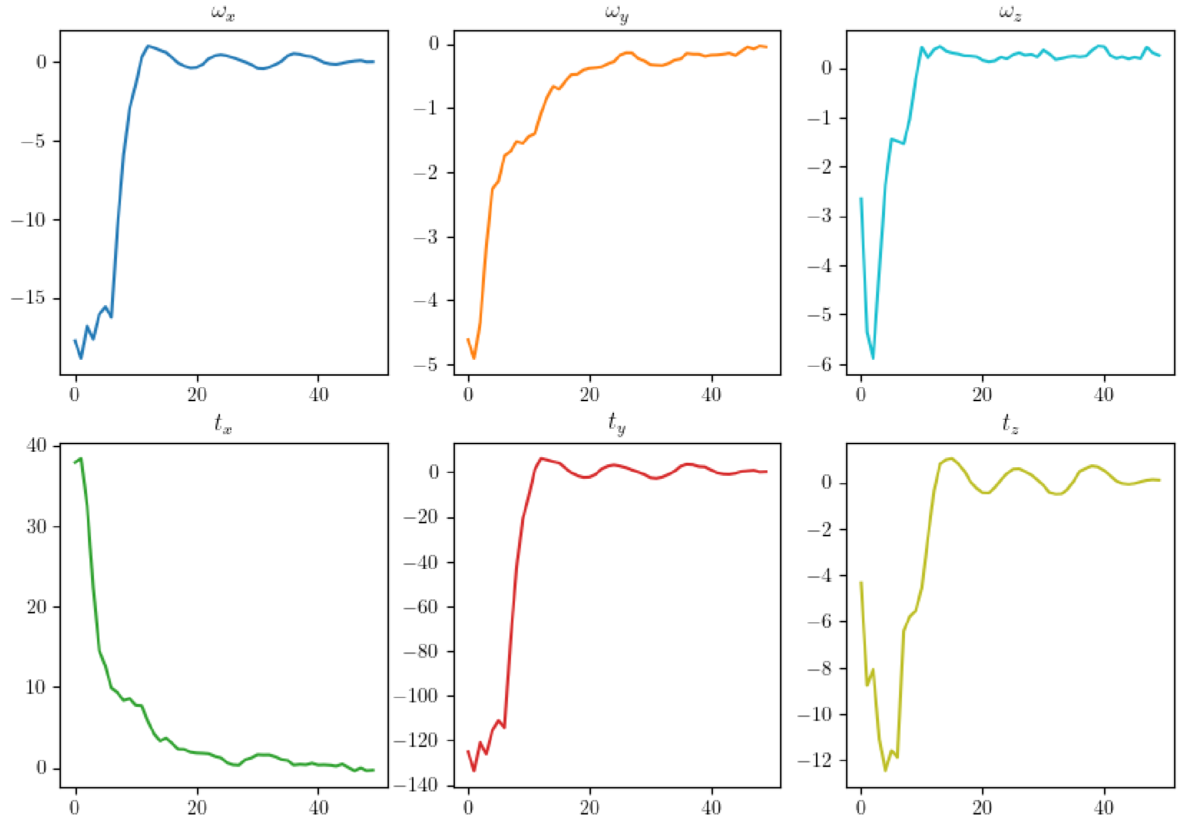


Figure 3.14: Parameters convergence over iterations. First row represents the angles (ω_x, ω_y, ω_z) [°], second row represents the translations(t_x, t_y, t_z) [mm].

3.2.2 Robust 3D/2D hybrid registration

The second method we propose relies on Image-Based Visual Servoing. It directly relies on the equations from the above Section 3.1.2. The critical aspect of IBVS lies in selecting the right features for registration. Similarly to *Iterative Inverse Perspective Matching*, we do not consider the entire projection, but rather focus solely on its contours. As observed in the previous method, a few well-chosen points provide sufficient information for accurate registration. While the entire image contains vast information, we limit ourselves again to a few points for algorithm robustness.

To binarise projections proves advantageous by allowing us to operate out of the greyscale framework. Consequently, precise knowledge of the source's energy or the exposure duration becomes unnecessary. This simplifies our ray casting simulation of projections significantly, focusing solely on detecting intersections between rays and the CAD mesh. Moreover, the CAD file is first decimated to minimise the number of elements [112]. This deliberate simplification effectively streamlines the entire process. The segmentation process of experimental projections exhibits resilience against potential edge blurring and scattering, ensuring a distinct delineation of the object's contours. In an industrial setting, the exterior shape is usually highly contrasted and typically suffices for the majority of objects' registration. However, in instances where object features are carved inside (such as a cylinder with internal grooves), this method must be adapted to greyscale.

Having opted to work with binary images, the next step involves selecting features for the registration process. Several types of points were considered initially. The Harris detector was one of the first attempts, yet the outcomes proved unsatisfactory [116]. The number of points detected was often insufficient, as this detector is not universally suitable for all geometries and types of objects, despite the method requiring robustness across various geometries. SIFT keypoints were also under consideration, yet their relevance was limited in the context of binary images; this detector will be given precedence in the greyscale version of visual servoing. Eventually, the decision was made to use the points forming the contours exclusively. They consistently exist, are easy to detect, and occur in sufficient numbers.

The contour points of the experimental images will be denoted as $\mathbf{s}^*$, while those of the simulated projections will be represented as $\mathbf{s}$. The matching step will be performed using the ICP algorithm. Each $\mathbf{s}^*$ will be assigned a counterpart in the simulated image, allowing to use visual servoing with the previously defined interaction matrix for points. If multiple views are employed, the multicamera equation 3.22 must be applied.

In practice, we have observed a significant dependency of the method on the quality of the ICP matching step. The rotation around the vertical axis O_y presents a particular challenge due to its unobservable nature in a direct frontal view during a circular trajectory, and it is the position of this axis that dictates whether the simulated projection shape aligns with the experimental one (Translations are generally manageable, merely requiring the repositioning of centroids). To address this challenge, a solution was proposed involving simulated annealing for the rotation around the vertical axis O_y. Simulated annealing is generally used for circumventing local minima. In our method, there is the double effect of avoiding local minima but also of improving the robustness of the method. Periodically, every ten iterations, two extra positions around the current orientation are evaluated. The algorithm selects the angle that minimises the discrepancy between matched points using ICP. It is important to note that while this metric does not precisely gauge the overall registration quality, it primarily assesses the alignment of the vertical axis O_y. Although this metric may lead to an increase in loss functions (i.e. the global overlap), it significantly enhances the method's robustness.

With the introduction of the simulated annealing strategy, the method exhibits greater robustness and improved results, yet the matching process may still present errors [12]. To demonstrate and quantify the benefits of simulated annealing, an additional study is provided in Appendix A.4. Another solution to this issue has been proposed: image moment-based servoing. Servoing the system with moments involves utilising global (rather than local) features of the image, thereby eliminating the matching step directly. Furthermore, moment-based servoing is particularly well-suited for binary images [117].

Image moments serve as mathematical descriptors in image analysis and computer vision, providing a comprehensive understanding of an image's intrinsic properties. They play a fundamental role in characterising crucial aspects such as shape, orientation, size, and intensity distribution. In a 2D image, the moment of order $(i + j)$ is defined as:

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y), \tag{3.39}$$

where $I(\cdot, \cdot)$ signifies the intensity of the pixel located at coordinates (x, y).

Zeroth-order moments represent the aggregate pixel values within an image, akin to the total intensity or coverage area. These moments offer a measure of the overall brightness

present in the image.
First-order moments denote the centroid or centre of mass of an image, indicating its position concerning both horizontal and vertical axes. For instance, the centroid (x_g, y_g) can be computed using $x_g = \frac{M_{10}}{M_{00}}$ and $y_g = \frac{M_{01}}{M_{00}}$. Central moments μ determine an image's orientation and scale concerning its centroid. They remain consistent irrespective of translations and are instrumental in computing variance, skewness, and other statistical properties:

$$\mu_{ij} = \sum_x \sum_y (x - x_g)^i (y - y_g)^j I(x, y). \tag{3.40}$$

Normalised moments ν, derived from central moments, offer scale-invariant and rotation-invariant features. They hold substantial significance in pattern recognition, object detection, and applications that demand robustness against scale and rotation effects:

$$\nu_{ij} = \frac{\mu_{ij}}{\mu_{00}^{\left(1+\frac{i+j}{2}\right)}}. \tag{3.41}$$

Computed across various orders, image moments capture diverse facets of an image's features and traits. Their widespread application spans across computer vision and image analysis tasks, facilitating critical functionalities like feature extraction, object segmentation, shape recognition, and feature-based matching.

The computation of moments is straightforward and resilient to noise since it relies on the entire image. However, the primary challenge in moment-based control lies in selecting the most appropriate features. While the positivity condition for the interaction matrix exists, its practical implementation poses difficulties. Most studies aim to decouple the interaction matrix concerning primitives. For instance, the image's centre of gravity is pertinent for translations along the vertical O_y and horizontal O_x axes, while the zeroth-order moment is relevant for depth control. The image orientation parameter α is particularly adapted to govern rotations around the source-detector axis O_z:

$$\alpha = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \tag{3.42}$$

To have an interaction matrix of rank six, one needs to use moments of order higher than three since symmetries cannot be properly handled by moments of order two. Features invariant to scale, translation or rotation, can be used to decouple $\mathbf{L_e}$.

To enhance performance and adaptability, we have devised a feature selection strategy. The algorithm initiates with point-based servoing using contours, bringing the object close to its actual position. But at each iteration, moments are computed. After the point-based servoing phase, we identify converged and unconverged moments. The latter are particularly crucial for our system, being more discriminative. Thus, we define a broad set of moments but restrict our consideration during the process to those deemed most relevant, adhering to the earlier stated criteria. We ensure the inclusion of both symmetric and anti-symmetric moments, spanning low and high orders, such as Hu moments. Moreover, for translations the mass and centroids are always used. The algorithm is summed up in Appendix A.4. Our complete set of image moments comprises:

$$\begin{cases} m_{00}, x_g, y_g \\ \mu_{02}, \mu_{20}, \mu_{11} \\ \nu_{02}, \nu_{20}, \nu_{11} \\ \alpha = \frac{1}{2}\arctan\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right) \\ \mathrm{hu}_0 = \nu_{20} + \nu_{02} \\ \mathrm{hu}_1 = (\nu_{20} - \nu_{02})^2 + 4\nu_{11}^2 \\ \phi_1 = \frac{\mu_{11}^2-\mu_{20}\mu_{02}}{(\nu_{20}-\mu_{02})^2+4\mu_{11}^2} \\ \phi_2 = \frac{(\mu_{03}-3\mu_{12})^2+(3\mu_{21}-\mu_{03})^2}{(\mu_{03}+\mu_{12})^2+(\mu_{21}+\mu_{03})^2} \end{cases} \tag{3.43}$$

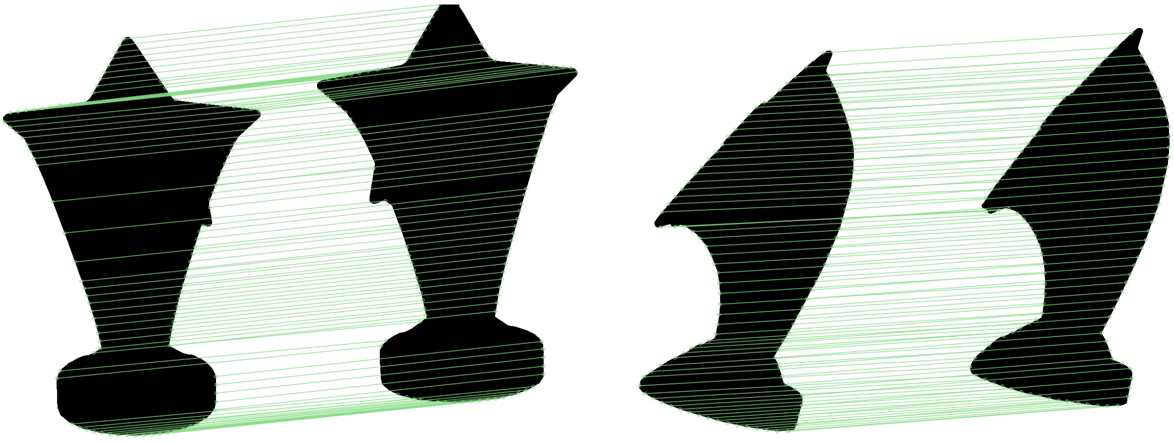
Results on an experimental part

The method was tested on simulated data in [12]. We tested the method on a variety of objects. The results are satisfactory in terms of quality, clearly demonstrating the benefits of simulated annealing. Again, this time we will test it on experimental projections of the star-shaped trophy. Similar to the previous tests, we selected the same two perpendicular projections as used in the registration via inverse perspective matching. The steps of binarisation and contour detection are the same. The results were already illustrated in Figure 3.10. The algorithm initiates with a simulated annealing step resulting in a change in orientation. This facilitates the matching stage. Figure 3.15 shows the ICP results for both views, displaying consistent correspondences. The algorithm executed 70 iterations in point-based servoing and 70 iterations using moment-based servoing. The whole procedure takes about 1 minute, but the speed depends greatly on the projection operator, which in our case can be greatly improved. The selected moments where (m_{00}, x_g, y_g, hu_0, ν_{11}, hu_1). The final overlay result is presented in Figure 3.16. As before, the target is represented in black, CAD model is in white and the overlap of the two is shown in grey. The registration is almost perfect. The imperfections observed can be attributed to the discrepancies between the 3D model and the actual object, as well as to positional errors of the source between the two projections.

The results are better than those obtained with the previous method, albeit at the expense of a slower registration. A combination of the inverse perspective matching with the moment-based registration appears to be a promising approach to balance speed and accuracy. Table 3.2 presents the quantitative results on the Hausdorff distance and the accuracy.

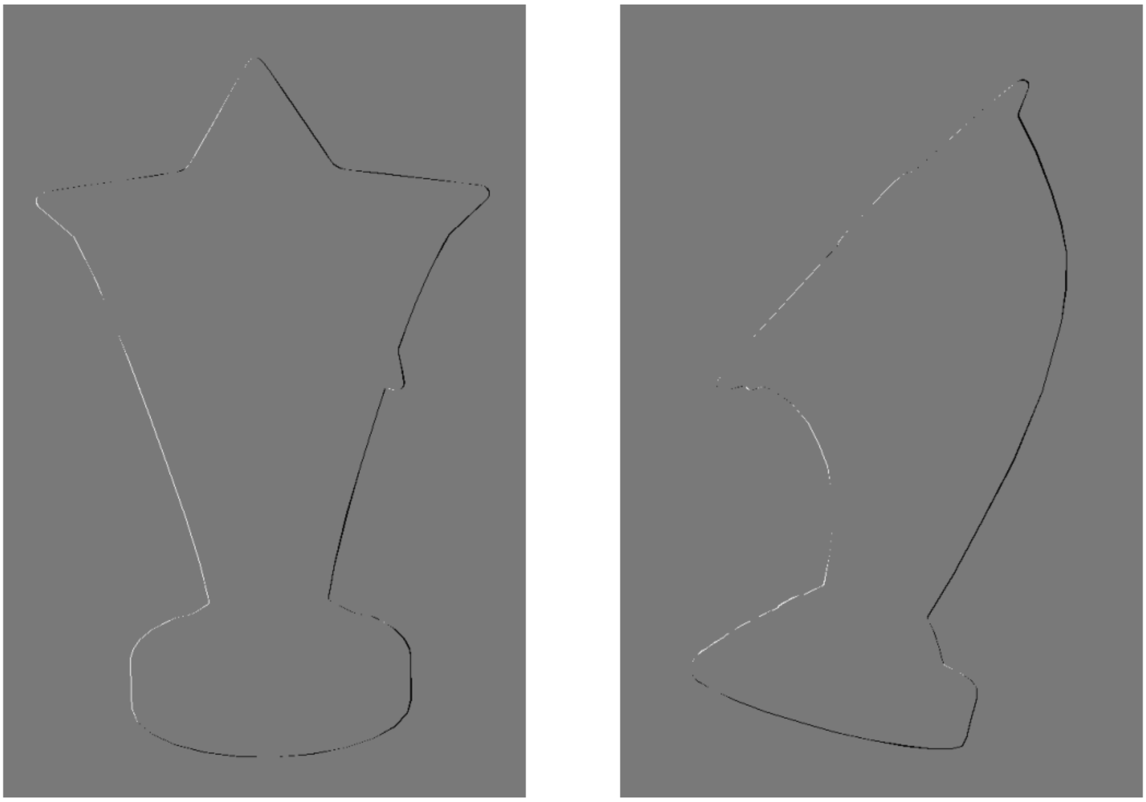
Table 3.2: Quantitative registration results for the star trophy

Metric	Front View	Transversal View
Hausdorff distance [pixel]		
Before registration	181.1	218.8
After registration	2.000	2.236
Accuracy [%]		
Before registration	63.32	69.55
After registration	99.64	99.67


(a) Front view

(b) Transverse view

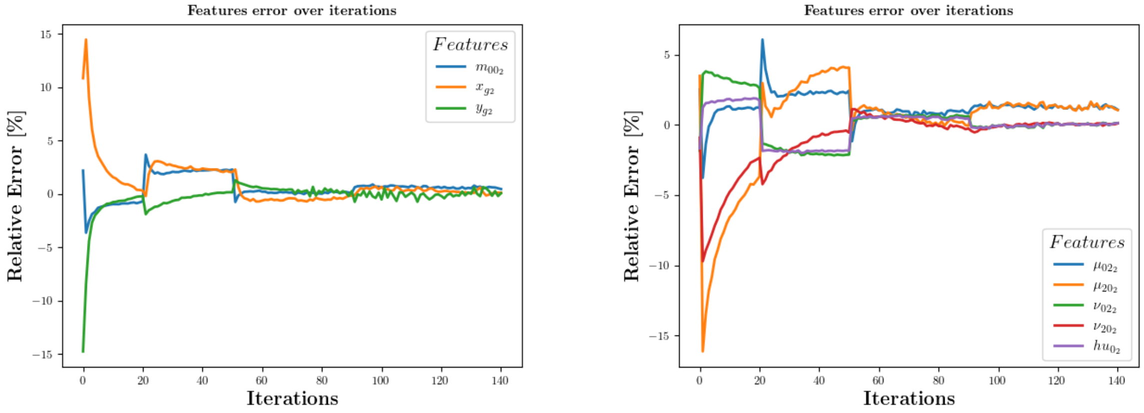
Figure 3.15: Illustration of the ICP matching results between binarised experimental images (left) and simulations (right) for the front and transverse views.


(a) Front view final overlap

(b) Transverse view final overlap

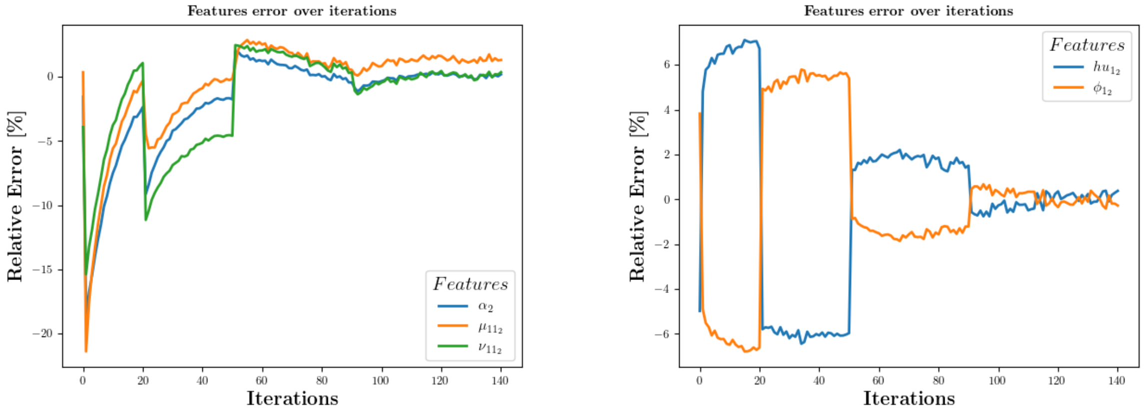
Figure 3.16: Final overlaps after 70 point-based iterations and 70 moments-based iterations.

We will now focus on the convergence during the second phase of the algorithm with image moments. We categorise moments into four groups: low-order moments (centroid and center of mass), symmetric moments (μ_{02}, μ_{20}, ν_{02}, ν_{20}, and hu_0), antisymmetric moments (μ_{11}, ν_{11}, α), and high-order moments (hu_1 and ϕ_1). The low-order moments handle translations, and they converge easily and quickly (see Fig. 3.17a). After 70 iterations, they had already reached their final values. Symmetric moments exhibit higher sensitivity but still converge reasonably rapidly (see Fig. 3.17b). Centred image moments μ show less convergence. This could be assumed to result from higher sensitivity and thus render them valuable as features, but this is mainly due to their lack of invariance regarding translation and scale, which are already governed by the 0^{th} and 1^{th} order moments. We prefer to analyse normalised moments, especially the 0^{th} Hu moment, as they are invariant to translation, scale, and even rotation for Hu. Rotation in the focal plane is accounted for by the parameter α. Besides, the antisymmetric moments display very similar curves and are nearly interchangeable (see Fig. 3.17c). However, we prefer the normalised moment ν_{11} due to its invariances and because, unlike α, it is defined when $\mu_{02} = \mu_{20}$. As for high-order features, their behaviours are identical except in sign. They exhibit higher sensitivity than low-order moments, thus being crucial for precise convergence (see Fig. 3.17d).



(a) Low order moments

(b) Symmetric moments

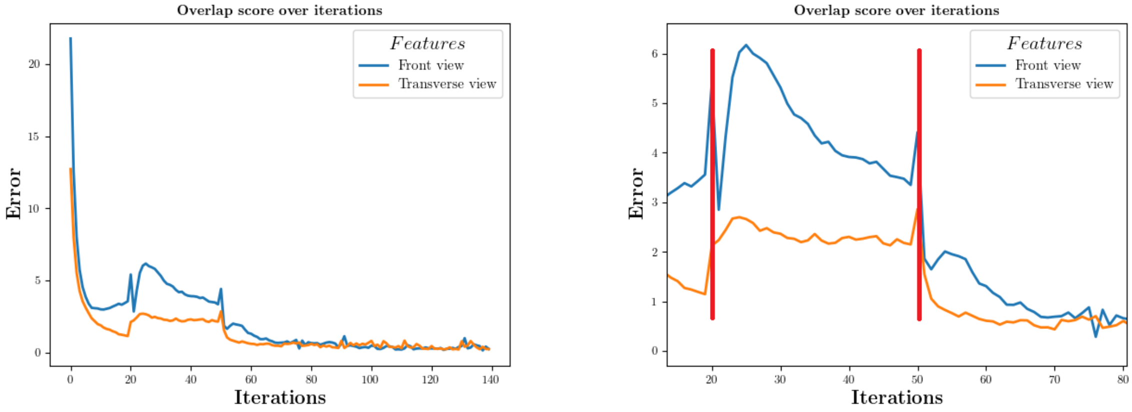


(c) Antisymmetric moments

(d) High order moments

Figure 3.17: Image moments convergence for the transverse view

Figure 3.18 depicts the error, defined here as $(100 - \text{accuracy})$ [%], over iterations. Point servoing is notably robust, but it is the moment-based servoing that truly ensures quality registration. As depicted in Figure 3.18b, the simulated annealing steps, represented by red bars, induced orientation changes. Notably, these alterations often initially increase the metric, but it generally rapidly descends to lower values than before. Figure 3.19 shows the orientation parameters convergence. Top row represents the orientation



(a)

(b)

Figure 3.18: Convergence of the error rate over iterations for the front (blue) and transverse (orange) views. Red vertical bars show the simulated annealing.

(ω_x,ω_y,ω_z), bottow row represents the translations (t_x,t_y,t_z). The interest and strength of the simulated annealing are evident at iteration 20, where the rotation along the vertical axis, notoriously difficult to converge, significantly benefits from the annealing process. At each ω_y discontinuity, we observe that the simulated annealing process has helped converge its value, even if the ICP selection criterion might not be immediately evident.

To support our study on simulated annealing, we performed the same registration (initially using points and subsequently moments) but without the intermediate steps of simulated annealing. The results are presented in Appendix A.4. The final outcomes are considerably inferior. The overlap between the final and target images is markedly less satisfactory, achieving only 96% accuracy in both views. More critically, the Hausdorff distance reaches up to 42 for the front view. In the point-based servoing phase, the error convergence curve declines similarly to the version with annealing but becomes trapped in a local minimum. Conversely, the moment-based servoing is significantly less effective without annealing; the relative differences are excessively large, resulting in incorrect movements that sometimes diverge from the target.

We proposed and evaluated two registration methods: Convex Hull Iterative Inverse Perspective Matching and Robust 3D/2D Hybrid Registration. Both methods demonstrate excellent results, even when there is a significant discrepancy between the model and the object. The Inverse Perspective Matching method has the advantage of not requiring any projections or special parameters. It is very simple to apply and fast. Due to its speed, it can be used with various initialisations to improve robustness and accuracy. Additionally, it can serve as an initialisation for the Robust 3D/2D Hybrid Registration method, which is more precise thanks to moment-based control that eliminates the need for a matching step, thereby reducing potential sources of error. Although satisfactory, these methods still need to be developed further to accommodate greyscale projections for registering objects without external physical features. Future work will also aim to adapt these methods to handle truncated projections.

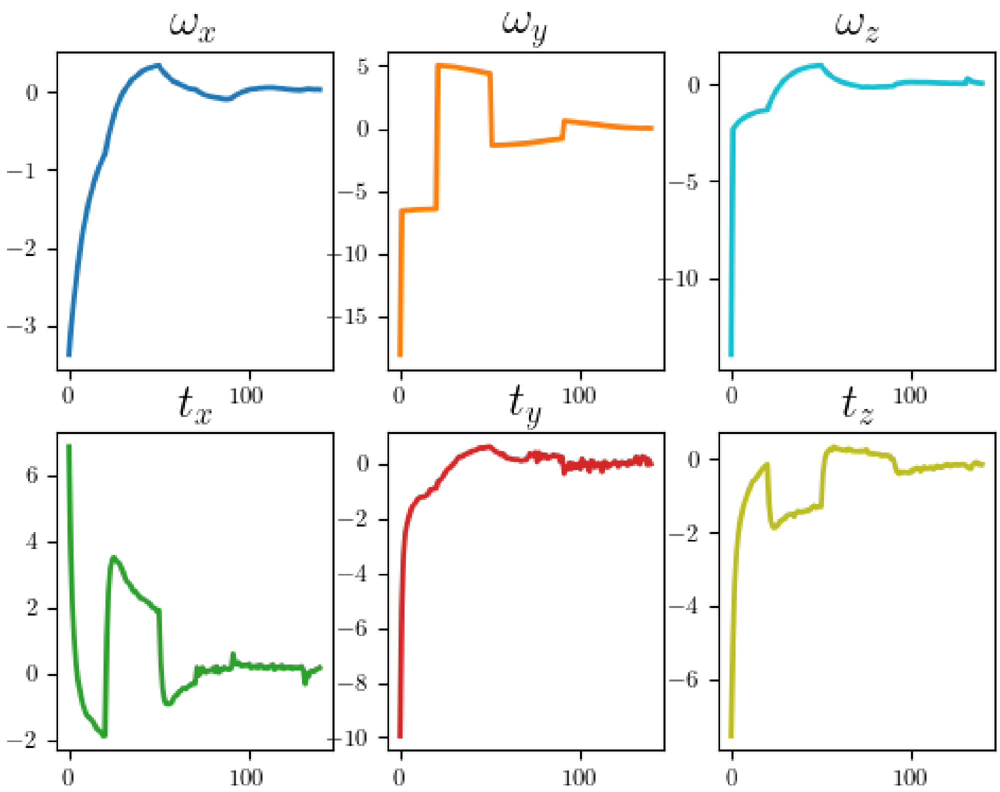


Figure 3.19: Parameters convergence. First row represents the angle (ω_x, ω_y, ω_z) [°], second row represents the translations(t_x, t_y, t_z) [mm].

Chapter 4

Discretisation of the reconstruction model

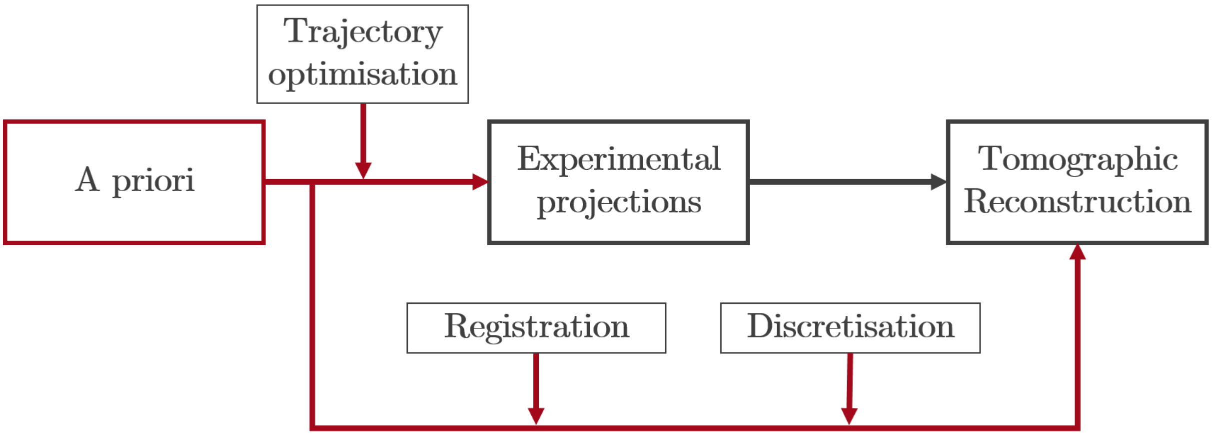


Figure 4.1: Diagram of the tomographic process. In this section, we study different methods to represent the object.

In the previous steps, we selected an optimised data acquisition trajectory, and we registered our CAD model. Now, we have the experimental projections as well as a coordinate system that aligns with our *a priori* information. We can therefore proceed to the reconstruction phase. Traditionally, reconstruction is performed on a voxel grid, but there are other methods to discretise the space and mathematically describe the reconstruction process.

For both analytical and iterative approaches, each volume $\boldsymbol{\mu}$ must be approximated and represented using a finite number N of discretising elements or basis functions $\boldsymbol{\Phi}_i$, weighted by their respective attenuation values μ_i:

$$\boldsymbol{\mu} = \sum_{i}^{N} \mu_i \boldsymbol{\Phi}_i. \tag{4.1}$$

Despite, the existence of many basis functions of varying sizes, dimensions, and complexities -implying discretisation on an irregular or adaptive grid- the most common basis function, $\boldsymbol{\Phi}_i$, is the indicator function of a cell. This choice leads to the classical discretisation on a regular cartesian grid of pixels/voxels:

$$\mathbf{\Phi}_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in i^{th} \text{ voxel,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

This representation is favoured due to its ease of use concerning both data encoding employing classical pointer manipulation for data storage and display capabilities, allowing visualisation of reconstructed images or volume sections using common software tools. However, these basis functions necessitate the use of a cartesian regular grid for representation. Consequently, to achieve high spatial resolution in reconstruction, very small cell sizes are required. As a result, large homogeneous areas in the volume are described by an excessive number of cells. These areas lead to increased memory consumption for storing reconstructed volumes, for only marginal improvement in overall reconstruction quality. Furthermore, these basis functions do not adequately represent curvilinear objects, resulting in aliasing artefacts at the boundaries of such domains.

A second possible set of basis functions are splines. They offer an alternative representation scheme that addresses some of the limitations associated with voxels. Splines provide a more adaptable and precise representation of the attenuation coefficients within the reconstruction volume [118]. Unlike voxels, beta splines allow for the flexible adjustment of the representation grid, enabling a finer and more accurate portrayal of complex geometries without requiring uniformly sized elements throughout the entire volume. Splines excel in handling data with variable resolution and can conform to the intricate boundaries and shapes of objects present in the reconstructed volume. This adaptability ensures a more efficient use of memory and computational resources while maintaining high reconstruction accuracy. However, their use may involve more complex computational methods compared to voxels, which might impact processing time and resource requirements.

Radial bases, belonging to the Gaussian family, named *blobs* are also utilised for image representation and reconstruction. They possess superior space-frequency localisation properties compared to pixels, and various operations, including the X-ray transform, gradient calculations, or interpolation, and can be analytically evaluated. Therefore, there is no necessity for discretisation or approximation of the X-ray projector [119].

The consideration of utilising mesh for tomographic applications presents an alternative approach. The employed mesh can either be regular or adaptive, necessitating a delicate balance between precision and computational cost. A regular mesh resembles a grid-like structure where elements (such as pixels in a 2D image or voxels in a 3D image) are uniformly arranged. While effective in various applications, this method might fall short in capturing intricate details or the complexity inherent in an image or object's structure. Conversely, an adaptive mesh adjusts its structure based on the discretised signal content [120]. This adaptation better represents significant features of the image or object [121]. For instance, in image compression, this adaptive approach concentrates representation on crucial image areas, thereby reducing redundancy and file size. The sole structure of the mesh already aids in interpreting the image content. In the realm of finite element simulations, mesh adaptation to each spatial domain stands pivotal in yielding precise and high-quality outcomes. By adjusting mesh density where necessary, simulations can aptly capture local variations in geometry or physical properties.

Moreover, hierarchical structures such as trees, notably quadtrees and octrees, offer many perspectives. In the realm of 2D images, a quadtree represents an arboreal structure where each node denotes an image area. The root encapsulates the entire image, and subsequent subdivisions divide this area into four equal parts. This recursive division persists until specific conditions are met, like achieving sufficiently low variance in pixel values. Octrees extend the concept of quadtrees to three-dimensional representations (volumetric images or 3D renderings) [122]. Each node within an octree represents a region in 3D space. The root symbolises the entirety of space, and subsequent subdivisions partition this region into eight equal parts. This recursive subdivision continues. A hierarchical representation mitigates oversampling issues prevalent in regular representations. However, this structure relies on square or cubic cells, posing challenges in representing objects with curved or non-regular interfaces. While beneficial in concentrating representation on pivotal zones and reducing details in homogenous areas, these methods also prove useful in lossless compression for simplification of image representation. In specific applications, like visualising large objects, they enable variable precision in representing volumetric structures across different zones [123]. Nevertheless, their limitation lies in inadequately representing objects with curved or non-regular shapes due to their use of square or cubic cells. Efforts to address these limitations using a Deep Learning framework have been proposed. Particularly noteworthy are the works of Rückert et al. [124], which combine octrees and machine learning to propose a hierarchical neural representation.

4.1 Reconstruction on a Mask

4.1.1 Benefits of masking the reconstruction

The question of the optimal representation for X-ray tomography has thus arisen, prompting numerous studies. In this thesis, we have first opted to persist with the conventional voxel-based approach. Before looking for another representation base, we wanted to optimise this one. Our contribution lies not in altering the basis of representation but rather in constraining it. We add to Equation 4.1 the condition that only the voxels belonging to a so-called 'mask' will be considered in the reconstruction:

$$\boldsymbol{\mu} = \sum_{\substack{i \\ i \in \text{mask}}}^{N} \mu_i \boldsymbol{\Phi}_i. \tag{4.3}$$

Mathematically speaking, a mask represents a constraint on the basis of the function representing the volume to be reconstructed, but in practice, it is a boolean array whose *True/1* values represent the voxels on which the reconstruction will be performed. Voxels with the *False/0* label will simply be ignored. This process allows us to reconstruct only those voxels that could potentially contain matter, and to eliminate those that we are sure only contain air. Thus, the number of basis functions is reduced. In tomographic reconstruction, the use of a mask offers multifaceted benefits to the reconstruction process and subsequent analysis.

One of the primary advantages lies in the reduction of variables number to solve. By removing empty areas and concavities from the object under study, many voxels are set to zero. The number of variables to be resolved in the tomographic system is reduced, which improves conditioning. Reconstructing the object becomes possible with fewer projections. Instead of increasing the number of projections for a constant number of voxels, we decrease the number of voxels for a constant number of projections. In both cases, the quality of the tomographic reconstruction is improved, but in the latter case, we find ourselves in a sparse-view strategy. Figure 4.2 shows a circular object in blue, enclosed by its pixel mask in grey. The reconstruction is performed only on the greyed pixels. This immediately reduces the number of unknowns. For the same ray, the backprojection will involve fewer pixels, ensuring a higher-quality backprojection operator for solving the tomographic problem.

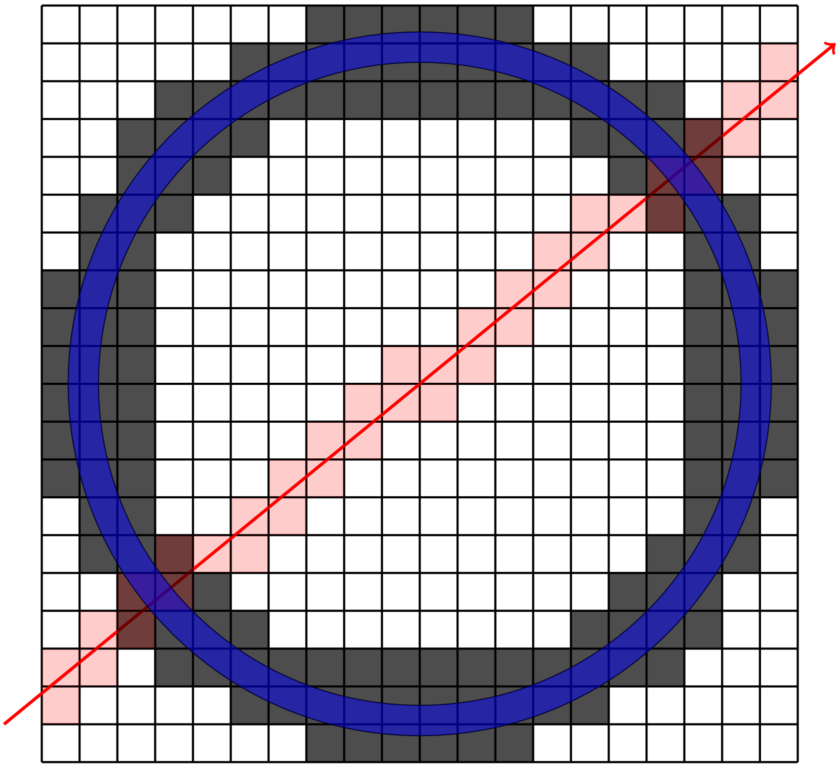


Figure 4.2: Illustration of a 2D mask on a circular object. The blue object is contained within a mask of voxels in grey. When a ray passes through, the red voxels are taken into account. With the conventional method, all red voxels are considered for projection and backprojection. However, with the mask, only the red voxels included in the mask are considered.

A direct benefit of employing masks is the minimisation of artefacts. By excluding irrelevant or problematic areas during the reconstruction process, masks effectively mitigate these artefacts, enhancing the overall reliability and accuracy of the reconstructed output. Moreover, the integration of masks allows focusing reconstruction efforts exclusively on specific areas and yields heightened clarity and precision within these regions. By allocating more computational resources to the specified areas and disregarding non-essential or less critical data, the reconstructed output achieves sharper and more refined details. Furthermore, it contributes to accelerated processing and computational efficiency during reconstruction. By confining the denoising processes to selected regions of interest, computational resources are optimally utilised, reducing the computational load and processing time compared to comprehensive volume reconstructions. Additionally, masks enable tailored analysis by facilitating the extraction and analysis of specific features or structures within the volume. This capability empowers detailed investigations.

4.1.2 Voxelisation of the 3D Model

To construct the mask, the CAD can be voxelised directly. There are several techniques [6], most of which assume that the surface of the model is closed. To ensure that all the material is included in the mask, it is important to expand it. Indeed, without doing so, incorrect alignment, roughness or deviation between the target and the final shape can cause the reconstructed part to come out of the mask. The mask must therefore be expanded to ensure that the entire part is included. A good practice is to expand by at least as much as the Hausdorff distance of the error on the registration step over the magnification factor. It prevents problems in case of mistakes during the registration.

Voxelisation is a frequently discussed topic within the realm of graphic rendering, primarily due to its potential to enhance performance in various applications such as fast ray tracing [125], shadow representation [126], and visibility analysis. However, in many algorithms, voxelisation can act as a performance bottleneck, leading to the development of numerous techniques aimed at leveraging the full potential of GPUs. Unfortunately, most of these methods focus solely on voxelising the surface of a mesh [127]. In typical computer graphics, the interior of objects is not depicted, as representing the surface suffices for applications like collision detection and is more computationally efficient. For our purposes, however, we require a type of solid voxelisation where the interior is entirely filled with voxels. While less common, solid voxelisation still has applications, such as in ambient occlusion [128]. Figure 4.3 illustrates different types of voxelisations. Conservative and 6-separating methods voxelise only the surface of the model. In contrast, solid voxelisation fills the interior of the volume, making it more suitable for tomographic applications.

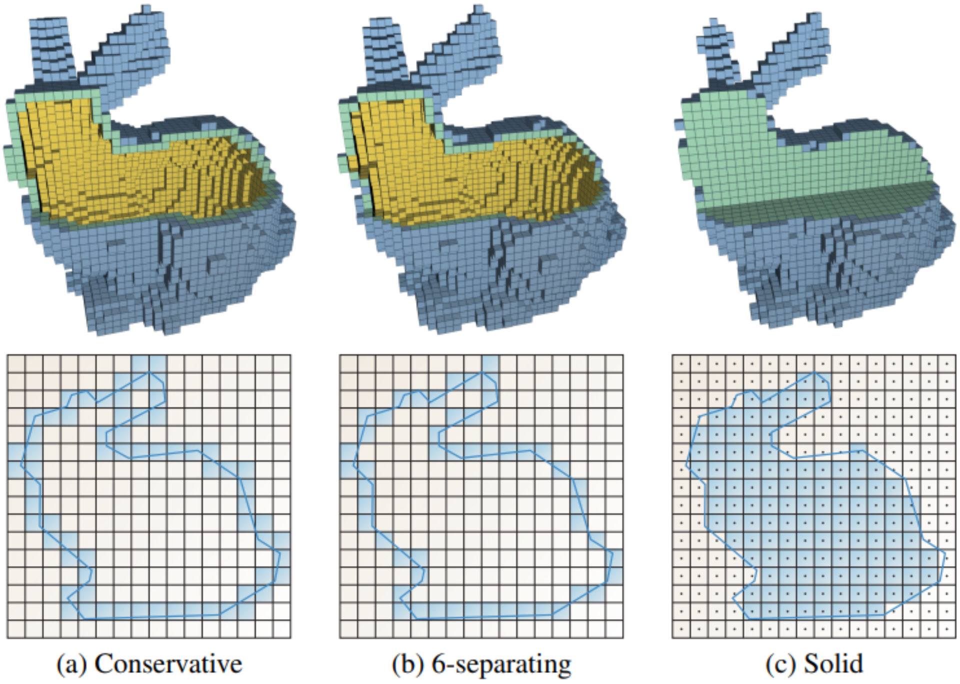

(a) Conservative (b) 6-separating (c) Solid

Figure 4.3: Different types of voxelisations. Image from [6].

Voxelised objects can have diverse configurations like holes, cavities and overlap, necessitating careful handling during the voxelisation process. Preserving object connectivity is crucial to maintaining discrete units that faithfully represent the original vector object. Voxelisation algorithms must enforce various topological constraints, including connectivity, separation and coverage. Connectivity, for instance, involves delineating N-paths between voxel pairs within an object, ensuring the resulting component remains N-connected and prevents disconnection. It offers insight into the interlinking of voxels, providing a measure of the voxelisation's "thin" or "thick" representation [7, 8]. Reciprocally, the concept of separation refers to a set of N-path voxels that divide two distinct sets of voxels. In 2D, this concept is articulated through 4- and 8-separations, while in 3D, one may encounter 6-, 18-, or 26-separations. These terms elucidate whether two voxels sharing a face, an edge, or a corner are connected. They delineate the relative thickness of the voxelised object and hold significant topological implications. Figure 4.4 illustrates the principles of 4- and 8- connectivity/separation in 2D and 6-, 18- and 26- connectivity in 3D. Figure 4.5 illustrates the concept of separation. The separation of the two sets of white pixels requires different line shapes.

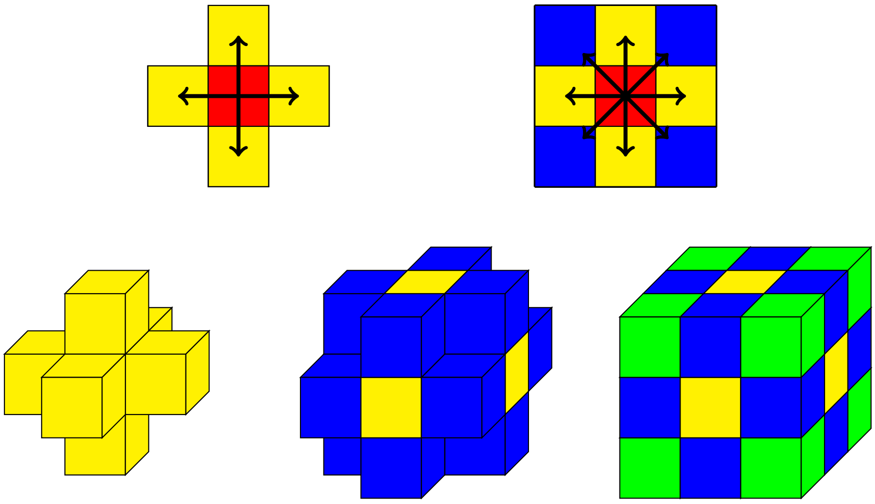


Figure 4.4: Illustration of the connectivity notion. On the top line: 4-connectivity and 8-connectivity. On the bottom line from left to right: 6-connectivity, 18-connectivity and 26-connectivity.

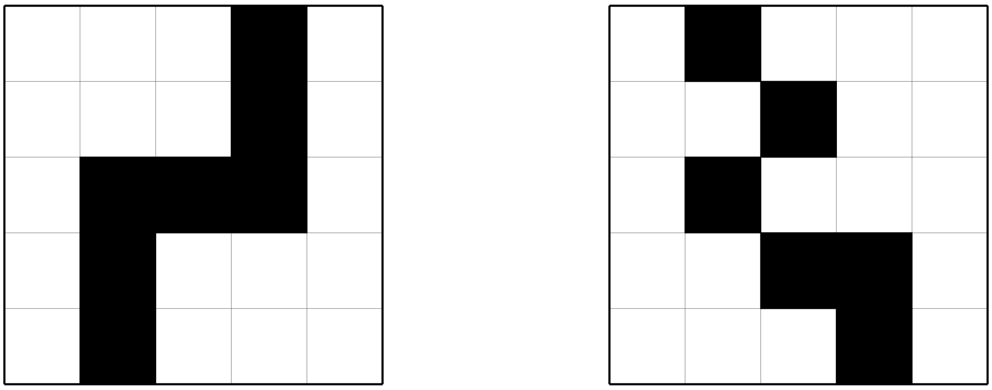


Figure 4.5: Illustration of the separation notion. On the left, the black pixels 8-separates the white set. On the right, the black set 4-separates the white set [7].

Coverage in voxelisation refers to quantifying the representation thickness for voxelised lines or surfaces. Within this concept, three primary variations exist: cover, supercover, and partial cover. A cover is a set of voxels that encompasses every point belonging to an object within the voxels. In 2D, an 8-connectivity pattern results in a cover set. On the other hand, a supercover set includes all voxels that either contain or touch the object points. In 2D, 4-connectivity generates a supercover, also known as conservative voxelisation. A partial cover represents the finer voxelisation. This approach preserves the object's shape but might not voxelise all points of the object. Figure 4.6 illustrates coverage in 2D.

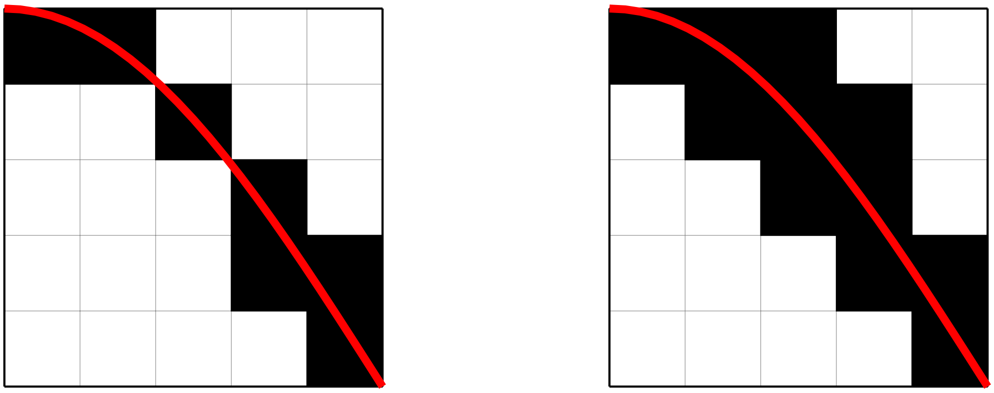


Figure 4.6: Representation of a cover and a supercover [8].

In our case, where we aim to create a mask for tomography, we are not striving to generate an optimal mask. We are not attempting to create a mask that would perfectly match geometrically with the object under study. We aim to exercise caution and encompass all the material without error. Furthermore, the mask will be dilated. Therefore, we allow ourselves to use any solid voxelisation method. Regardless of its behaviour on connectivity and separation, it will be erased by the dilation step. All these intricacies[1], in our specific scenario, are negligible.

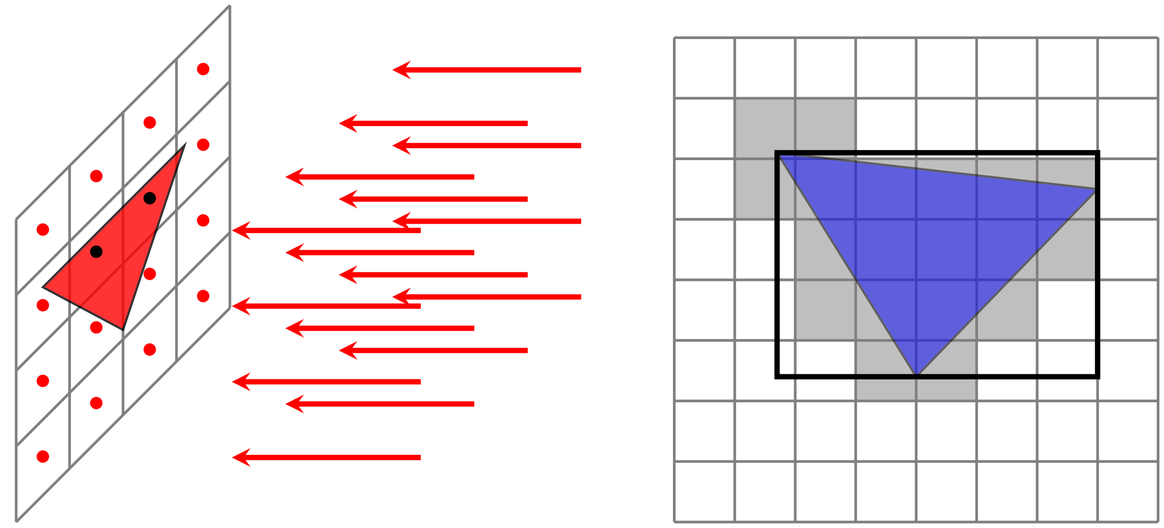
Since most CAD models can be converted into a stereolithography file (.stl), which remains the most common format, our focus will be solely on voxelising a surface made of triangles. Moreover, we focus on solid voxelisation which fills the interior too. There are two main algorithms to perform a solid voxelisation: ray casting and rasterisation.

Ray casting involves projecting rays onto a voxel grid and computing ray/triangle intersections. The rays are typically cast not from a fixed point, as in the perspective camera model, but from a plane, akin to the orthogonal camera model (see Figure 4.7a). The approach aims to simultaneously cast a ray along each row of the voxel grid. At each intersection, the voxel values on the line behind the intersection are updated (0 becomes 1 and vice versa). If the model is watertight, this process ensures voxelisation. To accommodate finer details, the method can be iterated by casting rays from different planes. Ray casting encounters implementation challenges when the model lacks proper definition or sealing. An odd number of intersections along a ray results in streaks during voxelisation. These artefacts can be addressed using a morphological opening operation.

Rasterisation, another common technique for voxelisation, involves parallelisation where each thread handles a triangle. Triangles are generally sorted according to their depth

---

[1] yet very interesting

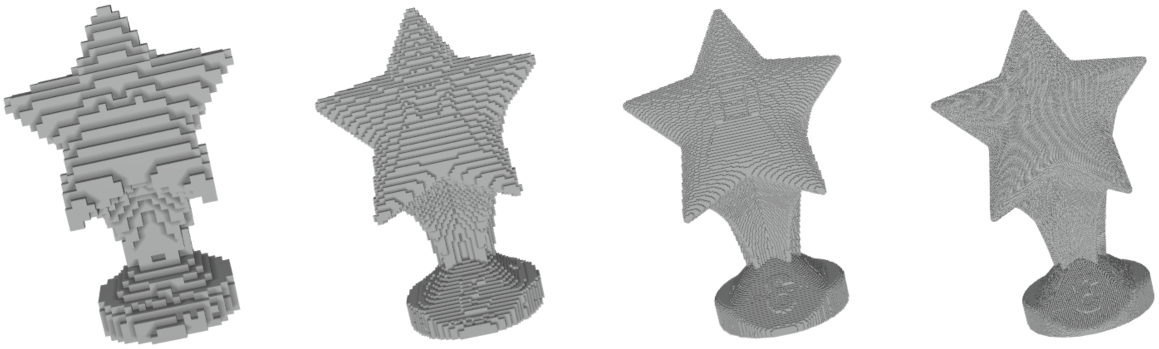


(a) Uniform ray casting. A ray is sent on each row of the voxel grid. Only two rays will detect an intersection (symbolised by black bullets).

(b) Rasterisation. For each voxel in the bold bounding box, we can determine which voxels contain the triangle.

Figure 4.7: Triangle voxelisation

direction and processed sequentially. For every triangle, the first step involves establishing the bounding box in a predefined plane and deriving the range of the encompassed voxel centres or just a part of the voxel in this plane (see Figure 4.7b where grey pixels intersect the triangle). When this range is not empty, a loop iterates over the columns of contained voxels. Each column's centre undergoes a test against the projection of the triangle on the plane. If the test is successful and an intersection occurs, similar to previous methods, the voxels behind it need to be flipped [6]. Implementing rasterisation remains challenging. It may result in cracks or holes, depending on the depth gradient of the rendered triangle. When the depth gradient is too high, a voxelised plane may exhibit "fissures" oriented perpendicular to the depth direction. Figure 4.8 displays the result of rasterisation-based voxelisation of our star trophy at various resolutions.



(a) 32x46x23 (b) 75x110x54 (c) 167x245x119 (d) 748x1102x534

Figure 4.8: Voxelisation of the star-shaped trophy at different resolutions.

Nonetheless, a small remark if the 3D model is not available, the binary mask can still be computed from a coarse reconstruction. Even if the reconstruction is not precise, the hull is generally visible and a simple thresholding can generate an approximation of the mask. The mask can even be generated gradually throughout the iterations. Interesting methods have also been developed to create the mask in parallel with the acquisition. In their work, Liu et al. [129] propose to use an optical scan to determine a mesh of the object's surface and then use it as a mask.

4.1.3 Masked reconstruction

To incorporate the mask, a revision of the reconstruction algorithm becomes necessary. To account for the mask, we need to alter our operators: the projection and the backprojection. Incorporating the mask during projection is relatively straightforward. Applying the binary mask to the volume and using the conventional tools already implemented suffices for projection. However, backprojection poses more challenges as the length over which each ray is backprojected depends on the mask.

There exist two strategies for implementing the mask, we denote them as the negative and positive approaches. The "negative" approach involves excluding volume's voxels outside the mask. For the projection, as explained above, except for the application of the mask on the volume, nothing changes. For backprojection, determining the intersecting voxels of a ray involves finding their intersection with the mask and backprojecting only these voxels. For each ray, the crossed voxels of the mask and their cumulative total length are calculated to make the backprojection. This strategy aligns closely with the mathematical formulation, requiring additional steps and thereby consuming more time than conventional unmasked operators.

Conversely, the "positive" approach aims for greater simplicity. Projections follow the same methodology as the negative approach. The binary mask is applied to the volume, which is then projected. However, during backprojection, instead of altering the volume, the backprojected value is modified. Backprojecting a value onto a smaller length (negative approach) is equated to backprojecting a larger value over the entire volume (positive approach) and only after applying the mask. Corrective values are determined by the ratio of the projections of a volume consisting of ones and the mask. Portions of the projection where the division is undefined are disregarded. Both strategies are illustrated by the equation:

$$\frac{\mathbf{r}_i}{\|\tilde{\mathbf{A}}_i\|^2} = \frac{\tilde{\mathbf{r}_i}}{\|\mathbf{A}_i\|^2}, \tag{4.4}$$

where $\mathbf{r}_i$ is the i^{th} element of the residual $(\mathbf{p} - \mathbf{A}\boldsymbol{\mu})$ to backproject, and $\|\mathbf{A}_i\|$ is the norm of the i^{th} column of the projection matrix. Masked operators are symbolised with a tilde. The left side corresponds to the negative approach and the right side to the positive one. However, we notice the emergence of a few inconsistent values, and overestimated points due to edge effects of the mask. If a point on the screen is not illuminated during the mask projection, the aforementioned ratio tends towards infinity. We will ensure to remove non-physical values before backprojection.

An adaptation of the SART algorithm that incorporates masking has been previously proposed in [130] for laminography. In our paper [13], we go into further details and present results with SIRT algorithm. We will outline the methodology of the article using SIRT too, but by redefining projection and backprojection operators, we can employ any reconstruction method.

As a reminder, the original SIRT algorithm operates iteratively to reconstruct the object $\boldsymbol{\mu}$ according to Equation 1.45. The equation is rewritten here:

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \lambda \mathbf{N_c} \mathbf{A}^{\mathsf{T}} \mathbf{N_r} (\mathbf{p} - \mathbf{A}\boldsymbol{\mu}^{(k)}). \tag{4.5}$$

Here, $\mathbf{N_r}$ and $\mathbf{N_c}$ are diagonal matrices that normalise the data. Their diagonal elements were defined as the sums of the lines and columns of the system matrix. Incorporating the mask results in a reduction in the dimensions of $\boldsymbol{\mu}$ and $\mathbf{A}$. Eliminating the inactive voxels from $\boldsymbol{\mu}$, along with the corresponding columns in $\mathbf{A}$, ensures that inactive voxels are not taken into account. Consequently, we define lower-dimensional operators, denoted as $\tilde{\boldsymbol{\mu}}$ and $\tilde{\mathbf{A}}$. Similarly, $\tilde{\mathbf{N_r}}$ and $\tilde{\mathbf{N_c}}$ are redefined using $\tilde{\mathbf{A}}$ in Eq. 1.46 and Eq. 1.47. Adjustments are also made to the vector $\mathbf{p}$. To avoid division by zero, rays such that $\sum_j \tilde{\mathbf{A}}_{ij} = 0$ are removed from the projection matrix. Correspondingly, these same rays are eliminated from $\mathbf{p}$ to make $\tilde{\mathbf{p}}$.

The update step for the masked-SIRT remains similar to the original SIRT but employs the modified low-dimensional operators and vectors:

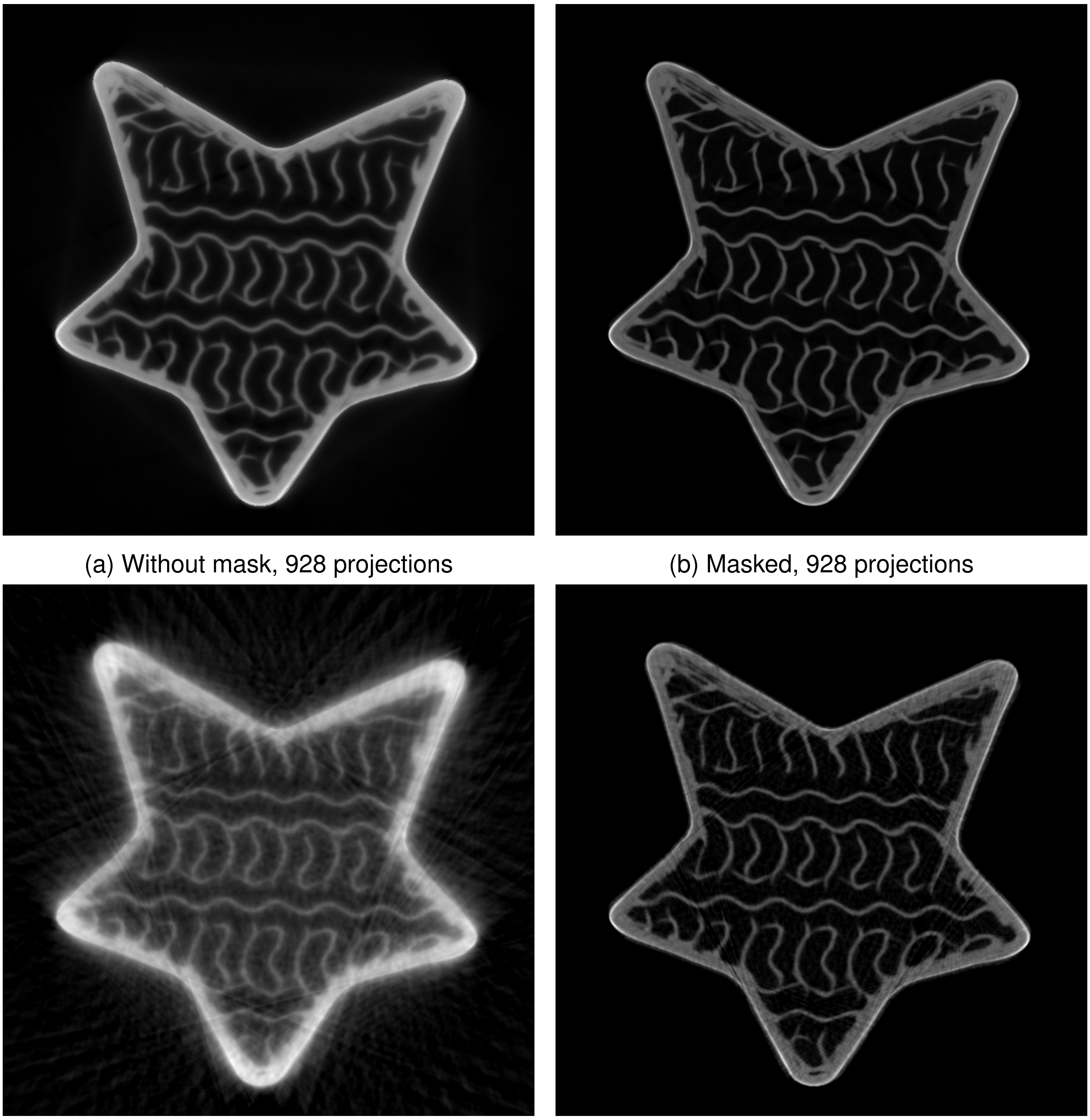
$$\tilde{\boldsymbol{\mu}}^{(k+1)} = \tilde{\boldsymbol{\mu}}^{(k)} - \tilde{\mathbf{N_c}} \tilde{\mathbf{A}}^{\mathsf{T}} \tilde{\mathbf{N_r}} (\tilde{\mathbf{p}} - \tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}^{(k)}) \tag{4.6}$$

This modification ensures compatibility with the mask, allowing for the reconstruction algorithm to effectively handle reduced dimensions in both the object and projection operator, thereby enabling accurate reconstructions despite the masked elements.

4.1.4 Experimental results

Simulated results have already been presented in our paper [13]. In this article, registration and masks were tested on simulated projections. A defect was inserted in the object and thanks to the mask, the probability of detection went from 3.576% to 16.66%. The use of the mask considerably improves the quality of the reconstruction. Here, we have tested masked reconstruction on the same experimental object as before: the star-shaped trophy. The reconstruction was performed using the masked version of SIRT over 150 iterations. The volume size was 1772 × 1204 × 1204 voxels. After the robust 3D/2D hybrid visual servoing registration, the Hausdorff distances for registration errors were measured at 2.000 and 2.236 pixels for the frontal and transverse views, equivalent to 0.254 and 0.284 mm, respectively, considering a magnification factor of 1.71 (detector pixels are $127\mu\text{m}^2$). Geometrically, the mask should be dilated by 0.167 mm. Our cubic voxels measure $0.074\ \text{mm}^3$, suggesting a precautionary dilation of at least three voxels in each direction. To ensure safety margins, we chose to dilate it by 12 voxels.

Figure 4.9 shows the same cross-section for the unmasked and masked 3D reconstructions with 100 and 928 projections. Whether it involves a sparse or dense trajectory, the mask significantly enhances image clarity. The reconstructed image exhibits significantly reduced blurring, enabling the perception of finer details within the object's walls. Additive manufacturing imperfections become evident, revealing uneven filling of the object's walls, manifesting as plastic streaks. While the masked reconstruction presents some artefacts—such as faint lines in the continuity of the star's branches—these artefacts are considerably less pronounced than those observed in the conventional reconstruction.


(a) Without mask, 928 projections

(b) Masked, 928 projections

(c) Without mask, 100 projections

(d) Masked, 100 projections

Figure 4.9: Comparison of SIRT reconstructions with and without a mask for 928 and 100 projections.

In Figure 4.10, we have displayed the profiles of the horizontal lines passing through the centre of the star cross-section for the classical and masked reconstruction. The profile is much sharper thanks to the mask. The contrast between the empty areas and the material is better defined. The outer layer of varnish, which might be responsible for the brightness at the edge of the star in Figure 4.9, is clearly visible here.

We also analysed the convergence of the algorithm with the mask. We have generated a plot showing the residual norm across the iterations. Figure 4.11a depicts the results for reconstructions with 100 projections. It shows the residual norm of the traditional SIRT, masked SIRT, and a variant which we will refer to here as POCS (Projection Onto Convex Sets), in which we simply apply the mask to the ongoing reconstruction at the end of each

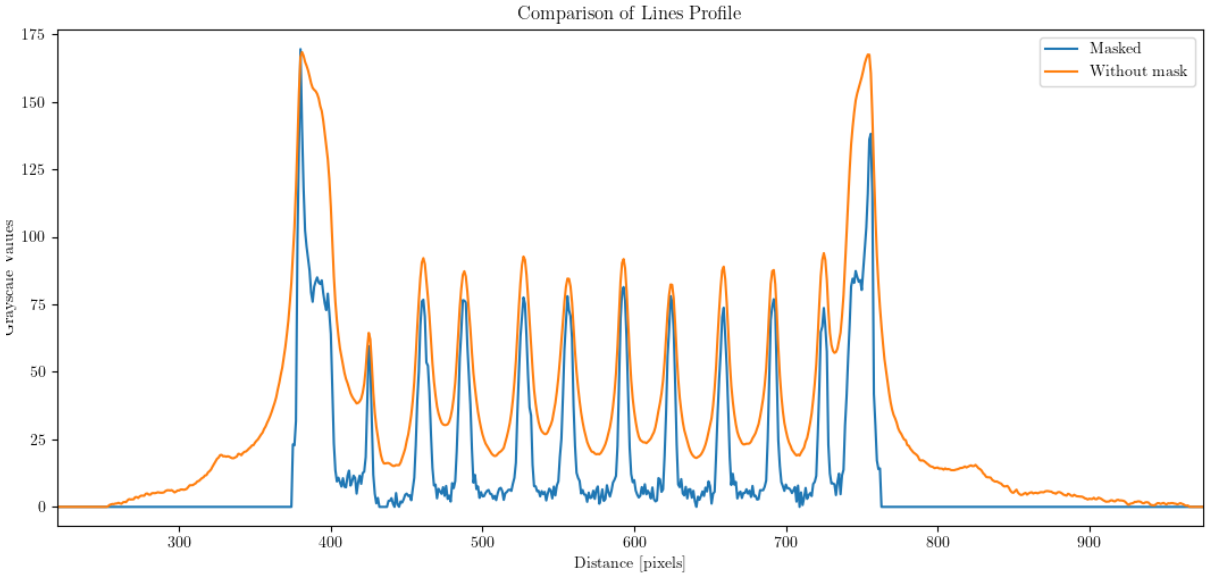
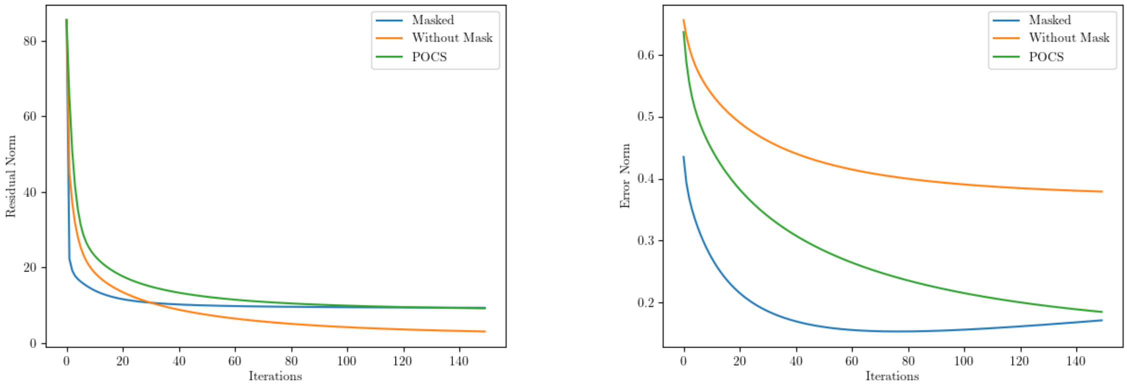


Figure 4.10: Comparison of the profiles of the horizontal lines passing through the centre of the star cross-section for the classical and masked reconstruction for 928 projections with SIRT

SIRT iteration [131]. The objective of the latter is to evaluate the usefulness of creating the new masked operators defined earlier. It appears that convergence requires less iterations compared to the conventional SIRT approach. As depicted in Figure 4.11a, the unmasked SIRT required around 150 iterations to converge while the masked version requires only 40. We think that is mainly due to the dimension reduction of the tomographic problem. The masked version converges more rapidly than POCS too. Fewer iterations are needed for convergence. The same residual norm is achieved with fewer computations. Figure 4.11b shows mean square root errors between the ongoing reconstruction and a reference made with all projections. The masked version achieved again better reconstruction quality with fewer iterations.

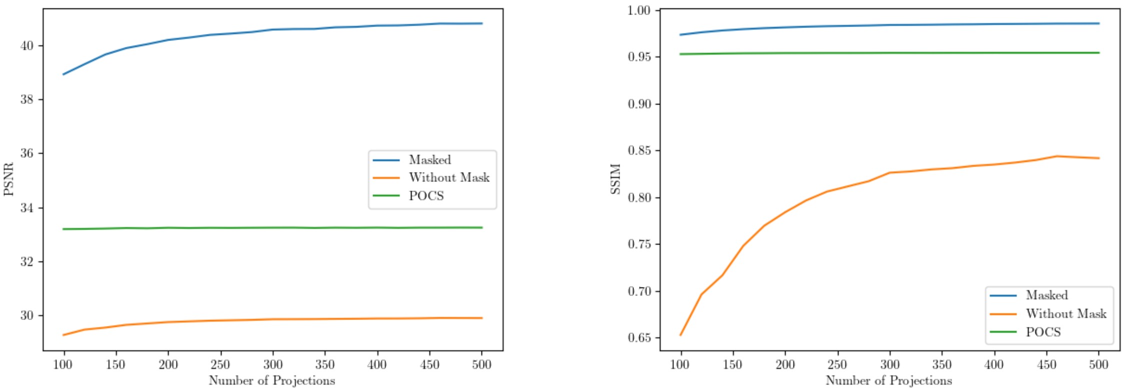


(a) Residual norm over iterations

(b) Error norm over iterations

Figure 4.11: Comparison of the convergence of algorithms with and without masks

Continuing with the same setup, we demonstrate here the impact of the mask on the reconstruction quality for various numbers of projections when combined with a SIRT reconstruction algorithm. Figure 4.12 depicts the reconstruction quality of the studied cross-section in terms of PSNR and SSIM for 60 iterations. The masked reconstruction exhibits significantly better reconstruction quality regardless of the number of projections used. The number of required projections stabilises quickly when the mask is utilised, plateauing at around 150 projections. The corresponding curves exhibit an almost horizontal trend. The mask effectively reduces the number of projections required for tomographic reconstruction.



(a) PSNR

(b) SSIM

Figure 4.12: Reconstruction quality of the studied slice for different projections numbers

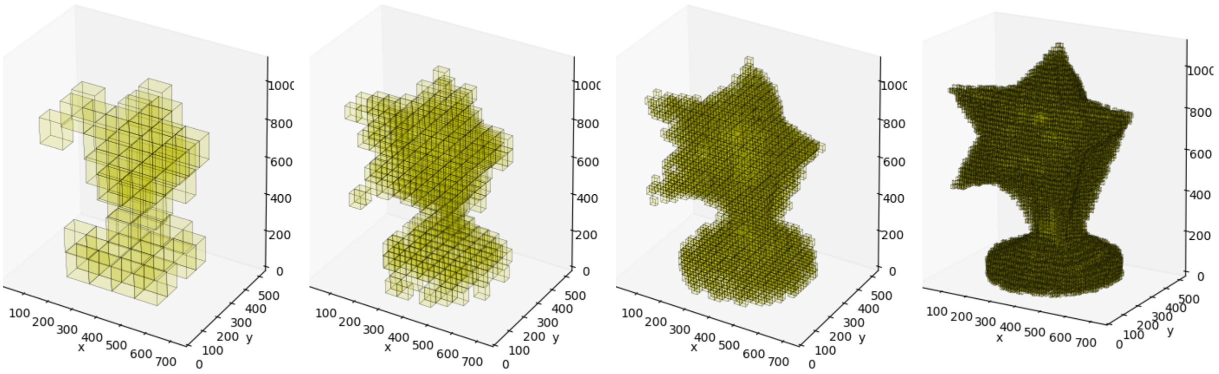
Following this result, we wanted to go further in reducing the number of variables and use a data structure that would allow us to incorporate the mask, but also to be able to describe homogeneous zones with a single variable.

4.2 Hierarchical structures

4.2.1 Octrees

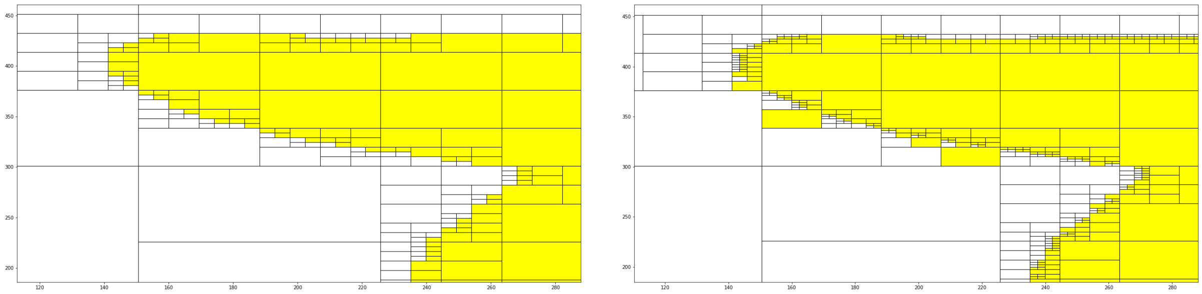
The first structure we examined to reduce the number of variables was the octree [122]. Widely regarded in the literature as an optimal method for representing a voxel grid, the octree allows for efficient storage of empty grid regions as null pointers, resulting in a *Sparse Voxel Octree*. Due to its simplicity and extensive use, this was our logical starting point. The process began by constructing an octree from a CAD model, which required a methodology akin to solid voxelisation for accurately filling the internal volume. Figure 4.13 demonstrates the octree structure at different depths for the star trophy, showing that while greater depth improves representation fidelity, it also significantly increases the number of tree leaves.

We initially experimented with the voxelisation technique proposed by [6]. This approach facilitates the creation of a sparse voxel octree without necessitating a complete intermediate solid voxelisation, thereby enabling GPU-based voxelisations of unprecedented scale. However, the complexity of its implementation led us to adopt the implementation



(a) Depth 3 (b) Depth 4 (c) Depth 5 (d) Depth 6

Figure 4.13: Octree's depths



(a) (b)

Figure 4.14: Examples of two quadtrees generated from the CAD mask of the star trophy. The difference between the two quadtrees arises from a slight translation of the image.

provided by Jeroen Baert [132], which allows us to voxelise complex meshes. We successfully voxelised the part and converted it into an octree [133]. This process, however, was not straightforward and resulted in several artefacts. For instance, Figure 4.14 shows two quadtrees generated from the cross-section of the mask used in the previous section. While the first example (Fig 4.14a) accurately represents the star, the second example (Fig 4.14b) illustrates a significantly different quadtree, with deeper leaves near the edges and coarser regions elsewhere. The only difference between the two was a translational adjustment of the mask. This issue indicates that translations or misalignments pose a substantial challenge, requiring additional nodes and increased complexity to maintain optimal mask contour descriptions. Thus, continuous tree structure adjustments during reconstruction became necessary.

Initialising the tree with a CAD model, while a feasible starting point, proved suboptimal due to the extensive structural modifications caused by real object misalignment. Alternatively, one could consider initialising the tree with a quick FDK reconstruction to align it correctly. However, this strategy diverges from our primary objectives.

We have implemented a ray tracing algorithm on an octree as described by Revelles et al. [134][2]. This octree traversal method is termed 'top-down,' where intersections between a ray and each node of the octree are computed parametrically while descending through the tree. The algorithm leverages the properties of the octree, such as dividing the space into two equal zones along each axis at each node, to accelerate the calculation of intersections. For each leaf node, the traversed length can be computed. Although this algorithm is reputed for its speed and parallelisability, our implementation did not achieve satisfactory performance for reconstructions of the desired size. Our code, written in Python, was hindered by the lack of pointers, which impeded efficient tree management. Furthermore, the use of GPU parallelisation and just-in-time compilation tools proved difficult to apply to octrees within Python. Due to these challenges, we opted to change our strategy and explored an alternative structure.

4.2.2 Meshes & Bounding Volume Hierarchies

In our research, we aimed to explore a more versatile and adaptable framework beyond voxels. Meshes emerged as an appealing option due to their inherent flexibility. However, previous investigations have indicated constrained outcomes with meshes, primarily due to slow convergence caused by the computational intensity of projection and backprojection operators. Conversely, tree structures have been fine-tuned for efficient ray-tracing. Trees boast rapid algorithms developed for graphics rendering and computer vision. Our goal is to harness the strengths of both structures to optimise our reconstruction process. Utilising a mesh that we will integrate into a hierarchical structure, we can expect the speed advantages of trees with the adaptable nature of meshes. This hybrid approach has the potential to significantly improve reconstruction efficiency and quality.

However, the octree has the drawback of consistently dividing space into two equal parts. The space partitioning, therefore, relies on the object's position. As an alternative, we suggest using a more general tree structure known as the Bounding Volume Hierarchy (BVH). A BVH is a data structure used primarily in computer graphics and computational geometry to accelerate ray tracing and collision detection algorithms. It organises objects in a scene or a set of geometric shapes into a hierarchical tree structure. The fundamental idea behind a BVH is to group objects together within hierarchical bounding volumes to facilitate efficient spatial queries, such as ray intersection tests, then the time complexity can be lowered to logarithmic in the number of objects. Each node in a BVH tree represents a bounding volume that encapsulates a set of objects or other bounding volumes. Bounding volumes can be represented by any shape but the most common are Axis-Aligned-Bounding-Boxes (AABB). Figure 4.15 illustrates a scene containing six objects. These objects are grouped into bounding boxes, which are then recursively grouped into larger bounding boxes. The hierarchical structure of these bounding boxes is represented by nodes in the BVH tree.

Regarding meshes, choices had to be made for our study as well. Since most CAD models are in the STL format, which represents triangles forming a surface, we first needed a method to convert a CAD model into a volumetric mesh. For this purpose,

---

[2] In their paper, Revelles et al. [134] discussed extending the quadtree algorithm to octrees. Their Table 1 gives indications to select the first node. This selection is implemented using bitwise operators. There is an error in the operations concerning the Z-axis: the rightmost bit must be altered (0) rather than the leftmost bit (2), as the leftmost bit pertains to x-axis operations. Correcting this by switching the 0's and 2's results in the accurate table.

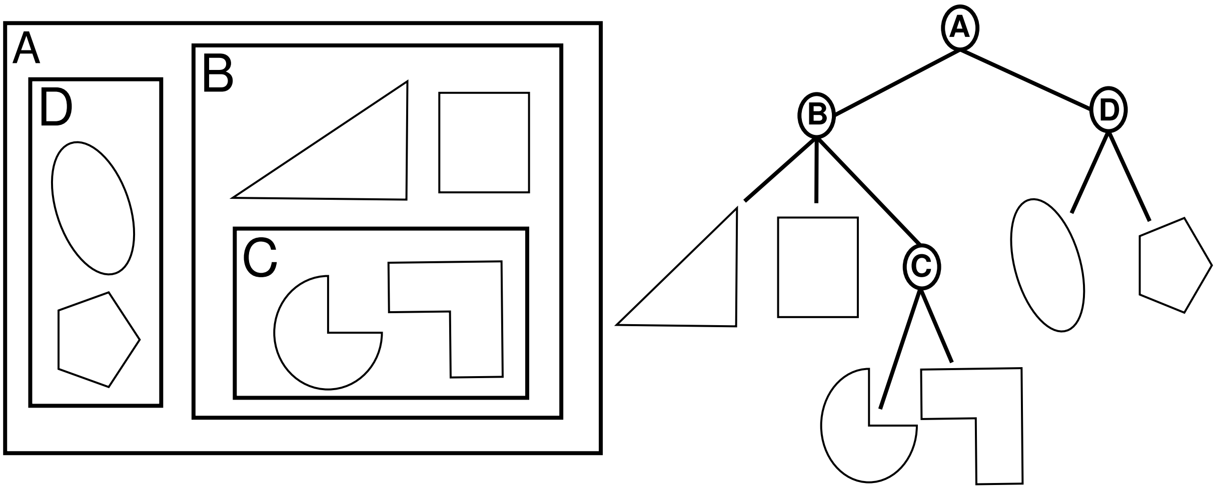


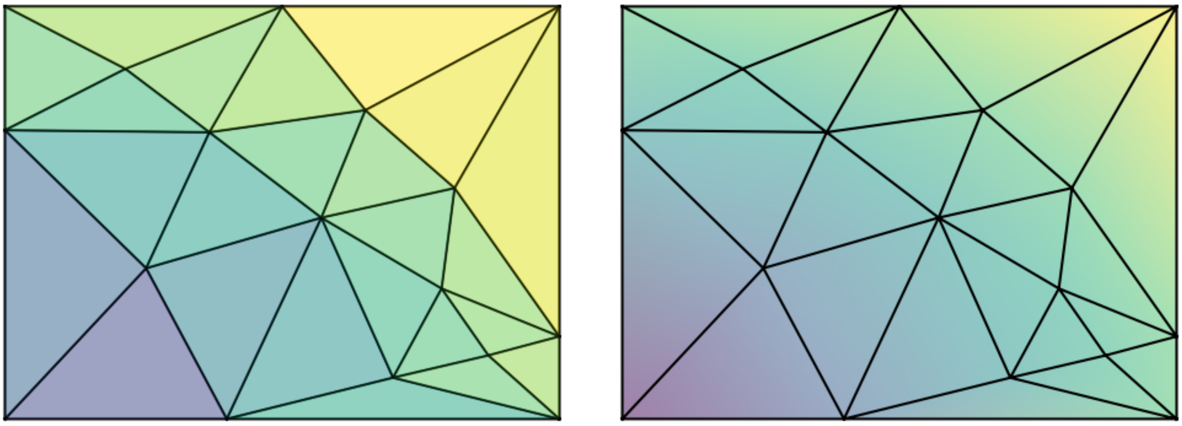
Figure 4.15: Representation of a BVH. On the left, the scene consists of six objects hierarchically distributed within bounding boxes (A, B, C, D). On the right, the BVH tree of the same scene is depicted. Each bounding box contains either other bounding boxes, the objects in the scene, or both.

we employed the method described in [135], which generates tetrahedra from an STL model. The advantage of using tetrahedra lies in their simplicity and compatibility with existing and efficient tools for our algorithms. For example, we were able to utilise the Intel® Embree library on the CPU. Intel® Embree is an open-source high-performance ray tracing library widely used in the field of computer graphics [136]. However, this library originally provides intersections between surface intersectables (i.e. surfaces) and rays. Consequently, the library was slightly modified to calculate intersections involving tetrahedra, by decomposing tetrahedra into triangles. Note that, for radiology, it is possible to make a projection directly from a surface description of the part, as is the case with CAD. Embree can be used directly without mesh processing. Triangles from an STL file can be directly integrated into a BVH for ray tracing. Absorption will depend only on the penetration length (monomaterial hypothesis). While Embree can offer barycentric information and ray tracing with nodal encoding is feasible, for simplicity, we opted for a Siddon-type ray tracing on cellular encoding.

When reconstructing using meshes, there are two methods to store the data: *cellular* and *nodal* [121]. Similar to voxels, cellular encoding saves a density value for each mesh cell, resulting in a piecewise constant representation of the reconstructed object (see Figure 4.16a). However, even with adapted discretisations, some cells might intersect object interfaces while others encompass heterogeneous regions within the object. Consequently, the cellular encoding might not remain optimal due to these intersecting cells and heterogeneous regions.

On the other hand, nodal encoding stores values for each vertex of the mesh. The density value is computed by interpolating between the vertices of the cell. This approach permits a smoother density gradient in the image (see Figure 4.16b). However, it tends to reduce the visibility of discontinuities, replacing sharp attenuation values on either side of an interface with a gradual transition between these two values.

Ray tracing also adapts according to the chosen encoding. With cellular encoding, a constant value is assigned to a cell. Thus, to determine a cell's absorption, only the



(a) Cellular encoding: each cell has a constant value

(b) Nodal encoding: the values in a cell are interpolated from the values at the vertices

Figure 4.16: Nodal and cellular encoding

traversed length is necessary. The classical ray tracing approaches (see Fig. 4.17a) can be employed for the reconstruction in this scenario. The value of the i^{th} ray is given by:

$$\mathbf{p}_i = \sum_j \boldsymbol{\mu}_j \ell_{ij}. \tag{4.7}$$

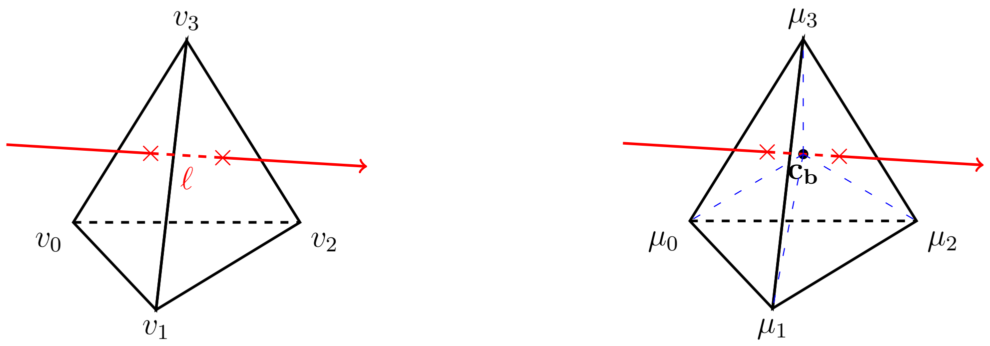
The nodal encoding does not possess the same piecewise constancy property. Therefore, the density value must either be integrated along the ray or averaged (see Fig. 4.17b). The latter is of course simpler and will be used most of the time for nodal encoding. For a tetrahedral mesh, if we note $\mathbf{c_b}$ the centre of the intersection between a tetrahedron and the ray ($\mathbf{c_b}$ is often expressed in the barycentric coordinates of the tetrahedron) and by denoting $\boldsymbol{\mu}_j = (\mu_0, \mu_1, \mu_2, \mu_3)$ as the vector representing nodal attenuations of the j^{th} cell, the projected value for the i^{th} ray is then expressed as:

$$\mathbf{p}_i = \sum_j \langle \mathbf{c_b} | \boldsymbol{\mu}_j \rangle \times \ell_{ij}. \tag{4.8}$$

Nonetheless, it is worth noting that the dual graph of our mesh can also be considered during reconstruction. Reconstructing the object on its dual grid provides supplementary information and potentially enables the interpolation of reconstructed values onto the final result.

4.2.3 Results

Initially, we assessed the projection and backprojection capabilities with BVH. For this purpose, we implemented two ray casting approaches. The first one, "handmade", hierarchically seeks the intersection of a ray with each bounding box of the BVH. The second approach relies on Embree. The Embree version is faster, taking 5.44 seconds for one projection while our implementation took 294 seconds (54 times slower). This highlights the relevance of this optimised library, even though it does not leverage GPU acceleration.



(a) Cellular coding. The contribution of this cell to the attenuation of the ray corresponds to μ times the travelled length ℓ. The absorption is constant within the tetrahedron.

(b) Nodal coding. The contribution of this cell to the attenuation of the ray corresponds to the travelled length ℓ times the attenuation value obtained by interpolation.

Figure 4.17: Cellular and Nodal Ray Tracing

All projections and backprojections will be made using the Embree kernel. Figure 4.18 shows our first 1880×1212 pixels projection of the meshed trophy CAD model. In this example, all tetrahedra are filled with the same constant value.

Although the initial projection appears satisfactory, a closer examination reveals a few anomalies within the projection values. The zoomed-in section at the bottom right of the image highlights these issues, showing black dots aligned along the edges of the inner cells. These anomalies are attributable to small tetrahedra. Embree's ray casting is primarily designed for graphic rendering, where only the first intersection of a ray with an object is significant, typically for opaque objects. To address this, we relaunch a ray in the same direction after each ray/triangle intersection, originating slightly behind the detected intersection. However, when multiple tetrahedra share a point, they tend to be thin, making the slight offset from the first intersection potentially too large, causing the ray to miss subsequent cell interactions. This reliance on cell continuity for calculating traversed thickness leads to errors.

A straightforward correction for this artefact involves launching the same ray but in the opposite direction. In this approach, the first detected intersection in the reverse pass will be the one missed in the initial pass. Figure 4.19 illustrates this method. After the first intersection '1/2', the red ray, relaunched from the third zone, misses the second zone. Conversely, the orange ray, travelling backwards, detects the '2/3' intersection but misses the '1/2'. By using both rays, all intersections are detected. While ideally, minimising the offset would resolve this issue, in practice it did not due to the floating-point precision. This bidirectional approach ensures that any missed intersections are accounted for, thereby enhancing projection accuracy and eliminating black dots.

Although our projection and back-projection operations are now correctly implemented, they remain quite computationally intensive. Consequently, we have only tested the reconstruction on meshes for a subset of our data. In this example, we have only considered the central four rows of the experimental projections. However, we first had to create a complete tetrahedral mesh from the registered CAD model and then trim this mesh to include only the part to be reconstructed. Similar to the mask, the CAD model of the part was slightly dilated to ensure all the material was encompassed. We tested

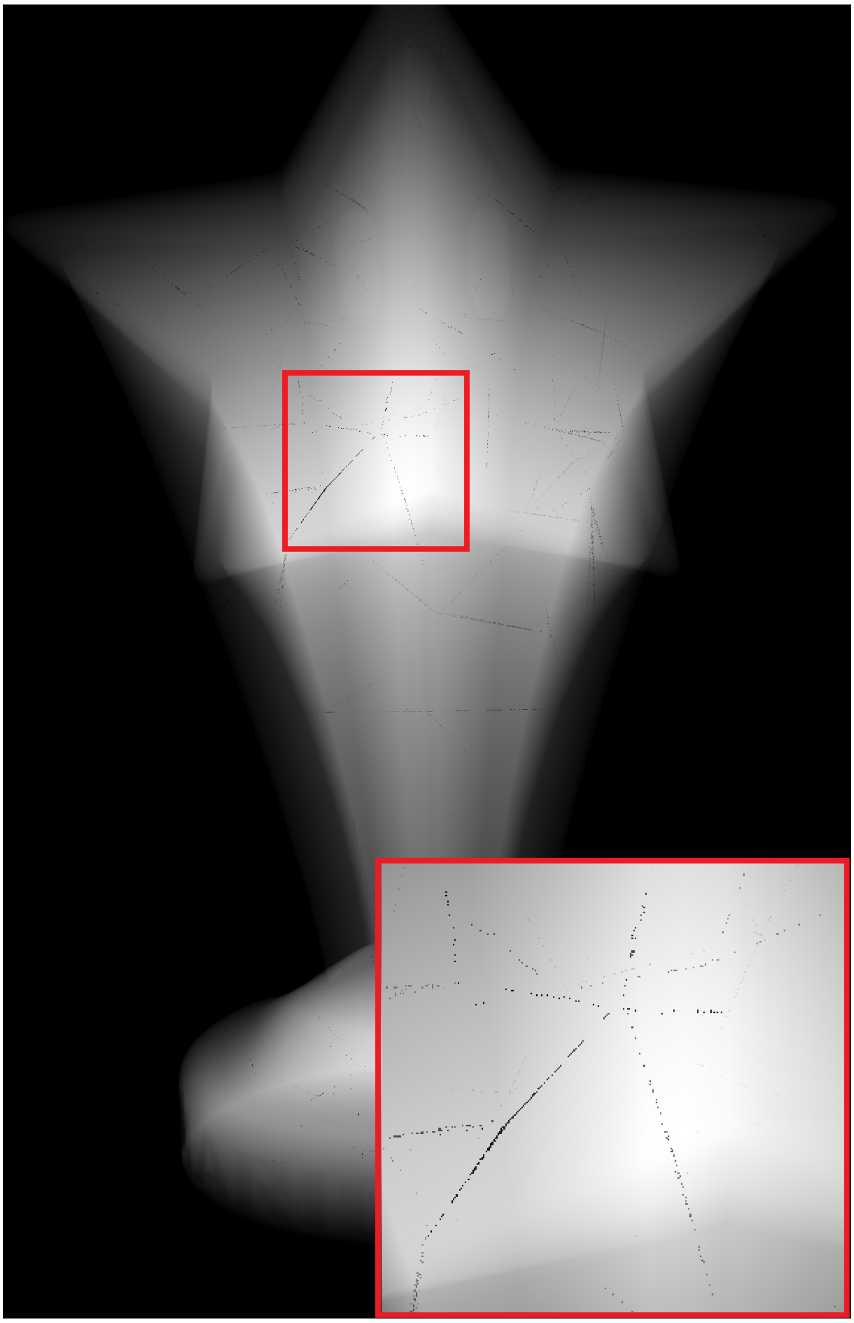


Figure 4.18: Simulated projection on tetrahedra using the meshed trophy model and a zoom showing black dots along the mesh's edges.

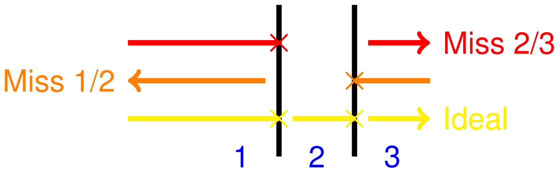
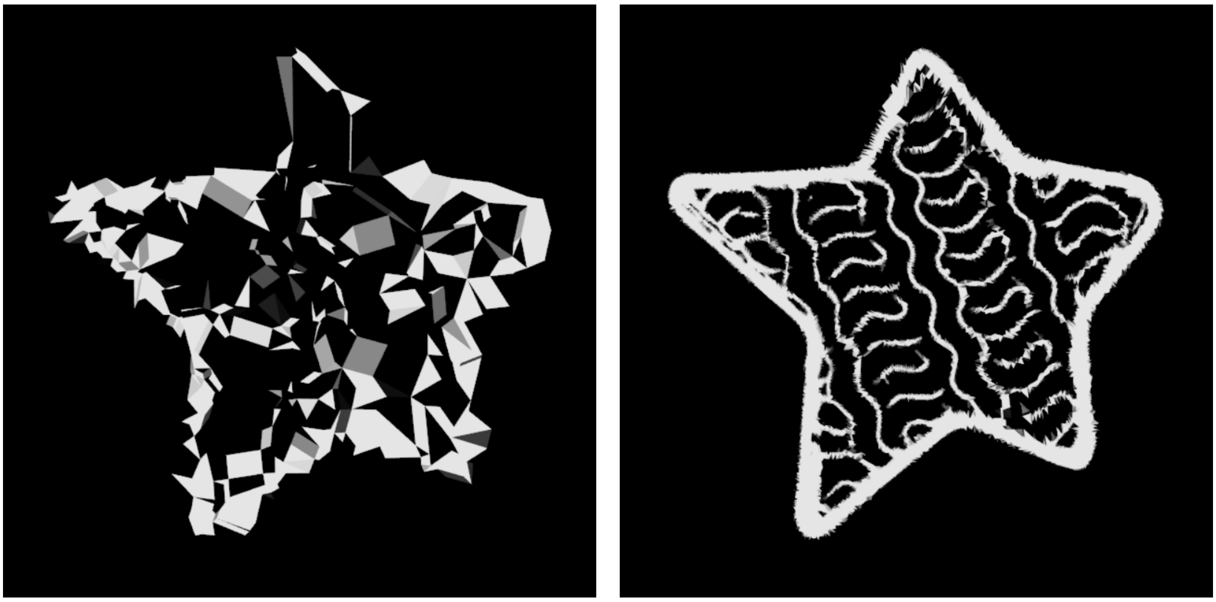


Figure 4.19: Explanation and proposed solution for black dots in projections. The offset between each relaunch of the ray can lead to intersections being missed. By casting a ray in the reverse direction, we aim to detect any missing intersections.

two sets of parameters for the tetrahedral size.

In our initial attempt, the cell sizes were too large. The mesh contained 1257 tetrahedra for a volume that previously corresponded to 4x1204x1204 voxels, of which 643.926 were included in the mask (representing only 0.2% of the variables compared to the voxel-based description). Figure 4.20a shows the cross-section of the reconstruction with this mesh. The reconstruction was performed using SART with 10 iterations on 300 projections.

The mesh was too coarse. The shape of the object is barely recognisable. The tetrahedra are too large and irregular, giving the representation a very polygonal appearance, which is both unpleasant and inaccurate. Consequently, we conducted another reconstruction, significantly increasing the number of tetrahedra to 2.172.096 (37% of the variables compared to the voxel-based description). Each tetrahedron is divided into four smaller tetrahedra and one octahedron, which is then further divided into eight tetrahedra. This operation was repeated several times to achieve the new mesh. In this implementation, an upper limit on cell size was imposed to ensure small cells even in empty regions, meaning the number of cells can still be significantly reduced. Figure 4.20b shows the reconstruction, which looks better but is still not satisfactory. The shape of the figurine is more discernible, with better-defined material and void areas. There seem to be fewer streak artefacts compared to the voxel-based reconstruction, and a better-defined background, but the display with triangles creates a visual effect that makes comparison by eye difficult.



(a) Reconstruction on a coarse mesh. The central slices of the volume were reconstructed on an unsuited mesh. The size of the tetrahedra needs to be reduced.

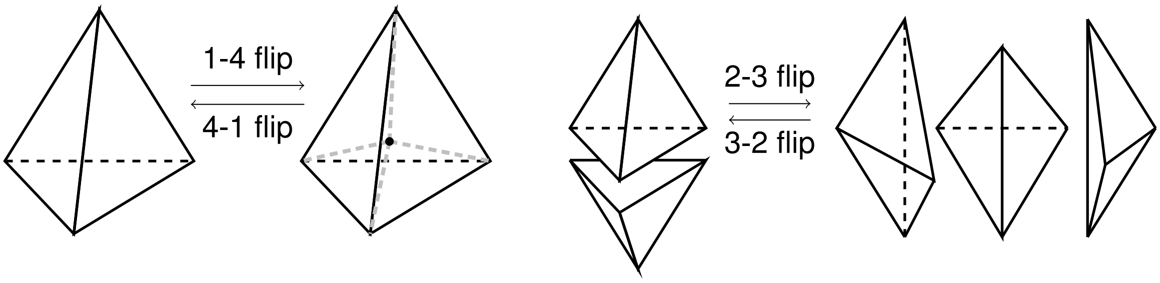
(b) Reconstruction on a refined mesh. The size of the tetrahedra seems visually correct, but the orientation of the triangles creates an undesirable visual effect.

Figure 4.20: Comparative reconstructions on different meshes.

4.2.4 Perspectives

An unfinished development of this method pertains to mesh updating. To achieve an optimally fitting mesh for the reconstructed object, it is necessary to introduce a strategy for mesh evolution throughout the reconstruction iterations.

The first strategy we aimed to experiment with was simply calculating the gradient across the volume. If the gradient is too steep, the cell needs to be subdivided. This operation can be achieved using topological operations. Any mesh modification can be decomposed into a series of elementary moves, called bistellar flips or Pachner moves [135]. For example, Figure 4.21 illustrates the '1-4' flip and the '2-3' flip for tetrahedra. There are also operations such as splits and edge removals, among others [137].



(a) '1-4' and '4-1' flips

(b) '2-3' and '3-2' flips

Figure 4.21: Pachner moves. The tetrahedra are subdivided and merged to realign them according to the gradient of the reconstructed volume.

The advantage is that we can increase the number of cells if the gradient is too strong and conversely, reduce the number of cells if the gradient is weak. Additionally, '2-2' flips can be used to align the face normals with the gradient. Subsequently, vertices can also be moved to improve mesh quality. Following each iteration of vertex adjustments, attenuation values within each tetrahedron can be revised or updated (Merckx et al. [14]). We began developing these methods between the iterations of SART, but unfortunately, we were unable to complete our experiments.

Figure 4.22 shows the ongoing developments. The plotted cross-section is obtained for a reconstruction with mesh modification between each iteration. The final mesh contains 84.212 tetrahedra, i.e. 13% of the variables compared to the voxel-based description. The results visually demonstrate a better representation of the object with fewer tetrahedra than before. Tetrahedra appear more regular, creating the impression of a pleasant, uniform tiling. The non-optimised mesh version exhibits spikes and vertices shared by numerous tetrahedra, resulting in a less aesthetic representation. Each iteration took approximately 5 minutes. Figure 4.23 also includes a zoomed view of the cross-section with and without mesh optimisation. The optimisation clearly enhances the quality of the reconstruction. The jagged appearance of some tetrahedra is eliminated in favour of a representation with more balanced tetrahedra. The faces of the tetrahedra are better aligned with the contours, allowing for a clearer delineation. The improvement in meshing enhances the study of connectivity. With the optimised mesh, the horizontal branch on the right is not connected to the vertical one.

Currently, the mesh appears satisfactory, and the reconstruction seems to be of good quality. However, compared to the voxel-based description, the image contains fewer

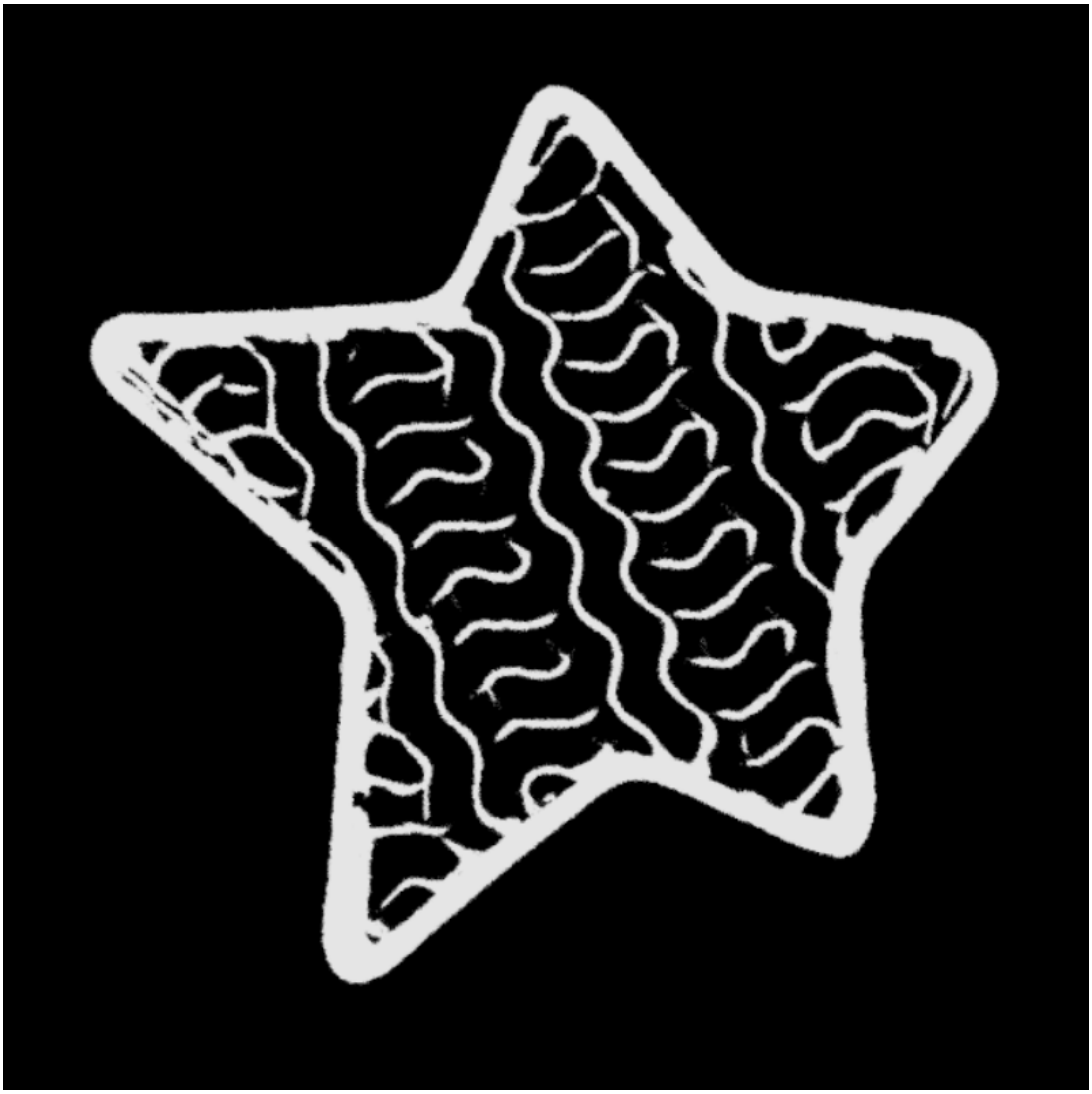


Figure 4.22: Reconstruction on an optimised mesh

shades of grey and exhibits a more binary aspect: material/void. This result is encouraging as it enables better segmentation of the object and a clearer definition of its contours despite the reduced number of tetrahedra. The goal is to further optimise the mesh and to use a differentiation tool, such as the one from [138], to more optimally reposition the vertices. Additionally, we are not fully leveraging the potential of BVH. In this thesis, we have limited ourselves to the use of tetrahedra, but it is essential to consider advancing the representation with more complex meshes. The CAD model provides an excellent starting point for meshed reconstruction, and if it contains an analytical representation (not just triangles, for instance), it could be feasible to optimise curved elements. This would also change the manner in which the real object is compared to the CAD model.

The issue of representing the reconstructed volume is inherently complex. Through our results, we observed that using masks, octrees, and meshes indeed helps reduce the number of variables, thereby enhancing the reconstruction process. Mask-based representations offer significant improvements without requiring complete modification of our existing codes. However, moving away from the traditional voxel-based representation necessitates the reimplementation of numerous components. Hierarchical structures and mesh representations demand considerable effort, as we need to rewrite our computation

(a) Unoptimised but refined mesh

(b) Optimised mesh

Figure 4.23: Aspects of the different meshes. Optimising the mesh throughout the iterations results in a structure that is better suited to representing the image and avoids the jagged visual effect.

kernels, visualisation functions, and post-processing routines. Nonetheless, the promising outcomes of our studies lead us to believe that overcoming these challenges is worthwhile, as it could greatly expand the possibilities for the future of X-ray tomography.

Chapter 5

Dictionary Learning and Sparse Coding

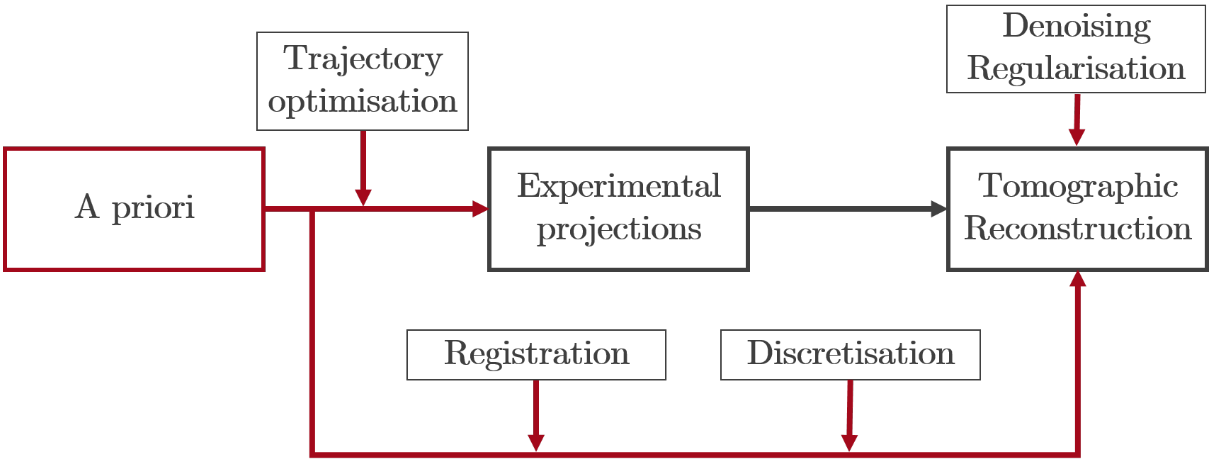


Figure 5.1: Diagram of the tomographic process using a priori information. In this section, we will study specific denoising and regularisation methods for sparse-view CT.

Within the framework of the sparse-view strategy, the lack of information inevitably leads to artefacts and diminishes reconstruction quality. While all the methods presented earlier help enhance reconstruction quality, the most effective approach is to directly act on the volume to be reconstructed through denoising or regularisation functions. To incorporate *a priori* knowledge during the reconstruction, one must use a *Statistical Iterative Reconstruction* (SIR). In sparse-view CT imaging, where data statistics significantly influence image reconstruction, SIR methods prove advantageous. Unlike analytical methods such as FBP, these methods employ iterative strategies to derive the most probable image based on projection measurements. Due to significant improvements in balancing the number of views (or radiation dose) against image quality, SIR methods are pivotal in low-dose and sparse-view CT imaging. Various approaches fall under the SIR category [139].

The mathematics underlying SIR methods are based on maximum *a posteriori* estimation. In addition to a data fidelity term to be minimised, these methods include a prior term, also known as the regularisation or penalty term. This term encapsulates prior knowledge about the object being studied and additional physical effects not included in the model [140].

Regularisation terms can vary widely in their formulation. They range from spatially independent priors, which, while preventing value divergence, are often deemed ineffective, to more sophisticated models like Markov random fields that enforce local properties such as smoothness by considering pixel neighbourhoods. Other techniques include compressed sensing approaches like Total Variation, anisotropic TV, and wavelets, as well as more recent methods leveraging deep learning [141, 142]. The literature on regularisation through priors is extensive and well-documented [139].

Additionally, the diversity of regularisation terms has fostered the development of *Plug-and-Play* methods. Plug-and-Play methods are reconstruction techniques that allow the incorporation of any denoising/priors term. Generally, two classes of algorithms are distinguished: the classical P^3 where the denoiser D replaces the proximal operator ($D = \mathrm{prox}_{\tau g}$), and the RED (Regularisation by Denoising) algorithms where the denoiser replaces the gradient descent operator ($D = I - \nabla g$) [143, 144].

Among all these techniques, we focused our interest on one particular family: *patch-based techniques* that consider patches (or blocks in 3D) instead of individual pixels. It is increasingly observed that methods utilising patches and signal redundancy are effective for denoising. The concept behind patch-based methods is to process only small portions of the images to be reconstructed. Each patch is treated independently or in conjunction with other similar patches. The patches are then re-aggregated to form the complete image. Techniques based on patches are becoming increasingly prevalent due to their ability to partition images into manageable segments, thereby overcoming limitations imposed by digital resources. They successfully mitigate the curse of dimensionality, as learning a model to solve a problem is always simpler on smaller signals. Additionally, processes are generally faster and more efficient. However, patches often overlap, which increases computational time. Thus, patch-based strategies require careful consideration regarding implementation and parallelisation of computations. But at the same time, overlapping patches ensure a sufficient number of samples to identify patterns and leverage the redundancy in the signal, significantly enhancing denoising efficacy. In imaging, the redundancy and patterns that we are looking for are typically the texture of the images, and finding them allows us to perform a tailored denoising [145].

One of the most famous patch-based denoising techniques is the *Non-Local Means* (NLM), due to its ability to leverage the inherent redundancy present in natural images. Unlike traditional denoising methods that primarily rely on local information, NLM exploits the self-similarity within an image by comparing and averaging non-local patches. In the NLM algorithm, the denoised value of a pixel is computed as a weighted average of all pixels in the image, where the weights are determined based on the similarity between the patches centred around the pixels. This approach allows NLM to preserve fine details and textures effectively while reducing noise. By operating on patches rather than individual pixels, NLM can capture more contextual information, leading to superior denoising performance, especially in images with repetitive structures or textures. We also utilised a variant of this technique that employs priors as a regularisation term. In [11], we used a simulated reconstruction for measuring patch similarity. This technique allowed us to eliminate numerous artefacts in sparse-view reconstruction. Thanks to rapid implementation, the patches of the ongoing reconstruction were compared to those in the vicinity in the reference simulated reconstruction, allowing the registration between the two reconstructions to be approximate. Another famous technique is the *Block-Matching and 3D Filtering* (BM3D) [146]. By dividing the image into overlapping patches, BM3D searches for similar patches within a predefined search window. Once similar patches are identified, a collaborative filtering approach is employed to estimate the clean patch based on the information from similar patches. This collaborative filtering strategy ensures that noise is effectively removed while preserving image details. BM3D has been particularly successful in denoising tasks due to its ability to exploit both local and non-local similarities in the image. It has been widely adopted in various imaging applications, demonstrating its effectiveness as a patch-based denoising technique. NLM and BM3D will be used as references to compare our results.

In our case, we operate within a context where we have one or two samples already scanned and analysed on a sparse or dense trajectory. We assume that some *a priori* knowledge about the image texture is available, but we lack sufficient samples to employ advanced learning methods such as Deep Learning. We aim to propose a method capable of learning and denoising a specific texture with minimal information. Following our literature review, we focused on *sparse coding* and *dictionary learning* based methods. The versatility and flexibility of these methods appeared well-suited to our use case.

5.1 Sparse Coding concepts

5.1.1 Sparse coding

The concept of sparse coding is handy in industrial tomography because the *a priori* information necessary for developing methods by sparse coding is relatively undemanding. The number of samples for training is low. The methods generalise well on different data types and remain robust even when faced with images different from those on which the learning was done. The concept of sparsity is already well-established in the X-ray community thanks to the compressed sensing framework. However, in its formulation, the signal is usually taken as a whole [147, 148].

Conversely, the idea behind sparse coding is that each small image patch from the signal is sparse on a tailored basis. Working with patches instead of the whole image can have numerical advantages in handling large reconstructions. Furthermore, sparse coding is closely linked to *dictionary learning*. The first concept consists of writing each patch of our complete image as a linear sparse sum of elements, called *atoms*, and the second consists of building those elements. Dictionary Learning is the preliminary step in which a redundant over-complete basis, called a *dictionary*, is built to favour the sparsity of a specific kind of signal. Figure 5.2 illustrates the principles of sparse coding. On the left-hand side, we observe a cross-section of the 3D reconstructed star trophy, which in this example is the signal we are aiming to approximate. The right-hand side exhibits the atoms of our dictionary, comprising individual 8×8×8 patches. Notably, each patch within our signal is approximated through a sparse sum of these dictionary atoms, this is the *sparse coding*.

Let's note $\mathbf{x} \in \mathbb{R}^N$ an image, $\mathbf{D} \in \mathbb{R}^{m \times n}$ a dictionary, i.e. an over-complete basis. We aim to find how to encode each patch $\mathbf{x}_s \in \mathbb{R}^m$ extracted from the signal $\mathbf{x}$. The code $\mathbf{z}_s \in \mathbb{R}^n$ should be sparse if possible. Usually, $n > 10m$, and patches are squares or cubes of side 8 or 10 pixels/voxels. The primary challenge of sparse coding lies in the computation of sparse representations, which entails identifying the optimal solution from the infinite set of solutions within an under-determined linear system. For every patch, one has to solve:

$$\min_{\mathbf{z_s} \in \mathbb{R}^p} \|\mathbf{z_s}\|_0 \text{ subject to } \|\mathbf{x_s} - \mathbf{D}\mathbf{z_s}\|_2^2 \leq \epsilon^2, \tag{5.1}$$

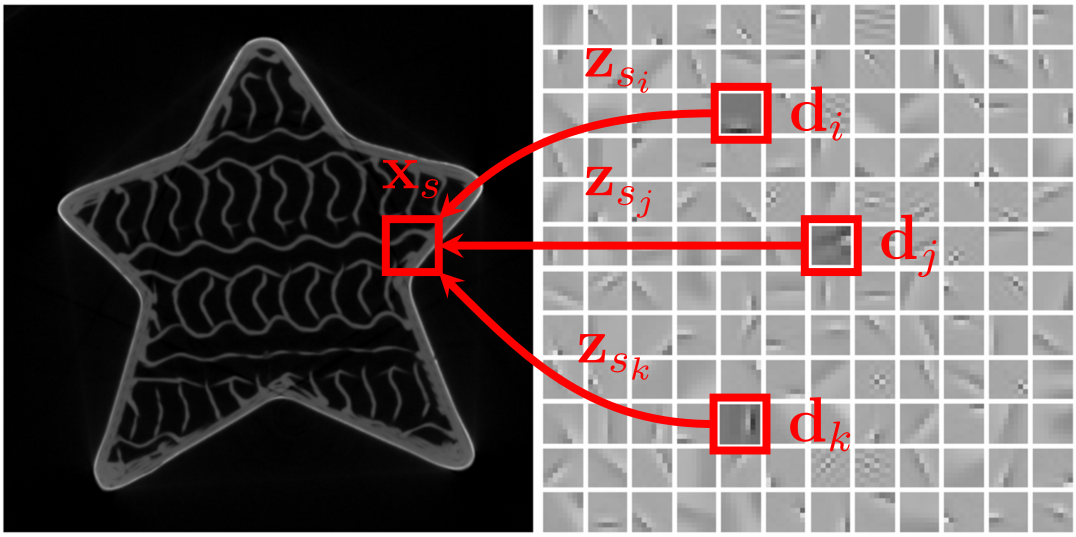


Figure 5.2: Principle of Sparse Coding. On the left, the image, denoted as $\mathbf{x}$, represents our signal to be locally approximated by patches. On the right, the dictionary atoms serve as the fundamental building blocks for our sparse approximation. The patch $\mathbf{x}_s$ of the complete signal $\mathbf{x}$ is approximated by a sparse sum of three atoms ($\mathbf{d}_i$, $\mathbf{d}_j$, $\mathbf{d}_k$) weighted by the code ($\mathbf{z}_{s_i}$, $\mathbf{z}_{s_j}$, $\mathbf{z}_{s_k}$).

where ϵ is a small positive constant. The equality $\mathbf{x_s} = \mathbf{D}\mathbf{z_s}$ is difficult to achieve because of noise in the signal. Therefore, this notation, which ensures a small gap between the signal and its approximation, is preferred. Due to the over-completeness of the dictionary, the solution is not unique, but the additional criterion of sparsity helps resolve the degeneracy introduced. This problem possesses a combinatorial nature. There are two classes of algorithms to solve the problem: greedy methods and convex relaxation techniques. The formers are usually formulated with a ℓ_0 pseudo-norm on the code, while the latter use a ℓ_1 approximation. In the literature, the expressions *matching pursuit* and *basis pursuit* (BP) are used to describe respectively those two formulations, which can be mathematically written:

$$\hat{\mathbf{z}_s} = \underset{\mathbf{z}_s}{\operatorname{argmin}} \; \|\mathbf{D}\mathbf{z}_s - \mathbf{E}_s\mathbf{x}\|_2^2 + \lambda\|\mathbf{z}_s\|_{0/1}, \tag{5.2}$$

where λ is a parameter that balances the data fidelity and sparsity terms, $\mathbf{E}_s$ represents an operator to extract the block $\mathbf{x}_s$ from the whole signal $\mathbf{x}$. $\mathbf{E}_s$ represents a diagonal matrix comprising solely zeros, except at the indices corresponding to the voxels of patch s, where ones are situated. Consequently, when multiplied by $\mathbf{x}$, the resulting product yields $\mathbf{E}_s\mathbf{x} = \mathbf{x}_s$. The used norm can be either ℓ_0 or ℓ_1 according to the chosen formulation. Unlike equation 5.1, the Lagrangian formulation is used here. The choice of norm is crucial in general. It depends not only on what we aim to model but also on the associated solving techniques. The pseudo-norm ℓ_0 enables the discovery of sparser representations compared to norm ℓ_2 when solving systems. However, solving with the pseudo-norm ℓ_0 poses a non-convex problem, which is particularly challenging in optimisation (see Figure 5.3). The compromise lies in utilising norm ℓ_1, which is convex and yields sparse results [147, 148].

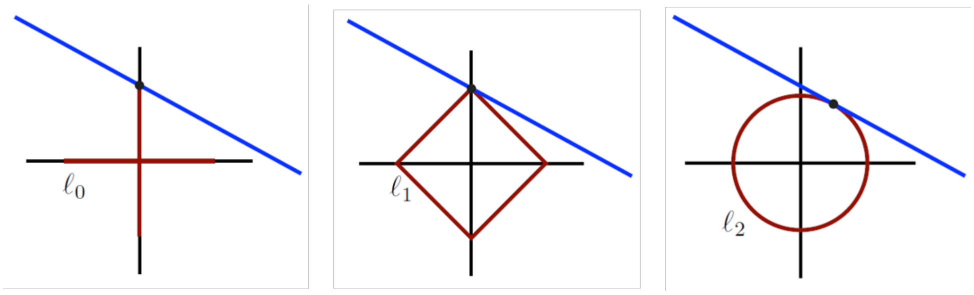


Figure 5.3: Representation of isonorms for the ℓ_0 pseudo-norm, ℓ_1 and ℓ_2 norms. If we consider that the solutions of the linear system belong to the blue line, the position of the solution with the smallest norm is sparse with the norms ℓ_0 and ℓ_1.

5.1.2 Matching Pursuit

The matching pursuit problem involves finding a sparse linear signal approximation. The simplest algorithm to do so is the *Iterative Hard Thresholding*. A hard threshold $\mathcal{H}_\tau$, which keeps only the τ largest entries in modulus, is applied at each iteration of a classic optimisation scheme to ensure sparsity. The hard threshold function of parameter one is shown in Figure 5.4. The algorithm produces the sequence:

$$\mathbf{z}^{(k+1)} = \mathcal{H}_\tau(\mathbf{z}^{(k)} - \mathbf{D}^\mathsf{T}(\mathbf{D}\mathbf{z}^{(k)} - \mathbf{x})). \tag{5.3}$$

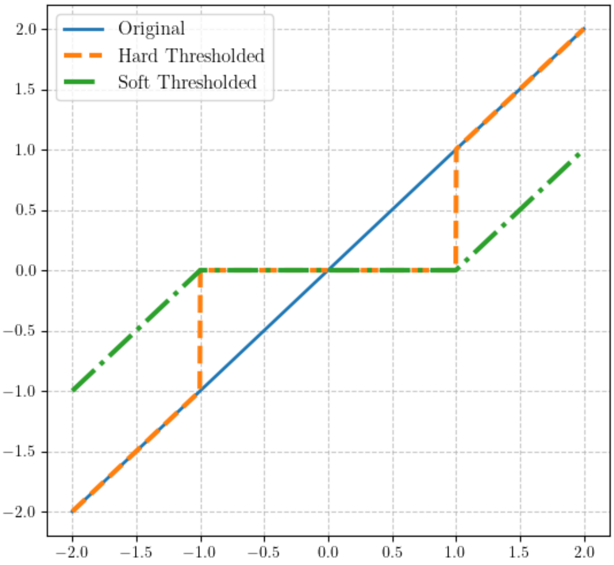


Figure 5.4: Soft and Hard Threshold

There are theoretical guarantees that this procedure succeeds [149, 150], but they are weaker than other common procedures. This is why despite its simplicity and speed of execution, the most famous matching pursuit algorithm is, without a doubt, the *Orthogonal Matching Pursuit* (OMP), which iteratively adds elements to the representation. Algorithm 8 shows how OMP builds the signal representation's support $\mathcal{S}$. At each iteration, OMP selects the atom most correlated to the residual, adds it to the support, and updates the code by solving a least squares problem. Note that all previous coefficients are updated too. The name "*Orthogonal*" Matching Pursuit comes from the orthogonality between the residual and the span of the atoms already selected, which comes from the least squares solution. This property also guarantees that the selected atom $\mathbf{d}_k$ belongs to the dictionary $\mathbf{D}$ restricted to the elements not in $\mathcal{S}$. The algorithmic complexity of OMP is reasonable, especially when the sparsity criterion is strong. The calculation of the index k is $\mathcal{O}(nm)$, the least squares problem step is the most consequent in $\mathcal{O}(m\tau^2)$ and the calculation of the residual $\mathcal{O}(m\tau)$. Overall complexity is about $\mathcal{O}(\tau m(\tau^2 + n))$.

Algorithm 8: Orthogonal Matching Pursuit.

```
Data:   Dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$
        Signal $\mathbf{x} \in \mathbb{R}^m$
        Sparsity level $\tau$
        Convergence criteria $\epsilon$
Result: Representation support $\mathcal{S}$
        Sparse reconstruction $\mathbf{z} \in \mathbb{R}^n$
Initialise $\mathcal{S} = \varnothing$, $\mathbf{r} = \mathbf{x}$
while $\|\mathbf{r}\|_2 > \epsilon$ and $\|\mathbf{z}\|_0 < \tau$ do
    Find $\mathbf{d}_k \in \mathbf{D}$ with maximum inner product $|\langle \mathbf{r}, \mathbf{d}_k \rangle|$
    Update support $\mathcal{S} \leftarrow \mathcal{S} \cup \{k\}$
    Update code $\mathbf{z} \leftarrow (\mathbf{D}_{\mathcal{S}}^{\mathsf{T}} \mathbf{D}_{\mathcal{S}})^{-1} \mathbf{D}_{\mathcal{S}}^{\mathsf{T}} \mathbf{x}$
    Compute new residual $\mathbf{r} \leftarrow \mathbf{x}$ - $\mathbf{D}_{\mathcal{S}} \mathbf{z}$
end
```

Note that there are other, more clever OMP implementations including Batch-OMP which allow the calculation time to be accelerated when the number of signals to be processed is large [151]. Another famous variation is Orthogonal Least Squares, which selects the index k by solving another least-squares problem, adding more complexity. Subspace Pursuit allows the removal of some elements during the process [152]. Stagewise OMP (StOMP) adds several columns to the support $\mathbf{S}$ at each iteration instead of considering a single column. StOMP is faster but often less accurate [153]. In order to understand the variations in these algorithms and the common basis of these methods, a 'classic' Matching Pursuit algorithm is given in Appendix 13.

5.1.3 Basis pursuit

Basis pursuit uses another family of methods, generally closer to those used in convex optimisation. The *Fast Iterative Shrinkage Thresholding Algorithm* (FISTA) is undoubtedly the most popular. FISTA is a powerful optimisation technique widely used for solving convex optimization problems with a ℓ_1 norm penalty. FISTA can be intuitively explained as an evolution of the ISTA algorithm, which is itself a generalisation of other optimisation methods. For example, the simplest method to minimise a continuous differentiable function $f(\mathbf{z})$ is the gradient descent, which solves the problem with the sequence:

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - t_k \nabla f(\mathbf{z}^{(k)}), \tag{5.4}$$

where t_k is a positive stepsize. But the gradient descent can also be interpreted as a proximal regularisation of the first-order approximation of f at $\mathbf{z}^{(k)}$[154], which can be written :

$$\mathbf{z}^{(k+1)} = \underset{\mathbf{z}}{\operatorname{argmin}} \left\{ f(\mathbf{z}^{(k)}) + \langle \mathbf{z} - \mathbf{z}^{(k)} | \nabla f(\mathbf{z}^{(k)}) \rangle + \frac{1}{2t_k} \|\mathbf{z} - \mathbf{z}^{(k)}\|^2 \right\}. \tag{5.5}$$

The advantage of this interpretation over the previous one is that proximal descents generalise to non-differentiable functions, which means that the ℓ_1 norm penalty term can be included here [152]:

$$\mathbf{z}^{(k+1)} = \underset{\mathbf{z}}{\operatorname{argmin}} \left\{ f(\mathbf{z}^{(k)}) + \langle \mathbf{z} - \mathbf{z}^{(k)} | \nabla f(\mathbf{z}^{(k)}) \rangle + \frac{1}{2t_k} \|\mathbf{z} - \mathbf{z}^{(k)}\|^2 + \lambda \|\mathbf{z}\|_1 \right\}. \tag{5.6}$$

Using the separability of the ℓ_1 norm, computing $\mathbf{z}^{(k)}$ reduces to solving a 1-dimension problem for each of its coordinates, whose solution is given by:

$$\mathbf{z}^{(k+1)} = \mathcal{T}_{\lambda t_k}(\mathbf{z}^{(k)} - t_k \nabla f(\mathbf{z}^{(k)})), \tag{5.7}$$

where $\mathcal{T}_\alpha : \mathbb{R} \to \mathbb{R}$ is the *shrinkage operator / soft threshold*, defined as:

$$\mathcal{T}_\alpha(\mathbf{z})_i = \max(|\mathbf{z}_i| - \alpha, 0) sign(\mathbf{z}_i). \tag{5.8}$$

Figure 5.4 shows the soft threshold function with $\alpha = 1$. Just by iteratively applying the sequence, it is possible to solve the basis pursuit, this method is called ISTA. For our problem, we have a ℓ_1 penalty constraint and $f(\mathbf{z}) = \|\mathbf{D}\mathbf{z} - \mathbf{x}\|_2^2$. So ISTA, can be written as:

$$\mathbf{z}^{(k+1)} = \mathcal{T}_{\lambda t_k}(\mathbf{z}^{(k)} - 2t_k \mathbf{D}^\mathsf{T}(\mathbf{D}\mathbf{z}^{(k)} - \mathbf{x})). \tag{5.9}$$

FISTA enhances the efficiency of ISTA by introducing acceleration through Nesterov's extrapolation scheme. Nesterov acceleration speeds up convergence by introducing a momentum term derived from extrapolating previous iterations' variables (see the last operations in the Algorithm 9). This combination of gradient-based updates, proximal operator steps, and acceleration makes FISTA particularly effective in tackling optimisation problems in large dimensions. FISTA successfully keeps the simplicity of ISTA while improving the asymptotic convergence from $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$. FISTA's algorithmic complexity is low, mainly due to matrix-vector multiplication in $\mathcal{O}(mn)$. This complexity must be multiplied by the number of iterations.

Although FISTA is a reference, we have used ADMM extensively in this thesis. We therefore give the ADMM algorithms for basis pursuit in Appendix A.6.

5.1.4 Dictionary Learning

Dictionary Learning is an optimisation problem close to sparse coding. There are many methods for building a dictionary tailored to a signal, but it is also important to note that a dictionary can be learned or fixed. Fixed dictionaries have the advantage of an analytical formulation, so it must not be stored, and there can be computational savings for some operations. The Discrete Cosine Transform (DCT) is often used because it offers those advantages. On the other hand, Learned dictionaries do not have a structure but are more efficient for a specific application [152].

```
Algorithm 9: FISTA
Data:  Dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$
       Signal $\mathbf{x} \in \mathbb{R}^m$
       Sparsity trade-off constant $\lambda \in \mathbb{R}^+$
       Initial guess $\mathbf{z}^{(0)} \in \mathbb{R}^n$
       Convergence criteria $\epsilon$
Result: Sparse reconstruction $\mathbf{z} \in \mathbb{R}^n$
Initialise $t_0 = 1$
for k=0,1,2,... do
    Compute an ISTA step $\mathbf{z}^{(k+1/2)} = \mathcal{T}_{\lambda t_k}(\mathbf{z}^{(k)} - 2t_k \mathbf{D}^\intercal(\mathbf{D}\mathbf{z}^{(k)} - \mathbf{x}))$
    Update stepsize $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
    Compute new approximation $\mathbf{z}^{(k+1)} = \mathbf{z}^{(k+1/2)} + \frac{t_k - 1}{t_{k+1}}(\mathbf{z}^{(k+1/2)} - \mathbf{z}^{(k)})$
end
```

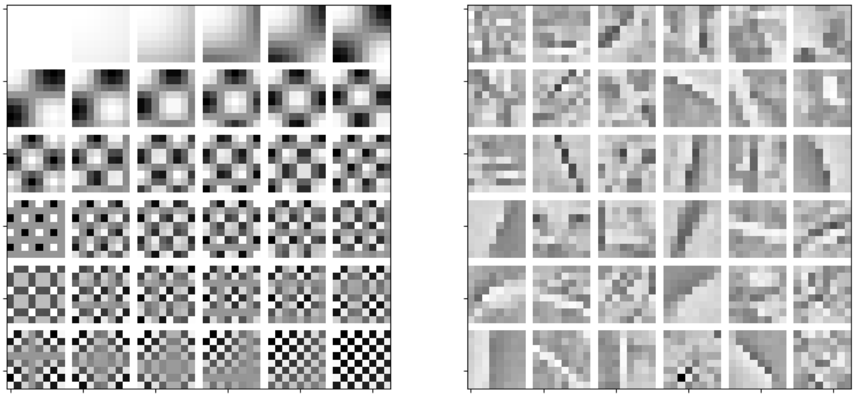


Figure 5.5: DCT dictionary (left) and a learned dictionary (right)

Figure 5.5 presents two dictionaries. The first is analytical, specifically the 2D Discrete Cosine Transform. The second is a dictionary learned using the Alternating Direction Method of Multipliers on standard photographs[1]. Dictionaries learned from photographs often exhibit this recognisable appearance: the patches seem to capture edges, contours, and patterns that are commonly observed in natural images. Certain atoms exhibit somewhat random structures; these are often less frequently utilised during sparse coding.

The Dictionary Learning problem can be written as:

$$\hat{\mathbf{z}}, \hat{\mathbf{D}} = \underset{\mathbf{z},\mathbf{D}}{\operatorname{argmin}} \frac{1}{2} \sum_s \|\mathbf{D}\mathbf{z}_s - \mathbf{E}_s\mathbf{x}\|_2^2, \tag{5.10}$$

$$\|\mathbf{z}_s\|_0 \leq \tau, \tag{5.11}$$

$$\|\mathbf{d}_i\|_2 = 1, \ \forall i \in \{1, \dots, n\}. \tag{5.12}$$

---

[1] The used pictures are taken from the Signal and Image Processing Institute https://sipi.usc.edu/database/database.php?volume=misc&image=13#top

Here $\mathbf{x}$ represents all the training data. We note τ the sparsity level. Usually, we have several images sharing the same properties, but we can also imagine learning a dictionary from a single image as long as there are enough patches. In addition, it is generally advisable to add a normalisation constraint to the dictionary. This improves the robustness of the algorithms, makes errors less dependent on the atoms used and, above all, enables a signal to be characterised and analysed directly by its code (which is particularly recommended for classification tasks) [152].

The most successful algorithms to solve the Dictionary Learning problem are the Method of Optimal Directions (MOD) and K-SVD [155]. The latter is prevalent and often described as the generalisation of the K-means algorithm. The procedure is mainly separated into two steps: sparse coding and dictionary update. For the sparse coding, any matching pursuit technique can be used. For the dictionary update, K-SVD relies on an interesting observation. The approximation error $\mathbf{R} = \mathbf{X} - \mathbf{DZ}$ can be rewritten by separating a term k: $\left\|\mathbf{X} - \sum_{j=1}^{n} \mathbf{d}_j \mathbf{z}_j^\mathsf{T}\right\|^2 = \left\|\left(\mathbf{X} - \sum_{j\neq k} \mathbf{d}_j \mathbf{z}_j^\mathsf{T}\right) - \mathbf{d}_k \mathbf{z}_k^\mathsf{T}\right\|^2 = \|\mathbf{R}_k - \mathbf{d}_k \mathbf{z}_k^\mathsf{T}\|^2$. If we consider that we fix all the atoms of the dictionary except the k^{th}, we obtain an equation consisting of finding the matrix of rank 1, which best approximates $\mathbf{R}_k$. The SVD decomposition gives the solution of this equation. By reiterating this step for each column of $\mathbf{D}$, we can create a dictionary tailored to the training base. Algorithm 10 shows the procedure for K-SVD. In the implementation, a size reduction step with ω_k guarantees the decomposition sparsity and speeds up the algorithm execution. K-SVD overall complexity is $\mathcal{O}(K(n\tau^2 + 2mn))$ [151].

Algorithm 10: K-SVD [155]

```
Data:   Initial normalised dictionary $\mathbf{D}^{(0)} \in \mathbb{R}^{m\times n}$
          Signals $\mathbf{X} \in \mathbb{R}^{m\times K}$
          Sparsity level $\tau$
Result: Learned dictionary $\mathbf{D} \in \mathbb{R}^{m\times n}$
while the convergence criteria have not been met do
    Solve $\min_{\mathbf{Z}} \|\mathbf{DZ} - \mathbf{X}\|_2$ with $\|\mathbf{Z}_i\|_0 \leq \tau$, $\forall i$, using any matching pursuit algorithm
    for each column $k \in \{1, \dots, n\}$ in do
        Define the group of samples that use this atom,
         $\omega_k = \{i \mid 1 \leq i \leq N, \mathbf{Z}_T^k(i) \neq 0\}$
        Compute the error matrix $\mathbf{R}_k = \mathbf{X} - \sum_{j\neq k} \mathbf{d}_j \mathbf{z}_j^\mathsf{T}$
        Restrict $\mathbf{R}_k$ by choosing only the columns in $\omega_k$ to obtain $\mathbf{R}_k^{'}$
        Apply SVD factorization, $\mathbf{R}_k^{'} = \mathbf{U\Sigma V}^\mathsf{T}$
        Update $\mathbf{d}_k = \mathbf{u}_1$, $\mathbf{Z}_{k,\omega_k} = \sigma_1 \mathbf{v}_1$
    end
end
```

In the same way that sparse coding has a formulation with a convex relaxation, here too, we can find the same formulation with a ℓ_1 norm on the sparsity term:

$$\hat{\mathbf{z}}, \hat{\mathbf{D}} = \underset{\mathbf{z},\mathbf{D}}{\operatorname{argmin}} \frac{1}{2} \sum_s \|\mathbf{D}\mathbf{z}_s - \mathbf{E}_s \mathbf{x}\|_2^2 + \lambda \sum_s \|\mathbf{z}_s\|_1. \tag{5.13}$$

To solve the relaxed problem, the most used algorithms are Least Angle Regression (LARS) and Alternating Directions Method of Multipliers (ADMM). Current approaches give satisfying results in most situations, yet there are inherent issues with these dictionary learning methods. The algorithmic complexity of these methods can become an obstacle when the size of the operators increases.

An alternative probabilistic interpretation opens when we adopt a Bayesian standpoint and conceptualise sparse coding as a generative model. The objective lies in identifying a set of basis feature vectors, $(d_i)_{0\leq i\leq n}$, that align the distribution of images, $P(\mathbf{x}|\mathbf{d})$, as closely as possible with the empirical distribution of our input data, $P^*(\mathbf{x})$. To achieve this, one approach involves minimising the Kullback-Leibler (KL) divergence between $P^*(\mathbf{x})$ and $P(\mathbf{x}|\mathbf{d})$, which can be written:

$$D_{\mathrm{KL}}(P^*(\mathbf{x}\|P(\mathbf{x}|\mathbf{d})) = \int P^*(\mathbf{x}) \log(\frac{P^*(\mathbf{x})}{P(\mathbf{x}|\mathbf{d})}d\mathbf{x}. \tag{5.14}$$

As the empirical distribution $P^*(\mathbf{x})$ is constant, minimising KL divergence is equivalent to maximising the log-likelihood of $P(\mathbf{x}|\mathbf{d})$. The dictionary learning problem becomes:

$$\mathbf{d}^* = \underset{\mathbf{d}}{\operatorname{argmin}} - \sum_i \log(P(\mathbf{x}_i|\mathbf{d})). \tag{5.15}$$

To calculate this term, the conditional density function of $\mathbf{x}$ knowing $\mathbf{D}$ can be further decomposed by assuming Gaussian white noise ν with standard deviation σ. The conditional distribution $P(\mathbf{x}|\mathbf{z},\mathbf{d})$ takes the form $P(\mathbf{x}|\mathbf{z},\mathbf{d}) \propto \exp\left(-\frac{(\mathbf{x}-\sum_i \mathbf{d}_i z_i)^2}{2\sigma^2}\right)$. Then, the prior $P(\mathbf{z})$ must also be specified. Usually, Laplace distribution is chosen as it promotes sparsity [156], with the parameter λ it can be written as: $P(\mathbf{z}) = \exp(-\lambda\|\mathbf{z}\|_1)$. Having established $P(\mathbf{x}|\mathbf{z},\mathbf{d})$ and $P(\mathbf{z})$, the probability of the image knowing the dictionary can be expressed as:

$$P(\mathbf{x}|\mathbf{d}) = \int P(\mathbf{x}|\mathbf{z},\mathbf{d})P(\mathbf{z})d\mathbf{z}, \tag{5.16}$$

$$\approx \max_{\mathbf{z}} P(\mathbf{x}|\mathbf{z},\mathbf{d})P(\mathbf{z}). \tag{5.17}$$

This last approximation is possible thanks to the high peak of the Laplace distribution and allows us to get rid of the intractable integration over the code space. This probability can now be plugged in Equation 5.15, giving the final objective function:

$$\mathbf{d}^* = \underset{\mathbf{d}}{\operatorname{argmin}} \|\mathbf{x} - \sum_i \mathbf{d}_i\mathbf{z}_i\|_2^2 + 2\sigma\lambda\|\mathbf{z}\|_1. \tag{5.18}$$

5.1.5 Convolutional Sparse Coding

One of the fundamental flaws of dictionary learning as presented here is the *non-continuity* of the signal. Each patch is processed independently of the others. The continuity property is disregarded, and the information it provides is therefore lost. This has several consequences for the final result and the execution of the methods. Firstly, artefacts may appear when the patches are aggregated after the basis pursuit. This effect can easily be reduced by considering overlapping patches and then averaging the results. However, this will significantly increase the number of patches. For example, if in 3D the image is a cube of N voxels length and blocks are length m, the number of blocks is $(N-m+1)^3$ instead of $(\frac{N}{m})^3$, so there are approximately m^3 more operations ($N \gg m$).

To overcome these limitations, *Convolutional Sparse Coding* (CSC), alternatively referred to as shift-invariant sparse coding, introduces a novel formalism that effectively incorporates the inherent continuity of the image [157]. In this new formalism, the decomposition of the signal $\mathbf{x}$ becomes:

$$\mathbf{x} \approx \sum_j \mathbf{d}_j * \mathbf{z}_j, \tag{5.19}$$

where $\mathbf{d}_j$ represent the filters of a *local* dictionary, $\mathbf{x}$ represent the image, and $\mathbf{z_j}$ are the *coefficient maps*. In CSC there is no longer any question of reconstructing patches. The map coefficients have the same dimension as the signal to be reconstructed. Considering that the coefficient maps are sparse by definition, they can be stored easily. Yet, the patch notion remains present in the filters. The filters $\mathbf{d}_j$ are the size of the patches and slide along the coefficient maps $\mathbf{z}_j$ to approximate $\mathbf{x}$.

CSC can also be interpreted as a traditional sparse coding method where a banded circulant matrix of a small dictionary replaces the global dictionary. This structure imposes that the blocks are a superposition of the small local dictionaries $\mathbf{D}_L$. This is useful for deducing algebraic properties about CSC.

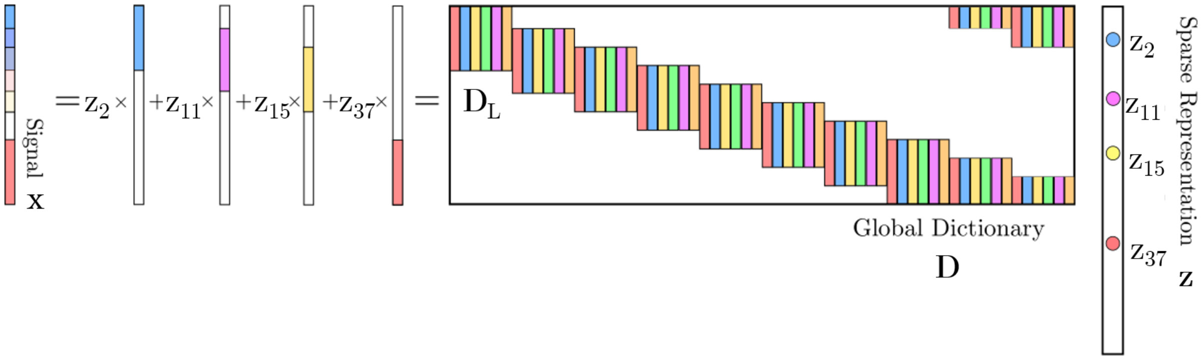


Figure 5.6: Structure of the Convolutional Sparse Coding Paradigm. Image adapted from Wikipedia. The global dictionary is constructed from shifted versions of a local dictionary.

What is particularly interesting about this interpretation is that it reveals that only a few coefficients are needed to represent a patch. If we consider only the rows of the global dictionary $\mathbf{D}$ which express the elements of the patch $\mathbf{x}_s$, this is equivalent to applying the patch extractor operator $\mathbf{E}_s$. We then notice that there are only $(2m-1)n$ columns of the dictionary which are involved in the representation of this same patch. So we can also reduce the number of dictionary columns and the number of code elements with the columns selection operator $\mathbf{S}_s^\mathsf{T}$. The resulting dictionary is called the *stripe dictionary* $\Omega \in \mathbb{R}^{m \times (2m-1)n}$, which is independent of the patch. The stripe dictionary can be seen in:

$$\begin{aligned}
\mathbf{x}_s &= \mathbf{E}_s \mathbf{D} \mathbf{z} && (5.20)\\
&= \mathbf{E}_s \mathbf{D} \mathbf{S}_s^\mathsf{T} \cdot \mathbf{S}_s \mathbf{z} && (5.21)\\
&= \boldsymbol{\Omega} \cdot \mathbf{z}_s. && (5.22)
\end{aligned}$$

While all the ideas that drive the concept of convolutional sparse coding are identical to the traditional sparse coding, there are some major consequences to this new formalism.

First, CSC allows the use of smaller operators, which is the first limitation of dictionary learning. All the filter translations can be used to describe the signal, thus considerably reducing the number of required atoms. Moreover, image continuity is finally considered, and there is no aggregation step. The sparse representation is now obtained by solving the following equation:

$$\underset{\mathbf{z}}{\operatorname{argmin}} \frac{1}{2}\Bigg|\Bigg| \sum_{j=1}^{n} \mathbf{d}_j * \mathbf{z}_j - \mathbf{x} \Bigg|\Bigg|_2^2 + \lambda \sum_{j=1}^{n} \|\mathbf{z}_j\|_1. \tag{5.23}$$

In the CSC framework, dictionary learning and basis pursuit are generally done using ADMM in the Fourier domain [158, 157]. More details about the algorithm are given in Algorithm 16. Considering CSC as sparse coding whose dictionary is a banded circulant matrix, we can use the theoretical results of compressed sensing in this new framework. New theoretical guarantees for stability and uniqueness can be derived for the stripe dictionary. In [159], the authors show that recovering signals can be done with weaker assumptions. They show that the same number of non-zeros coefficients is permitted for each stripe instead of the whole signal. Moreover, they propose an analysis based on shifted stripe dictionaries, giving more interesting theoretical results.

This notion of sparsity can be extended to the very representations themselves, leading to a cascading sequence of sparse representations. In this framework, each sparse code is defined through a small subset of atoms derived from a pre-defined set of dictionaries. Building upon this premise, an additional extension called multi-layer convolutional sparse coding (ML-CSC) is possible [160]. If we note a set of dictionaries $\{\mathbf{D}_i\}_i$ so that the product $(\prod_i \mathbf{D}_i)$ is defined, we can also define:

$$\mathbf{x} = \mathbf{D}_1\mathbf{z}_1 = \mathbf{D}_1\mathbf{D}_2\mathbf{z}_2 = (\prod_{i=1}^{k} \mathbf{D}_i)\mathbf{z}_k, \tag{5.24}$$

where one would expect $(\prod_{i=j}^{k} \mathbf{D}_i)\mathbf{z}_k$ to be sparse as well $\forall j$. In addition to all the signal recovery and stability properties of the algorithms. ML-CSC makes it possible to make an interesting link between dictionary learning and convolutional neural networks. In [160], they notice that the hard threshold (with a non-negative constraint) $\mathcal{H}_s$ used in matching pursuits is similar to the famous ReLU function, often used in Deep Learning to bring non-linearity to the network. The formulations are similar:

$$\mathbf{z}_2 = \operatorname{ReLU}\big(\mathbf{W}_2^{\mathsf{T}} \operatorname{ReLU}(\mathbf{W}_1^{\mathsf{T}}\mathbf{x})\big) \tag{5.25}$$

$$\mathbf{z}_2 = \mathcal{H}_s\big(\mathbf{D}_2^{\mathsf{T}} \mathcal{H}_s(\mathbf{D}_1^{\mathsf{T}}\mathbf{x})\big) \tag{5.26}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ represents the weights of the network layers. The first equation represents the Deep Learning network, the second the ML-CSC. The similarity makes it easier to analyse convolutional neural networks from the point of view of sparse coding and to deduce stability and noise conditions for signal recovery.

5.2 Applications and Results

5.2.1 Denoising

The simplest application of sparse coding is image denoising, which is straightforward thanks to the optimisation methods mentioned above. If training data are available, first, a dictionary is learnt from them. For traditional dictionary learning, the dictionary is usually learned on another tomographic image, obtained with the same acquisition parameters and reconstructed with the complete set of projections. Otherwise, if there is no available data, we will choose a fixed dictionary adapted to the type of image to be studied. Then, to use it on a test image, we simply perform a basis pursuit. The sparse approximation found will include the essential elements of the image while omitting potential noises. Denoising techniques based on the truncation of a description on another basis (dictionaries, wavelets, PCA) generally remove high frequencies, which makes it possible to get rid of the roughness of an image. A common practice is to separate high and low frequencies using a Tikhonov filter. As the size of the patterns to be reconstructed can not be adapted to the size of the filters, it may be more efficient to split high and low frequencies of the volume separately. In [161], only the high frequencies are denoised by CSC and then added to the low frequencies.

When comparing denoising methods using classical and convolutional dictionaries for a sparse reconstruction slice of the additive manufacturing star, distinct performance characteristics were observed. Various denoising algorithms were applied to a sparse reconstruction with 200 projections, and their results were compared to a dense reconstruction with 928 projections of the same object. For each algorithm, PSNR and SSIM were computed between the denoised sparse reconstruction and the dense reconstruction. The basis pursuits stop when the loss function reaches a plateau. Both classical and convolutional dictionaries were assessed, with different types and sizes of dictionaries tested. All methods employed patches of size 8x8. The '*learned*' dictionary was calculated using classical pictures (4 images), and the '*tailored*' dictionary was trained on a CT reconstruction of another plastic additive manufacturing part (4 slices).

Table 5.1 compares classical and convolutional sparse coding for different numbers of atoms, famous denoising methods are also given for a comparison baseline. The findings indicate that in CSC, the number of atoms has minimal impact on denoising performance. Reducing the number of atoms does not significantly degrade performance, thereby enhancing computational efficiency. Conversely, in classical sparse coding, the number of atoms critically affects the outcome. A redundant dictionary is necessary to ensure a robust sparse representation and effective denoising. Moreover, Table 5.2 shows the comparison in denoising for 128 atoms dictionaries. The dictionary's importance in representation is further highlighted. The analytical dictionary (DCT) produces inferior results compared to others. The dictionary learned from pictures shows satisfactory results, especially considering that no CT data was utilised. The tailored dictionary demonstrates superior results, achieving convergence more swiftly and effectively.

Additional results on denoising using both classical and convolutional dictionary learning have been previously published in [14] (Bussy et al.). In this study, we employed additively manufactured aluminium cubes to denoise reconstructions obtained from a limited number of projections. All results were conducted in three dimensions. The particularity of these parts was that small changes in the manufacturing process induced significant differences in the final texture of the cube. In addition to evaluating and comparing the sparse coding methods, we also observed the capacity of these methods to generalise to textures not observed during the dictionary learning phase.

Table 5.1: Comparison of Classical and Convolutional Dictionary-Based Denoising Methods for a 200 projections sparse reconstruction slice of the Additive Manufacturing Star

-	PSNR [dB]	SSIM	Time [s]
Before image denoising	37.16	0.8261	-
Wavelet denoising	38.44	0.8614	0.290
TV denoising	40.19	0.8927	0.204
NLM	38.95	0.8672	0.028
BM3D	39.37	0.9086	6.65
BP (8 atoms)	38.94	0.9039	1.37 (11it)
CSC BP (8 atoms)	41.41	0.9439	3.76 (34it)
BP (16 atoms)	39.65	0.9087	1.86 (16it)
CSC BP (16 atoms)	**41.49**	0.9440	10.51 (48it)
BP (32 atoms)	39.89	0.9096	3.11 (18it)
CSC BP (32 atoms)	41.42	0.9440	24.17 (56it)
BP (64 atoms)	40.31	0.9115	8.59 (23it)
CSC BP (64 atoms)	41.25	0.9436	210.87 (250it)
BP (128 atoms)	40.47	0.9121	8.37 (26it)
CSC BP (128 atoms)	41.44	0.9440	245.92 (143it)
BP (640 atoms)	41.25	**0.9430**	234.90 (153it)

Table 5.2: Comparison of denoising methods for different classical and convolutional dictionaries on a 200 projections sparse reconstruction slice of the additive manufacturing star

(128 atoms)	PSNR [dB]	SSIM	Time [s]
Before image denoising	37.16	0.8261	-
BP on DCT dictionary	37.19	0.8895	81.07 (250it)
BP with a learned dictionary	40.20	0.9097	8.16 (25it)
BP with a tailored dictionary	40.47	0.9121	8.37 (26it)
CSC BP	41.44	0.9440	245.92 (143it)

5.2.2 Joint dictionaries

An alternative methodology for signal denoising involves joint dictionaries, as expounded upon by [162]. In this approach, two dictionaries are jointly learned, based on the sparse and dense reconstructions extracted from a single sample. The constraint imposed in this process is that the two reconstructions should share identical code or mapping coefficients, albeit on distinct bases. This results in the alignment of each filter within the formed '*sparse*' dictionary denoted as $\mathbf{D}_s$, and the '*dense*' dictionary denoted as $\mathbf{D}_d$.

To expound further, the outcomes of a basis pursuit within the '*sparse*' basis can be used within the '*dense*' basis. A notable extension of this concept is the work by [163], where a super-resolution scheme is proposed through reconstructions of varying dimensions. There are no restrictions on the size of patches in dictionaries. A small patch in the '*sparse*' dictionary can therefore be associated with a larger patch in the '*dense*' dictionary. In this case, the terms 'low' and 'high' resolutions are used instead of sparse and dense to describe the reconstructions.

There are several methods for learning coupled dictionaries. Some are straightforward, such as [162], which build the two dictionaries by extracting patches at the same positions in the sparse and dense reconstructions. This method is quick but lacks robustness and effectiveness. The results depend significantly on the selected patches. Some areas have more noise than others, and there is no guarantee that the built dictionaries are well-adapted to the signals and assure coefficient maps sparsity. Furthermore, even when applying a mask to the image to select only regions of interest, it is possible, as shown in Figure 5.7, to obtain completely homogeneous atoms. This is generally not an issue as long as the number of atoms remains high. Figure 5.7 displays the first 30 atoms of the joint dictionaries. Correspondences between atoms are visible. It is particularly noticeable that the atoms representing an interface are sharper in the dense version.

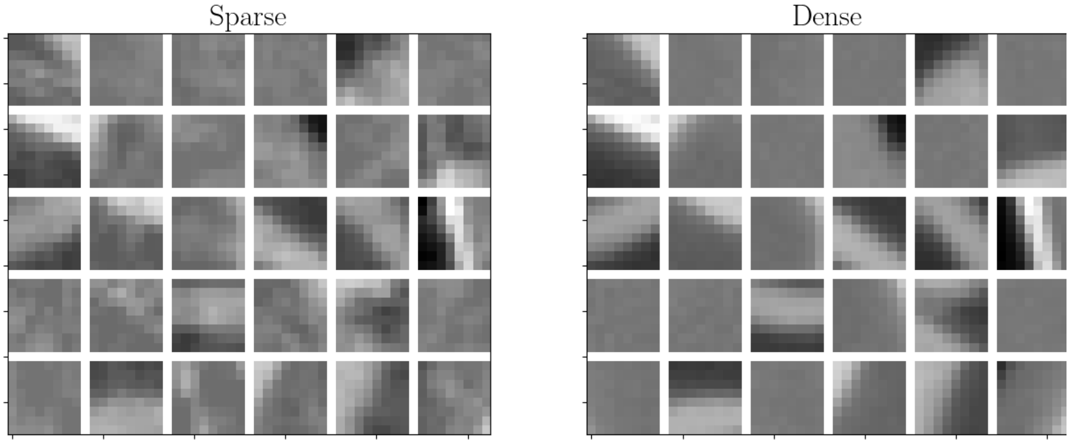


Figure 5.7: Joint dictionaries made by directly extracting blocks (8×8×8) from a sparse and dense reconstruction of the star-shaped trophy. Blocks are normalised after extraction.

In other works, the high-resolution dictionary is learned beforehand on a high-quality reconstruction. Then, this dictionary is downsampled and smoothed to create the low-resolution dictionary [164]. This method is fast, and the dictionary usually has good properties. This technique is suitable for super-resolution but less for denoising sparse reconstructions. Indeed, with this last technique, the mapping between the two dictionaries does not consider the granularity and the artefacts due to the lack of projections.

A mapping that has been trained on both low and high-quality data is preferable. To do this, one can rewrite the dictionary learning problem by concatenating the low-quality images $\mathbf{X}_s$ and high-quality training images $\mathbf{X}_d$. The coupled dictionary learning problem can then be formulated as:

$$\underset{\mathbf{D},\mathbf{Z}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\| + \lambda(\frac{1}{m_d} + \frac{1}{m_s})\|\mathbf{Z}\|_1, \tag{5.27}$$

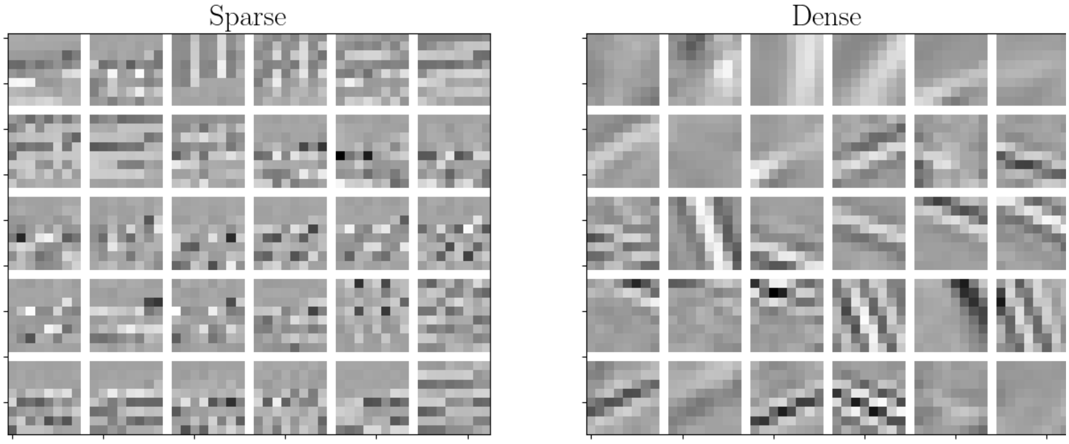
$$\mathbf{X} = \begin{pmatrix} \frac{1}{\sqrt{m_s}}\mathbf{X}_d \\ \frac{1}{\sqrt{m_d}}\mathbf{X}_s \end{pmatrix}, \mathbf{D} = \begin{pmatrix} \frac{1}{\sqrt{m_s}}\mathbf{D}_d \\ \frac{1}{\sqrt{m_d}}\mathbf{D}_s \end{pmatrix}, \tag{5.28}$$

where m_d and m_s are, respectively, the size of the '*dense*' and '*sparse*' patches. With this formulation, one can use the previous dictionary learning algorithms without any change. However, the large size of the operators can be a significant limitation. An 'online' dictionary learning resolution can circumvent this problem but does not solve the problem of calculation times.

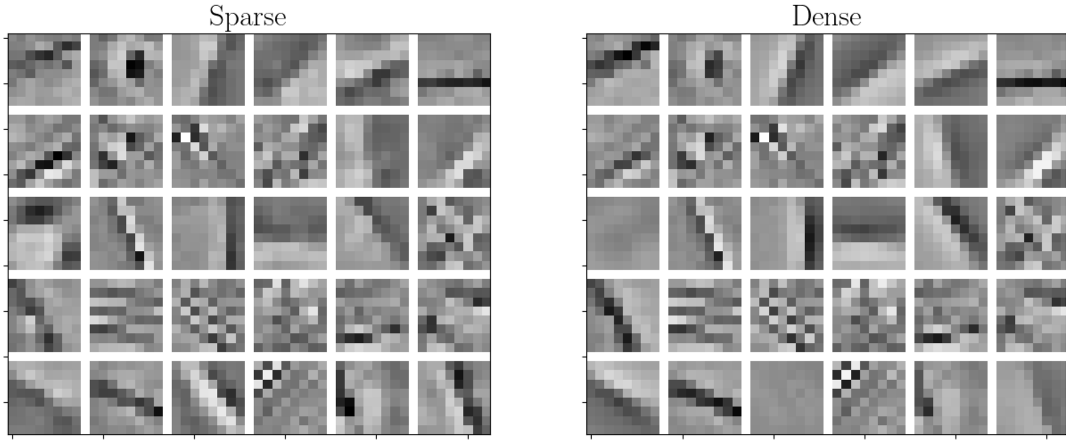
This last remark on the size of the matrices is all the more critical since, to increase the discrimination of the dictionaries, it is common to add other image features. For example, in [162], the first and second derivatives of the filters are added to $\mathbf{X}_s$ and $\mathbf{D}_s$. Thus, the gradient of the image and its Laplacian must also be approximated, increasing discrimination between the different dictionary atoms. On the other hand, there is no need to increase the size of $\mathbf{D}_d$. To denoise with coupled dictionaries, basis pursuit is only done on low-quality signals. Figure 5.8 illustrates the results for the construction of joint dictionaries for classical and convolutional sparse coding. The vectors of $\mathbf{X}$ included '*dense*' and '*sparse*' blocks, as well as the gradient and Laplacian of the '*sparse*' blocks. The atoms of the '*dense*' dictionary appear more refined, displaying interfaces as well as streaks and patterns with different frequencies. The correspondence with the '*sparse*' dictionary is less obvious than before for the classical version (Figure 5.8a), but it is important to note that the correspondence is also established through the gradient and Laplacian. Correspondences are clearer for the convolutional version in Figure 5.8b.

For our study on additive manufacturing of plastic and metal, involving both sparse and dense reconstructions, we proposed a method for CSC, called *hybrid*, that bridges the gap between analytical and direct methods. Thanks to an ADMM optimiser, a dictionary $\mathbf{D}_s$ is learned on the sparse data. Then, each filter is expressed as a linear combination of blocks extracted from the sparse volume. We create a '*dense*' version of our initial dictionary using the same linear combination and blocks extracted at the exact locations in the dense reconstruction. This technique takes advantage of the benefits of the analytical method, which finds the underlying signal structures, but also of the simplicity and computational speed of the direct method. The proposed method can also use image features other than voxels. It also requires few numerical resources, unlike the analytical method, which requires solving a high-dimensional problem. The method is recapped in Algorithm 11, and Figure 5.9 shows matched filters in $\mathbf{D}_s$ and $\mathbf{D}_d$. In the '*dense*' dictionary, atoms corresponding to interfaces and other common patterns in the image are clearly visible.

To illustrate the results, Figure 5.10 shows two denoising performed on the star-shaped trophy reconstructed with 200 projections. Figure 5.10a shows the results of a basis pursuit on the 'sparse' dictionary while Figure 5.10b shows its equivalent in the 'dense' basis, each is 1920 atoms. Even if the difference isn't obvious visually, the values are more accurate and it's especially in the empty areas between the materials that the dense



(a) Joint dictionaries



(b) Joint convolutional dictionaries

Figure 5.8: Joint dictionaries made analytically by solving Eq.5.27 and its convolutional equivalent (8×8×8)

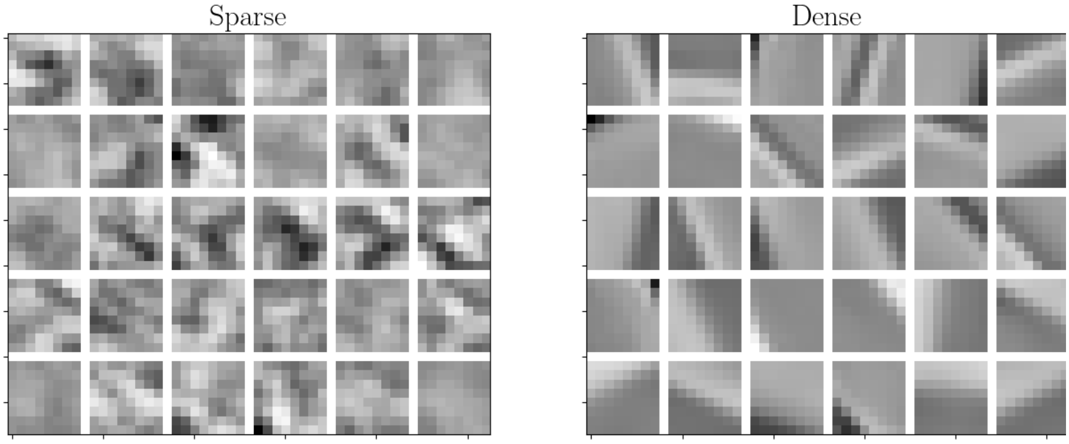


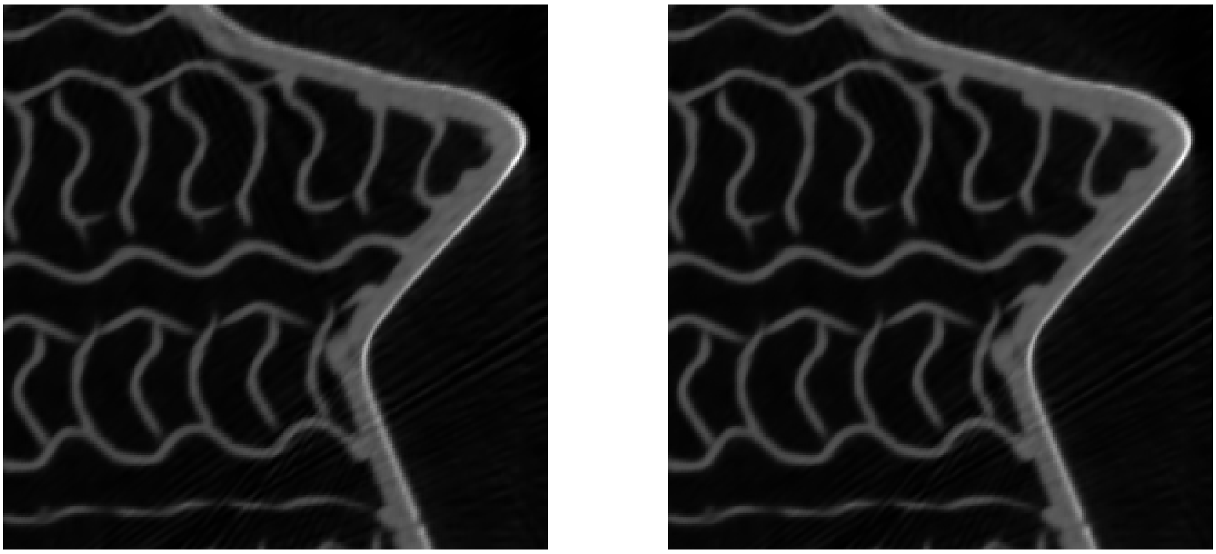
Figure 5.9: Joint dictionaries made with the proposed hybrid method (8×8×8)

Algorithm 11: Proposed coupled dictionaries learning algorithm.

```
**Data:** Sparse reconstruction $\mathbf{X}_s$, Dense reconstruction $\mathbf{X}_d$
**Result:** Coupled dictionaries $\mathbf{D}_s$ and $\mathbf{D}_d$
Compute $\mathbf{D}_s$ by solving eq.5.12 with $\mathbf{X}_s$
Extract blocks $\mathbf{B}_s$ and $\mathbf{B}_d$ at the same positions in $\mathbf{X}_s$ and $\mathbf{X}_d$
Concatenate its 1$^{st}$ and 2$^{nd}$ order derivatives to $\mathbf{B}_s$
Solve $\mathbf{B}_s\mathbf{Y} = \mathbf{D}_s$
$\mathbf{D}_d \leftarrow \mathbf{B}_d\mathbf{Y}$
```

version brings more homogeneity. The same study was carried out, in Figure 5.11, on convolutional dictionaries of 32 filters constructed using the analytical method. This time we can see directly that the streak artefacts are reduced and the clarity of the image improved. Finally, we also show the result of the same approach, but with convolutional dictionaries made using our 'hybrid' method in Figure 5.12. Here too, the difference is clear. The reconstruction seems less blurred.

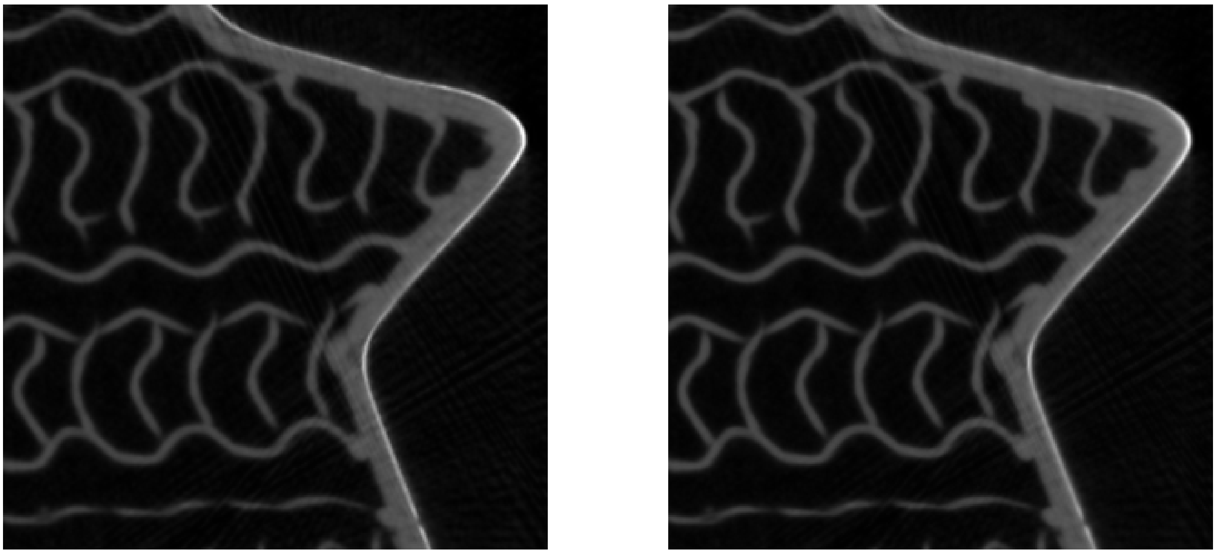
Table 5.3 summarises the results in terms of PSNR and SSIM of the comparison between classical and convolutional sparse coding for the analytical, hybrid and direct joint dictionaries. Results also show the comparison between a basis pursuit denoising and joint dictionaries. Unsurprisingly, the method with joint dictionaries offers better performance.



(a) Denoising with the 'sparse' dictionary

(b) Equivalent with the 'dense' dictionary

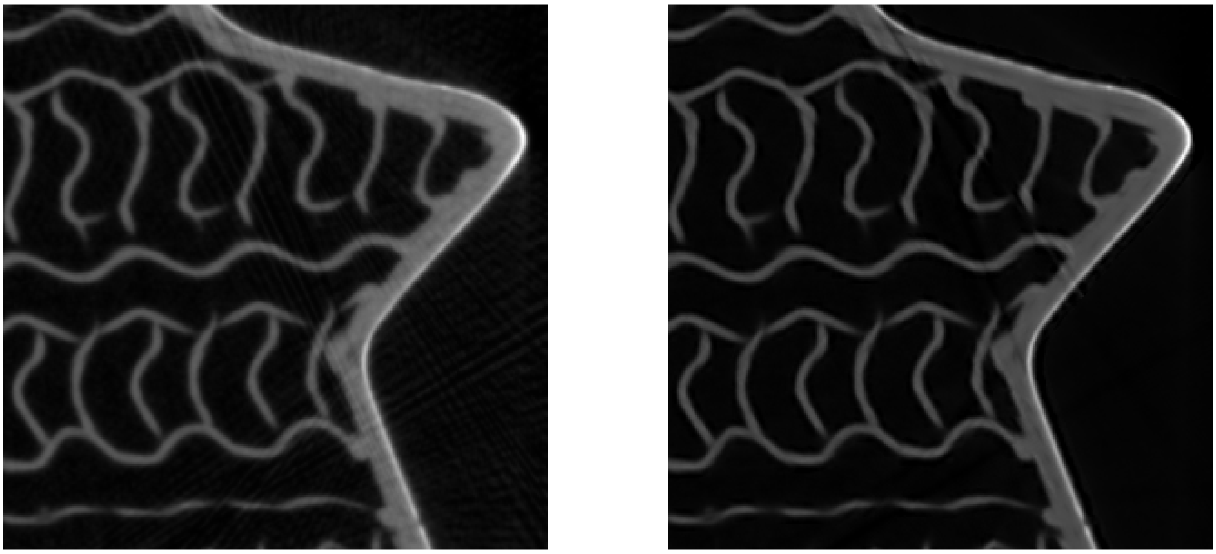
Figure 5.10: Results of denoising by sparse coding without and with joint dictionaries made by the direct method



(a) CSC Denoising

(b) Joint Dictionary CSC Denoising

Figure 5.11: Results of denoising by convolutional sparse coding without and with joint dictionaries made by the analytical method



(a) CSC Denoising

(b) Joint Dictionary CSC Denoising

Figure 5.12: Results of denoising by convolutional sparse coding without and with joint dictionaries made by the hybrid method

5.2.3 Regularisation

The sparsity term on patches can be extended for regularisation. In [165], Xu et al. were the first to propose incorporating a dictionary-learning approach into a statistical X-ray tomographic reconstruction. In their work, they propose two strategies, GDSIR and ADSIR (Global/Adaptive Dictionary Statistical Iterative Reconstruction), depending on whether the dictionary is updated during the reconstruction. The objective function is:

$$\min_{\mathbf{x},\mathbf{z},(\mathbf{D})} \sum_i \frac{\omega_i}{2}([\mathbf{A}\mathbf{x}]_i - \mathbf{p}_i)^2 + \lambda(\sum_s \|\mathbf{E}_s\mathbf{x} - \mathbf{D}\mathbf{z}_s\|_2^2 + \sum_s \nu_s \|\mathbf{z}_s\|_0) \tag{5.29}$$

Table 5.3: Comparison of denoising methods in terms of PSNR and SSIM for different classical and convolutional dictionaries on a 200 projections sparse reconstruction slice of the additive manufacturing star

PSNR[dB]/SSIM	BP denoising	Joint dictionaries denoising
CSC Analytical (32 atoms)	**41.42/0.9440**	**42.62**/0.9620
SC Analytical (1920 atoms)	39.89/0.9096	40.12/0.9600
CSC Hybrid (64 atoms)	41.25/0.9436	42.05/**0.9662**
SC Hybrid (1920 atoms)	40.11/0.9164	41.32/0.9580
CSC Direct (64 atoms)	40.95/0.9422	41.84/0.9609
SC Direct (1920 atoms)	40.09/0.9122	41.15/0.9570

where $\mathbf{A}$ is the projection operator, $\boldsymbol{\omega}$ weights the contribution of each ray to the tomographic problem, $\mathbf{p}$ are the experimental projections, $\mathbf{E}_s$ is still the operator to extract patch 's' and ν trades off patches approximation and sparsity. Here $\mathbf{x}$ represents the volume to be reconstructed, such as $\boldsymbol{\mu}$ before, but we will keep for consistency. $\mathbf{D}$ is a variable to optimise in ADSIR. For GDSIR, the dictionary is learned beforehand on similar images. To minimise the objective function, we have used a common solution strategy which consists of alternating the optimisations on the different variables while fixing the others. Then, reconstruction is made by alternating minimisation on $\mathbf{x}$, then $\mathbf{D}$ and $\mathbf{z}$, knowing that the latter is a dictionary learning / matching pursuit step. To update $\mathbf{x}$, Xu et al. use the separable paraboloid alternative method [165]:

$$\mathbf{x}_j^{(k+1/2)} = \mathbf{x}_j^k - \frac{\sum_i a_{ij}\omega_i([\mathbf{A}\mathbf{x}^{(k)}]_i - \mathbf{p}_i) + 2\lambda \sum_s \sum_n e^s_{nj}([\mathbf{E_s}\mathbf{x}^{(k)}]_n - [\mathbf{D}\mathbf{z_s}]_n)}{\sum_i (a_{ik}\omega_i \sum_j a_{ij}) + 2\lambda \sum_s \sum_n e^s_{nk} \sum_j e^s_{nj}}, \ \forall j \tag{5.30}$$

Then, the resulting image is re-encoded in the dictionary with a matching pursuit. For ADSIR, the procedure is the same, except that instead of a matching pursuit, a dictionary learning algorithm is used to update both the code and the dictionary:

$$\mathbf{z}, (\mathbf{D}) = \underset{\mathbf{z},(\mathbf{D})}{\operatorname{argmin}} \sum_s \|\mathbf{E}_s \mathbf{x}^{(k+1/2)} - \mathbf{D}\mathbf{z}_s\|_2^2 + \sum_s \nu_s \|\mathbf{z}_s\|_0. \tag{5.31}$$

Reconstructed patches are then aggregated to reconstruct the whole image:

$$\mathbf{x}_j^{(k+1)} = \frac{\sum_s \mathbf{E}_s^\top \mathbf{D}\mathbf{z}_s}{\sum_s \sum_n e^s_{nk} \sum_j e^s_{nj}} \tag{5.32}$$

As CSC is faster than traditional dictionary techniques, it is now possible to consider large 3D applications. The objective function is rewritten within the CSC framework [166]. All terms are similar to Equation 5.29, except that the sum of patches is replaced by the convolutional encoding. The last term is slightly simplified with a constant ν for each filter.

$$\min_{\mathbf{x},\mathbf{z},(\mathbf{D})} \sum_i \frac{\omega_i}{2}([\mathbf{A}\mathbf{x}]_i - b_i)^2 + \lambda(\|\sum_m \mathbf{d}_m * \mathbf{z}_m\|_2^2 + \nu \sum_m \|\mathbf{z}_m\|_1) \tag{5.33}$$

The optimisation scheme remains the same as in [165]. Alternating optimisation of $\mathbf{x}$ and then the dictionary and coefficient maps. Equation 5.30 is transformed to:

$$\mathbf{x}_j^{(k+1)} = \mathbf{x}_j^{(k)} - \frac{\sum_i a_{ij}\omega_i([\mathbf{A}\mathbf{x}^{(k)}] - b_i) + \lambda([\sum_m \mathbf{d}_m * \mathbf{z}_m]_j - \mathbf{x}_j^{(k)})}{\sum_i (a_{ik}\omega_i \sum_j a_{ij}) + \lambda} \tag{5.34}$$

As before, the next step is to encode the result in the dictionary basis. We made it with ADMM. Figure 5.13 shows four reconstructions with the different regularisation methods and compared them to the dense reference and a SART reconstruction.

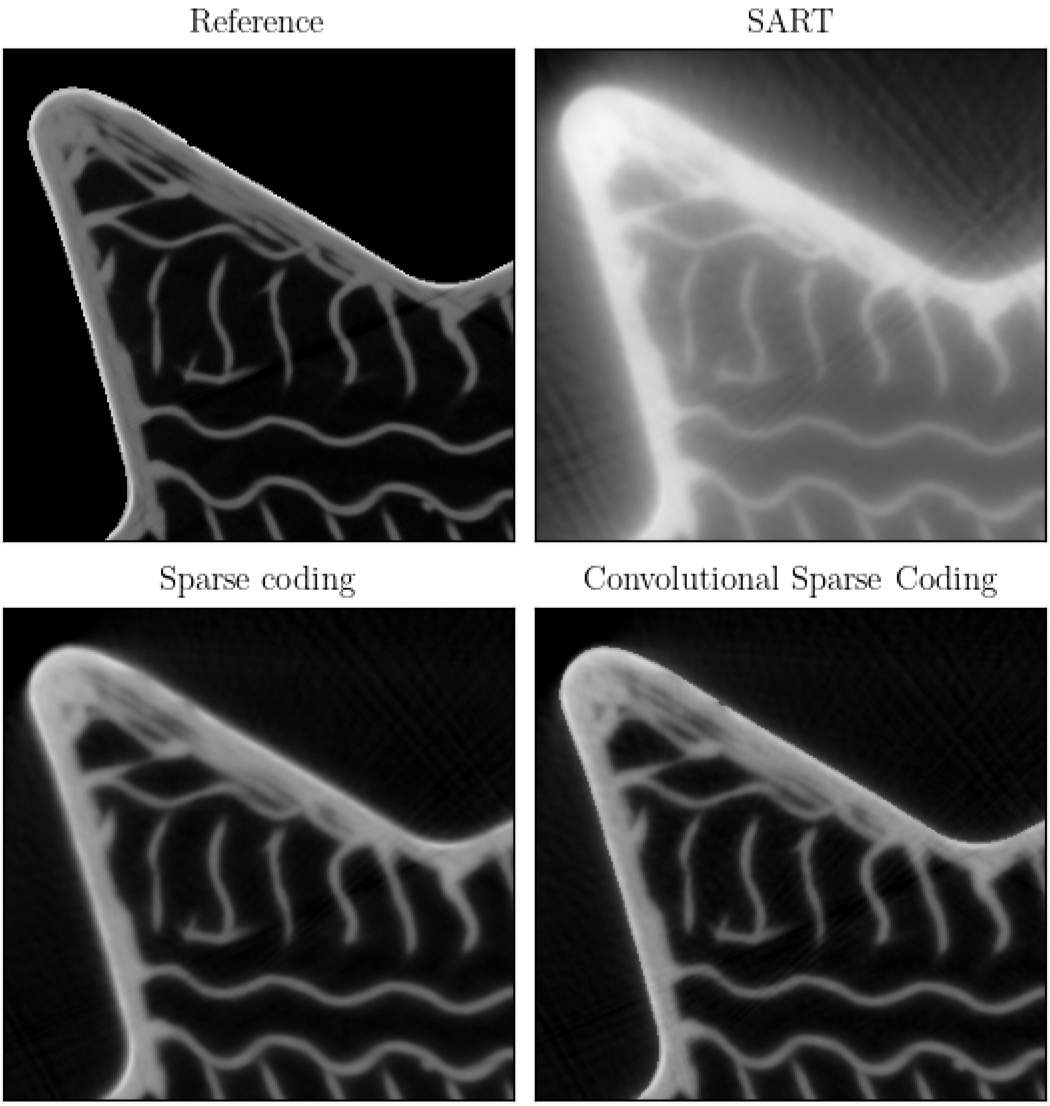


Figure 5.13: Reconstruction with different regularisation methods. Top left is the reference reconstruction made with 928 projections and the mask. Top right is made with SART with 200 projections and 50 iterations. Bottom left is the results of the described regularised reconstruction with classical sparse coding (Eq. 5.29). Bottom right is a regularised reconstruction on 200 projections with CSC 64 atoms (Eq. 5.34).

5.2.4 Classification

The strength of dictionary learning lies in its ability to extract essential features from a signal to find the main patterns that constitute the signal. This ability to extract the most representative features can also be used for classification tasks. One can make a discriminant dictionary by building a structured dictionary whose atoms are dedicated to describing one specific type of signal. A given signal is decomposed in a discriminant dictionary, and according to the atoms that have been used, it is possible to determine the class of the input signal. Generally, the global dictionary is made up of sub-dictionaries, each representing a class. We can then look at the atoms used in the representation of the input signal. If a class of atoms is more involved in the signal code, the signal will be classified in that class. We can also compare the reconstruction errors of each sub-dictionary and deduce the most appropriate dictionary for the signal and therefore its class. It is also possible to determine physical properties or abnormalities from the signal code. In [167], the Reynolds number of a flow around a cylinder is identified using the sparse representation of the flow. In [168], authors adapt classification and dictionary learning to detect faults and leakages in distributed water networks. The main difficulty with this method is learning a dictionary whose atoms are discriminant from one class to another. It is important to note that other strategies have been proposed, including some that use *analysis dictionary learning* (as opposed to the previous methods called *synthesis*) which directly transforms a signal to a sparse feature basis by multiplying the signal [169], but the challenges and difficulties remain similar. Both strategies can be applied at the same time for better performances [170].

Early attempts to create a discriminative dictionary for classification tasks used training samples directly to create the atoms of the dictionary [171]. This method is referred to as *sparse representation-based classifier* (SRC). Subsequently, existing algorithms for dictionary learning were adapted for classification, such as [172], who was the first to propose to learn jointly a dictionary for representation and a classifier. The idea of adding a classifier was then retaken many times. Many methods, such as the *Discriminative-KSVD*, include a classifier in its objective function. *Label Consistent-KSVD* goes even further by incorporating a discriminative sparse-code error and a classification error term in its goal function. However, the underlying idea behind these algorithms was to determine which sub-dictionary best described the signal without looking at the similarities and differences between these classes. As a result, the common characteristics between the classes led to non-discriminating decompositions and, therefore, to insufficient results. One of the solutions to this issue is to add a coherence penalty to the representation error [173], this method is referred to as *dictionary learning with structured incoherence* (DLSI). For K different classes, the objective function to minimise becomes:

$$\min_{\mathbf{Z},\mathbf{D}} \sum_{k=1}^{K} \|\mathbf{X} - \mathbf{D}\mathbf{Z}_k\|^2 + \lambda \sum_{k=1}^{K} \sum_{j\neq k} \|\mathbf{D}_k^{\mathsf{T}} \mathbf{D}_j\|_2^2 \tag{5.35}$$

Afterwards, concepts from Fisher discriminant analysis emerged in the newly proposed methods [174, 175]. The objective function now has a discriminative fidelity term r:

$$r(\mathbf{X}_i, \mathbf{Z}, \mathbf{D}) = \|\mathbf{X}_i - \mathbf{D}\mathbf{Z}\|^2 + \sum_{k\in\mathcal{C}(i)} \|\mathbf{X}_i - \mathbf{D}_k\mathbf{Z}_k\|^2 + \|\sum_{k\notin\mathcal{C}(i)} \mathbf{D}_k\mathbf{Z}_k\|^2 \tag{5.36}$$

The first term represents the overall fidelity with every element from the dictionary. The second term represents the fidelity only using the dictionary dedicated to $\mathbf{X}_i$'s class $\mathcal{C}(i)$. The last term reduces the participation of the rest of the dictionary in the representation.

But Fischer discriminant analysis also adds a discriminative coefficient term, called g, made of $S_W(\mathbf{Z})$ which minimises the within-class scatter of the code and $S_B(\mathbf{Z})$ which maximises the between-class scatter.

$$S_W(\mathbf{Z}) = \sum_{i=1}^{c} \sum_{z_j \in \mathcal{C}_i} (\mathbf{z}_j - \mathbf{m}_i)(\mathbf{z}_j - \mathbf{m}_i)^\mathsf{T} \tag{5.37}$$

$$S_B(\mathbf{Z}) = \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^\mathsf{T} \tag{5.38}$$

$$g(\mathbf{Z}) = \mathrm{trace}(S_W(\mathbf{Z})) - \mathrm{trace}(S_B(\mathbf{Z})) + \eta \|\mathbf{Z}\|_2^2 \tag{5.39}$$

where $\mathbf{m}_i$ and $\mathbf{m}$ represents the mean vectors of $\mathbf{Z}_i$ and $\mathbf{Z}$, and n_i the number of samples in the i^{th} class. The whole objective function becomes:

$$\underset{\mathbf{Z},(\mathbf{D})}{\mathrm{argmin}} \sum_i r(\mathbf{X}_i, \mathbf{Z}, \mathbf{D}) + g(\mathbf{Z}) + \lambda \|\mathbf{Z}\|_{1,1} \tag{5.40}$$

This type of optimisation problem is often solved by fixing the coefficients and the dictionary alternately. The discriminant terms introduced have greatly improved the results, but they cannot prevent some atoms from being very similar in different classes. Another solution was to add a common sub-dictionary $\mathbf{D}_0$ for patterns shared by the different signal classes [176, 177]. It will then be necessary to modify the optimisation and classification scheme to take this sub-dictionary into account. $\mathbf{D}_0$ is used for the representation but not the classification, this method is referred to as COPAR. Figure 5.14 shows the importance of previously mentioned fidelity terms, as well as the usefulness of a shared dictionary.

Low-Rank Shared Dictionary Learning, called LRSDL, takes things a step further by adding all those ideas and forcing the shared dictionary to be low-rank so it does not represent class-specific features [178]. The authors suggest the following objective function:

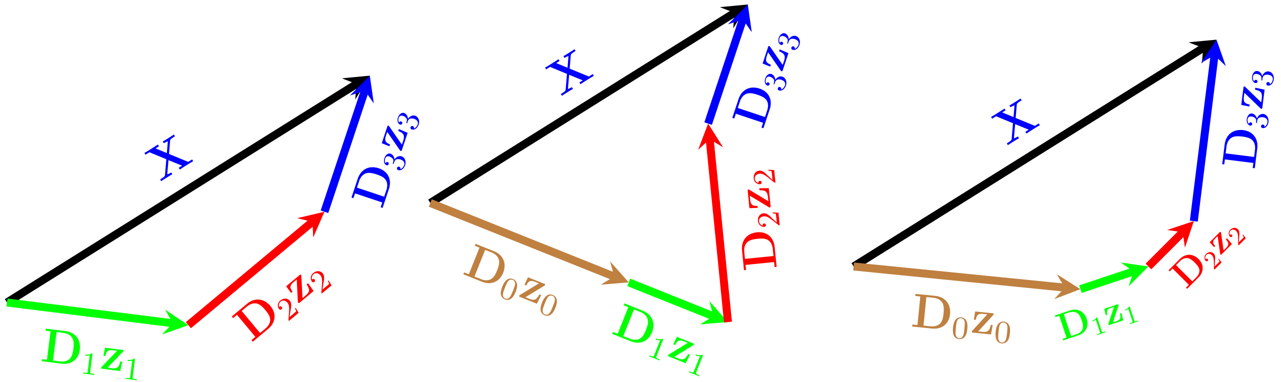
$$\underset{\mathbf{Z},\bar{\mathbf{D}}}{\mathrm{argmin}} \sum_i r_i(\mathbf{X}_i, \bar{\mathbf{Z}}, \bar{\mathbf{D}}) + g(\mathbf{Z}) + \lambda \|\mathbf{Z}\|_{1,1} + \eta \|\mathbf{D}^0\|_* \tag{5.41}$$

$$r_i(\mathbf{X}_i, \mathbf{Z}, \mathbf{D}) = \|\mathbf{X}_i - \bar{\mathbf{D}}\bar{\mathbf{Z}}_i\|^2 + \|\mathbf{X}_i - \mathbf{D}_c\mathbf{Z}_i^c - \mathbf{D}_0\mathbf{Z}_i^0\|^2 + \sum_{k>1,k\neq c} \|\mathbf{D}_k\mathbf{Z}_i^k\|^2 \tag{5.42}$$

$$g(\bar{\mathbf{Z}}) = \|\mathbf{Z}_0 - \mathbf{M}_0\|_2^2 + \sum_{c=1}^{C} (\|\mathbf{Z}_c - \mathbf{M}_c\|_2^2 - \|\mathbf{M}_c - \mathbf{M}\|_2^2) + \|\mathbf{Z}\|_2^2 \tag{5.43}$$

where c is the class of $\mathbf{X}_i$, the i^{th} sample. $\bar{\mathbf{D}} = [\mathbf{D}\ \mathbf{D}_0]$ denotes the concatenation of the class-specific dictionaries and the shared one. The same notation is used for $\bar{\mathbf{Z}} = [\mathbf{Z}^\mathsf{T}\ \mathbf{Z}_0^\mathsf{T}]^\mathsf{T}$. $\mathbf{M}_i = [\mathbf{m}_i, \ldots, \mathbf{m}_i]$ and $\mathbf{M} = [\mathbf{m}, \ldots, \mathbf{m}]$ are the mean matrices. The classification scheme has also been adapted to take account of the specific features of the shared dictionary.

$$\bar{\mathbf{z}} = \underset{\bar{\mathbf{z}}}{\mathrm{argmin}} \frac{1}{2} \|\mathbf{x} - \bar{\mathbf{D}}\bar{\mathbf{z}}\|_2^2 + \lambda \|\mathbf{z}^0 - \mathbf{m}^0\|_2^2 + \lambda_1 \|\bar{\mathbf{z}}\|_1 \tag{5.44}$$



(a) No shared dictionary nor constraints on classes

(b) Shared dictionary $\mathbf{D}_0$ but no constraints

(c) Shared dictionary with constraints on classes to which $\mathbf{X}$ does not belong

Figure 5.14: Illustration of the benefits of the shared dictionary and constraints on bad classes. $\mathbf{X}$ belongs to the third class. Without dictionary constraints, $\mathbf{X}$ may have features belonging to different classes, making classification difficult (left). A shared dictionary containing features that the different classes have in common overcomes this problem, but there is no guarantee that a linear combination of atoms can mimic a feature from a particular class (middle). We will therefore try to reduce the impact of the other classes when learning to be sure of creating atoms capable of describing this feature sparsely (right).

$$\underset{c}{\operatorname{argmin}}\, \omega \|\mathbf{x} - \mathbf{D}_c \mathbf{z}_c\|_2^2 + (1-\omega)\|\mathbf{z}^0 - \mathbf{m}^0\|_2^2 \tag{5.45}$$

where $\omega \in [0,1]$ balances the contributions of the two terms. This method showed better results than the previous ones but still suffers from the same disadvantages. Many atoms look similar in the same class, and most of the time, there are translated versions of one another. Moreover, for large dimensions, dictionary learning suffers the curse of dimensionality.

We propose to adapt dictionary learning to its convolutional version for classification. This would reduce the size of the dictionary as there would not be the shifted versions of a filter. Also, as CSC works with blocks, the size of the samples is not so relevant anymore. Learning and testing can be done on images of different sizes. For our case, we have focused on the latter *Fischer Discrimination Dictionary Learning* (FDDL), but the convolutional formalism made several terms irrelevant. First, the discriminative coefficient term g has been removed as CSC has no equivalent. We propose the following objective function $\forall k$:

$$\underset{\mathbf{z},\mathbf{d}}{\operatorname{argmin}} \left|\left| \mathbf{x}^k - \sum_{j=0}^{n} \mathbf{d}_j * \mathbf{z}_{k,j} \right|\right|_2^2 + \|\mathbf{x}^k - \sum_{c \in \mathcal{C}(k)} \mathbf{d}_c * \mathbf{z}_{k,c}\|_2^2 + \left|\left| \sum_{\substack{j \geq 1, \\ j \notin \mathcal{C}(k)}}^{n} \mathbf{d}_j * \mathbf{z}_{k,j} \right|\right|_2^2 + \lambda \|\mathbf{z}\|_{1,1} \tag{5.46}$$

The first term is a fidelity term on the k^{th} sample of $\mathbf{x}$. The second term is a fidelity term restrained on the classes to which $\mathbf{x}_k$ belongs. The third term is a constraint on the contribution of the classes $\mathbf{x}_k$ does not belong to. The last term promotes sparsity in the code. We use a General Consensus ADMM to optimise the function. The classification scheme is the same as LRSDL, described by Equations 5.44 and 5.45.

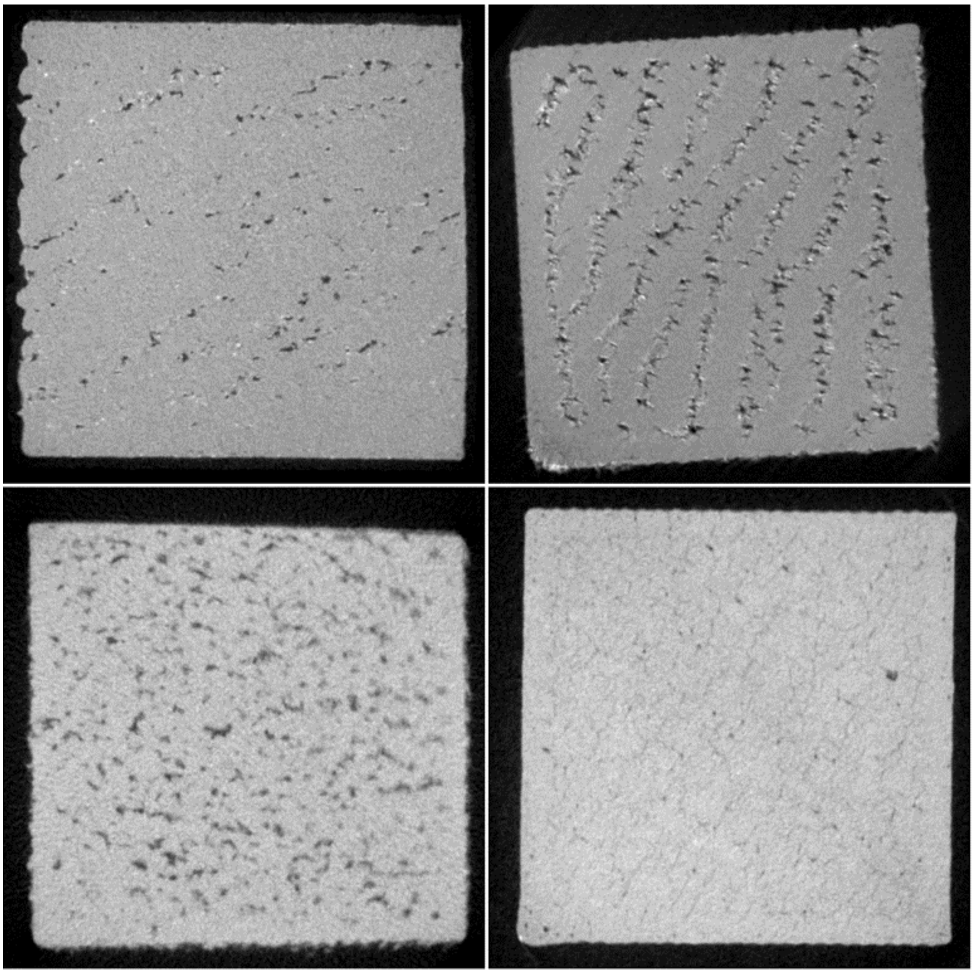


Figure 5.15: Additive manufactured aluminium cubes used for classification.

In this section, we needed several samples to classify so we did not use the usual trophy. The experimental data considered belong to a set of Aluminium cubes created by Laser Powder Bed Fusion (LPBF) additive manufacturing process using different process parameters (laser power and speed). The dense acquisition was made on a circular trajectory with 900 projections. The voxel size is 20.8μm, voltage is 140 kV, and current is 100μA. So far, we have only been able to test our results on four samples. Initially, we reconstructed four cubes using all available projections to learn dictionaries from these volumes. However, in the second phase, sparse reconstructions with 100 projections will be used to test our classification methods. The dense samples are depicted in Figure 5.15. The features to be represented are diverse, including cracks, porosities, bubbles, inclusions, and pores. Nevertheless, within a single sample, the texture remains fairly consistent. The intra-class variance is acceptable, while the inter-class variance is higher.

We tested our method and compared it to other state-of-the-art techniques. Unfortunately, due to implementation time constraints, the state-of-the-art methods are in 2D, while our implementation is in 3D. For each method, 8×8 blocks were classified into one of four classes. For each method, the percentage of exactitude is reported. It is noted that the percentage is relatively low for each method. This is due to the substantial contribution of the shared dictionary and the areas without distinctive features. Most of the misclassified blocks are homogeneous regions.

Table 5.4: Results and comparison of classification of aluminium cubes blocks

Methods	Exactitude [%]
DLSI [173]	31
SRC [171]	38
COPAR [176]	41
LRSDL [178]	53
Our method	61

Currently, the results are rather unsatisfactory. It would be prudent to test the methods on a different dataset to determine if the results vary. The classification step needs modification to address the issue of the shared dictionary, which describes almost all the energy in patches. Perhaps these cases should be excluded from classification, and such patches ignored. It would also be wise to study the impact of patch size, as certain features, like large bubbles, extend beyond the window of a single patch.

Chapter 6

Conclusion

The objective of this thesis was to propose a methodology for integrating *a priori* information into the X-ray tomographic reconstruction process in an industrial context. We considered the sparse-view scenario, where the number of available projections is limited. Constraints related to acquisition time or radiation dose often prevent us from obtaining a sufficient number of projections, making sparse-view increasingly common. However, this is a challenging problem because the lack of information introduces significant artefacts during reconstruction, severely reducing its quality. Therefore, finding solutions to this challenge is essential. In this thesis, we did not aim to propose a single method but rather a methodology, as it is crucial to adapt each process to the studied part and the industrial needs. We presented a series of techniques and ideas that can be selected to leverage the available *a priori* information that the industry possesses. We assumed that the industry has access to a CAD model of the inspected object and a reconstruction of an object similar to the one of interest. We are not in a situation where we have enough information to train a model or learn a distribution. The *a priori* information we have is almost exclusively the "theoretical" geometry of the object. This is where the difficulty of sparse-view lies. The available *a priori* information, such as the CAD, does not provide information of interest, since it is precisely the deviations from this theoretical design that we seek to evaluate with tomography. Therefore, it is essential to propose methods that are robust to the variations present in the actual object.

Our work begins with selecting the best projections for tomographic reconstruction, as described in Chapter 2. Not all projections provide the same amount of information, and it is important to choose the views that complement each other best, especially when they are limited in number. We proposed methods based on the *Empirical Interpolation Method*, testing several variants: DEIM, QDEIM, and CQDEIM. These methods are fast, simple to implement, and have shown a real gain in sparse-view reconstruction, with improvements of up to several dB in PSNR. Projection selection is particularly promising and should be further explored with the integration of more physical effects, as well as the consideration of regions of interest. We aim to develop methods tailored to specific types and locations of known defects, in order to propose a trajectory that is even more specific to the experimenter's needs. We would like to further refine the selection of trajectories by incorporating variable magnification factors and tilted detectors.

Next, we proposed registration methods in Chapter 3. Registration and tracking are well-known problems in optics, so we drew inspiration from work in these fields to propose a Point Visual Servoing method and an Image-Based Visual Servoing method.

The first, Convex Hulls Iterative Inverse Perspective Matching, is based on detecting the convex hull points of the CAD and experimental projections. These points are matched to propose a registration. In the second IBVS-based method, projections are simulated from the CAD and compared to experimental projections to suggest a movement that corrects the position of the CAD to match the real object. Both proposed methods are complementary: the first is faster, while the second is more precise. Ultimately, we achieve object registration with an accuracy of a few pixels. Having the CAD and the real object in the same reference frame is particularly useful as it allows us to directly use the CAD for volume reconstruction. The method must be extended to register greyscale features. This would allow the alignment of parts with interval grooves. Furthermore, extensions of our methods need to be developed to enable registration based on truncated projections of the object.

In Chapter 4, we leverage this registration to reduce the volume to be reconstructed using a mask. The mask is obtained by voxelizing the CAD mesh, allowing us to select only the voxels that potentially contain material during reconstruction. The mask significantly improves the reconstruction, reducing artifacts, providing sharper discontinuities, and smoother homogeneous areas, and increasing the PSNR by several dB. The reconstruction on the mask can be adapted to all algorithms by defining new masked projection and back-projection operators. We went even further and experimented with performing a reconstruction on a mesh using BVH to accelerate ray tracing. We transformed the CAD into a volumetric tetrahedral mesh and implemented ray tracing on tetrahedra. It is quite challenging to create a mesh directly from the CAD that is well-suited for the object to be reconstructed, so we attempted to develop an adaptive meshing method. Although we could not fully complete our developments, the initial results are very encouraging. We achieved high-quality reconstruction with fewer unknowns than in the conventional voxelized version. Currently, we are limited by computation times, but reconstruction on a mesh presents a promising potential. This work must be continued in order to explore all the potential opportunities it presents.

Subsequently, in Chapter 5, we focused on the reconstruction step and post-processing to improve sparse-view reconstruction. Since we cannot effectively learn an image texture from a CAD and generally do not have access to databases, we used sparse coding and dictionary learning. Dictionary learning allows for effective denoising of small blocks of the volume to be reconstructed by assuming that these blocks are sparse in a redundant dictionary. This method has proven effective for image denoising, and we tested it on tomographic reconstructions. However, this method is computationally expensive and does not account for the continuity between blocks. We proposed convolutional sparse coding to address these challenges. In this thesis, we proposed an efficient implementation of CSC that enables not only denoising but also super-resolution and regularisation. We even show some results on a classification problem. The CSC shows promising results, but there are still many areas for improvement. By incorporating regularisation terms into the basis pursuit, using blocks of variable sizes, and possibly exploring multi-layered approaches, the results could be further enhanced.

X-ray tomography is a vast and multidisciplinary field. It is necessary to advance it by combining methods and integrating as much *a priori* information as possible. By combining the improvements of each method, it will become possible to achieve sparse-view reconstructions with good quality, potentially in a fully automated manner.

References

[1] Simone Carmignato, Wim Dewulf, and Richard Leach. Industrial X-Ray Computed Tomography. Springer, Gewerbestrasse 11, 6330 Cham, Switzerland, 2018.

[2] Philip J Withers, Charles Bouman, Simone Carmignato, Veerle Cnudde, David Grimaldi, Charlotte K Hagen, Éric Maire, Marena Manley, Anton Du Plessis, and Stuart R Stock. X-ray computed tomography. Nature Reviews Methods Primers, 1, February 2021.

[3] Thorsten Buzug and Dimitris Mihailidis. Computed tomography from photon statistics to modern cone-beam ct. Medical Physics, 36:3858–, 08 2009.

[4] Victor Bussy, Caroline Vienne, and Valérie Kaftandjian. Fast algorithms based on empirical interpolation methods for selecting best projections in sparse-view x-ray computed tomography using a priori information. NDT&E International, 2022.

[5] Xinrong Mao, Kaiming Liu, and Yanfen Hang. Feature extraction and matching of slam image based on improved sift algorithm. volume 2020 2nd Symposium on Signal Processing Systems (SSPS), Guangdong, China, July 2020, pages 18–23, 07 2020.

[6] Michael Schwarz and Hans-Peter Seidel. Fast parallel surface and solid voxelization on gpus. ACM Trans. Graph., 29(6), December 2010.

[7] Daniel Cohen-Or and Arie Kaufman. Fundamentals of surface voxelization. Graphical Models and Image Processing, 57(6):453–461, 1995.

[8] Mitko Aleksandrov, Sisi Zlatanova, and David Heslop. Voxelisation algorithms and data structures: A review. Sensors, 21:8241, 12 2021.

[9] Victor Bussy, Caroline Vienne, Julie Escoda, and Valérie Kaftandjian. Best projections selection algorithm based on constrained qdeim for sparse-views x-ray computed tomography. In 12th Conference on Industrial Computed Tomography (iCT) 2023, 27 February - 2 March 2023 in Fürth, Germany. e-Journal of Nondestructive Testing Vol. 28(3), 2023.

[10] F. Bauer, C.U. Große, and F. Pfeiffer. Trajectory Optimization for Sparsely Sampled Computed Tomography. Universitätsbibliothek der TU München, 2022.

[11] Victor Bussy, Caroline Vienne, Julie Escoda, and Valerie Kaftandjian. Méthodologie optimisée pour la reconstruction tomographique avec ajout d'informations a priori pour l'inspection par rayons x. e-Journal of Nondestructive Testing, 28, 09 2023.

[12] Victor Bussy and Caroline Vienne. Robust 3D/2D hybrid registration for integrating a priori CAD model into X-ray Computed Tomography. In ORASIS 2021, Saint Ferréol, France, September 2021. Centre National de la Recherche Scientifique [CNRS].

[13] Victor Bussy, Caroline Vienne, Julie Escoda, and Valérie Kaftandjian. Sparse-View X-Ray CT Reconstruction using CAD Model Registration. volume 2022 49th Annual Review of Progress in Quantitative Nondestructive Evaluation of Quantitative Nondestructive Evaluation, page V001T13A001, 07 2022.

[14] Chuan Huang, Paul Vaska, Yongfeng Gao, Shaojie Chang, Thomas Wesley Holmes, Amir Pourmorteza, and Jerome Liang. Proceedings of the 17th international meeting on fully 3d image reconstruction in radiology and nuclear medicine, 2023.

[15] Michel Defrise and Pierre Grangeat. Analytical Methods, chapter 2, pages 21–62. John Wiley & Sons, Ltd, 2009.

[16] Daniele Panetta. Advances in x-ray detectors for clinical and preclinical computed tomography. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 809:2–12, 2016. Advances in detectors and applications for medicine.

[17] Jiang Hsieh. Computed Tomography: Principles, Design, Artefacts and Recent Advances. Wiley Interscience, Bellingham, Washington USA, second edition, 2009.

[18] Thierry LEMOINE. Imagerie médicale par rayons x détecteurs de rayons x. Techniques de l'ingénieur Technologies pour la santé, base documentaire : TIP597WEB.(ref. article : med201), 2015. fre.

[19] David V. Finch. Cone beam reconstruction with sources on a curve. SIAM Journal on Applied Mathematics, 45(4):665–673, 1985.

[20] Frédéric Noo, Rolf Clackdoyle, and Jed D Pack. A two-step hilbert transform method for 2d image reconstruction. Physics in Medicine & Biology, 49(17):3903, aug 2004.

[21] Avinash C. Kak and Malcolm Slaney. Principles of Computerized Tomographic Imaging. Society for Industrial and Applied Mathematics, 2001.

[22] Gengsheng Lawrence Zeng. Medical Image Reconstruction. Springer Berlin, Heidelberg, 2010. `https://link.springer.com/book/10.1007/978-3-642-05368-9`.

[23] Nicolas Gindrier. Reconstruction cone-beam à partir de projections avec troncations transverses. PhD thesis, Université Grenoble Alpes, 2022. Thèse de doctorat dirigée par Desbat, Laurent et Clackdoyle, Rolf Mathématiques appliquées Université Grenoble Alpes 2022.

[24] M. Defrise and R. Clack. A cone-beam reconstruction algorithm using shift-variant filtering and cone-beam backprojection. IEEE Transactions on Medical Imaging, 13(1):186–195, 1994.

[25] Terry M. Peters. Algorithms for fast back- and re-projection in computed tomography. IEEE Transactions on Nuclear Science, 28:3641–3647, 1981.

[26] W. Zhuang, S.S. Gopal, and T.J. Hebert. Numerical evaluation of methods for computing tomographic projections. IEEE Transactions on Nuclear Science, 41(4):1660–1665, 1994.

[27] Robert L. Siddon. Fast calculation of the exact radiological path for a three-dimensional ct array. Medical physics, 12 2:252–5, 1985.

[28] Gengsheng Zeng and G.T. Gullberg. A ray-driven backprojector for backprojection filtering and filtered backprojection algorithms. In Journal of Nutritional Biochemistry - J NUTR BIOCHEM, pages 1199 – 1201, 12 1993.

[29] Peter M. Joseph. An improved algorithm for reprojecting rays through pixel images. IEEE Transactions on Medical Imaging, 1(3):192–196, 1982.

[30] Hengyong Yu and Ge Wang. Finite detector based projection model for high spatial resolution. Journal of X-ray science and technology, 20:229–38, 01 2012.

[31] Bruno De Man and Samit Basu. Distance-driven projection and backprojection in three dimensions. Physics in medicine and biology, 49:2463–75, 07 2004.

[32] Gabor Herman. Fundamentals of Computerized Tomography. 01 2010.

[33] Richard Gordon, Robert Bender, and Gabor T. Herman. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. Journal of Theoretical Biology, 29(3):471–481, 1970.

[34] Jens Gregor and Thomas Benson. Computational analysis and improvement of sirt. IEEE Transactions on Medical Imaging, 27(7):918–924, 2008.

[35] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. IEEE Transactions on Medical Imaging, 1(2):113–122, 1982.

[36] Antonin Chambolle, Matthias J Ehrhardt, Peter Richtarik, and Carola-Bibiane Schönlieb. STOCHASTIC PRIMAL-DUAL HYBRID GRADIENT ALGORITHM WITH ARBITRARY SAMPLING AND IMAGING APPLICATIONS. SIAM Journal on Optimization, 28(4):2783–2808, October 2018.

[37] Antonin Chambolle and Pock Thomas. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 2011.

[38] Ge Wang, Yi Zhang, Xiaojing Ye, and Xuanqin Mou. Machine Learning for Tomographic Imaging. 2053-2563. IOP Publishing, 2019.

[39] Thierry LEMOINE. Imagerie médicale par rayons x qualité d'image. Techniques de l'ingénieur Technologies pour la santé, base documentaire : TIP597WEB.(ref. article : med205), 2015. fre.

[40] Lu Zhang. Numerical Observers for the Objective Quality Assessment of Medical Images. Theses, Université d'Angers, November 2012.

[41] Hussein Banjak. X-ray computed tomography reconstruction on non-standard trajectories for robotized inspection. PhD thesis, Université de Lyon, 2016. Thèse de doctorat dirigée par Kaftandjian, Valérie Traitement du signal et de l'image Lyon 2016.

[42] Lee Feldkamp, L. C. Davis, and James Kress. Practical cone-beam algorithm. J. Opt. Soc. Am, 1:612–619, 01 1984.

[43] Heang K. Tuy. An inversion formula for cone-beam reconstruction. SIAM Journal on Applied Mathematics, 43(3):546–552, 1983.

[44] Alexander Katsevich. Image reconstruction for the circle and line trajectory. Physics in medicine and biology, 49:5059–72, 12 2004.

[45] Zhicong Yu, Günter Lauritsch, Frank Dennerlein, Yanfei Mao, Joachim Hornegger, and Frédéric Noo. Extended ellipse-line-ellipse trajectory for long-object cone-beam imaging with a mounted c-arm system. Physics in Medicine & Biology, 61(4):1829, feb 2016.

[46] Frédéric Noo, Rolf Clack, Timothy A White, and Timothy J Roney. The dual-ellipse cross vertex path for exact reconstruction of long objects in cone-beam tomography. Physics in Medicine & Biology, 43(4):797, apr 1998.

[47] Tingliang Zhuang, Joseph Zambelli, Brian Nett, Shuai Leng, and Guang-Hong Chen. Exact and approximate cone-beam reconstruction algorithms for c-arm based cone-beam ct using a two-concentric-arc source trajectory. Proceedings - Society of Photo-Optical Instrumentation Engineers, 6913:691321, 02 2008.

[48] Haiquan Yang, Meihua Li, Kazuhito Koizumi, and Hiroyuki Kudo. Closed sinusoid trajectory for c-arm ct imaging. In 2006 IEEE Nuclear Science Symposium Conference Record, volume 6, pages 3480–3484, 2006.

[49] Fabian Bauer, Matthias Goldammer, and Christian Grosse. Selection and evaluation of spherical acquisition trajectories for industrial computed tomography. Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences, 477, 06 2021.

[50] F. Noo, R. Clack, and M. Defrise. Cone-beam reconstruction from general discrete vertex sets using radon rebinning algorithms. IEEE Transactions on Nuclear Science, 44(3):1309–1316, 1997.

[51] Gabriel Herl, Andreas Maier, and Simon Zabler. X-ray ct data completeness condition for sets of arbitrary projections. In Proceedings of the SPIE, Volume 12304, id. 123040C 7 pp, page 23, 10 2022.

[52] Chris Jacobsen. Relaxation of the crowther criterion in multislice tomography. Opt. Lett., 43(19):4811–4814, Oct 2018.

[53] F. Natterer. The Mathematics of Computerized Tomography. Society for Industrial and Applied Mathematics, 2001.

[54] Eric Todd Quinto. Singularities of the x-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$. SIAM Journal on Mathematical Analysis, 24(5):1215–1225, 1993.

[55] Ziyi Zheng and Klaus Mueller. Identifying sets of favorable projections for few-view low-dose cone-beam ct scanning. In In 11th Fully3D meeting, 314–7, pages 314–317, 2011.

[56] Marcelo Venere, Hstau Liao, and Alejandro Clausse. Genetic algorithm for adaptive tomography of elliptical objects. Signal Processing Letters, IEEE, 7:176 – 178, 08 2000.

[57] Mohammad Haque, M. Omair Ahmad, M.N.s Swamy, Md Kamrul Hasan, and Soo Lee. Adaptive projection selection for computed tomography. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 22, 08 2013.

[58] G. Placidi, M. Alecci, and A. Sotgiu. Theory of adaptive acquisition method for image reconstruction from projections and application to epr imaging. Journal of Magnetic Resonance, Series B, 108(1):50–57, 1995.

[59] Matz A., Holub W., and Schielein R. Trajectory optimization in computed tomography based on object geometry. 11th Conference on Industrial Computed Tomography, Wels, Austria (iCT 2022), 2022.

[60] Baodong Liu, J. Bennett, Ge Wang, B. De Man, K. Zeng, Z. Yin, P. Fitzgerald, and Hengyong Yu. Completeness map evaluation demonstrated with candidate next-generation cardiac ct architectures. Medical physics, 39 5:2405–16, 2012.

[61] Gabriel Herl, Jochen Hiller, and Andreas Maier. Scanning trajectory optimisation using a quantitative tuybased local quality estimation for robot-based x-ray computed tomography. Nondestructive Testing and Evaluation, 35(3):287–303, 2020.

[62] Grace J. Gang, Johns H. Siewerdsen, and J Webster Stayman. Non-circular ct orbit design for elimination of metal artifacts. In International Society for Optical Engineering, 2020.

[63] Christoph Heinzl, Johann Kastner, Artem Amirkhamov, Eduard Gröller, and Christian Gusenbauer. Optimal specimen placement in cone beam x-ray computed tomography. NDT & E International, 09 2012.

[64] S. Kasperl, M. Franz, S. Reisinger, U. Schmid, and J. Hiller. imulation-based planning of optimal conditions for industrial computed tomography. In International Symposium on Digital Industrial Radiology and Computed Tomography, 20-22 June 2011, Berlin, Germany, 2011.

[65] N. Grozmani, Robert H. Schmitt, and Andrea Buratti. Investigating the influence of workpiece placement on the uncertainty of measurements in industrial computed tomography. In Proceedings of 9th Conference on Industrial Computed Tomography (iCT), 2019.

[66] Artem Amirkhanov, Christoph Heinzl, Michael Reiter, and Eduard Gröller. Visual optimality and stability analysis of 3dct scan positions. IEEE Transactions on Visualization and Computer Graphics, 16(6):1477–1486, 2010.

[67] Sepideh Hatamikia, Ander Biguri, Gernot Kronreif, Tom Russ, Joachim Kettenbach, and Wolfgang Birkfellner. Short scan source-detector trajectories for target-based cbct. In Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, volume 2020, pages 1299–1302, 07 2020.

[68] Nick Brierley, C. Bellon, and Borja Lazaro Toralles. Optimized multi-shot imaging inspection design. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 474:20170319, 08 2018.

[69] J. Siewerdsen. Task-based trajectories in iteratively reconstructed interventional cone-beam ct. In In 12t Fully3D meeting, pages 257–260, 2013.

[70] Andreas Fischer, Tobias Lasser, Michael Schrapp, Jürgen Stephan, and Peter Noël. Object specific trajectory optimization for industrial x-ray computed tomography. Scientific Reports, 6:19135, 01 2016.

[71] Fabian Bauer, Daniel Forndran, Thomas Schromm, and Christian Grosse. Practical part-specific trajectory optimization for robot-guided inspection via computed tomography. Journal of Nondestructive Evaluation, 41:54, 07 2022.

[72] Mareike Thies, Jan-Nico Zäch, Cong Gao, Russell Taylor, Nassir Navab, Andreas Maier, and Mathias Unberath. A learning-based method for online adjustment of c-arm cone-beam ct source trajectories for artifact avoidance. International Journal of Computer Assisted Radiology and Surgery, 15, 08 2020.

[73] Linda-Sophie Schneider, Mareike Thies, Christopher Syben, Richard Schielein, Mathias Unberath, and Andreas Maier. Task-based generation of optimized projection sets using differentiable ranking. arXiv.org, abs/2303.11724, 2023.

[74] Zlatko Drmač and Serkan Gugercin. A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions. SIAM Journal on Scientific Computing, 38(2):A631–A648, jan 2016.

[75] Steven L. Brunton and J. Nathan Kutz. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press, 2 edition, 2022.

[76] Yvon Maday, Ngoc Nguyen, Anthony Patera, and George Shu Heng Pau. A general multipurpose interpolation procedure: The magic points. Communications on Pure and Applied Analysis, 8, 10 2008.

[77] Yvon Maday and Olga Mula. A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation. In Analysis and Numerics of Partial Differential Equations, pages 221–235. Springer Milan, 2013.

[78] G. Golub. Numerical methods for solving linear least squares problems. Numer. Math., 7:206–216, 1965.

[79] Wikipedia. Qr decomposition, 2022. Accessed on January 14, 2022.

[80] Alok Sharma, Kuldip Paliwal, Seiya Imoto, and Satoru Miyano. Principal component analysis using qr decomposition. International Journal of Machine Learning and Cybernetics, 4, 12 2013.

[81] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, and Steven L. Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. IEEE Control Systems Magazine, 38(3):63–86, 2018.

[82] Paola Sebastiani and Henry Wynn. Maximum entropy sampling and optimal bayesian experimental design. Journal of the Royal Statistical Society Series B, 62:145–157, 02 2000.

[83] Yuji Saito, Taku Nonomura, Keigo Yamada, Kumi Nakai, Takayuki Nagata, Keisuke Asai, Yasuo Sasaki, and Daisuke Tsubakino. Determinant-based fast greedy sensor selection algorithm. IEEE Access, 9:68535–68551, 2021.

[84] Matthias Seeger, Hannes Nickisch, R. Pohmann, and Bernhard Schölkopf. B.: Optimization of k-space trajectories for compressed sensing by bayesian experimental design. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, 63:116–26, 11 2009.

[85] Pere-Pau Vázquez Alcocer, Miquel Feixas, Mateu Sbert, and Wolfgang Heidrich. Automatic view selection using viewpoint entropy and its application to image-based modeling. Computer Graphics Forum, 22:689–700, 12 2003.

[86] Xavier Bonaventura Brugués. Perceptual information-theoretic measures for viewpoint selection and object recognition. PhD thesis, Universitat de Girona, 2015.

[87] Helin Dutagaci, Chun Pan Cheung, and Afzal Godil. A benchmark for best view selection of 3d objects. In Proceedings of the ACM Workshop on 3D Object Retrieval, 3DOR '10, page 45–50, New York, NY, USA, 2010. Association for Computing Machinery.

[88] Christos Boutsidis, Michael W. Mahoney, and Petros Drineas. An improved approximation algorithm for the column subset selection problem. In Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '09, page 968–977, USA, 2009. Society for Industrial and Applied Mathematics.

[89] I G Kazantsev. Information content of projections. Inverse Problems, 7(6):887–898, 12 1991.

[90] Emily Clark, Travis Askham, Steven Brunton, and J. Kutz. Greedy sensor placement with cost constraints. IEEE Sensors Journal, PP:1–1, 12 2018.

[91] G.P. Penney, J. Weese, J.A. Little, P. Desmedt, D.L.G. Hill, and D.J. hawkes. A comparison of similarity measures for use in 2-d-3-d medical image registration. IEEE Transactions on Medical Imaging, 17(4):586–595, 1998.

[92] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš. A review of 3d/2d registration methods for image-guided interventions. Medical Image Analysis, 16(3):642–661, 2012. Computer Assisted Interventions.

[93] Frank Langlotz Xuan Zhang, Guoyan Zheng and Lutz-Peter Nolte. Assessment of spline-based 2d–3d registration for image-guided spine surgery. Minimally Invasive Therapy & Allied Technologies, 15(3):193–199, 2006. PMID: 16785186.

[94] Xin Chen, Martin Varley, Lik-Kwan Shark, Glyn Shentall, and Mike Kirby. An extension of iterative closest point algorithm for 3d-2d registration for pre-treatment validation in radiotherapy. In International Conference on Medical Information Visualisation - BioMedical Visualisation (MedVis'06), pages 3–8, 01 2006.

[95] Martin Groher, Frederik Bender, Ralf Hoffmann, and Nassir Navab. Segmentation-driven 2d-3d registration for abdominal catheter interventions. In Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, volume 10, pages 527–35, 02 2007.

[96] Domenico Iuso, Pavel Paramonov, Jan De Beenhouwer, and Jan Sijbers. Practical multi-mesh registration for few-view poly-chromatic x-ray inspection jo - journal of nondestructive evaluation. Journal of Nondestructive Evaluation, 2024.

[97] Vincent Lepetit and Pascal Fua. Monocular model-based 3d tracking of rigid objects: A survey. Foundations and Trends in Computer Graphics and Vision, 1, 01 2005.

[98] Richard Hartley and Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2 edition, 2004.

[99] I.E. Sutherland. Three-dimensional data input by tablet. Proceedings of the IEEE, 62(4):453–461, 1974.

[100] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(4):376–380, 1991.

[101] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24(6):381–395, jun 1981.

[102] Eric Marchand, Hideaki Uchiyama, and Fabien Spindler. Pose estimation for augmented reality: A hands-on survey. IEEE Transactions on Visualization and Computer Graphics, 22, 01 2016.

[103] Daniel DeMenthon and Larry Davis. Model-based object pose in 25 lines of code. In International Journal of Computer Vision, volume 15, 02 1998.

[104] Philip David, Daniel DeMenthon, Ramani Duraiswami, and Hanan Samet. Softposit: Simultaneous pose and correspondence determination. International Journal of Computer Vision, 59, 09 2003.

[105] Steven Gold, Anand Rangarajan, Chien-Ping Lu, Suguna Pappu, and Eric Mjolsness. New algorithms for 2d and 3d point matching: pose estimation and correspondence. Pattern Recognition, 31(8):1019–1031, 1998.

[106] François Chaumette and Seth Hutchinson. Visual servo control. i. basic approaches. Robotics & Automation Magazine, IEEE, 13:82 – 90, 01 2007.

[107] François Chaumette and Seth Hutchinson. Visual servo control. ii. advanced approaches [tutorial]. Robotics & Automation Magazine, IEEE, 14:109 – 118, 04 2007.

[108] David Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60:91–, 11 2004.

[109] S. Chen, Youfu Li, J. Zhang, and W. Wang. Active Sensor Planning for Multiview Vision Tasks. 01 2008.

[110] Ezio Malis, François Chaumette, and Sylvie Boudet. $2\frac{1}{2}$d visual servoing. Robotics and Automation, IEEE Transactions on, 15:238 – 250, 05 1999.

[111] Peter I. Corke and Seth A. Hutchinson. A new partitioned approach to image-based visual servo control. IEEE Transactions on Robotics and Automation, 17(4):507–515, August 2001.

[112] William Schroeder, Jonathan Zarge, and William Lorensen. Decimation of triangle meshes. SIGGRAPH Comput. Graph., 26:65–70, 06 1997.

[113] Patrick Wunsch and Gerd Hirzinger. Registration of cad-models to images by iterative inverse perspective matching. Proceedings of 13th International Conference on Pattern Recognition, 1:78–83 vol.1, 1996.

[114] Peter J. Huber. Robust Statistics, pages 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[115] Alfred Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. American Journal of Mathematics, 76(3):620–630, 1954.

[116] Christopher G. Harris and M. J. Stephens. A combined corner and edge detector. In Alvey Vision Conference, 1988.

[117] François Chaumette. Image moments: A general and useful set of features for visual servoing. Robotics, IEEE Transactions on, 20:713 – 723, 09 2004.

[118] Alireza Entezari, Masih Nilchian, and Michael Unser. A box spline calculus for the discretization of computed tomography reconstruction problems. IEEE Transactions on Medical Imaging, 31(8):1532–1541, 2012.

[119] Han Wang. Méthodes de reconstruction d'images à partir d'un faible nombre de projections en to
Theses, Université de Grenoble, Oct 2011.

[120] Jakeoung Koo, Anders B. Dahl, J. Andreas Bærentzen, Qiongyang Chen, Sara Bals, and Vedrana A. Dahl. Shape from projections via differentiable forward projector for computed tomography. Ultramicroscopy, 224:113239, 2021.

[121] Anthony Cazasnoves. Représentation par maillage adapté pour la reconstruction 3D en tomographie par rayons X. PhD thesis, Université de Lyon, 2015. Thèse de doctorat dirigée par Peyrin, Françoise Traitement du signal et de l'image Lyon, INSA 2015.

[122] Hanan Samet. An overview of quadtrees, octrees, and related hierarchical data structures. In Rae A. Earnshaw, editor, Theoretical Foundations of Computer Graphics and CAD, pages 51–68, Berlin, Heidelberg, 1988. Springer Berlin Heidelberg.

[123] Seungki Kim, Akitsune Sakane, Yutaka Ohtake, Hiromasa Suzuki, Yukie Nagai, Katsutoshi Sato, Hiroyuki Fujimoto, Makoto Abe, Osamu Sato, and Toshiyuki Takatsuji. Efficient iterative ct reconstruction on octree guided by geometric errors. volume 6th Conference on Industrial Computed Tomography (iCT) 2016, 9-12 February 2016, Wels, Austria of e-Journal of Nondestructive Testing Vol. 21(2)., 2016.

[124] Darius Rückert, Yuanhao Wang, Rui Li, Ramzi Idoughi, and Wolfgang Heidrich. Neat: Neural adaptive tomography, 2022.

[125] John Amanatides and Andrew Woo. A fast voxel traversal algorithm for ray tracing. Proceedings of EuroGraphics, 87, 08 1987.

[126] Ben Gorte, K. Zhou, Corné Sande, and C. Valk. A computationally cheap trick to determine shadow in a voxel model. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-4:67–71, 09 2018.

[127] Samuli Laine. A topological approach to voxelization. Computer Graphics Forum, 32, 07 2013.

[128] Christoph K. Reinbothe, Tamy Boubekeur, and Marc Alexa. Hybrid ambient occlusion. In Eurographics, 2009.

[129] Yu Liu, Andreas Beyer, Philipp Schuetz, Juergen Hofmann, Alexander Flisch, and Urs Sennhauser. Cooperative data fusion of transmission and surface scan for improving limited-angle computed tomography reconstruction. NDT & E International, 83:24–31, 2016.

[130] Christian Schorr. Exploitation of geometric a priori knowledge for limited data reconstruction in nondestructive testing. 06 2013.

[131] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Physics in Medicine & Biology, 53(17):4777, aug 2008.

[132] Jeroen Baert. Cuda voxelizer: A gpu-accelerated mesh voxelizer. `https://github.com/Forceflow/cuda_voxelizer`, 2017.

[133] Renben Shu and Mohan S Kankanhalli. Efficient linear octree generation from voxels. Image and Vision Computing, 12(5):297–303, 1994.

[134] Jorge Revelles, Carlos Ureña, M. Lastra, Dpt Lenguajes, Sistemas Informaticos, and E. Informatica. An efficient parametric algorithm for octree traversal. 05 2000.

[135] Célestin Marot and Jean-François Remacle. Quality tetrahedral mesh generation with HXT. CoRR, abs/2008.08508, 2020.

[136] Ingo Wald, Sven Woop, Carsten Benthin, Gregory S. Johnson, and Manfred Ernst. Embree: A kernel framework for efficient cpu ray tracing. ACM Trans. Graph., 33(4), jul 2014.

[137] Paolo Cignoni, Leila De Floriani, Paola Magillo, Enrico Puppo, and Roberto Scopigno. Selective refinement queries for volume visualization of unstructured tetrahedral meshes. IEEE transactions on visualization and computer graphics, 10:29–45, 01 2004.

[138] P. Paramonov, N. Francken, J. Renders, D. Iuso, T. Elberfeld, J. De Beenhouwer, and J. Sijbers. Cad-astra: a versatile and efficient mesh projector for x-ray tomography with the astra-toolbox. Opt. Express, 32(3):3425–3439, Jan 2024.

[139] Zhang H, Wang J, Zeng D, Tao X, and Ma J. Regularization strategies in statistical image reconstruction of low-dose x-ray ct: A review. Medical physics, 45(10):886–907, 2018.

[140] Johan Nuyts, Bruno De Man, Jeffrey Fessler, Wojciech Zbijewski, and Freek Beekman. Modelling the physics in the iterative reconstruction for transmission computed tomography. Physics in medicine and biology, 58:R63–R96, 06 2013.

[141] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. IEEE Transactions on Medical Imaging, 38(2):394–405, 2019.

[142] Steven Diamond, Vincent Sitzmann, Felix Heide, and Gordon Wetzstein. Unrolled optimization with deep priors, 2018.

[143] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red), 2017.

[144] Samuel Hurault. Convergent plug-and-play methods for image inverse problems with explicit and Theses, Université de Bordeaux, November 2023.

[145] Davood Karimi. Sparse and redundant signal representations for x-ray computed tomography, 2019.

[146] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing, 16(8):2080–2095, 2007.

[147] David Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. Comm. Pure Appl. Math, 59, 01 2006.

[148] Emmanuel J. Candes and Yaniv Plan. A probabilistic and ripless theory of compressed sensing, 2010.

[149] Simon Foucart. Hard thresholding pursuit: An algorithm for compressive sensing. SIAM Journal on Numerical Analysis, 49(6):2543–2563, 2011.

[150] Simon Foucart. Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants. Springer Proceedings in Mathematics, 13, 01 2012.

[151] Ron Rubinstein, Michael Zibulevsky, and Michael Elad. Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit. CS Technion, 40, 01 2008.

[152] Bogdan Dumitrescu and Paul Irofti. Dictionary learning algorithms and applications. Springer, Gewerbestrasse 11, 6330 Cham, Switzerland, 2018.

[153] David Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. IEEE Transactions on Information Theory, 58:1094–1121, 02 2012.

[154] Neal Parikh. Proximal algorithms. Foundations and Trends in Optimization, 1:127–239, 01 2014.

[155] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing, 54(11):4311–4322, 2006.

[156] Y. Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35:1798–1828, 08 2013.

[157] Brendt Wohlberg. Efficient algorithms for convolutional sparse representations. IEEE Transactions on Image Processing, 25(1):301–315, 2016.

[158] Cristina Garcia-Cardona and Brendt Wohlberg. Subproblem coupling in convolutional dictionary learning. In Proceedings of IEEE International Conference on Image Processing (ICIP), pages 1697–1701, Beijing, China, September 2017.

[159] Vardan Papyan, Jeremias Sulam, and Michael Elad. Working locally thinking globally: Theoretical guarantees for convolutional sparse coding. IEEE Transactions on Signal Processing, 65(21):5687–5701, nov 2017.

[160] Vardan Papyan, Yaniv Romano, and Michael Elad. Convolutional neural networks analyzed via convolutional sparse coding, 2016.

[161] S. V. Venkatakrishnan and Brendt Wohlberg. Convolutional dictionary regularizers for tomographic inversion. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 7820–7824, 2019.

[162] Yang Lu, Jun Zhao, and Ge Wang. Few-view image reconstruction with dual dictionaries. Physics in medicine and biology, 57:173–89, 12 2011.

[163] Changhui Jiang, Qiyang Zhang, Rui Fan, and Zhanli Hu. Super-resolution ct image reconstruction based on dictionary learning and sparse representation. Scientific Reports, 8, 06 2018.

[164] Toshiaki Omori, Shoi Suzuki, Katsuyoshi Michibayashi, and Atsushi Okamoto. Super-resolution of x-ray ct images of rock samples by sparse representation: applications to the complex texture of serpentinite. Scientific Reports, 13:6648, 04 2023.

[165] Qiong Xu, Hengyong Yu, Xuanqin Mou, and Ge Wang. Dictionary learning based low-dose x-ray ct reconstruction. IEEE transactions on medical imaging, 11 2014.

[166] Peng Bao, Huaiqiang Sun, Zhangyang Wang, Yi Zhang, Wenjun Xia, Kang Yang, Weiyan Chen, Mianyi Chen, Yan Xi, Shanzhou Niu, Jiliu Zhou, and He Zhang. Convolutional sparse coding for compressed sensing CT reconstruction. IEEE Transactions on Medical Imaging, 38(11):2607–2619, nov 2019.

[167] Ido Bright, Guang Lin, and J. Kutz. Compressive sensing based machine learning strategy for characterizing the flow around a cylinder with limited pressure measurements. Physics of Fluids, 25, 11 2013.

[168] Paul Irofti and Florin Stoican. Dictionary learning strategies for sensor placement and leakage isolation in water networks. 07 2017.

[169] Chao Zhang and Mirko Van Der Baan. Signal processing using dictionaries, atoms, and deep learning: A common analysis-synthesis framework. Proceedings of the IEEE, 110(4):454–475, 2022.

[170] Meng Yang, Weiyang Liu, Weixin Luo, and Linlin Shen. Analysis-synthesis dictionary learning for universality-particularity representation based classification. Proc. AAAI Conf. Artif. Intell., 30:2251–2257, 03 2016.

[171] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2):210–227, 2009.

[172] Duc-Son Pham and Svetha Venkatesh. Joint learning and dictionary construction for pattern recognition. 2008 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, 2008.

[173] Ignacio Ramirez Paulino, Pablo Sprechmann, and Guillermo Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 3501 – 3508, 07 2010.

[174] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang. Fisher discrimination dictionary learning for sparse representation. In Proceedings of the IEEE International Conference on Computer Vision, pages 543–550, 11 2011.

[175] Tiep Huu Vu, Hojjat Seyed Mousavi, Vishal Monga, Ganesh Rao, and U. K. Arvind Rao. Histopathological image classification using discriminative feature-oriented dictionary learning. IEEE Transactions on Medical Imaging, 35(3):738–751, 2016.

[176] Shu Kong and Donghui Wang. A dictionary learning approach for classification: Separating the particularity and the commonality. In Proceedings of the 12th European conference on Computer Vision - Volume Part I, volume 7572, pages 186–199, 10 2012.

[177] Shenghua Gao, Ivor Wai-Hung Tsang, and Yi Ma. Learning category-specific dictionary and shared dictionary for fine-grained image categorization. IEEE Transactions on Image Processing, 23(2):623–634, 2014.

[178] Tiep Vu and Vishal Monga. Fast low-rank shared dictionary learning for image classification. IEEE Transactions on Image Processing, PP, 10 2016.

APPENDICES

Appendix A

Appendix

A.1 Interaction matrix for visual servoing

To handle a general case with arbitrary features, we provide a method for computing the interaction matrix $\mathbf{L_c}$. Let us define P_s as one of the features composing the pattern in the scene, parameterised by $\mathbf{P}$. Similarly, we define p_i as one of the features forming the resulting image after projection, parameterised by $\mathbf{p}$. Consider a parametrizable geometric primitive in the scene, P_s, which can be generally described by the equation:

$$h(\mathbf{X}, \mathbf{P}) = 0, \tag{A.1}$$

where h defines the primitive and the parameter $\mathbf{P}$ represents one of its configurations. Our objective is to estimate the value of the parameters $\mathbf{P}$ in order to reconstruct and locate the primitive P_s, described by h (the type of primitive to be reconstructed is assumed to be known *a priori*). To achieve this, the representation of P_s by the set of parameters $\mathbf{P}$ is chosen such that it is complete and unique. Thus, for any configuration of the primitive, there exists only one unique set of values for $\mathbf{P}$ corresponding to that configuration.

Using the perspective projection equation, and after rearranging the terms, we obtain:

$$h(Z\mathbf{x}, \mathbf{P}) = 0 \quad \text{and} \quad h'(Z, \mathbf{x}, \mathbf{P}) = 0. \tag{A.2}$$

The implicit function theorem guarantees that under the condition $\frac{\partial h'}{\partial Z} \neq 0$ (which is satisfied in non-degenerate cases, such as when a line projects to a point), there exists a unique function f in a neighbourhood of the solution to EquationA.2 such that:

$$Z = f(\mathbf{x}, \mathbf{P}). \tag{A.3}$$

Since p_i is the projection of P_s onto the image plane, using the preceding equations, we can write:

$$h'(\mathbf{x}, \mathbf{P}) = 0, \tag{A.4}$$

which, after reparametrisation, can be rewritten as:

$$g(\mathbf{x}, \mathbf{p}) = 0. \tag{A.5}$$

The calculation of the interaction matrix can now be carried out under the assumption of rigidity of the geometric primitive:

$$\dot{g}(\mathbf{x}, \mathbf{p}) = 0, \quad \forall \mathbf{x} \in P_i, \tag{A.6}$$

leading to:

$$\frac{\partial g}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p})\dot{\mathbf{p}} = -\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{p})\dot{\mathbf{x}}. \tag{A.7}$$

By incorporating the optical flow equations into $\dot{\mathbf{x}}$, we obtain:

$$\frac{\partial g}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p})\dot{\mathbf{p}} = -\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{p})\mathbf{L_c}\mathbf{T}, \tag{A.8}$$

where $\mathbf{T}$ is the kinematic twist that represents the relative velocity of the scene with respect to the camera. To solve for $\mathbf{L_c}$ analytically, or through a system, consider:

$$\frac{\partial g}{\partial \mathbf{p}}(\mathbf{x_i}, \mathbf{p})\dot{\mathbf{p}} = \alpha_i^{\mathsf{T}}, \quad i \in [1, m], \tag{A.9}$$

$$-\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x_i}, \mathbf{p})\mathbf{L_c}^{\mathsf{T}} = \beta_i^{\mathsf{T}}, \quad i \in [1, m], \tag{A.10}$$

where m is the dimension of the parametrisation. Solving for $\mathbf{L_c}^{\mathsf{T}}$ yields:

$$\mathbf{L_c}^{\mathsf{T}} = (\alpha^{\mathsf{T}})^{-1}\beta^{\mathsf{T}}. \tag{A.11}$$

The matrix $\mathbf{L_c}$ has dimensions $m \times 6$. To convert from one parameterisation $\mathbf{p}'$ to another parametrisation $\mathbf{p}$, we use:

$$\mathbf{L}^{\mathsf{T}}(\mathbf{P}, \mathbf{p}') = \frac{\partial \mathbf{p}'}{\partial \mathbf{p}}\mathbf{L}^{\mathsf{T}}(\mathbf{P}, \mathbf{p}). \tag{A.12}$$

A.2 Projection of a convex hull

We will show that the points on the convex hull of the projection onto the detector are the projections of the points on the convex hull of the 3D object. Here, we model the CAD as a set of triangles, which is equivalent to considering the STL model. The camera or X-ray source is at the origin O. Let's consider a line $\overline{AB}$ of the convex hull of the projection. By definition, all the points on the convex hull are on the same side of $\overline{AB}$ (or belong to the line); conventionally, we say these points are 'beneath' the facet $\overline{AB}$ (opposed to 'beyond'). Now, let's consider the points A' and B' of the object from which A and B originate. We have $P(A') = A$ and $P(B') = B$ (A' and B' may not be unique, but their existence is guaranteed; otherwise, A and B would not be defined). Now, let's consider the plane $\Delta(A'B'O)$ and a point p on the 3D convex hull, denoted as $\mathcal{S}$. If for every p in $\mathcal{S}$, p is in or beneath the plane, then all the points in $\mathcal{S}$ are on the same side of the plane $\Delta(A'B'O)$, so the points A' and B' belong to the convex hull of the object. If there exists a p for which this property is not satisfied, then $P(p)$ is beyond $\overline{AB}$, so AB is not a side of the convex hull, which contradicts the definition of the points A and B.

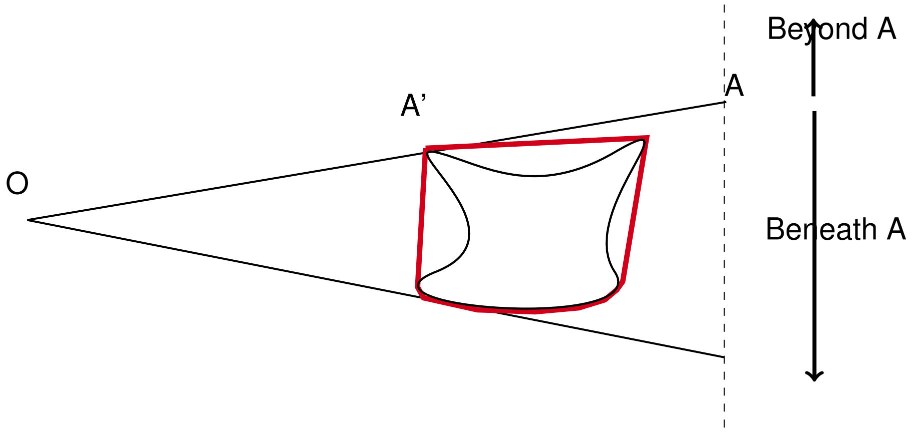


Figure A.1: Illustration of the projection of a convex hull point

A.3 Iterative Inverse Perspective Matching

We can demonstrate that the iterative inverse perspective matching converges monotonically towards a local minimum. We denote $\mathbf{Y}^0$ the initial set of image features, $\mathbf{R}^0 = \mathbf{I}_3$ the initial rotation, $\mathbf{t}^0 = \mathbf{0}$ the initial translation and $\mathbf{H^0}$ the initial homography made with the rotation and the translation. Given $\mathbf{Y}^k {=} \left\{\mathbf{Y}_i^k\right\}_i = \mathbf{H}^k(\mathbf{Y}^0)$, we can define the errors:

$$\mathbf{e}_k = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{X}_i^k - \mathbf{Y}_i^k\|^2 \tag{A.13}$$

$$\mathbf{d}_k = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{X}_i^k - (\mathbf{R}^k\mathbf{Y}_i^0 + \mathbf{t}^k)\|^2 \tag{A.14}$$

The error $\mathbf{e}_k$ represents the error at the start of the k^{th} iteration after the matching process, while $\mathbf{d}_k$ represents the error at the end. The relation $\mathbf{d}_k \leq \mathbf{e}_k$ holds, otherwise if $\mathbf{d}_k > \mathbf{e}_k$, then $(\mathbf{R}^k = \mathbf{I_3}, \mathbf{t}^k = \mathbf{0})$ would be a better solution to the least-squares problem, which is impossible by definition. During the iterations, the movement $\mathbf{H}^k$ is applied to the set $\mathbf{Y}^0$. If the matching does not change, then:

$$\mathbf{d}_k = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{X}_i^k - \mathbf{Y}_i^{k+1}\|^2 \tag{A.15}$$

However, if $\mathbf{X}^k$ changes:

$$\|\mathbf{X}_i^{k+1} - \mathbf{Y}_i^{k+1}\| \leq \|\mathbf{X}_i^k - \mathbf{Y}_i^{k+1}\| \tag{A.16}$$

Which leads to the following equations:

$$0 \leq \mathbf{d}_{k+1} \leq \mathbf{e}_{k+1} \leq \mathbf{d}_k \leq \mathbf{e}_k \ , \ \forall k \tag{A.17}$$

The procedure monotonically converges to a local minimum.

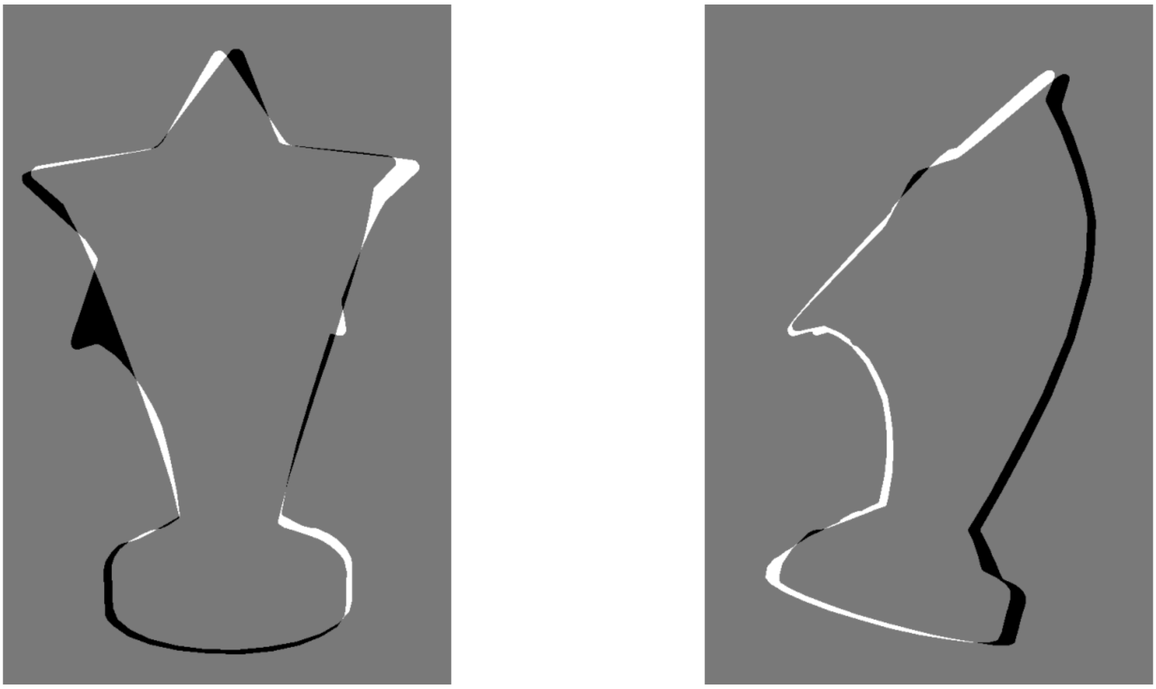
A.4 Effects of the simulated annealing on the Robust Hybrid Visual Servoing

The algorithm for the Robust Hybrid Visual Servoing method is summarised below:

Algorithm 12: Robust Hybrid method

```
Data: CAD file, Experimental projections $\mathbf{p}_{exp}$, Experimental setup, initial position
      $\mathbf{H_0}$, threshold $\epsilon$, step $\lambda$, timestep $\Delta t$
Result: $\mathbf{H}$
Compute projections $\mathbf{p}$ at $\mathbf{H_0}$
$\mathbf{H}$, $\mathbf{p} \leftarrow \mathbf{H_0}$, Proj$_{\mathbf{H_0}}$(CAD)
i = 0
$\mathbf{s}^*$ = contours points from $\mathbf{p}_{exp}$
$\mathbf{s}$ = contours points from $\mathbf{p}$
while visual error $> \epsilon$ do
    $\mathbf{v} = -\lambda \mathbf{L}_e^+(\mathbf{s} - \mathbf{s}^*)$
    $\mathbf{H} = \mathbf{H} + \mathbf{v} \cdot \Delta t$
    Recompute $\mathbf{p}$ and $\mathbf{s}$
    $\epsilon$ = $||\mathbf{s} - \mathbf{s}^*||$
    if i % 10 ==0 then
        Update $\mathbf{H}$ with simulated annealing
    end
    i = i +1
end
Choose moment features and compute the new $\mathbf{L}_e$
$\mathbf{s}^*$ = moments features from $\mathbf{p}_{exp}$
$\mathbf{s}$ = moments features from $\mathbf{p}$
while visual error $> \epsilon$ do
    $\mathbf{v} = -\lambda \mathbf{L}_e^+(\mathbf{s} - \mathbf{s}^*)$
    $\mathbf{H} = \mathbf{H} + \mathbf{v} \cdot \Delta t$
    Recompute $\mathbf{p}$ and $\mathbf{s}$
    $\epsilon$ = $||\mathbf{s} - \mathbf{s}^*||$
    if i % 10 ==0 then
        Update $\mathbf{H}$ with simulated annealing
    end
    i = i +1
end
return $\mathbf{H}$
```

To demonstrate the benefits of simulated annealing, we will show the result of a registration without this step. Figure A.2 shows the overlapping after registration. The results are not satisfactory, the algorithm has almost converged after the first half with points registration but then diverged with moments. The shown results are taken from the iteration with the best overlapping score. Table A.1 shows the Hausdorff distance and the accuracy with and without the simulated annealing. The simulated annealing is clearly essential to the method. This can also be seen in Figure A.3 and A.4.

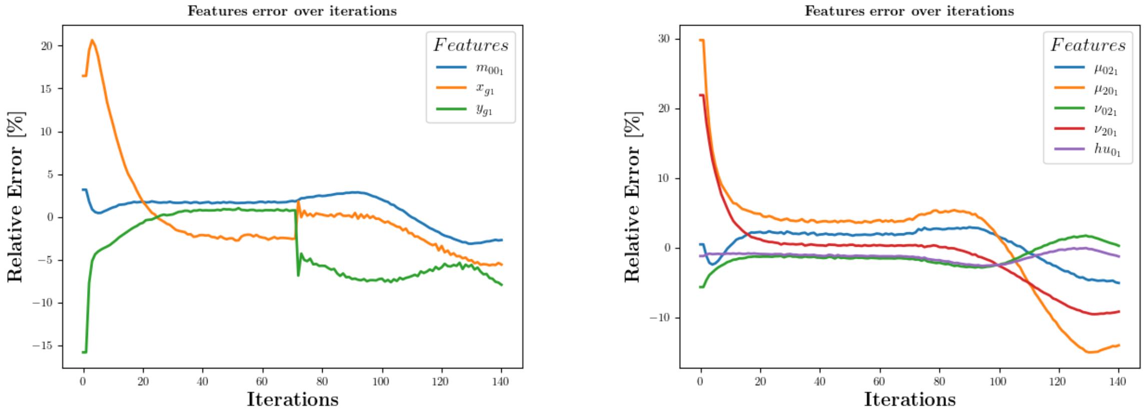


(a) Front view (b) Transverse view

Figure A.2: Illustration of the overlapping between binarised experimental images (left) and simulations (right) for the front and transverse views without simulated annealing.

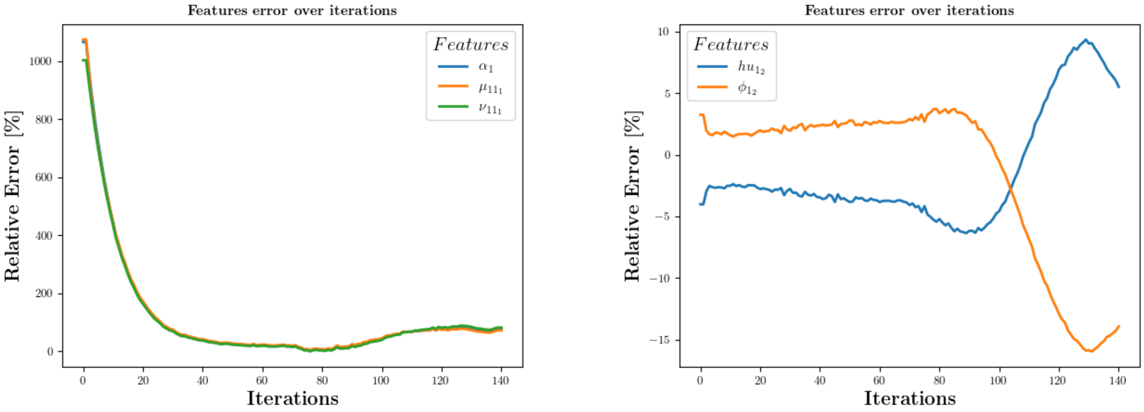
Table A.1: Quantitative registration results for the star trophy without simulated annealing

Metric	Front View	Transversal View
Hausdorff distance [pixel] Before registration After registration	 181.1 ~~2.000~~ →42.48	 218.8 ~~2.236~~ →21.58
Accuracy [%] Before registration After registration	 63.32 ~~99.64~~ → 96.49	 69.55 ~~99.67~~ → 96.22



(a) Low order moments

(b) Symmetric moments



(c) Antisymmetric moments

(d) High order moments

Figure A.3: Image moments convergence for the transverse view without simulated annealing

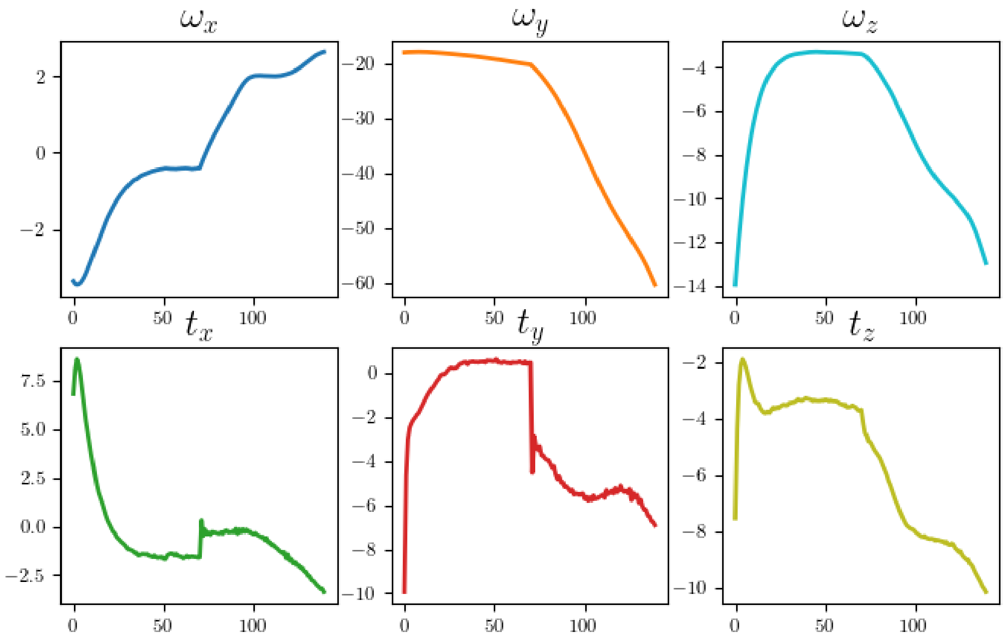


Figure A.4: Parameters convergence. First row represents the angle (ω_x, ω_y, ω_z) [°], second row represents the translations(t_x, t_y, t_z) [mm].

A.5 Matching Pursuit Algorithm

Algorithm 13: Matching Pursuit

```
Data:   Dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$
        Signal $\mathbf{x} \in \mathbb{R}^{m}$
        Sparsity level $\tau$
        Convergence criteria $\epsilon$
Result: Representation support $\mathcal{S}$
        Sparse reconstruction $\mathbf{z} \in \mathbb{R}^{n}$
Initialise $\mathcal{S} = \varnothing, \mathbf{R} = \mathbf{x}$
while $\|\mathbf{R}\|_2 < \epsilon$ and $\|\mathbf{z_s}\|_0 < \tau$ do
    Find $\mathbf{d_k} \in \mathbf{D}$ with maximum inner product $|\langle \mathbf{R}, \mathbf{d_k} \rangle|$
    Update support $\mathcal{S} \leftarrow \mathcal{S} \cup \{k\}$
    Update code $\mathbf{z}_k \leftarrow \langle \mathbf{R}, \mathbf{d_k} \rangle$
    Compute new residual $\mathbf{R} \leftarrow \mathbf{R}$ - $\langle \mathbf{R}, \mathbf{d_k} \rangle$
end
```

A.6 ADMM for basis pursuit

Algorithm 14: Scaled ADMM for linear system with ℓ_1 norm constraint

```
Data:   Initial guess $\mathbf{x}^0 \in \mathbb{R}^m$, $\rho$ parameter
          Signal $\mathbf{x} \in \mathbb{R}^m$
Result: Argmin $\mathbf{x} \in \mathbb{R}^m$
Initialize $\mathbf{z}^0 = \mathbf{x}^0, \mathbf{u}^k = \mathbf{0}$
while the convergence criteria have not been met do
    $\mathbf{x}^{k+1} = (A^TA + \rho I)^{-1}(A^Tb + \rho y - u)$
    $\mathbf{y}^{k+1} = \mathcal{T}_{\lambda/\rho}(x + y/\rho)$
    $\mathbf{u}^k = \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}$
end
```

Algorithm 15: Scaled ADMM for basis pursuit with ℓ_1 norm and constraint Ax=b

```
Data:   Initial guess $\mathbf{x}^0 \in \mathbb{R}^m$, $\rho$ parameter
        Signal $\mathbf{x} \in \mathbb{R}^m$
Result: Argmin $\mathbf{x} \in \mathbb{R}^m$
Initialize $\mathbf{z}^0 = \mathbf{x}^0, \mathbf{u}^k = \mathbf{0}$
while the convergence criteria have not been met do
    $\mathbf{x}^{k+1} = \mathcal{T}_{1/\rho}(y - u)$
    $\mathbf{y}^{k+1} = (I - A^T(A^TA)^{-1}A)(x + u) + A^T(AA^T)^{-1}b$
    $\mathbf{u}^k = \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}$
end
```

A.7 ADMM for CSC

Algorithm 16: ADMM CSC

```
Data: Initial dictionary $\{\mathbf{d}_j\}_{1\leq j\leq m}$ of $m$ filters
      K training images $\{\mathbf{x}_j\}_{1\leq j\leq K}$ of N voxels each
      Regularisation parameter $\lambda$
      Initial penalty parameter $\rho$
      Maximum iterations number $j_{max}$
Result: Dictionary $\{g_i\}$, coefficient maps $\{y_i\}$
Initialize $\forall(k,i)$: $\mathbf{y}_{k,i} = \mathbf{y}_{k,i}^{prev} = \mathbf{0}$,$\mathbf{u}_{k,i} = \mathbf{0}$
$\forall i$ $\mathbf{h}_i = \mathbf{0}$, $\mathbf{h}_i = \mathbf{0}$, $\mathbf{g}_i = \mathbf{g}_i^{prev} = \mathbf{d}_i\}$
$j = 1$
while $j < j_{max}$ do
    $\hat{\mathbf{g}}_i = FFT(\mathbf{g}_i)$ , $\forall i$
    $\hat{\mathbf{w}}_{k,i} = FFT(\mathbf{y}_{k,i} - \mathbf{u}_{k,i})$ , $\forall k, i$
    Solve $\underset{\hat{\mathbf{x}}_k}{\operatorname{argmin}} \frac{1}{2}\|\sum_{j=1}^{m}\hat{\mathbf{D}}_j\mathbf{z}_{k,j} - \hat{\mathbf{s}}_k\|_2^2 + \frac{\rho}{2}\sum_{j=1}^{m}\|\hat{\mathbf{z}}_{k,j} - \hat{\mathbf{w}}_{k,j}\|_2^2$, with
      Sherman-Morrison formula $\forall k$
    $\mathbf{z}_{k,i} = IFFT(\hat{\mathbf{z}}_{k,i})$ , $\forall i$
    $\mathbf{z}_{k,i}^{relax} = \alpha_z\mathbf{z}_{k,i} + (1-\alpha_z)\mathbf{y}_{k,i}$
    $\mathbf{y}_{k,i} = \mathcal{T}_{\lambda/\rho}(\mathbf{z}_{k,i}^{relax} + \mathbf{u}_{k,i})$, $\forall(k,i)$
    $\mathbf{u}_{k,i} = \mathbf{u}_{k,i} + \mathbf{z}_{k,i}^{relax} - \mathbf{y}_{k,i}^{relax}$, $\forall(k,i)$
    $\hat{\mathbf{y}}_{k,i} = FFT(\mathbf{y}_{k,i})$ , $\forall(k,i)$
    $\hat{\mathbf{w}}_i = FFT(\mathbf{g}_i - \mathbf{h}_i)$ , $\forall i$
    Solve $\underset{\hat{\mathbf{d}}}{\operatorname{argmin}} \frac{1}{2}\sum_{j=1}^{m}\|\hat{\mathbf{Z}}_j\mathbf{d}_j - \hat{\mathbf{x}}_k\Big\|_2^2 + \frac{\rho}{2}\sum_{j=1}^{m}\|\hat{\mathbf{z}}_j - \hat{\mathbf{w}}_j\|_2^2$, with Sherman-Morrison
      formula
    $\mathbf{d}_i = IFFT(\hat{\mathbf{d}}_i)$ , $\forall i$
    $\mathbf{d}_i^{relax} = \alpha_d\mathbf{d}_i + (1-\alpha_d)\mathbf{d}_i$ , $\forall i$
    $\mathbf{g}_i = \operatorname{prox}_{\iota_{C_{PN}}}(\mathbf{d}_i^{relax} + \mathbf{h}_i)$ , $\forall i$
    $\mathbf{h}_i = \mathbf{h}_i + \mathbf{d}_i^{relax} - \mathbf{z}_i$, $\forall i$
    $\mathbf{y}_i^{prev} = \mathbf{y}_i$, $\forall i$
    $\mathbf{g}_i^{prev} = \mathbf{g}_i$, $\forall i$
end
```

Appendix B

Résumé Etendu

B.1 Contexte

Cette thèse explore des sujets de recherche dans le domaine du contrôle non destructif industriel par rayons X (CND). Le CND est nécessaire puisqu'il permet d'évaluer les propriétés d'un matériau ou d'un système sans causer de dommages. Parmi les diverses méthodes de CND, on note le contrôle ultrasonique, la thermographie, les courants de Foucault, la radiographie et surtout la tomographie par rayons X (CT). Cette dernière se distingue comme un outil puissant pour caractériser ou localiser des défauts internes et vérifier la conformité géométrique d'un objet. L'application de la CT s'est considérablement étendue et son utilisation s'est intensifiée dans de nombreux secteurs industriels, notamment pour l'inspection de composants afin de détecter des défauts, mesurer des paramètres géométriques, et pour des applications d'ingénierie inversée. La CT peut être employée pour inspecter des composants pendant la fabrication, garantissant ainsi l'intégrité des produits. Bien que son utilisation principale fut longtemps l'imagerie médicale, l'inspection par CT est désormais bien établie dans le domaine industriel. Cependant, en raison des exigences croissantes et des contraintes sur les processus de contrôle, celle-ci se doit de constamment évoluer et s'adapter. Que ce soit en termes de qualité de reconstruction ou en temps d'inspection, la tomographie par rayons X est en constante progression, notamment dans ce qu'on appelle la stratégie de *vues éparses.* Cette stratégie consiste à reconstruire un objet en utilisant le minimum possible de projections radiologiques tout en maintenant une qualité de reconstruction satisfaisante. Cette approche réduit les temps d'acquisition et les coûts associés. La reconstruction en vues éparses constitue un véritable défi car le problème tomographique est mal conditionné, on le dit *mal posé.* De nombreuses techniques ont été développées pour surmonter cet obstacle, dont plusieurs sont basées sur l'utilisation d'informations *a priori* lors du processus de reconstruction. En exploitant les données et les connaissances disponibles avant l'expérience, il est possible d'améliorer le résultat de la reconstruction malgré le nombre réduit de projections.

Dans la majorité des cas, l'intégration d'informations *a priori* se fait via un terme de régularisation qui modélise des connaissances sur l'objet, souvent appris par apprentissage automatique à partir d'échantillons similaires à celui étudié. Cependant, l'utilisation d'informations *a priori* peut être étendue à de nombreuses autres étapes du processus tomographique. Qu'il s'agisse de l'expertise de l'expérimentateur dans le choix des paramètres optimaux de la machine, du positionnement précis des objets, de la trajectoire idéale, ou même de connaissances spécifiques sur le type de pièce inspectée : les informations *a priori* peuvent prendre différentes formes. Toutes ces connaissances peuvent être mises à profit pour améliorer la reconstruction en vues éparses.

Dans notre contexte industriel, par exemple, le modèle de conception assistée par ordinateur (CAO) de l'objet est souvent disponible, ce qui représente une information précieuse sur la géométrie de l'objet étudié. Néanmoins, il est important de noter que le modèle CAO ne fournit qu'une représentation approximative de l'objet. En CND ou en métrologie, ce sont précisément les différences entre un objet et son modèle CAO qui sont d'intérêt. Par conséquent, l'intégration d'informations *a priori* est complexe car ces informations sont souvent "approximatives" et ne peuvent pas être utilisées telles quelles. Ainsi, dans cette thèse, nous ne visons pas à incorporer des informations *a priori* obtenues à partir de vastes bases de données pour créer un terme de régularisation *a priori*. Nous proposons plutôt d'utiliser judicieusement les informations géométriques disponibles à partir du modèle CAO à chaque étape du processus. Nous ne proposons donc pas une méthode, mais une méthodologie pour l'intégration des informations géométriques *a priori* lors la reconstruction tomographique par rayons X.

Les différents chapitres de cette thèse proposent chacun des innovations pour chaque étape du processus tomographique. Mis bout à bout, ces étapes forment notre méthodologie optimisée. La méthodologie est résumée dans la figure B.1 ci-dessous. A travers les chapitres, nous montrerons à chaque fois des nouveaux résultats sur des données expérimentales d'une pièce en fabrication additive plastique. Cette pièce représente une figurine en forme de trophée étoile.

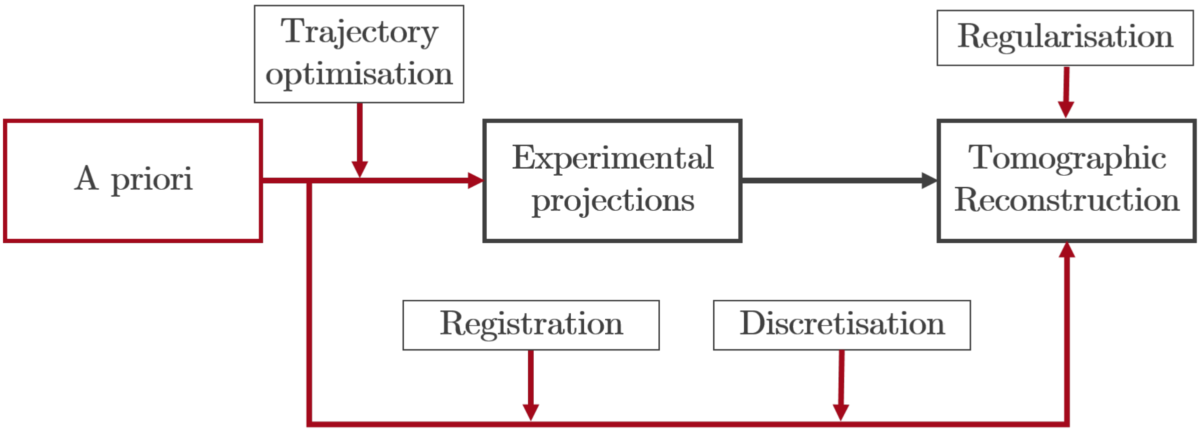


Figure B.1: Diagramme du processus tomographique. Les informations *a priori* sont utilisées à différentes étapes afin d'optimiser la qualité de la reconstruction en vue éparse.

Le chapitre 1 fournit une introduction générale à la tomographie, couvrant à la fois les composants physiques pour la génération et la détection des rayons X, ainsi que les composants mathématiques pour résoudre le problème tomographique.

Le chapitre 2 présente l'état de l'art des méthodes d'optimisation de trajectoire d'acquisition de projections ainsi que les méthodes que nous avons développées. Nos algorithmes, basés sur la méthode d'interpolation empirique (EIM), utilisent le modèle CAO comme information géométrique *a priori* pour définir les projections les plus pertinentes pour la reconstruction. Au fur et à mesure de l'avancement de nos recherches, nous avons proposé plusieurs variantes en fonction de l'algorithme. Les méthodes ont été testées sur de nombreux objets et certains résultats ont déjà été publiés.

Le chapitre 3 introduit les techniques de recalage 2D/3D. Le recalage est une étape préliminaire essentielle pour incorporer les informations *a priori*, car elles doivent souvent être positionnées dans le même référentiel que l'objet réel. Nous introduirons les concepts fondamentaux du recalage de modèles CAO sur des objets réels. Deux techniques de recalage ont été développées et sont présentées : le recalage hybride 3D/2D robuste par asservissement visuel et la correspondance itérative par perspective inverse sur enveloppe convexe. Chacune présente des avantages selon les besoins en qualité ou rapidité. La première méthode utilise l'asservissement visuel et les moments d'image, tandis que la seconde utilise un schéma itératif proche de l'ICP pour recaler le modèle CAO sur des projections expérimentales.

Le chapitre 4 propose d'utiliser nos informations géométriques *a priori* pour fournir une description mathématique adaptée à l'objet étudié. En utilisant le modèle CAO, il est possible d'adapter la discrétisation de l'objet pour réduire le nombre de variables et améliorer la résolution. Nous présenterons des résultats sur les reconstructions dites *masquées* ainsi que sur les structures hiérarchiques en arbres mêlées à des maillages.

Le chapitre 5 n'utilise pas d'informations géométriques *a priori* mais se concentre plutôt sur l'information de texture et l'apparence des objets reconstruits. Ce chapitre introduit les concepts fondamentaux du codage parcimonieux et de l'apprentissage de dictionnaire pour débruiter les reconstructions en vues éparses. Nous présentons une extension des méthodes classiques utilisant la convolution. Les résultats sur le débruitage, la régularisation et même la classification seront présentés.

Enfin, le chapitre 6 conclut nos travaux et présente les perspectives futures.

B.2 Contributions

Dans le cadre de mes travaux de recherche, j'ai eu la chance de pouvoir publier mes contributions.

- Dans un premier article, nous présentons des algorithmes d'optimisation des trajectoires d'acquisition basés sur l'interpolation empirique, exploitant les modèles CAO pour définir les projections les plus pertinentes, augmentant ainsi la qualité des reconstructions pour des objets complexes.
  V. Bussy, C. Vienne, and V. Kaftandjian. Fast algorithms based on empirical interpolation methods for selecting best projections in sparse-view x-ray computed tomography using a priori information. NDT&E International, 2022.

- Nous avons même pu complétrer nos travaux sur les trajectoires et rajouter une contrainte physique dans la méthode précédente
  V. Bussy, C. Vienne, J. Escoda, and V. Kaftandjian. Best projections selection algorithm based on constrained qdeim for sparse-views x-ray computed tomography. In 12th Conference on Industrial Computed Tomography (iCT) 2023, 27 February - 2 March 2023 in Fürth, Germany. e-Journal of Nondestructive Testing Vol. 28(3),2023.

- Par la suite, pour le recalage nous avons publié un article sur l'asservissement visuel en points et en moments d'images.
  Victor Bussy and Caroline Vienne. Robust 3D/2D hybrid registration for integrating a priori CAD model into X-ray Computed Tomography. In ORASIS 2021, Saint Ferréol, France, September 2021. Centre National de la Recherche Scientifique [CNRS].

- Cette dernière méthode de recalage a été complétée par un autre article pour un recalage rapide basée sur la correspondance en perspective inverse des points d'enveloppes convexes. Dans cet article, nous avons justement utilisé ce recalage afin d'introduire une notion de masque dans les reconstructions. Cette approche permet de réduire le nombre de variables et baisser considérablement les artefacts.
  Victor Bussy, Caroline Vienne, Julie Escoda, and Valérie Kaftandjian. Sparse-View X-Ray CT Reconstruction using CAD Model Registration. volume 2022 49th Annual Review of Progress in Quantitative Nondestructive Evaluation of Quantitative Nondestructive Evaluation, page V001T13A001, 07 2022.

- L'idée de réduire le nombre de variables a été poursuivie et nous avons chercher à utiliser des structures hiérarchiques en arbres mêlées à des maillages pour la reconstruction ce qui a fait l'objet d'un brevet.

- Nos travaux sur le codage parcimonieux convolutifs et ses applications en débruitage de volumes reconstruits en vues éparses ont été validés sur des pièces manufacturées et ont montré des résultats prometteurs, ouvrant la voie à des améliorations futures dans les domaines du contrôle non destructif et de la métrologie industrielle.
  Chuan Huang, Paul Vaska, Yongfeng Gao, Shaojie Chang, Thomas Wesley Holmes, Amir Pourmorteza, and Jerome Liang. Proceedings of the 17th international meeting on fully 3d image reconstruction in radiology and nuclear medicine, 2023.

- Nous avons déroulé notre méthodologie entière pour la reconstruction d'un objet métallique en fabrication additive. Cette approche a permis d'améliorer significativement la précision des reconstructions tout en réduisant les temps d'acquisition.
  Victor Bussy, Caroline Vienne, Julie Escoda, and Valerie Kaftandjian. Méthodologie optimisée pour la reconstruction tomographique avec ajout d'informations a priori pour l'inspection par rayons x. e-Journal of Nondestructive Testing, 28, 09 2023.

B.3 Introduction à la reconstruction tomographique par rayons X

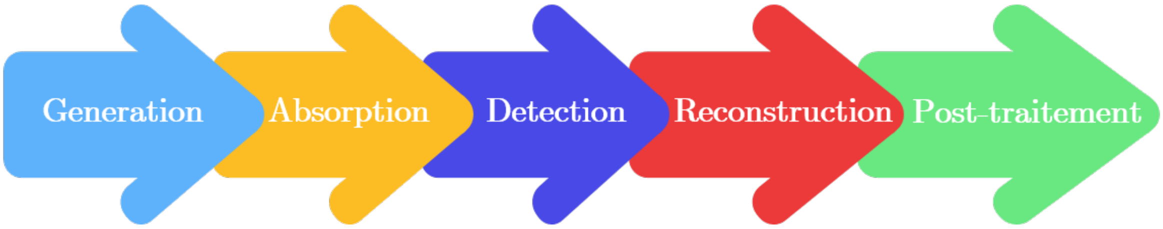


Figure B.2: Représentation schématique d'un processus de tomographie par rayons X

Pour introduire notre méthodologie optimisée pour la tomographie industrielle, il est primordial d'abord de comprendre l'ensemble du processus tomographique. Les rayons X sont générés par une source, absorbés par l'objet étudié puis détectés. Le signal sera ensuite traité pour reconstruire l'objet en 3 dimensions et inspecter l'intérieur de la matière. Le processus global est explicité dans le diagramme B.2. Les concepts abordés ici sont développés dans les références [1, 3, 15].

B.3.1 Génération des rayons X

Les rayons X ont été découverts par Wilhelm Conrad Röntgen, une découverte qui lui valut le prix Nobel en 1901. Leur capacité à pénétrer la matière organique permet de différencier nettement les tissus et les os sur les images radiographiques. Les rayons X ont une longueur d'onde variant entre 0,01 nm (rayons X durs) et 10 nm (rayons X mous), mais on les décrit plus souvent en termes d'énergie, allant de 100 eV à 1 MeV. Les dispositifs les plus courants pour générer des rayons X dans les domaines médical et industriel sont les tubes à rayons X, tels que le tube *Coolidge*. Ce dernier comporte une cathode et une anode ; la cathode est chauffée pour libérer des électrons par émission thermoïonique, et ces électrons sont ensuite accélérés vers l'anode sous l'effet d'une différence de potentiel. Lorsque les électrons entrent en collision avec l'anode, plusieurs interactions peuvent se produire :

- *Rayonnement de freinage (Bremsstrahlung)* : l'électron est dévié par le champ électromagnétique du noyau de l'atome et perd de l'énergie sous forme de rayonnement continu (voir Figure B.3a).
- *Radiation caractéristique* : l'électron incident éjecte un électron d'une couche interne de l'atome, ce qui libère de l'énergie sous forme de photons lorsque l'atome se réorganise (voir Figure B.3b).
- *Interaction directe électron-noyau* : l'électron transfère toute son énergie sous forme de photon, mais cette interaction est rare (voir Figure B.3c).

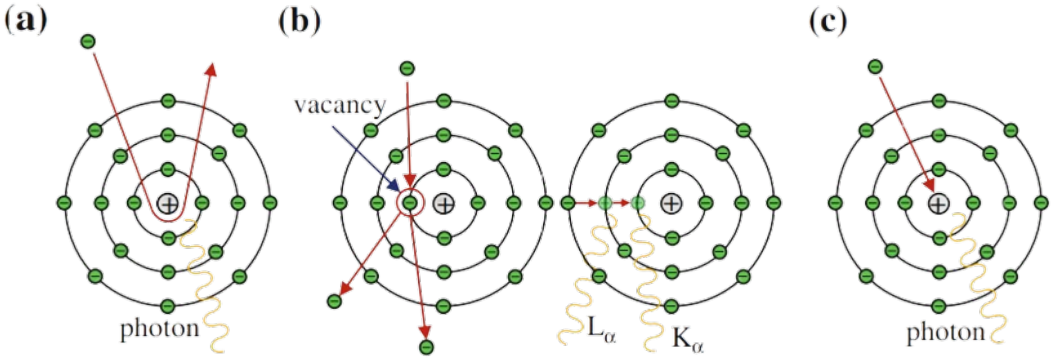


Figure B.3: Principe de génération des rayons X. Image adaptée de [1]. (a) Effet Bremsstrahlung. (b) Radiations caractéristiques. (c) Interaction directe électron/noyau.

B.3.2 Absorption des rayons X

Une fois générés, les rayons X pénètrent l'objet à étudier et subissent diverses interactions avec le matériau traversé, telles que l'effet photoélectrique, la diffusion Compton incohérente, la diffusion Rayleigh et la production de paires. L'effet photoélectrique prédomine à basse énergie, tandis que la diffusion Compton prend de l'importance à haute énergie. Ces interactions sont modélisées par le coefficient d'atténuation linéaire μ, mesuré en cm^{-1}, qui dépend de l'énergie du photon et du matériau traversé.

Le processus d'atténuation globale des rayons est décrit par la loi de Beer-Lambert, qui relie l'intensité transmise I à l'intensité initiale I_0 en fonction du coefficient d'atténuation μ et du chemin L parcouru par le faisceau. Si on note $\boldsymbol{\mu}_i$ l'atténuation du $i^{\text{ème}}$ pixel et ℓ_i la contribution de ce pixel, l'équation s'écrit :

$$\ln\left(\frac{I}{I_0}\right) = -\sum_i \boldsymbol{\mu}_i \ell_i. \tag{B.1}$$

B.3.3 Détection des rayons X

Une fois que le faisceau de rayons X a interagi avec l'objet, le rayonnement atténué est détecté par le système d'imagerie. Historiquement, les détecteurs de rayons X étaient basés sur la technologie des films photographiques. Cependant, ceux-ci ont été progressivement remplacés par des détecteurs numériques. Ces détecteurs de rayons X se répartissent en trois catégories principales : les détecteurs à ionisation gazeuse, les détecteurs à semi-conducteurs et les détecteurs à scintillation. Parmi eux, les détecteurs à scintillation prédominent dans les systèmes contemporains de tomographie industrielle. Ces détecteurs sont constitués d'un cristal scintillateur couplé à un photodétecteur. Lorsque les rayons X frappent la couche scintillante, ils sont convertis en photons de lumière visible. Cette lumière est ensuite absorbée par le photodétecteur, qui génère une charge électrique proportionnelle à l'énergie du rayonnement X [16, 18].

B.3.4 Reconstruction tomographique

Les méthodes de reconstruction tomographique peuvent être analytiques ou itératives. Parmi les méthodes analytiques, la transformation de Radon, qui représente l'intégrale de ligne de la fonction d'atténuation, est un concept clé. Elle est utilisée dans les algorithmes de reconstruction tels que la rétroprojection filtrée (FBP). Ces méthodes permettent de reconstituer l'image de l'objet à partir des projections radiographiques collectées sous différents angles.

Il existe de nombreux algorithmes itératifs pour résoudre le problème de la tomographie. Les plus connus sont les algorithmes de Landweber, souvent utilisés pour la reconstruction itérative des images en rayons X. Ils fonctionnent par descente de gradient en appliquant l'opérateur adjoint au résidu de la projection. Parmi ces méthodes, on retrouve l'*Algebraic Reconstruction Technique* (ART), mais aussi toutes ses variantes.

D'autres méthodes d'optimisation sont aussi couramment utilisées comme l'algorithme des gradients conjugués (CGLS) et l'algorithme *Maximum Likelihood Expectation Maximization* (MLEM). Mais récemment ce sont les méthodes proximales qui sont les plus populaires car elles permettent d'inclure du *Plug-and-Play*. L'*Alternating Direction Method* *of Multipliers* (ADMM) et l'algorithme *Primal-Dual Hybrid Gradient* (PDHG) sont les plus célèbres de ce genre.

B.3.5 Post-traitement de la reconstruction

Une fois la tomographie terminée, on peut directement passer au diagnostic et à l'étude de l'image reconstruite. Malheureusement, et particulièrement en vues éparses, lors de la reconstruction de nombreux bruits et différents types d'artefacts surviennent et réduisent la qualité de reconstruction et donc rendent difficile le diagnostic. A cause du durcissement de faisceaux, des effets de cône, du manque de projections ou d'un mauvais alignement de l'appareil de mesure, la qualité de reconstruction peut être compromise. Il faut alors quantifier la qualité de reconstruction. Il existe de nombreuses métriques comme l'erreur quadratique moyenne, pic du signal sur bruit, ratio signal-bruit, similarité de structure, etc. Ces dernières nécessitent une référence pour évaluer une image mais il existe d'autres méthodes lorsqu'aucune référence n'est disponible.

B.4 Optimisation de trajectoire en vues éparses

L'acquisition de données en imagerie par rayons X est souvent réalisée selon une trajectoire circulaire autour de l'objet, car celle-ci est facile à mettre en œuvre. Cependant, cette trajectoire ne respecte pas le critère de Tuy pour la complétude des données, qui exige que chaque plan contenant un point d'intérêt coupe la trajectoire de la source de manière non tangentielle. Cela a conduit à l'exploration de trajectoires alternatives, notamment les trajectoires hélicoïdales, utilisées en imagerie médicale et industrielle.

Le premier concept clé est l'importance des rayons tangents pour bien reconstruire les contours et capturer les discontinuités dans les images. Ce concept a été détourné selon plusieurs critères, comme : la richesse spectrale, l'entropie, la densité fréquentielle ou les *wavelets*, etc. Un critère de plus en plus populaire est l'indice de détectabilité, qui évalue la performance d'un ensemble de projections pour une tâche spécifique en utilisant des observateurs modèles.

Enfin, des méthodes inspirées par les méthodes de modèles d'ordre réduit (ROM) ont été explorées pour réduire le nombre de projections tout en maintenant une bonne qualité de reconstruction.

L'une de ces méthodes est l'*Empirical Interpolation Method* (EIM), utilisée pour l'échantillonnage parcimonieux et l'approximation d'opérateurs non-linéaires coûteux en calcul dans les équations différentielles. Elle est utilisée dans divers domaines comme les éléments finis, en particulier pour les systèmes complexes et de grande dimension, car son coût de calcul dépend du rang de la décomposition en bases orthogonales propres (POD), une technique couramment utilisée pour la réduction de dimensionnalité.

L'algorithme DEIM, pour *Discrete-EIM*, utilise la POD d'une matrice de "snapshots" pour sélectionner les emplacements d'échantillonnage les plus pertinents. L'intérêt pour les méthodes d'interpolation empirique a suscité de nombreuses études, donnant lieu à plusieurs variantes, comme l'algorithme Q-DEIM, qui utilise la décomposition QR pivotée pour sélectionner les meilleurs emplacements d'échantillonnage. Une variante du Q-DEIM applique d'abord une POD sur les snapshots, puis effectue une décomposition QR pour sélectionner les points d'échantillonnage à l'aide du pivot.

Les méthodes EIM sont prometteuses pour la sélection de projections, mais elles reposent uniquement sur des raisonnements algébriques et ne prennent pas en compte certains aspects physiques comme le bruit, l'absorption ou les artefacts. Cela peut poser des problèmes dans certains cas, car la vue la plus informative peut également être celle où le bruit est le plus important. Pour améliorer les résultats du Q-DEIM, il est proposé d'intégrer les aspects physiques des rayons X en imposant une contrainte sur le pivot de la décomposition QR, en attribuant un poids à chaque projection potentielle en fonction de la tâche spécifique (bruit, stabilité du bras robotique, etc.).

Des résultats simulés concernant ces méthodologies ont été publiés, notamment dans le cadre de l'application sur un modèle CAO d'une grille métallique fabriquée par fabrication additive. Les résultats ont montré que l'optimisation des vues améliore la reconstruction des détails, notamment avec une trajectoire sphérique plutôt que circulaire. La variante avec contraintes a également été testée, intégrant des paramètres physiques comme l'atténuation, l'espacement des projections et le bruit ; et a montré des résultats intéressants. Les trajectoires sphériques, avec une magnification constante, ont montré un plus grand potentiel d'amélioration, surtout avec des comparaisons de projections plus homogènes.

Dans cette section, nous avons présenté des algorithmes de sélection de trajectoires et montré les résultats de nos propres algorithmes. Les méthodes que nous avons proposées sont rapides et faciles à mettre en œuvre, car elles ne nécessitent aucun paramètre à régler ni de connaissance approfondie de l'objet étudié. Elles nous permettent d'améliorer la qualité de la reconstruction en utilisant une stratégie de vue éparse. Pour les développements futurs, nous visons à comparer nos méthodes avec les techniques de pointe et à développer de nouvelles fonctionnalités dans l'optimisation des trajectoires. Incorporer la notion de région d'intérêt dans la sélection des vues serait particulièrement bénéfique. Si une zone spécifique est sujette à des défauts ou à des caractéristiques nécessitant un diagnostic, améliorer la qualité de la reconstruction dans cette région devient crucial. Des développements supplémentaires sont prévus pour explorer des trajectoires au-delà de celles avec une amplification constante. De plus, il serait intéressant d'optimiser l'orientation du détecteur.

B.5 Recalage

Cette thèse vise à intégrer des informations *a priori* du modèle CAO vers l'objet réel. Pour se faire, il faut impérativement faire correspondre les référentiels de ces derniers. Disposer d'une CAO recalée permet non seulement de comparer l'objet à son modèle, mais également de simuler l'environnement expérimental, les projections, de corriger la position de la source de rayons X.

Pour aligner le modèle CAO avec l'objet réel, plusieurs stratégies sont possibles. L'option la plus couramment utilisée est le recalage 3D/3D : l'objet est reconstruit, puis le volume est recalé avec le modèle CAO. Cette procédure est fréquemment rencontrée dans les tests non destructifs, car elle permet de comparer directement la reconstruction avec le modèle CAO pour estimer l'état de surface. Une autre option, appelée recalage

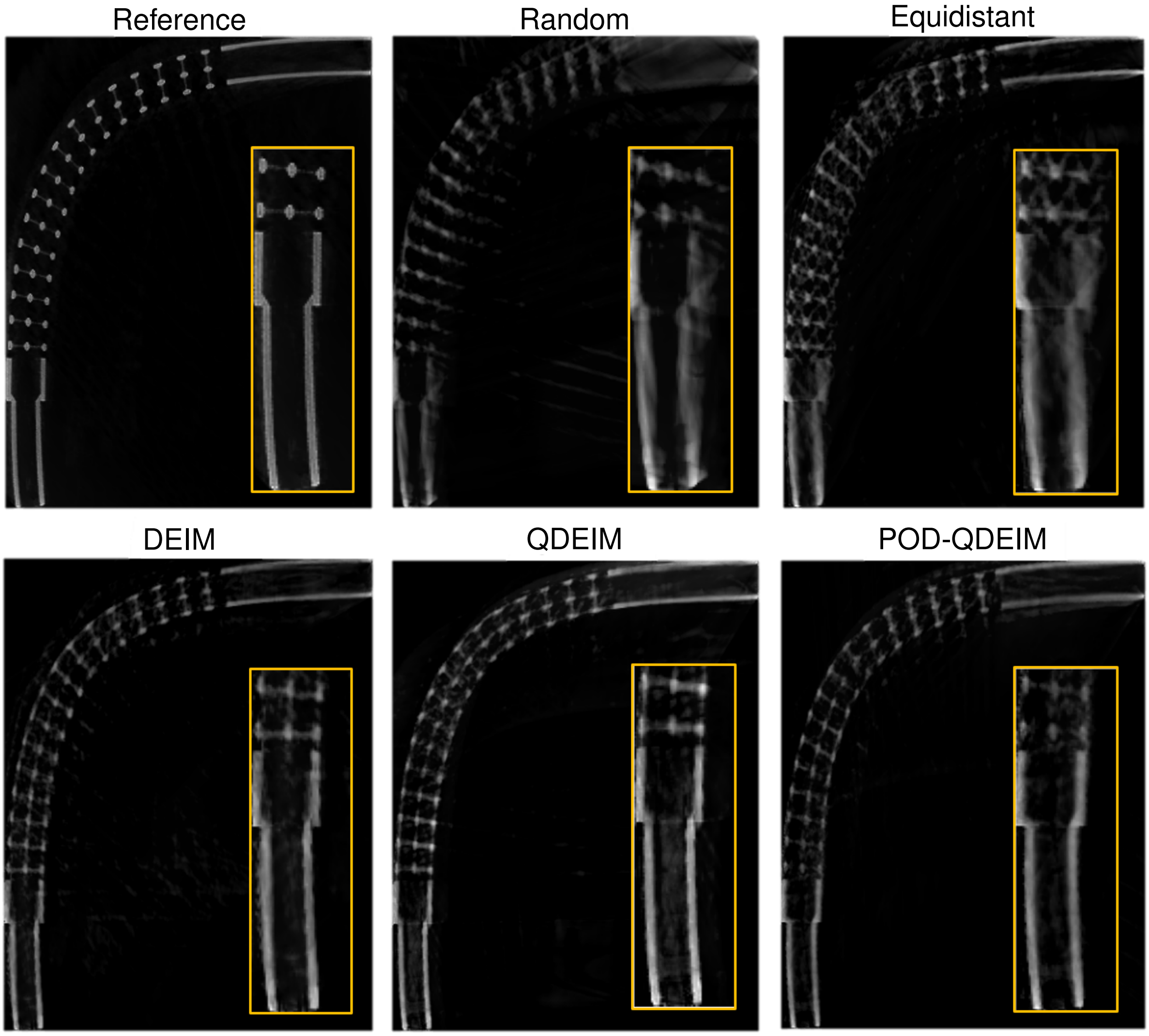
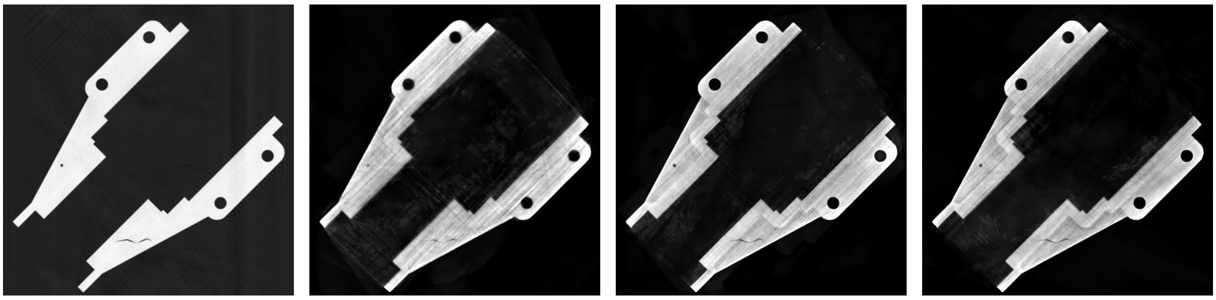


Figure B.4: Coupe transversale d'une grille reconstruit avec 10 projections selon une trajectoire sphérique utilisant différentes stratégies d'acquisition. Image tirée de [4]. L'optimisation de la trajectoire améliore considérablement la qualité de la reconstruction, en particulier dans le cas des trajectoires non planes.



(a) Reference (b) Uniform Sampling (c) QDEIM (d) Constrained QDEIM

Figure B.5: Comparaison d'une coupe transversale reconstruite pour différentes stratégies d'échantillonnage à 40 projections. Image tirée de [9].

3D/2D, consiste à aligner le modèle CAO à partir de quelques projections expérimentales. Le recalage 3D/2D ne nécessite pas de reconstruction tomographique. C'est cette dernière stratégie que nous avons choisi d'étudier dans nos travaux.

L'objectif du recalage est d'aligner le modèle CAO avec l'objet physique dans un système de coordonnées commun. Cela nécessite de déterminer une transformation entre le système de coordonnées CAD et le système de coordonnées *world*. Étant donné que la plupart des pièces examinées dans cette étude sont non déformables, nous nous concentrerons sur les transformations rigides. Les transformations rigides impliquent seulement six paramètres, représentant une translation 3D et des rotations autour de chaque axe.

Le recalage peut s'appuyer sur des marqueurs fiduciaires ou des cadres stéréotactiques. On parle alors de recalage *extrinsèque* basé sur des marqueurs est plus facile. En revanche, pour le recalage *intrinsèque*, les caractéristiques naturelles des images sont utilisées. Cela nécessite de trouver des primitives particulières dans les projections et l'objet 3D et de les faire correspondre. Les possibles primitives ne se basent pas uniquement sur l'intensité de l'image, elles peuvent être de toute nature. Dans cette thèse, nous avons choisi de ne pas ajouter de marqueurs à nos images. Toutes les techniques employées utiliseront exclusivement des primitives naturelles.

B.5.1 Recalage par Asservissement visuel

Dans cette thèse, nous avons développé deux méthodes de recalage. La première est basée sur l'asservissement visuel. Pour illustrer le fonctionnement de l'asservissement visuel, nous montrerons un exemple où les primitives utilisées pour le recalage sont des points. Comme nous comparerons des images expérimentales avec des images simulées, il s'agit d'un exemple dit *image-based*. Nous noterons d la distance source-détecteur, $\mathbf{x}$ les coordonnées d'un point sur le détecteur et $\mathbf{X}$ le point correspondant sur l'objet 3D. La projection perspective donne l'équation :

$$\mathbf{x} = \frac{d}{Z}\mathbf{X}. \tag{B.2}$$

Notez que la valeur de Z est approximative ; elle n'est pas supposée être connue au début de l'algorithme. De plus, la vitesse du point 3D peut être exprimée en fonction de la vitesse spatiale de la caméra $\mathbf{v} = (\mathbf{v}_c, \boldsymbol{\omega}_c)$ comme suit :

$$\dot{\mathbf{X}} = -\mathbf{v}_c - \boldsymbol{\omega}_c \times \mathbf{X}. \tag{B.3}$$

En utilisant les équations B.2 et B.3, nous obtenons :

$$\dot{x} = -\frac{d}{Z}v_x + \frac{x}{Z}v_z + \frac{xy}{d}\omega_x - d\left(1 + \frac{x^2}{d^2}\right)\omega_y + y\omega_z, \tag{B.4}$$

$$\dot{y} = -\frac{d}{Z}v_y + \frac{y}{Z}v_z + d\left(1 + \frac{y^2}{d^2}\right)\omega_x - \frac{xy}{d}\omega_y - x\omega_z. \tag{B.5}$$

En réarrangeant les termes, on obtient le résultat souhaité $\dot{\mathbf{x}} = \mathbf{L_e}\mathbf{v}$, avec :

$$\mathbf{L_e} = \begin{pmatrix} -\frac{d}{Z} & 0 & \frac{x}{Z} & \frac{xy}{d} & -\left(1 + \frac{x^2}{d^2}\right)d & y \\ 0 & -\frac{d}{Z} & \frac{y}{Z} & \left(1 + \frac{y^2}{d^2}\right)d & -\frac{xy}{d} & -x \end{pmatrix}. \tag{B.6}$$

La matrice d'interaction $\mathbf{L_e}$ dépend de la position de l'objet, un facteur qui reste inconnu. Par conséquent, une approximation, notée $\widehat{\mathbf{L_e}}$, est utilisée pour pallier cette limitation. La matrice d'interaction peut également être approximée par $\mathbf{L_{e^*}}$, mais ce choix impacte la trajectoire.

L'aspect critique de la commande visuelle basée sur l'image réside dans la sélection des bonnes primitives pour le recalage. Ici, nous ne considérerons pas l'intégralité de la projection, mais nous nous concentrerons uniquement sur ses contours. Nous avons observé que quelques points bien choisis fournissent suffisamment d'informations pour un recalage précis. Bien que l'image complète contienne une quantité importante d'informations, nous nous limitons à nouveau à quelques points pour assurer la robustesse de l'algorithme.

De plus, comme nous nous limitons aux contours, nous avons binariser les projections. La binarisation des projections s'avère avantageuse car elle rend le processus indépendant des niveaux de gris. Par conséquent, la connaissance précise de l'énergie de la source ou de la durée d'exposition n'est plus nécessaire. Cela simplifie considérablement la simulation de projections. Il suffit uniquement de détecter les intersections entre les rayons et la CAO. De plus, pour la rasterisation, le fichier CAO est d'abord décimé pour minimiser le nombre d'éléments [112]. Cette simplification délibérée accélère efficacement l'ensemble du processus.

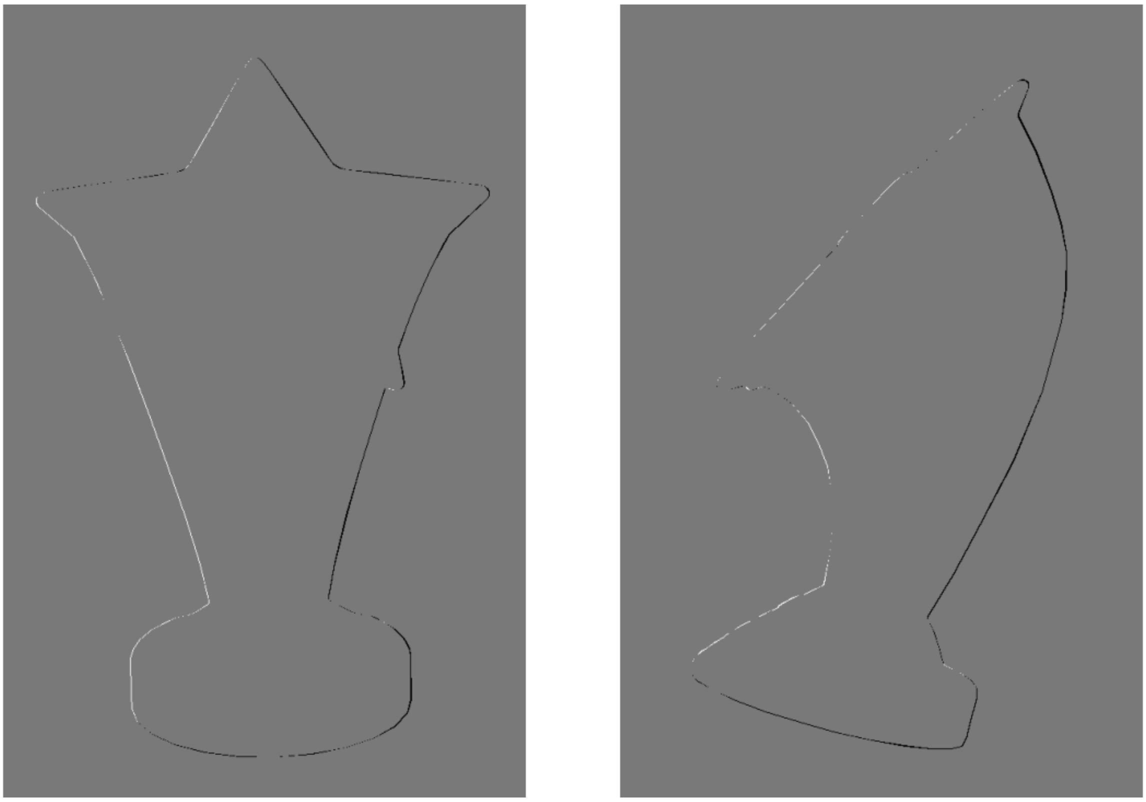
Ayant opté pour des images binaires, la prochaine étape consiste à sélectionner des primitives pour le processus de recalage. La décision a été prise d'utiliser exclusivement les points formant les contours. Ils existent toujours, sont faciles à détecter et sont en nombre suffisant.

Les points des contours des images expérimentales seront notés $\mathbf{s}^*$, tandis que ceux des projections simulées seront représentés par $\mathbf{s}$. L'étape de correspondance sera effectuée en utilisant l'algorithme ICP. Chaque $\mathbf{s}^*$ se verra attribuer un équivalent dans l'image simulée. Le déplacement du robot $\mathbf{v}$ est donnée selon :

$$\mathbf{v} = -\lambda \widehat{\mathbf{L_e}}^{-1}(\mathbf{s}^* - \mathbf{s}). \tag{B.7}$$

En pratique, nous avons observé une dépendance significative de la méthode à la qualité de l'étape de correspondance ICP. La rotation autour de l'axe vertical O_y présente un défi particulier en raison de son caractère inobservable dans une vue frontale directe lors d'une trajectoire circulaire. Pour remédier à ce problème, une solution a été proposée impliquant un recuit simulé pour la rotation autour de l'axe vertical O_y. Le recuit simulé est généralement utilisé pour contourner les minima locaux. Dans notre méthode, il a le double effet d'éviter les minima locaux et d'améliorer la robustesse de la méthode.

Avec l'introduction de la stratégie de recuit simulé, la méthode se montre plus robuste et donne de meilleurs résultats, bien que le processus de correspondance puisse encore présenter des erreurs. Une autre solution à ce problème a été proposée : les moments d'images. L'asservissement visuel par moments d'images n'impliquent pas d'étape de correspondances. Les primitives sont globales, plutôt que locales [117]. Les moments d'image servent de descripteurs mathématiques dans l'analyse d'images et la vision par ordinateur, fournissant une compréhension approfondie des propriétés intrinsèques d'une image. Ils jouent un rôle fondamental dans la caractérisation d'aspects cruciaux tels que la forme, l'orientation, la taille et la distribution d'intensité. La méthode a été testée sur des données simulées dans [12] et réelles dans cette thèse. Les résultats sont satisfaisants en



(a) Front view final overlap

(b) Transverse view final overlap

Figure B.6: Supoerposition finale après 70 itérations en points puis 70 itérations en moments.

termes de qualité, démontrant clairement les avantages du recuit simulé. Le résultat final de la superposition des images expérimentales et simulées est présentée dans la figure B.6. La cible est représentée en noir, le modèle CAO est en blanc et le chevauchement des deux est montré en gris. Le recalage est très satisfaisant. Les imperfections observées peuvent être attribuées aux différences entre le modèle 3D et l'objet réel, ainsi qu'aux erreurs de position de la source entre les deux projections.

B.5.2 Iterative Inverse Perspective Matching

L'appariement par perspective inverse a été initialement introduit par Wunsch et al. [113]. Cette méthode s'apparente à l'algorithme ICP, mais avec une approche distincte impliquant l'établissement de correspondances entre des lignes et des points. Le concept fondamental de l'appariement par perspective inverse est d'établir des correspondances entre des primitives 2D de l'image sur le détecteur et des primitives 3D du modèle CAO. Cependant, ces correspondances sont établies dans le domaine spatial 3D, plutôt que dans l'espace image 2D. On parle ici de recalage *pose-based*.

Pour chaque point caractéristique de la projection expérimentale, un rayon est rétro-projeté jusqu'à la source rayons X. Le processus d'appariement consiste à associer chaque point 3D du modèle CAO à son rayon le plus proche, et par conséquent à une caractéristique d'image. Chaque point du modèle CAO est ensuite projeté orthogonalement sur son rayon correspondant, créant ainsi des paires de points. Ce processus garantit qu'à chaque point du modèle CAO correspond un point 3D, transformant ainsi le problème initial en un recalage 3D/3D, généralement plus simple que le recalage 3D/2D.

De plus, afin d'améliorer la robustesse de la méthode, de réduire l'impact des valeurs aberrantes et d'assurer une stabilité numérique, la distance utilisée est la fonction M-estimateur de Huber. Le recalage peut être réalisé en utilisant l'algorithme de Kabsch-Umeyama [100], qui produit la matrice de rotation $\mathbf{R}$ et le vecteur de translation $\mathbf{t}$ minimisant la distance entre le modèle et les points rétro-projetés. Le schéma de contrôle et la mise à jour de la position sont relativement simples et efficaces sur le plan computationnel.

Cependant, la convergence pratique devient incertaine lorsque l'appariement par perspective itérative est appliqué tel quel. L'étape cruciale de l'établissement des correspondances influence fortement la précision des calculs de déplacement. Des associations incorrectes peuvent entraîner des mouvements erronés. Nous avons donc décidé d'utiliser deux projections pour améliorer la robustesse de la méthode, une stratégie reconnue pour son efficacité dans l'amélioration de la convergence [113].

De plus, notre implémentation se distingue par le fait que nous nous concentrons uniquement sur l'enveloppe convexe du modèle CAO 3D et l'enveloppe convexe de la projection expérimentale. Ces points sont particulièrement importants dans notre contexte. D'une part, ils sont relativement rares pour les pièces industrielles. D'autre part, afin de réduire encore le nombre de points, les points de l'enveloppe 3D sont eux-mêmes filtrés. Typiquement, de nombreux points de l'enveloppe 3D ne sont pas responsables de l'enveloppe convexe de la projection. Ainsi, l'enveloppe convexe 3D est projetée sur le détecteur à chaque itération. Seuls les points contribuant aux enveloppes convexes des projections sont conservés. Ce processus élimine les points incapables de générer des correspondances, améliorant la stabilité de la méthode en réduisant les appariements erronés, qui pourraient rallonger la phase de minimisation en raison d'erreurs importantes.

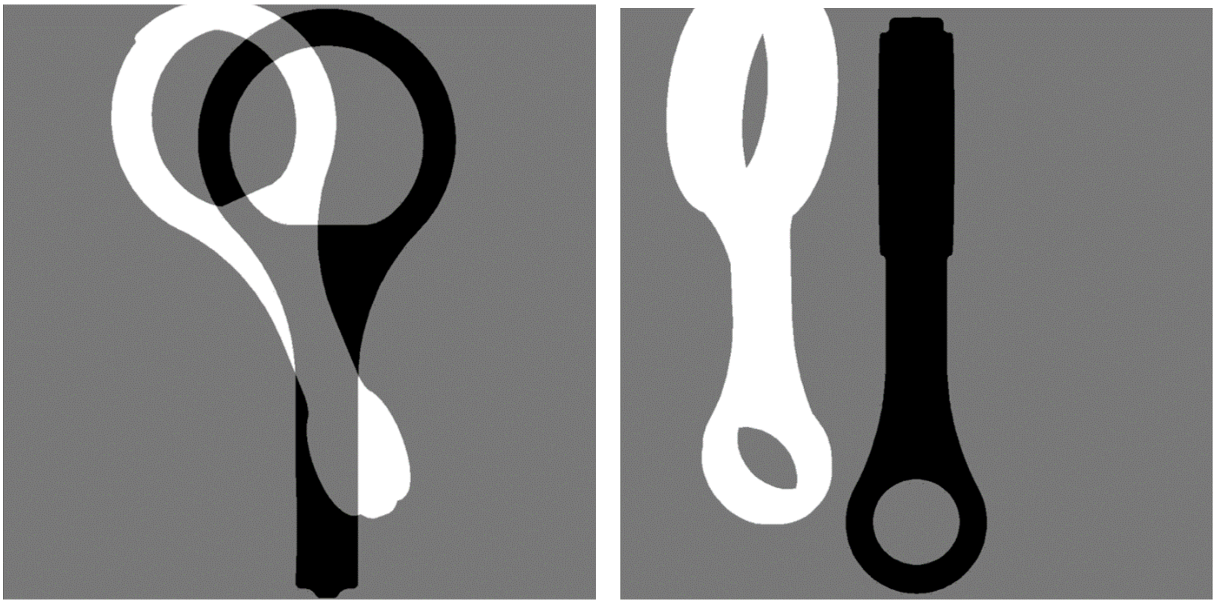
La méthode a d'abord été validée sur des données simulées et les résultats correspondants ont été présentés dans [13], visibles dans la Fig. B.7. Les résultats sont extrêmement encourageants, certains cas montrant des écarts d'environ un pixel sur le détecteur. En utilisant deux projections perpendiculaires, nous avons réussi à aligner presque parfaitement la bielle. Cet alignement a même été utilisé pour appliquer un masque lors de la reconstruction.

La méthode a également été testée sur une pièce réelle avec des projections expérimentales dans [11], où la pièce étudiée était très différente de son modèle 3D. La figure B.8 montre les superpositions initiales et finales après 50 itérations pour une pièce expérimentale recalée sur deux projections perpendiculaires. La cible est représentée en noir, tandis que la projection du modèle CAO est en blanc. La superposition des deux est en gris. L'objectif est de maximiser la zone grise. La première ligne représente la superposition initiale, et la seconde montre la superposition après le recalage. La vue transverse montre un alignement très précis, avec seulement une petite portion des contours non superposée. Le tableau B.1 présente la distance de Hausdorff en pixels et la précision du recalage.

B.6 Reconstruction sur un masque

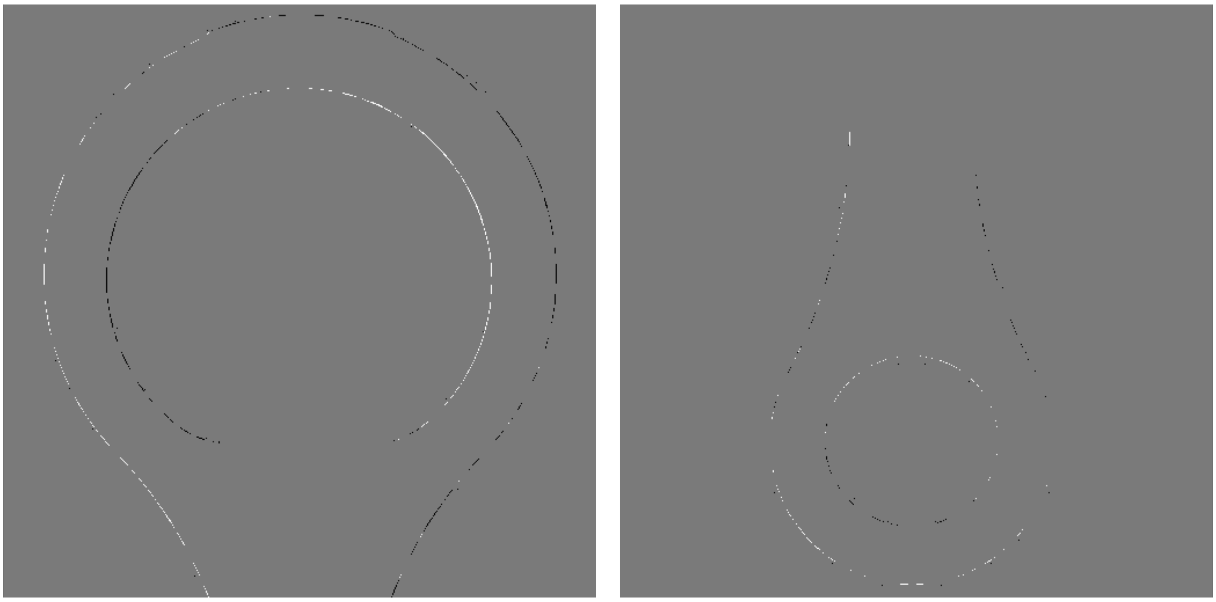
Dans l'étape précédente, nous avons aligné notre modèle CAO. Nous pouvons désormais incorporer directement nos *a priori* et passer à la phase de reconstruction. Traditionnellement, cette reconstruction est effectuée sur une grille de voxels, mais d'autres méthodes de discrétisation de l'espace existent.

Les méthodes de reconstruction, qu'elles soient analytiques ou itératives, nécessitent une discrétisation de l'espace. Bien que plusieurs types de discrétisations existent, la plus couramment utilisée est la fonction indicatrice sur une grille cartésienne régulière, ce qui donne les pixels/voxels habituels. Cependant, cette approche, bien que simple, consomme beaucoup de mémoire et ne permet pas de représenter efficacement des objets curvilignes, ce qui peut entraîner des artefacts d'aliasing. Nous avons d'abord conservé l'approche conventionnelle basée sur les voxels, mais en l'optimisant grâce à l'intégration d'un masque. Plutôt que de modifier la base de représentation, nous avons ajouté une



(a) Vue frontale initiale

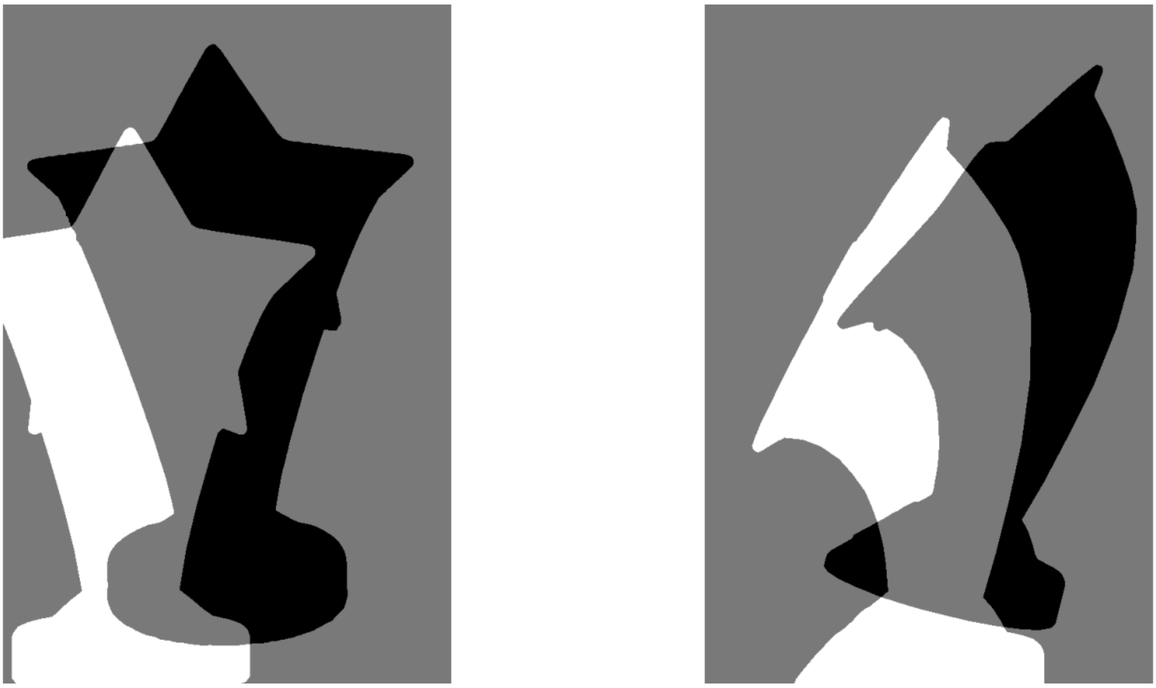
(b) Vue transverse initiale



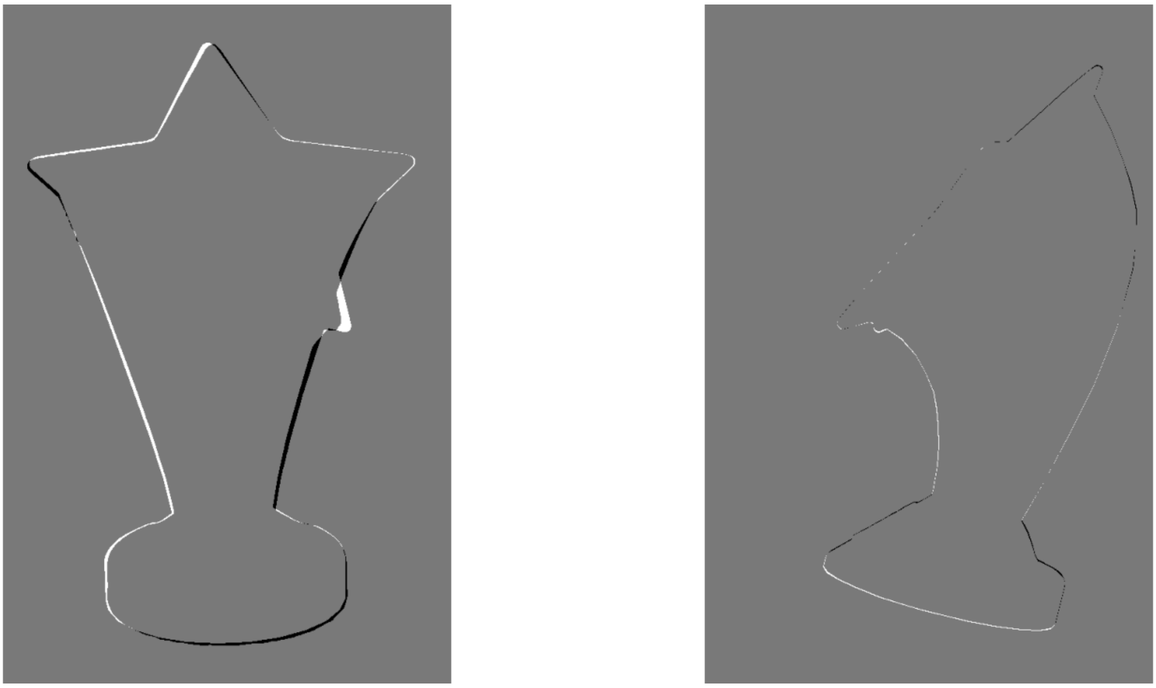
(c) Vue frontale finale (zoom)

(d) Vue transverse finale (zoom)

Figure B.7: Recalage de la bielle. Les zones bien recalées sont en gris, les projections cibles sont en noir et les projections du modèle CAD en blanc.



(a) Vue frontale superposition initiale (b) Vue transverse superposition initiale



(c) Vue frontale superposition finale (d) Vue transverse superposition finale

Figure B.8: Superpositions initiales et finales.

contrainte : seuls les voxels appartenant au "masque" seront pris en compte dans la reconstruction. Un masque est une matrice booléenne où les valeurs *True/1* représentent les voxels à reconstruire, tandis que les autres sont ignorés. Cela réduit le nombre de variables à résoudre dans le système tomographique, ce qui permet une reconstruction de qualité supérieure avec un nombre réduit de projections. Cette stratégie permet de réduire les artefacts, d'améliorer la précision dans les zones d'intérêt et d'accélérer le processus de reconstruction.

Table B.1: Résultats quantitatifs du recalage pour le trophée étoilé.

Métrique	Vue frontale	Vue transverse
Distance de Hausdorff [pixel]		
Avant recalage	181.1	218.8
Après recalage	13.15	4.000
Précision [%]		
Avant recalage	63.32	69.55
Après recalage	97.77	99.29

B.6.1 Reconstruction masquée

Pour incorporer le masque, il est nécessaire de modifier les opérateurs de projection et rétroprojection. La projection reste simple : on applique le masque binaire au volume avant d'utiliser les outils conventionnels. Cependant, la rétroprojection est plus complexe car elle dépend de la longueur du rayon traversant le masque.

Il existe deux approches : l'approche "négative", qui exclut les voxels hors du masque, et l'approche "positive", qui rétroprojette les valeurs sur tout le volume, puis applique le masque après coup. Les deux méthodes ajustent les valeurs rétroprojetées pour correspondre à la longueur parcourue dans le masque.

Les deux stratégies sont décrites par l'équation :

$$\frac{\mathbf{r}_i}{\|\tilde{\mathbf{A}}_i\|^2} = \frac{\tilde{\mathbf{r}}_i}{\|\mathbf{A}_i\|^2},$$

où $\mathbf{r}_i$ est le résidu à rétroprojetter, $\|\mathbf{A}_i\|$ est la norme de la matrice de projection, et le tilde désigne les opérateurs masqués. Des valeurs incohérentes peuvent apparaître aux bords du masque, et ces points non physiques doivent être retirés avant la rétroprojection.

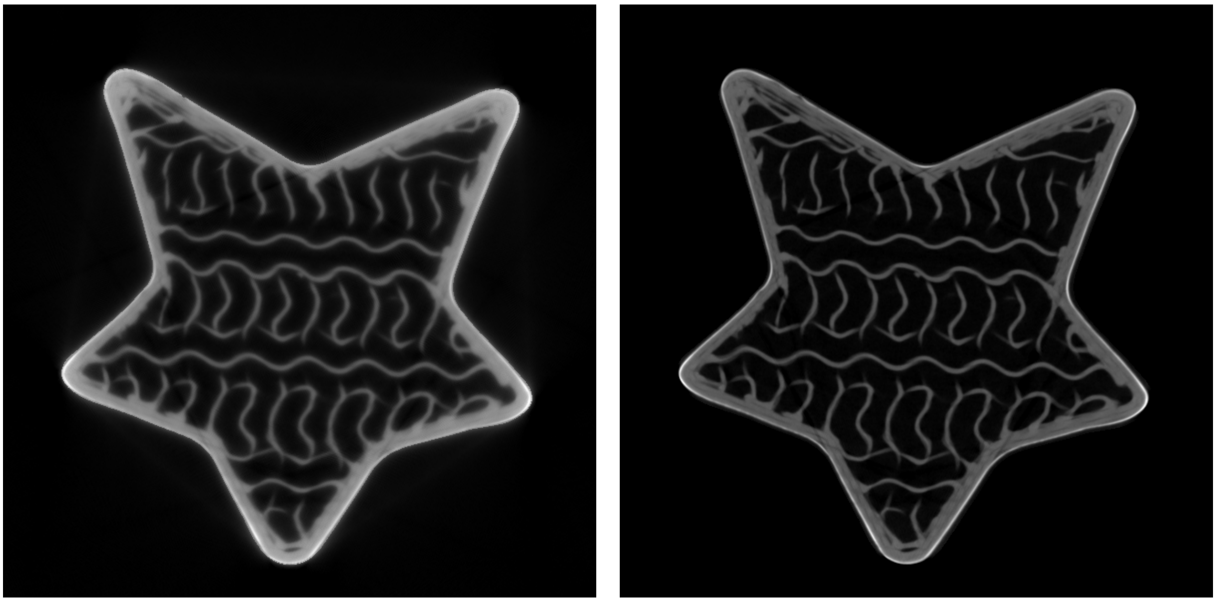
B.6.2 Résultats expérimentaux

L'efficacité de la reconstruction masquée a été validée sur des données simulées et réelles. Dans nos travaux précédents [13], l'utilisation du masque a permis d'améliorer la probabilité de détection d'un défaut de 3,576% à 16,66%. Nous avons également testé la reconstruction masquée sur un objet expérimental, en utilisant une version masquée du SIRT sur 150 itérations avec un volume de 1772 × 1204 × 1204 voxels.

Figure B.9 compare les reconstructions avec et sans masque pour 100 et 928 projections. Le masque améliore significativement la clarté des images, réduisant le flou et révélant des détails plus fins, comme les imperfections liées à la fabrication additive. Bien que certaines lignes discrètes subsistent dans les reconstructions masquées, les artefacts sont nettement moins marqués que dans les méthodes non masquées.

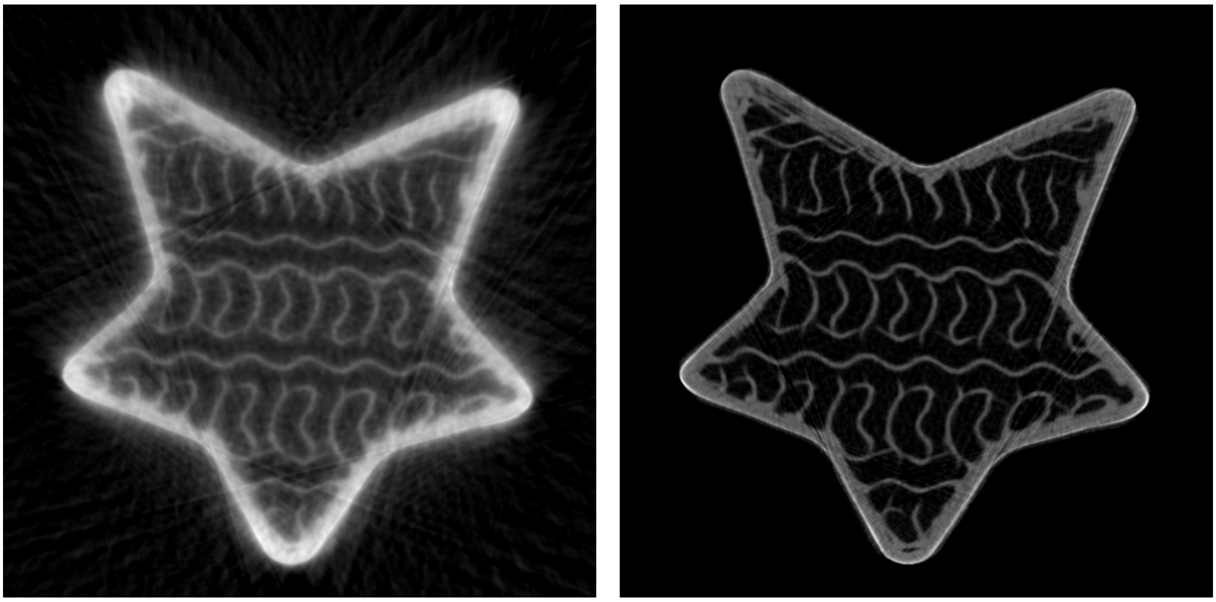
B.7 Reconstruction sur arbres et maillages

Dans notre recherche, nous avons cherché à explorer un cadre plus polyvalent et adaptable que celui basé sur les voxels. Les maillages se sont rapidement imposés comme une option



(a) Sans masque, 928 projections

(b) Masqué, 928 projections



(c) Sans masque, 100 projections

(d) Masqué, 100 projections

Figure B.9: Comparaison de la reconstructions SIRT avec et sans masque pour 928 et 100 projections.

attrayante en raison de leur flexibilité. Cependant, les études antérieures ont montré que les résultats obtenus avec les maillages sont souvent limités, principalement à cause de la lenteur de convergence due à l'intensité des calculs liés aux opérateurs de projection et rétroprojection. En parallèle, les structures arborescentes, comme les octrees, sont déjà optimisées pour le lancer de rayons grâce à des algorithmes rapides développés pour l'infographie et la vision par ordinateur. Notre objectif est de combiner les forces de ces deux structures afin d'optimiser notre processus de reconstruction. En intégrant un maillage dans une structure hiérarchique, nous escomptons bénéficier des avantages de vitesse des arbres tout en conservant la flexibilité des maillages.

La première structure que nous avons examinée pour réduire le nombre de variables a été l'octree [122]. Toutefois, l'inconvénient majeur de l'octree est qu'il divise systématiquement l'espace en deux parties égales, indépendamment de la répartition des objets. En alternative, nous proposons l'utilisation d'une structure plus générale, le Bounding Volume Hierarchy (BVH). Une BVH est une structure de données principalement utilisée en infographie et géométrie computationnelle pour accélérer les algorithmes de lancer de rayons et de détection de collisions. Elle permet d'organiser les objets ou les formes géométriques dans une hiérarchie de volumes englobants, facilitant ainsi les requêtes spatiales comme les tests d'intersection de rayons. Cette organisation hiérarchique permet de réduire la complexité temporelle de ces tests, la rendant logarithmique par rapport au nombre d'objets dans la scène. Chaque nœud dans un arbre BVH représente un volume englobant qui encapsule un ensemble d'objets ou d'autres volumes englobants.

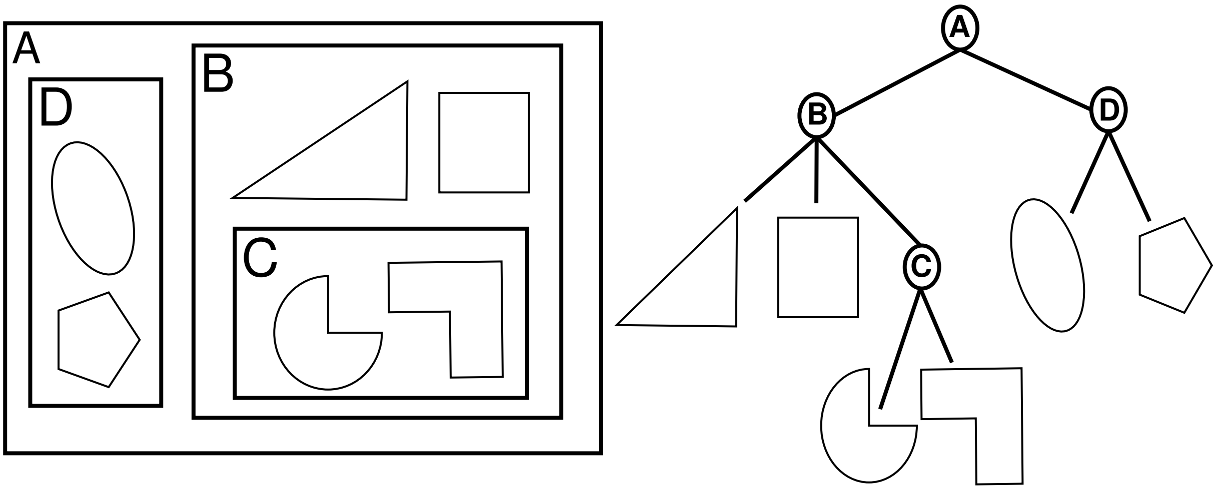


Figure B.10: Représentation d'une BVH. À gauche, la scène contient six objets distribués hiérarchiquement dans des boîtes englobantes (A, B, C, D). À droite, l'arbre BVH de la même scène est représenté. Chaque boîte contient soit des objets, soit d'autres boîtes englobantes.

En ce qui concerne les maillages, il a d'abord été nécessaire de convertir ces modèles en maillages volumétriques tétraédriques. L'avantage d'utiliser des tétraèdres réside dans leur simplicité et leur compatibilité avec les outils efficaces existants pour nos algorithmes.

Lors de notre première tentative de reconstruction, la taille des cellules était trop importante. Le maillage comportait 1 257 tétraèdres pour un volume qui correspondait auparavant à 4x1204x1204 voxels, dont 643 926 faisaient partie du masque (représentant seulement 0,2% des variables comparées à la description voxel). Le maillage était trop grossier, et la forme de l'objet était à peine reconnaissable. Les tétraèdres étaient trop grands et irréguliers, donnant une apparence très polygonale à la représentation, à la fois désagréable et imprécise. Nous avons donc procédé à une nouvelle reconstruction en augmentant considérablement le nombre de tétraèdres, atteignant 2 172 096 tétraèdres (soit 37% des variables par rapport à la description voxel). Chaque tétraèdre a été divisé en quatre tétraèdres plus petits et un octaèdre, qui a ensuite été subdivisé en huit tétraèdres. Cette opération a été répétée plusieurs fois pour affiner le maillage. Une limite supérieure à la taille des cellules a également été imposée afin d'assurer la présence de petites cellules, même dans les régions vides, permettant ainsi de réduire le nombre total de cellules. La Figure B.11a montre la reconstruction obtenue, qui est visuellement améliorée, bien que des imperfections subsistent. La forme de l'objet est plus discernable, avec des zones de matière et de vide mieux définies. Les streak artefacts semblent moins présents par rapport à la reconstruction voxel, et l'arrière-plan est mieux défini, mais l'affichage avec des triangles crée un effet visuel rendant la comparaison difficile à l'œil.

Un développement encore inachevé de cette méthode concerne la mise à jour du maillage au fil des itérations de reconstruction. Pour obtenir un maillage optimal adapté à l'objet reconstruit, il est nécessaire de mettre en place une stratégie d'évolution du maillage au cours du processus. La première stratégie que nous avons envisagée est de calculer simplement le gradient dans le volume. Si le gradient est trop important, la cellule doit être subdivisée. Cette opération peut être réalisée à l'aide d'opérations topologiques, telles que les mouvements bistellaires ou les flips de Pachner [135]. L'avantage de cette méthode est que nous pouvons augmenter le nombre de cellules si le gradient est trop fort, et inversement, réduire le nombre de cellules si le gradient est faible. De plus, les flips '2-2' peuvent être utilisés pour aligner les normales des faces avec le gradient. Par la suite, les sommets peuvent également être déplacés pour améliorer la qualité du maillage. Nous avons commencé à développer ces méthodes au cours des itérations de SART, mais nous n'avons malheureusement pas pu mener ces expériences à terme.

La version finale optimisée du maillage contient 84 212 tétraèdres, soit 13% des variables par rapport à la description voxel. Les résultats visuels montrent une meilleure représentation de l'objet avec moins de tétraèdres qu'auparavant. Les tétraèdres apparaissent plus réguliers, créant l'impression d'un pavage homogène et agréable. La version non optimisée du maillage présente des pointes et des sommets partagés par plusieurs tétraèdres, donnant une représentation moins esthétique. Chaque itération a pris environ 5 minutes. La Figure 4.23 montre un zoom sur la coupe transversale avec et sans optimisation du maillage. L'optimisation améliore clairement la qualité de la reconstruction. L'apparence irrégulière de certains tétraèdres est éliminée, au profit d'une représentation avec des tétraèdres plus équilibrés. Les faces des tétraèdres sont mieux alignées avec les contours, permettant une délimitation plus nette. L'amélioration du maillage favorise également l'étude de la connectivité. Avec le maillage optimisé, la branche horizontale à droite n'est pas connectée à la branche verticale.

À l'heure actuelle, le maillage semble satisfaisant et la reconstruction semble de bonne qualité. Cependant, comparée à la description voxel, l'image contient moins de nuances de gris et présente un aspect plus binaire : matière/vide. Ce résultat est encourageant, car il permet une meilleure segmentation de l'objet et une définition plus claire de ses contours, malgré le nombre réduit de tétraèdres. De plus, nous n'exploitons pas encore pleinement le potentiel du BVH. Dans cette thèse, nous nous sommes limités à l'utilisation de tétraèdres, mais il est essentiel de considérer des représentations plus complexes, incluant des éléments courbes. Le modèle CAO offre un excellent point de départ pour la reconstruction maillée, et s'il contient une représentation analytique (au-delà des simples triangles), il serait possible d'optimiser les éléments courbes. Cela modifierait également la manière dont l'objet réel est comparé au modèle CAO, et le lancer de rayons pourrait s'appuyer directement sur cette description. Nous pourrions alors reconstruire directement en termes d'éléments courbes, en remplaçant les rayons par des surfaces et les intersections par des tests analytiques. Cette approche semble prometteuse pour des recherches futures.

(a) Maillage non optimisé

(b) Maillage optimisé

Figure B.11: Aspects des différents maillages. L'optimisation du maillage au fil des itérations permet d'obtenir une structure mieux adaptée à la représentation de l'image et d'éviter l'effet visuel en dents de scie.

B.8 Régularisation et post-traitement par codage parcimonieux

Dans le cadre de la stratégie d'acquisition en vues éparses, le manque d'informations conduit inévitablement à des artefacts et dégrade la qualité de la reconstruction. Bien que les méthodes présentées précédemment permettent d'améliorer la qualité, l'approche la plus efficace consiste à agir directement sur le volume à reconstruire grâce à des méthodes de débruitage ou de régularisation.

Une famille de méthodes intéressante est celle basée sur des *patchs* (ou blocs en 3D), où chaque patch est traité indépendamment ou en conjonction avec d'autres patches similaires. Cela permet de surmonter les limitations liées aux ressources numériques, tout en réduisant la dimensionnalité et en accélérant les processus, tout en améliorant l'efficacité du débruitage grâce à la redondance du signal [145].

Dans notre cas, nous disposons d'un ou deux échantillons déjà scannés et analysons une trajectoire dense ou parcimonieuse. Nous supposons que des connaissances *a priori* sur la texture de l'image sont disponibles, mais nous ne disposons pas de suffisamment d'échantillons pour utiliser des méthodes avancées comme l'apprentissage profond. Ainsi, nous nous intéressons aux méthodes de *codage parcimonieux* et d'*apprentissage de dictionnaire*, bien adaptées à notre cas d'utilisation. Ils sont particulièrement utiles en tomographie industrielle car ils nécessitent peu d'informations *a priori* et généralisent bien sur des types de données variées.

Le codage parcimonieux est une technique couramment utilisée dans le traitement du signal et les problèmes inverses. Cette technique suppose qu'un signal $\mathbf{x} \in \mathbb{R}^m$ peut être reconstruit à partir de quelques éléments, appelés *atomes*, issus d'un dictionnaire redondant $\mathbf{D} \in \mathbb{R}^{m \times n}$. Dans ce cadre, chaque bloc $\mathbf{x}_s$ d'un signal $\mathbf{x}$ est encodé via une combinaison linéaire parcimonieuse $\mathbf{z}_s \in \mathbb{R}^m$ d'atomes du dictionnaire, telle que pour chaque bloc $\mathbf{x}_s \approx \mathbf{E}_s \mathbf{D} \mathbf{z}_s$, où $\mathbf{E}_s$ représente un opérateur permettant d'extraire le bloc $\mathbf{x}_s$. Le problème de trouver une représentation parcimonieuse est appelé *sparse pursuit*, et il constitue un point central dans toutes les techniques de codage parcimonieux. Différentes méthodes ont émergé pour résoudre ce problème. Parmi les plus utilisées, on trouve les algorithmes de *matching pursuit*, qui cherchent une solution de manière gloutonne, et des formulations convexes comme le *basis pursuit* qui résolvent le problème avec une contrainte de norme ℓ_1 sur la représentation [152]. L'objectif dans ce cas devient :

$$\hat{\mathbf{z}} = \underset{\mathbf{z},(\mathbf{D})}{\operatorname{argmin}} \frac{1}{2} \sum_s \|\mathbf{D}\mathbf{z}_s - \mathbf{E}_s \mathbf{x}\|_2^2 + \lambda \sum_s \|\mathbf{z}_s\|_1, \tag{B.8}$$

où λ est un paramètre équilibrant les termes de fidélité des données et de parcimonie. Le dictionnaire peut également être appris en même temps que la représentation parcimonieuse, rendant ainsi le problème dépendant de la variable $\mathbf{D}$, ce qu'on appelle l'apprentissage de dictionnaire.

Cependant, un inconvénient majeur des concepts précédents est la non-continuité du signal, chaque patch étant traité indépendamment. L'agrégation des patchs peut entraîner des artefacts, un problème que l'on peut atténuer en considérant des patchs chevauchants, bien que cela augmente les opérations de calcul. Pour surmonter ces limitations, le *codage parcimonieux convolutionnel* (CSC) propose une nouvelle formalisation qui prend en compte la continuité des images [157]. Il reformule le signal comme une somme de convolutions entre des filtres locaux et des cartes de coefficients, supprimant ainsi l'étape d'agrégation des patchs. Cette approche permet de réduire considérablement le nombre d'atomes nécessaires tout en maintenant la continuité de l'image, améliorant ainsi la qualité de la reconstruction [159]. La fonction objectif s'écrit désormais :

$$\underset{\mathbf{z},(\mathbf{d})}{\operatorname{argmin}} \frac{1}{2} \left\| \sum_j \mathbf{d}_j * \mathbf{z}_j - \mathbf{x} \right\|_2^2 + \lambda \sum_j \|\mathbf{z}_j\|_1, \tag{B.9}$$

B.8.1 Débruitage

L'application la plus simple du codage parcimonieux est le débruitage d'images, qui s'effectue de manière directe grâce à un *matching pursuit.* Si des données d'entraînement sont disponibles, un dictionnaire est d'abord appris à partir de celles-ci.

Lors de notre étude nous avons comparé les méthodes de débruitage utilisant des dictionnaires classiques et convolutionnels sur une coupe de reconstruction parcimonieuse d'une étoile de fabrication additive. Des différences de performance notables ont été observées. Les résultats indiquent que dans le CSC, le nombre d'atomes a un impact minimal sur la performance du débruitage. Réduire le nombre d'atomes ne dégrade pas significativement la performance, ce qui améliore l'efficacité computationnelle. En revanche, dans le codage parcimonieux classique, le nombre d'atomes affecte de manière cruciale le résultat. Un dictionnaire redondant est nécessaire pour garantir une représentation parcimonieuse robuste et un débruitage efficace. En outre, l'importance du dictionnaire dans la représentation est encore plus soulignée. Le dictionnaire analytique (DCT) produit des résultats inférieurs par rapport aux autres. Le dictionnaire appris à partir d'images montre des résultats satisfaisants, surtout compte tenu du fait qu'aucune donnée CT n'a été utilisée. Le dictionnaire personnalisé montre des résultats supérieurs, atteignant la convergence plus rapidement et de manière plus efficace.

B.8.2 Dictionnaires joints

Une approche supplémentaire pour le débruitage d'un signal consiste à utiliser des dictionnaires joints. Deux dictionnaires sont appris conjointement sur les reconstructions sparse et dense d'un même échantillon. En forçant les deux reconstructions à partager les mêmes coefficients de codage mais sur des bases différentes, chaque filtre dans le dictionnaire '*sparse*' D_s est associé à un filtre correspondant dans le dictionnaire '*dense*' D_d. Les résultats obtenus par un *basis pursuit* dans la base '*sparse*' peuvent ainsi être réutilisés dans la base '*dense*'.

Nous proposons une méthode intermédiaire pour la CSC entre les méthodes analytiques et directes. Grâce à un optimiseur ADMM, un dictionnaire D_s est appris à partir des données sparse, puis chaque filtre est exprimé comme une combinaison linéaire de blocs extraits du volume sparse. Une version dense du dictionnaire est créée en utilisant les mêmes blocs extraits aux mêmes emplacements dans la reconstruction dense. Cette technique combine les avantages de la méthode analytique (détection des structures sous-jacentes du signal) avec la simplicité et la rapidité de la méthode directe. La figure B.12 illustre les filtres associés dans $\mathbf{D}_s$ et $\mathbf{D}_d$, où les motifs communs, tels que les interfaces, sont clairement visibles dans le dictionnaire '*dense*'. Des résultats de débruitage sur un cube d'aluminium en fabrication additive sont montrées en figure B.13.

B.9 Conclusion

Les travaux menés dans cette thèse ont permis de proposer plusieurs méthodes innovantes pour améliorer la reconstruction tomographique à partir de vues éparses en intégrant des informations *a priori.*

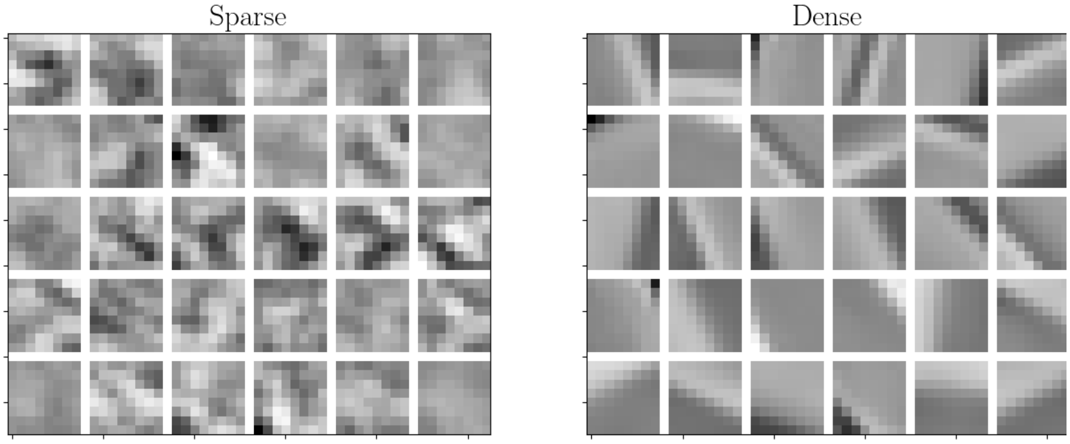


Figure B.12: Dictionnaires joints réalisés avec la méthode hybride proposée (8×8×8)

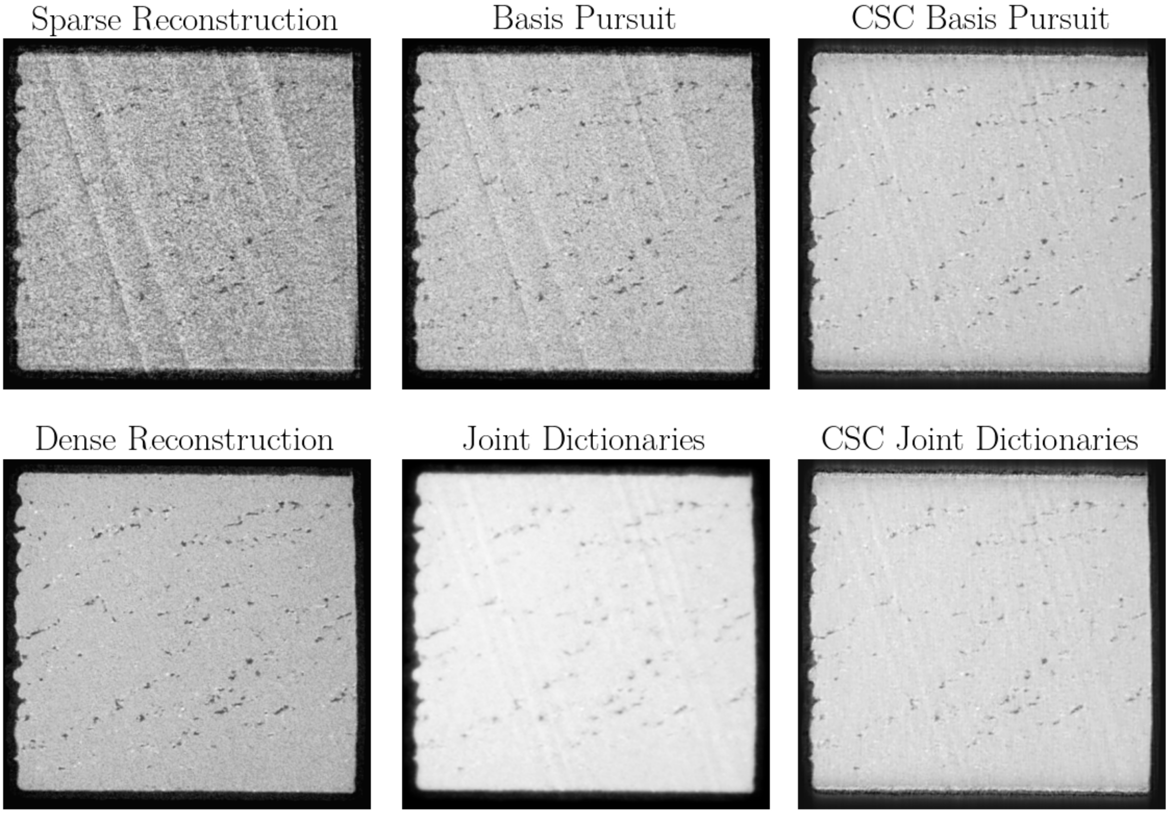


Figure B.13: Résultats des différentes techniques de débruitage sur un cube de fabrication additive.

Tout d'abord, la sélection des projections optimales a montré une amélioration significative de la qualité de reconstruction. Les méthodes basées sur l'*Empirical Interpolation Method* (EIM) et ses variantes (DEIM, QDEIM, et CQDEIM) ont permis d'augmenter la qualité de reconstruction dont le PSNR de plusieurs dB, prouvant ainsi leur efficacité dans un contexte de vues éparses. Ces méthodes sont rapides, simples à mettre en œuvre, et offrent une complémentarité des vues optimale.

Ensuite, les méthodes de recalage ont montré des résultats prometteurs. En particulier, l'algorithme basé sur les enveloppes convexes et la méthode par asservissement visuel ont permis d'atteindre un recalage de l'objet avec une précision de quelques pixels. Le recalage précis entre le modèle CAO et l'objet réel s'est avéré essentiel pour améliorer la reconstruction et l'utilisation des informations géométriques disponibles.

L'utilisation de masques obtenus par voxelisation du CAO a également montré une nette amélioration de la reconstruction, en réduisant les artefacts, en augmentant la netteté des discontinuités, et en rendant les zones homogènes plus lisses. Cette approche a permis de gagner plusieurs dB en PSNR, tout en réduisant le nombre de voxels à reconstruire. L'extension de cette idée vers la reconstruction sur un maillage volumétrique a démontré un potentiel important, bien que les temps de calcul actuels restent un frein à son utilisation.

Enfin, la reconstruction basée sur l'apprentissage de dictionnaires a permis de réduire le bruit dans les reconstructions et d'améliorer la résolution. L'implémentation efficace du Convolutional Sparse Coding a ouvert la voie à des améliorations non seulement en termes de débruitage mais aussi de super-résolution et de régularisation.

En conclusion, les méthodes proposées ont montré des gains significatifs en qualité de reconstruction dans des scénarios de vues éparses, avec des améliorations mesurables en termes de PSNR, de réduction d'artefacts, et d'amélioration de la résolution. Ces résultats ouvrent des perspectives prometteuses pour des applications industrielles où la reconstruction précise d'objets avec un nombre limité de projections est cruciale.

FOLIO ADMINISTRATIF

THESE DE L'INSA LYON, MEMBRE DE L'UNIVERSITE DE LYON

NOM : **BUSSY** DATE de SOUTENANCE : **04/12/2024**

Prénom : **Victor**

TITRE : **Integration of a priori data to optimise industrial X-ray tomographic reconstruction**

NATURE : **Doctorat** Numéro d'ordre : **2024ISAL0116**

École Doctorale : **Électronique, Électrotechnique et Automatique de Lyon**

Spécialité : **Traitement du signal et de l'image**

RÉSUMÉ :

Cette thèse explore des sujets de recherche dans le domaine du contrôle non destructif industriel par rayons X (CND). L'application de la tomographie CT s'est considérablement étendue et son utilisation s'est intensifiée dans de nombreux secteurs industriels. En raison des exigences croissantes et des contraintes sur les processus de contrôle, la CT se doit de constamment évoluer et s'adapter. Que ce soit en termes de qualité de reconstruction ou en temps d'inspection, la tomographie par rayons X est en constante progression, notamment dans ce qu'on appelle la stratégie de *vues éparses*. Cette stratégie consiste à reconstruire un objet en utilisant le minimum possible de projections radiologiques tout en maintenant une qualité de reconstruction satisfaisante. Cette approche réduit les temps d'acquisition et les coûts associés. La reconstruction en vues éparses constitue un véritable défi car le problème tomographique est mal conditionné, on le dit *mal posé*. De nombreuses techniques ont été développées pour surmonter cet obstacle, dont plusieurs sont basées sur l'utilisation d'informations *a priori* lors du processus de reconstruction. En exploitant les données et les connaissances disponibles avant l'expérience, il est possible d'améliorer le résultat de la reconstruction malgré le nombre réduit de projections.

Dans notre contexte industriel, par exemple, le modèle de conception assistée par ordinateur (CAO) de l'objet est souvent disponible, ce qui représente une information précieuse sur la géométrie de l'objet étudié. Néanmoins, il est important de noter que le modèle CAO ne fournit qu'une représentation approximative de l'objet. En CND ou en métrologie, ce sont précisément les différences entre un objet et son modèle CAO qui sont d'intérêt. Par conséquent, l'intégration d'informations *a priori* est complexe car ces informations sont souvent "approximatives" et ne peuvent pas être utilisées telles quelles. Nous proposons plutôt d'utiliser judicieusement les informations géométriques disponibles à partir du modèle CAO à chaque étape du processus. Nous ne proposons donc pas une méthode, mais une méthodologie pour l'intégration des informations géométriques *a priori* lors la reconstruction tomographique par rayons X.

MOTS-CLÉS :

Tomographie X robotisée, Problème inverse, Reconstruction, Algorithmes itératifs, Optimisation de trajectoire, Empirical Interpolation Method, Recalage 3D/2D, Asservissement Visuel, Moments d'images, Masque, Maillages, Codage Parcimonieux, Apprentissage de Dictionnaires, Dictionnaires convolutifs, Informations a priori

Laboratoires de recherche :
CEA-LIST, Département d'Imagerie et de Simulation pour le Contrôle (DISC)
INSA Lyon, Laboratoire Vibrations Acoustique (LVA)
Directeur de thèse : **(Directeur)**
Valérie KAFTANDJIAN

Président du Jury : **Laurent Desbat**

Composition du Jury :
Laurent Desbat, Jan Sijbers, Gabriel Herl, Yann Quinsat, Valérie Kaftandjian, Caroline Vienne, Julie Escoda