Tomographic Segmentation from Limited Projection Data

Koo, Jakeoung

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Tomographic Segmentation from Limited Projection Data

Jakeoung Koo

Kongens Lyngby 2021

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

Computed Tomography (CT) enables analyzing the internal structures of objects from projections data. The projections data measures the attenuation of penetrating radiations such as X-rays. The conventional CT pipeline includes reconstructing an image from projections and segmenting the reconstructed image for quantitative analysis. However, in challenging situations, when projection data are noisy or limited, the reconstructed images can be degenerate, which can lead to incorrect segmentation results.

Instead of reconstructing images, the main goal of the thesis is to develop direct tomographic segmentation methods from projections for homogeneous objects. This goal is achieved by representing objects using triangle meshes and deforming them to be aligned with the boundaries of the scanned objects. In this regard, we propose two direct segmentation methods. The first method addresses mesh deformation in 2D space with the advantage of topological adaptivity during deformation.

The second proposed method tackles 3D shape estimation directly from projections. We extend recent results on differentiable rendering to tomographic reconstruction and this extension enables optimizing 3D shapes from projections. The experimental results in electron tomography show the effectiveness of our method for reconstructing shapes of some nano-particles.

Also, we investigate another representation of objects using coordinate-based neural networks for tomographic reconstruction. Finally, the thesis studies a regularization term using a vectorial total variation norm for spectral CT. The proposed regularization term is demonstrated to benefit spectral CT data with the potential to be of practical use in security applications.

Summary (Danish)

Computed tomografi (CT) muliggør analyse af objekters interne strukturer ud fra projektionsdata. Projektionsdata måler dæmpningen af gennemtrængende stråling såsom røntgenstråler. Den konventionelle CT-pipeline inkluderer rekonstruktion af et billede fra projektioner og herefter segmentering af det rekonstruerede billede til brug for kvantitativ analyse. I udfordrende situationer, når projektionsdata er støjende eller begrænsede, kan de rekonstruerede billeder dog være degenererede, hvilket kan føre til forkerte segmenteringsresultater.

Målet med afhandlingen er at udvikle direkte tomografiske segmenteringsmetoder fra projektionsdata af homogene objekter som alternativ til at rekonstruere billeder. Dette mål blev opnået ved at repræsentere objekter ved hjælp af trekantsmesh og deformere dem så de tilpasser sig kanterne af de scannede objekter. I denne henseende foreslår vi to direkte segmenteringsmetoder. Det første arbejde adresserer deformation af trekantsmeshes i 2D med den fordel at topologien tilpasses under deformation.

Den anden metode tackler 3D-formestimering direkte fra projektionsdata. Vi udvider de seneste resultater om differentiable rendering til tomografisk rekonstruktion, og denne udvidelse muliggør optimering af 3D shapes fra projektionsdata. De eksperimentelle resultater fra elektrontomografi viser effektiviteten aff vores metode til rekonstruktion af shapes for nogle nanopartikler.

Vi undersøger også en anden repræsentation af objekter ved hjælp af koordinatbaserede neurale netværk til tomografisk rekonstruktion. Endelig studerer vi i afhandlingen et regulariseringsudtryk ved hjælp af en vektoriel totalvariationsnorm for spektral CT. Det foreslåede regulariseringsudtryk vises at det gavner spektral CT-data med potentialet til at være praktisk anvendelig i sikkerhedsapplikationer.

Preface

This Ph.D. thesis was prepared at the section of Visual Computing at the Technical University of Denmark (DTU) in fulfillment of the requirements for acquiring a Ph.D. degree in the Department of Applied Mathematics and Computer Science. The work presented in this thesis was funded by EU Horizon 2020 MSCA Innovative Training Network: MUltiscale, Multimodal and Multidimensional imaging for EngineeRING (MUMMERING) with Grant Number 765604.

The thesis has been supervised by Associate Professor Vedrana Andersen Dahl and co-supervised by Professor MSO Anders Bjorholm Dahl from DTU. The research has been carried out at DTU Compute at the section of Visual Computing.

The main topic of the thesis is tomographic reconstruction and segmentation and the thesis investigates different representations of objects such as mesh-based and neural network-based representation.

The thesis includes four original manuscripts in Contribution A to D which have been carried out during the project. I suggested the original ideas for Contribution B and C. For Contribution D, my contribution is in the methodology part.

Lyngby, 14-March-2021

Jakeoung Koo

List of contributions

Peer reviewed

- Paper A. Jakeoung Koo, Anders B. Dahl, Vedrana A. Dahl. DALM, Deformable Attenuation-Labeled Mesh for Tomographic Reconstruction and Segmentation. *IEEE Transactions on Computational Imaging (TCI)*, vol. 7, pp. 151-163, 2021.
- Paper B. Jakeoung Koo, Anders B. Dahl, J. Andreas Bærentzen, Qiongyang Chen, Sara Bals, Vedrana A. Dahl. Shape from Projections via Differentiable Forward Projector for Computed Tomography. *Accepted in Ultramicroscopy*, 2021. https://doi.org/10.1016/j.ultramic.2021.113239
- Paper C. Jakeoung Koo, Elise O. Brenne, Anders B. Dahl, Vedrana A. Dahl. A Tomographic Reconstruction Method using Coordinate-based Neural Network with Spatial Regularization. *Northern Lights Deep Learning Workshop*, 2021.

In preparation

- Paper D. Doniyor Jumanazarov, Jakeoung Koo, Henning F. Poulsen, Ulrik L. Olsen, Mihai Iovea, Material clasificiation from sparse spectral X-ray CT using vectorial total variation based on L infinity norm, 2021.

Contribution which is not included in the thesis

- Doniyor Jumanazarov, Jakeoung Koo, Matteo Busi, Henning F. Poulsen, Ulrik L. Olsen, Mihai Iovea. System-independent material classification through X-ray attenuation decomposition from spectral X-ray CT. *NDT & E International*, vol.116, 102336, 2020.

List of abbreviations

CT	Computed tomography
DART	Discrete algebraic reconstruction technique
DIB-R	Differentiable interpolation-based renderer
DSC	Deformable simplicial complexes
DTU	Technical University of Denmark
HGPD	Hybrid gradient primal dual
PCA	Principal component analysis
SIRT	Simultaneous iterative reconstruction technique
STEM	Scanning transmission electron microscope
TNV	Total nuclear variation
TV	Total variation
VTV	Vectorial total variation

Acknowledgements

First of all, I would like to thank greatly my supervisors, Vedrana Andersen Dahl and Anders Bjorholm Dahl. My supervisors' advice, feedback and support help me keep going on my PhD journey and focus on the research, especially during the pandemic. In particular, their positive energies are what inspire me and what I want to learn. I also appreciate Anders' leadership, as a member of the Visual Computing group.

I thank Ali Chirazi and Daniel Lichau for hosting my external stay in Thermo Fisher Scientific in the beautiful city, Bordeaux. I also thank Jakob Andreas Bærentzen for introducing me to the world of computer graphics and Sara Bals for kindly helping me and sharing her knowledge on electron tomography.

My special thanks go to MUMMERING colleagues. I must thank Elise for helping me keep going. I have greatly enjoyed working with Doniyor. I also thank Anuj, Robert, Sotiris, and so on for making good memories. I thank Poulami for her invitation to CWI, although it did not happen due to the pandemic. Lastly, I express my gratitude to all the people who make MUMMERING possible including Jens and Hanne.

I also thank the colleagues at the Visual Computing group including Tuan, Monica, Shihav, Martin, Patrick and Wail to name a few, for interesting conversations and discussions.

Finally, I owe a debt of gratitude to my mother for her unconditional support. I also thank Byung-Woo Hong for his sincere advice as always.

Contents

Chapter 1

Introduction

1.1 Scope of the thesis

Computed Tomography (CT) has emerged as a powerful imaging technique to enable investigating the internal structures of objects. CT has been successfully applied to many areas such as medical diagnostics, material science, and security applications. A workflow in CT begins with scanning an object from different angles by radiation sources such as X-rays to obtain data called projections. Projections measure how much radiations are attenuated after they pass through the object. Following that, the tomographic reconstruction procedure aims to recover the object from projections.

Over the past few decades, there have been many useful reconstruction methods developed, but it still remains challenging to reconstruct objects correctly when the projections data are noisy or limited. For instance, in electron tomography, the available range of projection angles is limited, or security applications require the reconstruction algorithm to be robust from a small number of noisy projection images.

The thesis focuses on developing tomograhpic reconstruction methods from such noisy or limited projections data. Instead of conventional image-based representation, we have investigated alternative representations of objects. These include

meshes to be deformed to align with the boundaries of the scanned objects, and a coordinate-based neural network to reconstruct a continuous function for the object. We also investigate an image reconstruction method from spectral CT data for security applications using a conventional voxel-based representation.

The works carried out in the thesis are part of a multi-disciplinary project: MUltiscale, Multimodal and Multidimensional imaging for EngineeRING (MUMMERING) funded by EU Horizon Marie Skłodowska-Curie Actions with the grant number 765604. MUMMERING covers the whole pipeline of tomography from data acquisition and reconstruction to finite element modeling. Many parts of the thesis are performed in close collaboration with some colleagues in MUMMERING working on electron tomography or security applications.

1.2 Motivation

The main theme of the thesis is to investigate tomographic reconstruction methods from limited projection data and we have developed four reconstruction methods in Contribution A - D. In this section, we discuss the motivations of the contributions.

1.2.1 Direct segmentation from projections

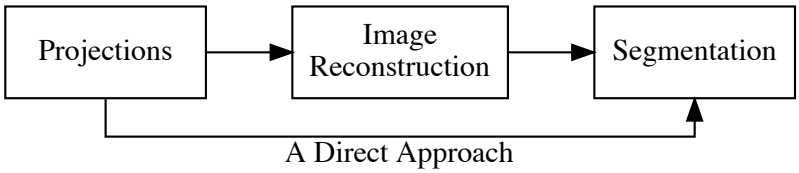


Figure 1.1: A conventional pipeline of tomography consists of image reconstruction from projections data and estimates the segmentation of the reconstructed image. On the contrary, a direct approach aims to obtain the segmentation directly from projections.

As shown in Fig. 1.1, the conventional pipeline of CT includes the reconstruction of an image from projections and the segmentation of the reconstructed image for quantitative analysis. However, in challenging situations, when projection data are noisy or limited, the reconstructed images can be degenerate, which would lead to incorrect segmentation results.

Instead of reconstructing images, in the thesis, we investigate other tomographic segmentation methods directly from projections for homogeneous objects. We consider the situation where the number of materials (mostly less than 5) is known and the objects are homogeneous. Exploiting this prior information, we propose two direct segmentation methods based on mesh deformation in Contribution A and B.

1.2.2 Alternative ways to represent objects

Most of existing tomographic reconstruction methods represent objects by images on a regular grid. That is, each voxel on the image is assigned with an attenuation coefficient. The attenuation coefficient is a material property to indicate how much a radiation is attenuated when penetrating the material. In many cases, the attenuation value is assumed to be constant within the voxel. Instead of using an image-based representation, in this thesis, we investigate three alternative ways to represent objects, which is summarized in Table 1.1.

Papers	Representation of objects
Contribution A	Triangular mesh with attenuation-labeled faces in 2D
Contribution B	Triangular surface mesh in 3D
Contribution C	Continuous function using an implicit neural network

Table 1.1: We investigate different ways to represent objects for tomographic reconstruction.

First, for reconstructing homogeneous objects in 2D, we use a triangle mesh to represent objects, where each face in the mesh is labeled with one material, so each face has the corresponding attenuation coefficient. We name such labeled-mesh as Attenuation-Labeled Mesh (ALM). The deformation of ALM is studied in Contribution A to align with object boundaries.

Second, we consider a surface mesh in 3D, which differs ALM in the sense that the surface mesh only considers the boundary of an object and has no information on the inside of the object. From this representation, we have suggested a differentiable forward model to enable optimization of the mesh in Contribution B.

Lastly, we investigate a continuous representation of attenuation functions. Although the measurements are finitely available, the underlying attenuation function can be considered as continuous. To reconstruct a continuous attenuation coefficient function, in Contribution C, we study a coordinate-based neural net-

work which takes spatial coordinates as input and outputs the attenuation coefficients at the coordinate positions.

1.2.3 Spectral X-ray CT

The conventional CT using X-ray sources relies on energy-integrating detectors. That is, conventional X-ray detectors measure the projections which are integrated over X-ray energies. In contrary, an advanced technique called spectral X-ray CT uses energy-dependent detectors which obtain energy-dependent projection data for some X-ray bin levels.

One challenge in spectral X-ray CT is that the quality of energy-dependent projection data become more noisy, as the number of X-ray bin levels increases (e.g. 15 energy bins). Moreover, security applications require the reconstruction algorithm to be robust from limited projections data, where the number of projection images is small (e.g. 7 or 12 projection images). To handle these challenging data, we study a regularization scheme in Contribution D.

1.3 Thesis objectives

During the PhD study, we have produced four academic papers of Contribution A - D and the objective of each contribution can be summarized as follows:

- Developing a mesh-based deformable model to reconstruct homogeneous objects from projections with the advantage of topological adaptivity during deformation (Contribution A)
- Developing a differentiable forward model for a triangle surface mesh to reconstruct 3D shapes of homogeneous objects (Contribution B)
- Investigating continuous representation of object, using coordinate-based neural networks (Contribution C)
- Studying a robust regularization term in iterative image reconstruction method from spectral CT data (Contribution D)

1.4 Thesis overview

The remainder of the thesis consists of four chapters. Chapter 2 reviews existing tomographic reconstruction methods from image-based reconstruction methods to geometric approaches for tomographic reconstruction. Section 2.4 derives a curve evolution equation which will be helpful to understand Contribution A. Chapter 3 is concerned with some technical backgrounds which are necessary to understand the contributions. In particular, Section 3.3 provides a detailed background on a computer graphics technique related to Contribution B. Chapter 4 gives an overview of the four contributions of the thesis with some motivational examples and Chapter 5 concludes the thesis. Finally, the thesis includes the three published papers (Contribution A, B, C) and one draft (Contribution D) in the appendix.

Chapter 2

Existing tomographic reconstruction works

Tomographic reconstruction is an inverse problem where we aim to reconstruct an object from the projections of the object. Tomographic reconstruction enables seeing the inside of the object without breaking it apart. The projections are obtained by the interaction of the object and the radiations such as X-rays in X-ray tomography or electrons in electron tomography. This chapter begins with explaining the process of obtaining data and discusses the reconstruction methods. For a more thorough treatment on X-ray tomography, one may refer to the book by Buzug [Buz08].

In X-ray tomography, the projection measurements can be modeled based on Lambert-Beer's law [Buz08]. According to this law, the X-ray intensity I passing through an object can be given as a function of running length l as follows:

$$I(l) = I(0) \exp\left(- \int_0^l \mu(\eta) \, \mathrm{d}\eta \right) \tag{2.1}$$

where $I(0)$ is the X-ray intensity before passing the object and μ is the attenuation coefficient function. This function characterizes a material property of how much the material can attenuate the X-rays. A more accurate model for μ

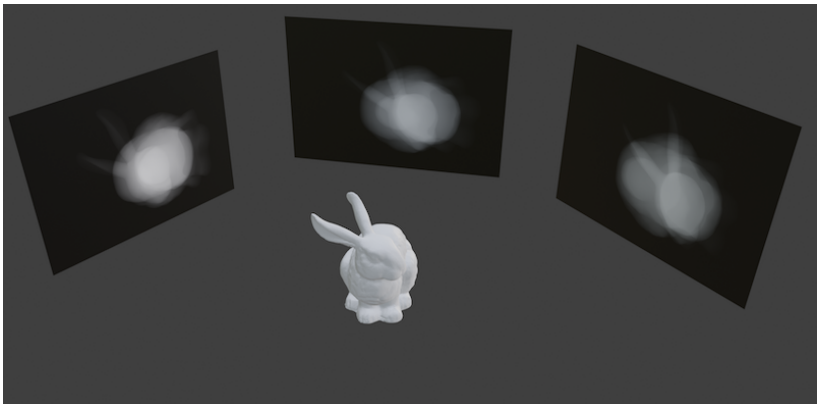


Figure 2.1: Illustration of the projections of a 3D object called bunnny, for 3 projection angles. The projections are shown on the black planes. From these projections, tomographic reconstruction aims to recover the object.

would be to include the dependency on X-ray energies as follows:

$$I(l) = \int_0^{\mathcal{E}_{max}} I(0) \exp\left(- \int_0^l \mu(\eta, \mathcal{E}) \, \mathrm{d}\eta \right) \mathrm{d}\mathcal{E}. \tag{2.2}$$

This energy-dependent model is considered in an advanced technique called spectral CT where the detectors can measure energy-dependent X-ray intensities. Spectral CT is considered in Contribution D, but from now on, we focus on the simplified form (2.1), which is considered in Contribution A, B, C.

X-ray detectors can detect the intensities $I(0)$ and $I(l)$, and these intensities can be converted to obtain the so-called projections. By arranging Eq. (2.1), we have

$$-\ln\left(\frac{I(l)}{I(0)}\right) = \int_0^l \mu(\eta) \, \mathrm{d}\eta. \tag{2.3}$$

and the projection value p is defined as follows:

$$p = \int_0^l \mu(\eta) \, \mathrm{d}\eta. \tag{2.4}$$

This equation represents one projection measurement p corresponding to one detector pixel. A detector comprises many detector pixels and, by rotating the object, the total number of measurements can be the number of detector pixels times the number of rotation angles. We will only consider the converted form p from now on. In Fig. 2.1, we provide an example of projections of a 3D object for 3 projection angles.

In electron tomography, the radiation sources are electrons rather than X-rays, and there are many different types of imaging techniques based on electron sources. Among them, we consider a technique called Scanning Transmission Electron Microscope (STEM) in Contribution B. In this case, STEM measures the projection value directly and we can still use Eq. (2.4) under a certain condition that the objects are not too thick [Haw92].

In summary, the goal of tomographic reconstruction is to recover the attenuation function μ from the projections which are the line integrals of the attenuation function along the rays. We have described how projections are obtained. We now review some existing reconstruction methods. For the simplicity of presentation, we consider 2D reconstruction problem whose goal is to reconstruct one slice of the object, but most of the described methods can generalize to 3D reconstruction as well.

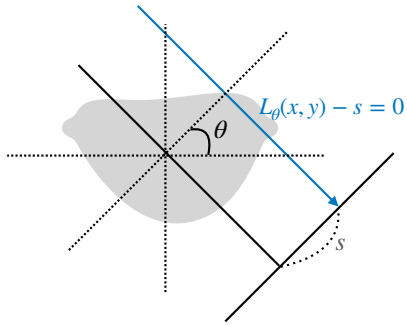


Figure 2.2: Illustration of a parallel beam ray for the projection angle θ and the detector position s. The blue line shows the line equation $L_\theta(x, y) - s = x\cos\theta + y\sin\theta - s = 0$. The line in the bottom right represents the detector.

2.1 Analytic image reconstruction

Analytic reconstruction methods are based on mathematical solutions using Fourier transforms. These methods are very efficient for an ideal case where projection data are noise-free and the available number of projection angles is large enough. We will review two well-known analytic methods and discuss their limitations.

For the simplicity of presentation, we restrict our attention to the parallel beam case in a 2D reconstruction domain. The parallel beam refers to the case where

radiation rays are perpendicular to the detector, as shown in Fig. 2.2. Let $p(\theta, s)$ denote the projection value obtained by the line integral of the attenuation function μ along the line $L_\theta(x, y) - s = 0$, where $L_\theta(x, y) = x \cos\theta + y \sin\theta$.

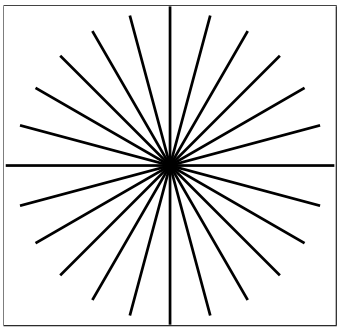


Figure 2.3: According to the Fourier slice theorem, for the 2D parallel beam case, projections can be interpreted as the measurements on the Fourier domain representation of an object. Each line corresponds to one projection angle and the line shows the positions on which frequency components of the object are available on the Fourier domain, where the origin is located at the center.

Analytic reconstruction methods in a 2D reconstruction domain are based on a surprising theorem called the Fourier slice theorem, which connects projections and frequency domain representation of the object. According to the theorem, the 1D Fourier transform $P_\theta(q)$ of projections $p(\theta, s)$ for a fixed angle θ is equivalent to the slice on the Fourier domain representation of μ through the origin along the angle θ. Specifically, we can write the theorem as

$$\mathcal{F}\{\mu\}(q\cos\theta, q\sin\theta) = P_\theta(q) := \mathcal{F}\{p(\theta, \cdot)\}(q), \tag{2.5}$$

where $\mathcal{F}\{\mu\}$ stands for 2D Fourier transform of the attenuation function μ and $\mathcal{F}\{p(\theta, \cdot)\}$ is the 1D Fourier transform of projections, given θ.

Hence, the projection data can be interpreted as the frequency components of the object on the radial lines. As shown in 2.3, these measurements in Fourier domain are concentrated near the origin, so we have dense low-frequency observations and sparse high-frequency observations. One analytic solution is to apply 2D inverse Fourier transform on such measures on Fourier domain. This direct inverse method requires an interpolation on the grid points.

Another analytic method is a classical method in computed tomography, called filtered backprojection [KS88]. This method does not require 2D inverse Fourier transform, but uses a set of 1D Fourier transforms. Filtered backprojection is based on a mathematical equation in a continuous setting. To derive the

equation, we represent the function μ by the 2D inverse Fourier transform in polar coordinates and with the Fourier slice theorem, we can derive the following reconstruction formula:

$$\mu(x,y) = \int_0^{\pi} \int_{-\infty}^{+\infty} P_\theta(q)\,|q|\,\mathrm{e}^{2\pi i q L_\theta(x,y)}\,\mathrm{d}q\,\mathrm{d}\theta \tag{2.6}$$

The outer integral is known as backprojection, as it maps back from projection domain to the reconstruction domain. The inner integral is involved with filtering $P_\theta(q)$ by the so-called ramp filter $|q|$, followed by taking 1D inverse Fourier transform. The ramp filter is a high pass filter, which suppresses low-frequency components.

Filtered backprojection is a discrete algorithm for Eq. (2.6) and can be summarized as follows: 1. Filter the projections by a high pass filter such as ramp filter, independently with respect to angles. 2. Backproject those filtered projections onto the reconstruction domain. This two-step method is fast, but heavily dependent on the quality of the data. This is because using high pass filter can amplify the noise on the data. Moreover, we need enough projection data to accurately approximate the original equation (2.6).

In the thesis, we mostly consider the case of noisy or incomplete data where the projection angles are sparsely given or limited to a small range. To such data, popular approaches are to use iterative methods based on gradient descent scheme, which is the topic of the next subsection.

2.2 Algebraic image reconstruction

Algebraic reconstruction methods represent objects by an image of the attenuation values, represented by a vector $\mathbf{u}$ and formulate the reconstruction problem based on a linear system of equations. To be specific, we can model the projection value for the i-th measurement as a weighted sum of pixel values u_j as follows:

$$\sum_j a_{ij} u_j \tag{2.7}$$

where j denotes the index for the pixels and a_{ij} is the weight determined by the contribution of the i-th ray and the j-th pixel. For example, as shown in Fig. 2.4, the weight can be set as the line segment length bounded by the ray and the pixel.

From Eq. (2.7), we can construct a system of linear equation for all the mea-

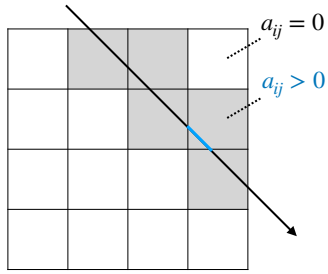


Figure 2.4: The projection for the i-th ray can be computed by a weighted sum of image values. The ray passes the shaded pixels where the weights are positive. The weight a_{ij} can be determined by the line segment (blue) bounded by the intersection of the ray i and the pixel j.

surements in a vectorized form:

$$\mathbf{A}\mathbf{u} = \mathbf{p} \tag{2.8}$$

where $\mathbf{A} \in \mathbb{R}^{I \times J}$ is a system matrix with elements a_{ij} and $\mathbf{p}$ is a vector representing the projection data. Solving such linear equation depends on the properties of the matrix $\mathbf{A}$. In tomography, the system matrix is typically not a square matrix and so we consider two cases: the overdetermined matrix when $I > J$ and the underdetermined matrix when $I < J$, where I is the number of rows and J is the number of columns of the matrix.

To explore how to deal with a non-square matrix, we begin with an ideal case where the matrix has the full rank and relax the condition later. A full rank matrix means that an overdetermined matrix has linearly independently columns or an underdetermined matrix has linearly independent rows. In this case, it is rather easy to solve the linear equation and we summarize the solutions in the following (see [BV18] for the derivation):

$\mathbf{A}$: overdetermined with full rank $\iff$ $\mathbf{A}^T\mathbf{A}$ is invertible.

$$\implies \mathbf{u}_{\text{LS}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{p} \iff \mathbf{u}_{\text{LS}} = \arg\min_{\mathbf{u}} \|\mathbf{A}\mathbf{u} - \mathbf{p}\|_2^2$$

$\mathbf{A}$: underdetermined with full rank $\iff$ $\mathbf{A}\mathbf{A}^T$ is invertible.

$$\implies \mathbf{u}^0_{\text{LS}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{p} \iff \mathbf{u}^0_{\text{LS}} = \arg\min_{\mathbf{u}} \|\mathbf{u}\|^2 \text{ such that } \mathbf{A}\mathbf{u} = \mathbf{p}$$

Hence, if the system matrix $\mathbf{A}$ has full rank, we can compute the inverse of $\mathbf{A}^T\mathbf{A}$ or $\mathbf{A}\mathbf{A}^T$ and obtain the closed-form solution in the sense of least-square error (for the overdetermine case) or minimum-norm least-square error (for the underdetermined case).

This approach, however, is not applicable to tomographic reconstruction, because the system matrix $\mathbf{A}$ in tomography typically has not full rank and the projection data $\mathbf{p}$ are noisy. To deal with these cases, iterative approaches are common for tomographic reconstruction and such iterative methods are also known as algebraic reconstruction methods. One simple solution would be to apply gradient descent to the least square norm, which leads to the update at the k-th step with the step size τ as follows:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \tau \mathbf{A}^T \left(\mathbf{p} - \mathbf{A}\mathbf{u}^k\right). \tag{2.9}$$

This algorithms is known as Landweber's method and its convergence to the least square solution is guaranteed [Byr14] under $\tau < 2/\|\mathbf{A}^T\mathbf{A}\|_2$, where $\|B\|_2$ denotes the largest eigenvalue of the square matrix B.

Another iterative method, known as Simultaneous Iterative Reconstruction Technique (SIRT) [AK84], is one of the most popular algebraic methods in tomography and known to converge faster than Landweber's method. SIRT update the solution u^k at the k-th step as follows:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \tau \mathbf{V}^{-1}\mathbf{A}^T\mathbf{W}^{-1}\left(\mathbf{p} - \mathbf{A}\mathbf{u}^k\right) \tag{2.10}$$

where $\mathbf{V}$ and $\mathbf{W}$ are the diagonal matrices consisting of the elements $\mathbf{A}^T\mathbf{1}$ and $\mathbf{A1}$, respectively. These diagonal matrices have some normalization effects on the system matrix and make the method converge faster. SIRT is known to converge in a sense [Byr14] under the condition of the step size $0 < \tau < 2$.

SIRT has a disadvantage that it is vulnerable to noise, as it only fits to the data. To overcome this drawback, one can add a regularization term, by imposing a prior information on the solution such as the smoothness on the image. A popular regularization scheme is total variation regularization which promotes the sparsity on the gradients of the solution, by considering the following (isotropic) total variation term:

$$\mathrm{TV}(\mathbf{u}) = \sum_j \sqrt{(\partial_x(u_j))^2 + (\partial(u_j))^2} \tag{2.11}$$

where ∂_x and ∂_y denote the partial derivative with respect to x and y axis, respectively. With this regularization term and the residual from the linear equation, one can consider the following optimization problem:

$$\min_{\mathbf{u}} \|\mathbf{A}\mathbf{u} - \mathbf{p}\|_2^2 + \mathrm{TV}(\mathbf{u}). \tag{2.12}$$

This minimization problem can be solved by an efficient convex optimization solver, which will be explained in Section 3.4.

2.3 Discrete tomography

While algebraic reconstruction methods can reconstruct any types of objects, discrete tomography methods assume that objects consist of a small number of materials. This assumption acts as an additional prior information to deal with some challenging cases such as the limited angle problem where a small range of projection angles is given. Similar to algebraic reconstruction methods, discrete tomography methods represent the object by an image $\mathbf{u}$, but the image is constrained to have a small number of attenuation values.

An influential work in discrete tomography is a method called Discrete Algebraic Reconstruction Technique (DART) [BS11]. DART combines an algebraic reconstruction method with a heuristic step which refines the boundary based on hard thresholding. This novel approach has been improved by many works. Among them, an optimization approach is suggested, called Total Variation Regularized Discrete Algebraic Reconstruction Technique (TVR-DART). TVR-DART turns the heuristic approach into an optimization problem, by using a differentiable thresholding function with total variation regularization.

The outcome of a discrete tomography method can be considered as a direct segmentation from projections, as such method assigns each pixel to a material. In terms of the direct segmentation, Contribution A and B also consider the similar assumption that the scanned object comprises a small number of materials. The major difference lies on the way how the object is represented. Contributions A and B represent the object based on meshes, while discrete tomography represent the object by a discrete image.

2.4 Level-set based reconstruction

Similar to discrete tomography, level-set based reconstruction methods are concerned with objects comprising a small number of materials. Most of these methods reconstruct a shape of the object and that shape can be represented by a zero level-set of an implicit function. Level-set based methods evolve such implicit function, to reconstruct the shape of the object in the end.

This section reviews some level-set based reconstruction methods which are related to Contribution A and B. The section begins with the derivation of curve evolution equation (2.21) in a general setting and from this equation the level-set based approach is explained. The evolution equation (2.21) is derived in detail, as it is helpful to understand Contribution A. Finally, another approach

based on parametric level-set methods is briefly mentioned, which is not related to the evolution equation (2.21).

Curve evolution equation. Consider a homogeneous object with a constant attenuation value μ and we denote the inside region of the object by R and its boundary by a curve C, *i.e.*, $C = \partial R$. For the simplicity of presentation, we assume that the attenuation coefficient μ is known.

Let θ denote a projection angle, s denote a detector position and $L_\theta(x, y) - s = 0$ be line equations for radiation rays. For example, as shown in Fig. 2.2, the line equation L_θ is given by $L_\theta(x, y) = x \cos\theta + y \sin\theta$ for the parallel beam case where the rays are perpendicular to the detector. Then, the projection value can be computed as an integration form as follows:

$$\hat{p}(\theta, s) = \mu \int_R \delta(L_\theta(x, y) - s) \, \mathrm{d}x \, \mathrm{d}y \tag{2.13}$$

where δ is the delta function. From the computed value $\hat{p}$, we aim to minimize the residual between $\hat{p}$ and the measurement p and define the objective function with respect to the curve C as follows:

$$E(C) = \frac{1}{2} \sum_\theta \int_S (p(\theta, s) - \hat{p}(\theta, s))^2 \, \mathrm{d}s. \tag{2.14}$$

where S denotes the detector line.

The goal is to find an evolution equation of the curve in the direction of minimizing the above objective function. There are several ways to derive the evolution equation [EW01, AY04]. I find the derivation based on Reynolds transport theorem [RBM03, ZJL20] clear, so the evolution equation will be derived based on this theorem. The Reynolds transport theorem states that for a time-dependent scalar function F on the region R, the following relation holds:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_R F(x, y, t) \, \mathrm{d}x \, \mathrm{d}y = \int_R \frac{\partial}{\partial t} F(x, y, t) \, \mathrm{d}x \, \mathrm{d}y + \int_{\partial R} \langle C_t, F(x, y, t) \, \mathbf{n} \rangle \, \mathrm{d}r \tag{2.15}$$

where t is the time parameter, C_t is the velocity on the boundary that explains the evolution of the boundary, $\mathbf{n}$ is the outward normal vector, and r is the arc length parameter. The theorem is stated in 2D case, but it holds for 1D case as well. Moreover, if the boundary is fixed, the second term on the right hand side vanishes.

We first apply the Reynolds transport theorem (2.15) to the 1D integral in (2.14)

and note that the boundary of the detector is fixed. Hence, we have

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{1}{2}\sum_{\theta}\int_{S}\frac{\partial}{\partial t}(p(\theta,s)-\hat{p}(\theta,s))^2\,\mathrm{d}s \tag{2.16}$$

$$= -\sum_{\theta}\int_{S}(p(\theta,s)-\hat{p}(\theta,s))\frac{\partial\hat{p}}{\partial t}\,\mathrm{d}s \tag{2.17}$$

To compute $\partial\hat{p}/\partial t$, we again apply the Reynolds theorem to (2.13) and since $\partial\delta/\partial t = 0$, we obtain:

$$\frac{\partial\hat{p}}{\partial t} = \mu\int_{\partial R}\langle C_t(r), \delta(L_\theta(x,y)-s)\,\mathbf{n}(r)\rangle\,\mathrm{d}r \tag{2.18}$$

where C_t denotes the evolution of the boundary. By plugging Eq. (2.18) into Eq. (2.17), the integral in Eq. (2.17) disappears due to the delta function $\delta(L_\theta(x,y)-s)$ and we have

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\mu\sum_{\theta}\langle C_t(r), (p(\theta,\tilde{s})-\hat{p}(\theta,\tilde{s})\,\mathbf{n}(r)\rangle\,\mathrm{d}r. \tag{2.19}$$

where $\tilde{s} = L_\theta(C(r))$ is the detector position on which the point $C(r)$ on the curve is projected. As a reminder, we aim to evolve the curve C in the direction that minimizes the objective function E. To evolve in the steepest direction, we make the right hand side in Eq. (2.19) minimum or make the inner product in Eq. (2.19) maximum, by setting C_t as follows:

$$C_t(r) = \mu\sum_{\theta}(p(\theta,\tilde{s})-\hat{p}(\theta,\tilde{s}))\,\mathbf{n}(r), \tag{2.20}$$

This is the curve evolution equation to minimize the objective function (2.14) which fits to data.

To deal with incomplete data and avoid degenerate solutions, one can add some regularization terms in Eq. (2.14) such as the length of a curve. Minimizing this length term leads to a well-known equation called mean curvature motion [ZY96], which is $C_t(r) = \kappa(r)\mathbf{n}(r)$, where κ is the curvature at the point $C(r)$. If we add this additional regularization term, the curve evolution equation becomes

$$C_t(r) = \mu\sum_{\theta}(p(\theta,\tilde{s})-\hat{p}(\theta,\tilde{s}))\,\mathbf{n}(r) + \lambda\kappa(r)\,\mathbf{n}(r), \tag{2.21}$$

where λ is the weighting parameter to balance the two terms.

The evolution equation 2.21 is derived in a continuous setting based on explicit representation. This explicit representation has two major issues when implemented numerically. One issue is how the points on the curve are sampled and

it might be handled by resampling the points uniformly at each iteration. The other issue is regarding the topological changes during evolution. For example, if we start one closed curve and need to split it into two closed curves, handling such splitting would require many additional efforts.

Level-set based reconstruction. Instead of explicit representation, one can consider an indirect representation which represents a curve implicitly. A numerical algorithm using implicit representation can handle the resampling issue and the topological changes automatically. To be specific, consider an implicit function $\phi(x, y\,; t)$ which changes in time t such that its zero level-set

$$\{(x, y) \in \mathbb{R}^2 : \phi(x, y\,; t) = 0\} \tag{2.22}$$

represents a curve. We now put the curve evolution equation (2.21) into the implicit function, by setting $\phi(C_t(r)\,; t) = 0$ and if we derivative ϕ with respect to the time parameter t, we have

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \phi_t + \nabla\phi \cdot C_t = 0, \tag{2.23}$$

which can be rearranged as

$$\phi_t = -\nabla\phi \cdot C_t. \tag{2.24}$$

This equation is known as level-set evolution equation and explains the evolution of the implicit function ϕ. This evolution equation provides a basis for some level-set based tomographic reconstruction works [EW01, WE02, AY04] when the curve evolution C_t in (2.21) is considered together. There are many numerical methods to solve level-set equations such as the fast marching method [OS88].

Parametrized level-set based reconstruction. Aghasi et al. [AKM11] introduced another type of level-set method called the parametric level-set method for solving inverse problems. This method parametrizes the implicit function ϕ by the aggregation of radial basis functions and estimates the parameters of these basis functions, not the level-set function. This parametrization reduces the number of unknowns significantly and enables using an efficient nonlinear least squares solvers such as Gauss-Newton method [NW99]. The parameterized level-set methods have been applied to tomographic reconstruction [AKM11, KvB18, EST20], to reconstruct a shape of a homogeneous object. As a limitation, such works are designed to reconstruct a single material, while Contribution A supports multiple materials and Contribution B can deal with a composite object.

2.5 Spectral CT image reconstruction

As mentioned in Section 2.1, spectral CT considers energy-dependent projection data and the goal is to reconstruct multi-channel images $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n_e}$ where n_e is the total number of X-ray energy bins considered. The spectral CT data are known to become more noisy, as the considering number n_e of X-ray channels increases [HSF12]. As the data are not reliable, we need a robust regularization scheme, by imposing a prior knowledge. One prior knowledge in spectral CT is that the multiple reconstructed images are expected to be highly correlated. Depending on how to model such correlation, different regularizers can be introduced.

A popular regularizer in spectral CT is Total Nuclear Variation (TNV) [RLR15, DMSC16]. TNV correlates the gradients of the reconstruction images such that those gradients align together in a sense over the multi-channel images. Such alignment can be quantified by the nuclear norm of the Jacobian of the image, which is defined as

$$\mathrm{TNV}(\{\mathbf{u}_1, ..., \mathbf{u}_{n_e}\}) = \sum_{j=1}^{J} \left\| \begin{pmatrix} \partial_x u_{1,j} & \partial_x u_{2,j} & \cdots & \partial_x u_{n_e,j} \\ \partial_y u_{1,j} & \partial_y u_{2,j} & \cdots & \partial_y u_{n_e,j} \end{pmatrix} \right\|_*, \tag{2.25}$$

where J is the number of pixels per each image channel, $u_{l,j}$ is the pixel value for the l-th energy on the j-th pixel and $\|\cdot\|_*$ is a nuclear norm of a matrix, which is the summation of the singular values of the matrix. The nuclear norm can be considered as the relaxation of the matrix rank, the number of nonzero singular values of the matrix.

In Eq. (2.25), if the non-zero gradients are all aligned in the sense of parallel or anti-parallel over the channels on a pixel, the matrix rank on the pixel will be 1, or otherwise 2. As TNV relaxes the matrix rank, TNV will favor the solutions whose gradients over the multiple channels are parallel or anti-parallel (when the gradient magnitudes are the same).

Contribution D studies TNV and another regularization scheme which is shown to be more effective than TNV in spectral CT.

Chapter 3

Technical background

This chapter describes some technical backgrounds necessary to understand the contributions of the thesis. In particular, Section 3.3 describes a rendering technique in detail, as it is important to understand Contribution B.

3.1 Triangular mesh

A triangular mesh is a collection of triangles with their connectivity information. The triangular mesh consists of a set of vertices, edges and faces. A vertex can be considered as a point in a Euclidean space; an edge connects two vertices; a face connects three vertices. Note that vertices contain geometric information, while edges and faces encode the connectivity information. The triangle mesh can be used to represent a surface in 3D. In Contribution B, we use triangle meshes to represent 3D surfaces of homogeneous objects, where the topology of objects is assumed to be known and kept fixed. When the connectivity of a mesh is fixed, a simple data structure to store edges and faces may be preferred, such as an adjacency list that collects vertex indices corresponding to the triangle faces.

In some situations, however, we often need to manipulate the mesh by changing

the connectivity information of the mesh. In this case, we need to traverse vertices, edges or faces efficiently and such efficient traverse can be achieved by using another data structure called half-edge data structure. In this data structure, each edge is augmented with two directional edges called half-edges, as shown in Fig. 3.1. Using these additional edges, a mesh based on half-edge structure, called a half-edge mesh, can manipulate the mesh efficiently and it is very useful when the connectivity information of the mesh needs to be updated frequently. A half-edge mesh is used in Contribution A.

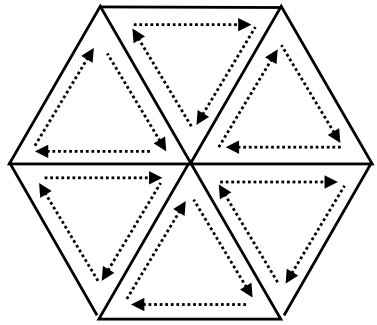


Figure 3.1: A half-edge mesh uses two directed edges called half-edges (dotted arrows), instead of one edge.

3.2 Deformable Simplicial Complexes (DSC)

As mentioned earlier in Section 2.4, when dealing with curve evolutions, explicit representation of a curve has issues with regards to sampling the points on the curve and handling topological changes. To overcome these issues, Misztal and Bærentzen [MB12] proposed a clever method called Deformable Simplicial Complexes (DSC). DSC represents the curve based on a mesh explicitly and deforms the curve, while improving the mesh and changing the connectivity information of the mesh. Specifically, DSC uses a half-edge mesh in 2D case, where each triangular face is labeled with a material and the interface is a set of edges whose two adjacent faces have different labels, as shown in Fig. 3.2.

With this labeled mesh, we assume that vertices on the interface have some destinations. For example, from the curve evolution equation 2.21, we can assign a displacement to each vertex on the interface. Towards these destinations, DSC deforms the mesh, by moving the interface vertices. Moving the vertices could make some neighboring triangles obtuse or degenerate. DSC fixes such degenerate triangles, while keeping the interface fixed during the refinement steps. The whole procedure of DSC is summarized in Algorithm 1.

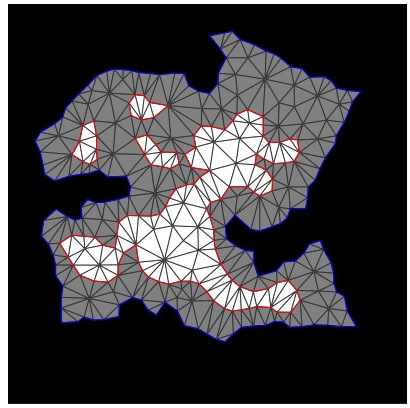


Figure 3.2: A triangular mesh in 2D with the labeled faces (gray or white) and the boundary edges (blue or red). The background faces are not visualized in the figure.

Algorithm 1 Deformation step by DSC in 2D

input Half-edge mesh $\mathcal{M}$, displacements on the interface vertices, average edge length l_1

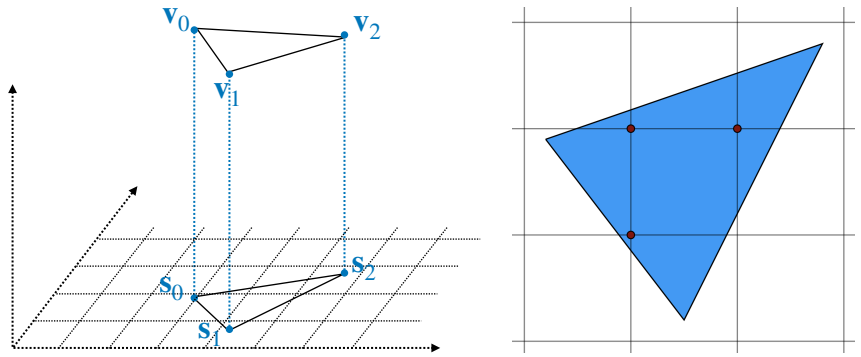
while the interface vertices do not reach their destinations **do**
  Move the interface vertices until they do not cross other triangles
  Improve the mesh while keeping the interface fixed
  Remove degenerate edges and triangles
end while
Resize the mesh, making its average edge length around l_1.

3.3 Mesh deformation based on differentiable rendering

This section addresses 3D triangular meshes and the deformation of the meshs based on an inverse process of a computer graphics procedure called rendering. Rendering in computer graphics refers to a process of generating 2D images from 3D object models and the information of visible lights and a camera. Note that this section addresses visible lights, instead of X-rays or electrons and deals with computer graphics, rather than tomography. Here, we restrict our attention to opaque objects represented by 3D triangular meshes in the sense that visible lights could not pass through such opaque objects. We first review a rendering process and explain how the rendering process can be made differentiable for the shape estimation problem.

3.3.1 Rasterization rendering technique

Among rendering techniques, we focus on a fast technique called rasterization. Rasterization rendering technique projects object geometry onto the camera plane and determines the pixels to update. As shown in Fig. 3.3 (a), the first step of the rasterization rendering technique projects a triangle onto an camera plane and, as shown in (b), the second step called *rasterization* finds the pixels which are covered by the projected triangle. The projection step for a triangle can be done by an affine transformation of the positions of the three vertices in the triangle and so is differentiable. The second step is a discrete sampling operation which is not straightforward to differentiate.



(a) projection (b) rasterization

Figure 3.3: Illustration of the rasterization rendering technique for the parallel projection case (also called orthographic projection in computer graphics). (a) shows the projection step which projects a triangle in a mesh onto an camera plane. (b) shows the rasterization step which determines the pixels (red) covered by the triangle.

After finding the visible pixels from the triangle, the rasterization rendering technique updates the image values by interpolating the value between three known attributes such as colors on the three vertices. Lastly, we need to choose the triangles visible from the camera. For opaque objects, we can only observe the nearest triangle to the camera, so we need to filter out far triangles. This can be achieved by using an array called *z-buffer*. For each pixel on the image, z-buffer aims to keep track of the nearest triangle index and we can update the buffer when we find a triangle closer to the camera than the current triangle in the z-buffer. Using this z-buffer, we can update the image value by the nearest triangle. The whole procedure of rasterization rendering process is summarized in Algorithm 2.

Algorithm 2 Rasterization rendering algorithm

```
Initialize the z-buffer array as infinity
for each triangle in the mesh do
  // projection step
  Project the triangle onto the camera plane
  for i = 1 to the number of pixels in the image do
    // rasterization step
    if the i-th pixel is covered by the projected triangle then
      z ← the distance between the i-th pixel and the triangle
      if z < z-buffer(i) then
        z-buffer(i) ← z
        Update the image value on the i-th pixel, by interpolating the value
        between known three vertex attributes
      end if
    end if
  end for
end for
```

3.3.2 Differentiable rasterization for shape estimation

In recent years, researchers have investigated a procedure of how to invert the rasterization rendering process so that we can estimate unknown parameters such as the shape of an object from 2D images. Such inverse analysis can be possible through a tool called differentiable rasterization. Differentiable rasterization connects the rasterization rendering process and optimization of unknown parameters, by estimating the derivatives of observation images with respect to unknown variables. Among differentiable renderers, we review a recent work called Differentiable Interpolation-based Renderer (DIB-R) [CGL+19], which is used in Contribution B.

Whereas previous methods approximate differentiable renderers [KUH18, LCLL19], DIB-R provides an analytic differentiation with respect to unknown variables such as the vertex positions $\mathbf{v}_k$ of the triangular mesh for the shape estimation problem. Consider a projected triangle on the camera plane with three vertices $\mathbf{s}_k$, as shown in Fig. 3.3. An image value I_i on the i-th pixel can be written in a barycentric coordinate form as

$$I_i = \sum_{k=0}^{2} w_k d_k \tag{3.1}$$

where d_k is an attribute of the vertex k (e.g., colors) and w_k is the barycentric weights such that $w_0 + w_1 + w_2 = 1$ and $w_k > 0$. The clever idea of DIB-R is

to rewrite the barycentric weights w_k as a function of $\mathbf{s}_k$ and the position $\mathbf{r}_i$ of the pixel i on the camera plane as:

$$w_k = w_k(\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \mathbf{r}_i). \tag{3.2}$$

For the exact formula, the reader is referred to [CGL+19, the supplementary, Eq. (7)].

Such reformulation enables computing the derivative of an objective function with respect to the vertex positions. Consider an objective function of the residual between the image observation $\hat{I}_i$ and the estimation I_i

$$L(\{\mathbf{v}_l\}) = \sum_i \|I_i - \hat{I}_i\|^2. \tag{3.3}$$

Then, its derivative with respect to the vertex position v_l reads

$$\frac{\partial L}{\partial \mathbf{v}_l} = \sum_i \frac{\partial L}{\partial I_i} \frac{\partial I_i}{\partial \mathbf{s}_l} \frac{\partial \mathbf{s}_l}{\partial \mathbf{v}_l}, \tag{3.4}$$

where the derivative $\partial L / \partial I_i$ is straightforward to compute and the second term $\partial I_i / \partial \mathbf{s}_l$ can be computed, using the reformulation in Eq. (3.2). The last term $\partial \mathbf{s}_l / \partial \mathbf{v}_l$ is also straightforward to compute, since the projected vertex position $\mathbf{s}_l$ is an affine transformation of the 3D vertex $\mathbf{v}_l$.

This differentiable rasterization algorithm is extended to tomographic reconstruction for estimating the shapes of homogeneous objects in Contribution B.

3.4 Convex optimization

This section touches on convex optimization and explains a specific solver called Hybrid Gradient Primal Dual (HGPD) which is used in Contribution A and D. Convex optimization is concerned with minimizing a convex function over convex sets. In particular, the thesis deals with minimizing convex functions F and G having the composite forms as follows:

$$\min_{\mathbf{u}} F(\mathbf{A}\mathbf{u} - \mathbf{p}) + G(\mathbf{D}\mathbf{u}). \tag{3.5}$$

where $\mathbf{A}$ is a linear operator such as the system matrix explained in Sec. 2.2, $\mathbf{p}$ is a vector for the projection data and $\mathbf{D}$ is a linear operator for a discrete gradient operator. In particular, we consider F to be the L_2 squared norm such that $F(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ and G be the L_1 norm function which is not differentiable.

Solving the problem (3.5) is not straightforward, as it is involved with the composite operators $\mathbf{A}$ and $\mathbf{D}$ and has the non-differentiable term G. To deal with such difficulty, there have been many first-order convex optimization methods developed. Among them, HGPD is a primal dual algorithm where the minimization problem is reformulated as a minimization-maximization problem and finds the saddle point there. The reformulation is based on a conjugate function f^* of a convex function f, defined by

$$f^*(y) = \sup_x \langle y, x \rangle - f(x). \tag{3.6}$$

It is well known [BC17, Theorem 13.37] that if f is a convex continuous function, it holds that $f^{**} = f$, where f^{**} is the conjugate function of f^*. Since the norm functions F and G are convex and continuous, we can identify them with F^{**} and G^{**}, respectively. Using this identification, we can rewrite (3.5) as

$$\min_{\mathbf{u}} \sup_{\mathbf{q_1},\mathbf{q_2}} \langle \mathbf{Au} - \mathbf{p}, \mathbf{q_1} \rangle - F^*(\mathbf{q_1}) + \langle \mathbf{Du}, \mathbf{q_2} \rangle - G^*(\mathbf{q_2}), \tag{3.7}$$

where the dual variables $\mathbf{q_1}$ and $\mathbf{q_2}$ are introduced. Then, HGPD updates the primal variable $\mathbf{u}$ by a procedure similar to the gradient descent method and the dual variables by a similar procedure to the gradient ascent method. Specifically, the updates of HGPD read [CP11]:

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \tau(\mathbf{A}^T\mathbf{q}_1^k + \mathbf{D}^T{\mathbf{q_2}}^k) \tag{3.8}$$

$$\bar{\mathbf{u}} = 2(\mathbf{u}^k - \mathbf{u}^{k+1}) \tag{3.9}$$

$$\mathbf{q}_1^{k+1} = \mathrm{prox}_{\sigma_1 F^*}(\mathbf{q}_1^k + \sigma_1(\mathbf{A}\bar{\mathbf{u}} - \mathbf{p})) \tag{3.10}$$

$$\mathbf{q}_2^{k+1} = \mathrm{prox}_{\sigma_2 G^*}(\mathbf{q}_2^k + \sigma_2\,\mathbf{D}\bar{\mathbf{u}}), \tag{3.11}$$

where k is the iteration number, τ and σ_1, σ_2 are the step sizes and the operator prox is defined as

$$\mathrm{prox}_{\tau f}(y) = \arg\min_x f(x) + \frac{1}{2\tau}\|x - y\|_2^2. \tag{3.12}$$

The solution to the proximal operator (3.12) can be interpreted as a solution to the implicit gradient method, when f is differentiable [PB14].

The convergence of the algorithm depends on the step sizes, which can be determined by the operator norms $\|\mathbf{A}\|$ and $\|\mathbf{D}\|$, the largest singular values of the operators. To be specific, according to [CP16, Lemma 5.5], the convergence of HGPD is guaranteed under the condition that $\tau = c/(\|\mathbf{A}\| + \|\mathbf{D}\|)$, $\sigma_1 = 1/(c\|\mathbf{A}\|)$ and $\sigma_2 = 1/(c\|\mathbf{D}\|)$ where c is a positive number. The choice of the number c might affect the balance of convergence between the primal variable and the dual variables updates. For a more detailed discussion on the step sizes, the reader is referred to [CP16, Example 5.7].

3.5 Coordinate-based neural networks

Coordinate-based neural network, also known as implicit neural network, is a neural network to enable the parametrization of a continuous function, by receiving the spatial coordinates as input and yielding the signals for the coordinates. Since a recent work called NeRF [MST+20] was developed for the novel view synthesis task in computer vision, the coordinate-based neural network has been successfully applied to many problems involving the computation of integrals.

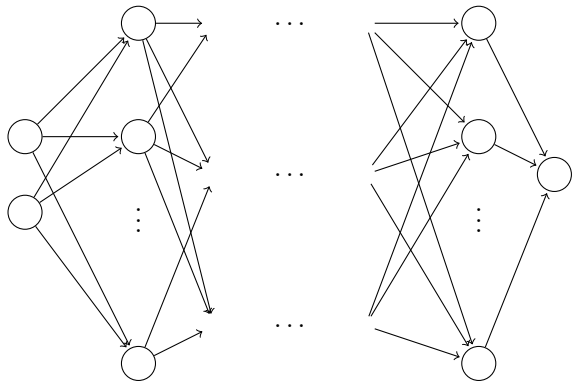


Figure 3.4: An illustration of a fully connected neural network which inputs a 2D coordinate and outputs the signal for the coordinate.

A neural network architecture called SIREN [SMB+20] is a fully connected neural network which uses the sine activation function and a principled initialization scheme. SIREN is shown to represent well the signals with the high frequency components. This neural network is used for tomographic reconstruction in Contribution D.

Chapter 4

Contributions

The goal of this chapter is to give an overview of the contributions of the thesis with some motivational examples. The chapter consists of four sections, where each section corresponds to one paper from Contributions A - D in the same order.

Section 4.1 and 4.2 describe the main contributions of the thesis: Contribution A and B, which aim to estimate the segmentation of homogeneous objects directly from projection data based on mesh deformation. In this regard, these works share some common features, but differ in several aspects. Contribution A addresses 2D object space and does not require the knowledge of the topology information of the objects. On the other hand, Contribution B considers 3D object space and needs the information on the topology of the objects as a priori. The difference of two works will be more discussed in Section 4.2.

Section 4.3 gives an overview of Contribution C which investigates a neural network-based representation for tomographic reconstruction. Section 4.4 describes Contribution D which studies a regularization scheme for spectral CT reconstruction. Note that Contribution C and D are not related to mesh deformation.

4.1 Multi-material segmentation from projections in 2D (Contribution A)

In Section 2.4, we have discussed a geometric approach to tomographic reconstruction and derived the curve evolution equation (2.21) for alinging a curve with the boundary of a homogeneous object from projections. There are two main ways to represent such a curve: an explicit and an implicit representation. A method based on explicit representation keeps track of the points on the curve and evolves the curve by displacing the points. The explicit representation was used in Dahl et al. [DDH18] to reconstruct a single homogeneous object, by extending a curve evolution method called Snakes [KWT88] to tomographic segmentation. This preliminary work has the limitations that it only supports a single material and could not handle the topological changes during deformation. These limitations are shared by many curve evolution methods based on explicit representation.

In Contribution A, we extend the preliminiary work [DDH18] to support multiple materials and the topological adaptivity during deformation, using Deformable Simplicial Complex (DSC). As explained in Section 3.2, DSC is a deformable model which supports topological changes during deformation and takes care of improving the quality of the mesh.

Fig. 4.1 shows an overview of the proposed method in Contribution A. From a given configuration of the mesh, we use a specific forward model to compute the projections. These computed projections are compared with projection data to estimate attenuation coefficients and the displacements of the interface vertices. The displacements are computed based on the curve evolution equation (2.21), but we extend it to multiple materials. From these estimated displacements, we use DSC to deform the mesh and repeat the procedure, until a stopping criterion is met.

In Fig. 4.2, a simple experiment shows an advantage of our work in terms of topological changes during deformation for reconstructing an object consisting of disconnected components. We compare our method with the previous work [DDH18] from the same initialization by one closed circle. As shown in the figure, the proposed method can split the initial circle into 5 separate components, which matches with the ground truth phantom. On the other hand, the previous work [DDH18] keeps the topology fixed from the beginning, which yields a degenerate result.

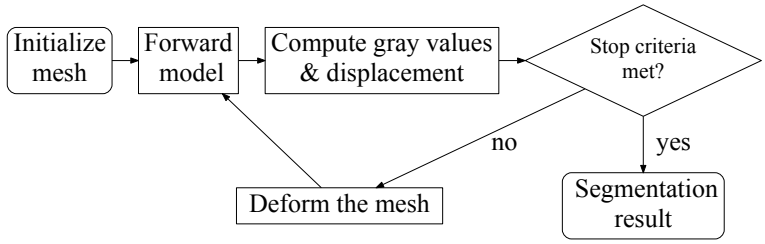
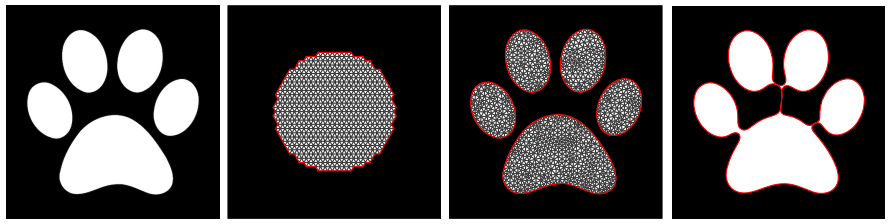


Figure 4.1: An overview of the proposed method in Contribution A.



Phantom | Initialization | Result of ours A | Result of [DDH18]

Figure 4.2: An example of reconstructing the phantom (left). The proposed method can change the topology from the initial sphere (middle-left) to 4 connected components (middle-right). On the other hand, the preliminary work [DDH18] keeps the initial topology fixed, which yields the degraded segmentation result (right).

4.2 Surface segmentation from projections in 3D (Contribution B)

In the previous section, we have seen 2D tomographic segmentation from projections. DSC, which we used in Contribution A], has also been implemented in 3D where it is based on tetrahedral mesh. This means that our 2D method might be extended to 3D segmentation. However, this would require expensive computational costs in terms of memory usage and running time required by DSC. Instead, this section addresses 3D surface reconstruction from projections in another way, by extending a differentiable rendering technique explained in Section 3.3 under the assumption that the topology of the objects is given. It means that we need to initialize meshes such a subdivided icosphere with the correct topology of the objects. In this section, we consider parallel beam geometry, where the rays are perpendicular to the detector plane.

In Contribution B, we extend the differentiable rasterization technique explained

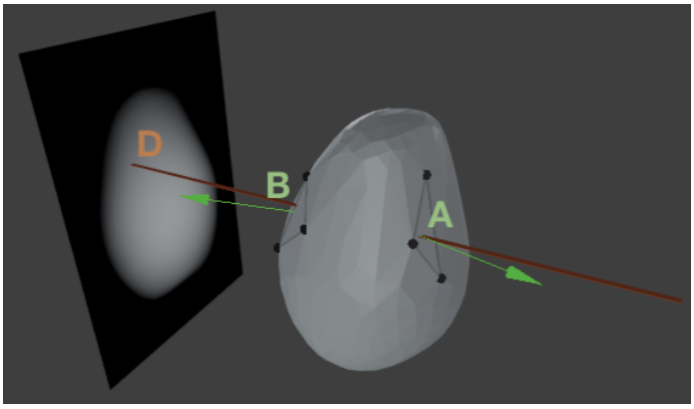


Figure 4.3: Forward projection of a homogeneous object represented by a triangular mesh for a given ray (darkred). The projection value for the ray can be computed by the signed sum: $\mu\,(+\overline{AD}-\overline{BD})$, where μ is the constant attenuation value, $\overline{AD}$ is the length between two points A and D and $\overline{BD}$ between B and D. Here, the signs for $\overline{AD}$ and $\overline{BD}$ are determined by the sign of the dot product of the ray vector and the normal vectors (green).

in Section 3.3.2 to tomographic reconstruction. I first explain the connection between the rasterization rendering technique and the forward model procedure in tomography, known as forward projection. Whereas the rendering process only considers the nearest triangle to the camera plane for an opaque object, the forward projection in tomography needs to consider all the triangles, as the ray passes through the object. The forward projection of a triangular mesh is illustrated in Fig. 4.3, where the object is assumed to be homogeneous. As shown in the figure, given an attenuation coefficient μ and a ray, the projection value for the ray can be computed by the multiplication of μ and the thickness of the object along the ray ($\overline{AD} - \overline{BD}$ in the figure).

How can we compute the length such as $\overline{AD}$ in the figure? One can compute $\overline{AD}$ by finding the exact point A, the intersection point of the ray and the corresponding triangle. However, we want to write down $\overline{AD}$ in terms of three vertices on the triangle containing A, so that the forward projection can be differentiable with respect to the vertex positions.

To represent $\overline{AD}$ in terms of the vertices, we first assign to each vertex on the triangle its distance to the detector. Then, the length $\overline{AD}$ can be given by the barycentric interpolation from three assigned distance values on the vertices. In a similar way, we can compute the length $\overline{BD}$ and so the thickness for the ray in the figure can be obtained by $\overline{AD} - \overline{BD}$. Here, we assign the positive sign for

$\overline{AD}$ and the negative sign for $\overline{BD}$. These signs are determined by the sign of the dot product of the ray direction vector and the normal vector of the triangle.

In summary, the forward projection of a triangular mesh for a homogeneous object can be interpreted in terms of a rendering process, where we need to consider all the triangles and assign to each vertex its distance to the detector, instead of color values as in the rendering technique. Based on this connection, we extend the differentiable rasterization work discussed in Section 3.3.2, to make the forward projection process differentiable in terms of the vertex positions. Using this differentiable forward projection, we propose a shape estimation method from projections in Contribution B.

The proposed method in Contribution B begins from a template mesh such as an icosphere and moves the vertices, but does not change the connectivity of the mesh. To deform the mesh, we use the gradient descent method with respect to the vertices of the mesh where the gradient is computed from the proposed differentiable forward projector. It is worthwhile to mention that in Contribution A, the displacements are computed only along the normal directions of the vertices. This is because, for example, the curve evolution equation (2.21) for the objective function (2.14) is based on the steepest descent method in a space of smooth closed curves, not in a Euclidean space [Sun11]. (Note that depending on the objective function, the curve evolution equation can have the tangent components [CKS95].) In contrast, in Contribution B, the unknowns are the positions of the vertices where the solution space is a Euclidean space. Hence, the method in Contribution B has no restriction on the direction of displacements.

Fig. 4.4 shows an example of deformation by the proposed method to estimate a surface with genus 1. This topology information is assumed to be known and the proposed method begins with a torus, as the torus is one of the simplest shapes with genus 1. In the beginning, the intermediate meshes can have few degraded triangles. To deal with such degenerate triangles, we add regularization terms and refine the meshes several times during the deformation. This mesh refinement also helps the method converge fast. As shown in the figure, our method yields a high-quality mesh around the iteration 360.

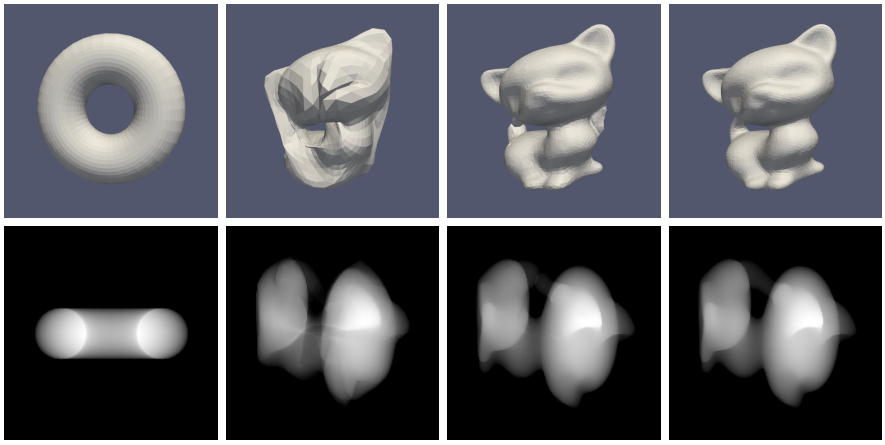


Figure 4.4: Deformation example for the kitten data model from a torus. The top row shows the intermediate meshes during deformation and the bottom row shows the corresponding computed projections. The mesh was refined at the iterations 60, 120 and 180.

4.3 Spatial regularization for tomography using coordinate-based neural networks (Contribution C)

As mentioned in Section 3.5, SIREN [SMB+20] is a fully connected neural network with sine activation functions and inputs spatial coordinates and outputs the signals for the coordinates. Contribution D applies SIREN to tomographic reconstruction such that the output of the network is the attenuation coefficient values on the coordinates.

Specifically, we use 3 hidden layers where an output of a hidden layer can be written as

$$h(\mathbf{z}) = \sin\left(\omega(\mathbf{W}\mathbf{z} + \mathbf{b})\right) \tag{4.1}$$

where $\mathbf{W}$ and $\mathbf{b}$ are the neural network weights to optimize, ω is a constant to control the frequency of the signal and sin is the sine function. We use the rectified linear unit (ReLU) as the activation function in the last layer, where ReLU maps negative input values to zeros. Using this network, we represent an object by a continuous function in the sense that we can evaluate the attenuation coefficient function for any spatial coordinates.

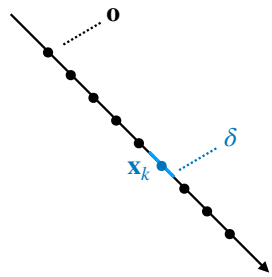


Figure 4.5: Illustration of uniform sampling for one ray with the origin **o** and the constant interval δ.

Given such continuous representation, the forward projection can be computed, using a numerical integration method, since we can evaluate the attenuation function at any spatial coordinate. We used a simple numerical integration method called the mid-point rule. As shown in Fig.4.5, in the mid-point rule, we sample the points $\mathbf{x}_k$ uniformly with a constant interval δ. Then, the estimated projection for a ray is given by $\sum_k f(\mathbf{x}_k)\delta$ where f represents the neural network function. With such forward model, we aim to minimize the misfit between our estimation and projection data in the L_2 sense.

To deal with noisy data and obtain high-quality results, we consider a regularization term. Inspired by total variation regularization, in Contribution C, we propose a spatial smoothness term along only one direction of the ray. This way, we impose a prior that the network outputs of neighboring locations should be similar, but it is only imposed on along one direction for computational efficiency.

Fig. 4.6 shows some intermediate results for reconstructing a phantom with different frequency parameters ω, while the object space is normalized to $(-1, 1)^2$. We can clearly see the effect of ω, where the result from the low value $\omega = 10$ could not capture the fine details. We observe that in most cases, the value $\omega = 30$ gives the best performance, while the parameter ω with higher values can yield some artifacts.

4.4 Vectorial total variation based on L_∞ norm for spectral CT (Contribution D)

As discussed in Section 2.5, one important question for iterative methods in spectral CT is how to model a robust regularization term to correlate the

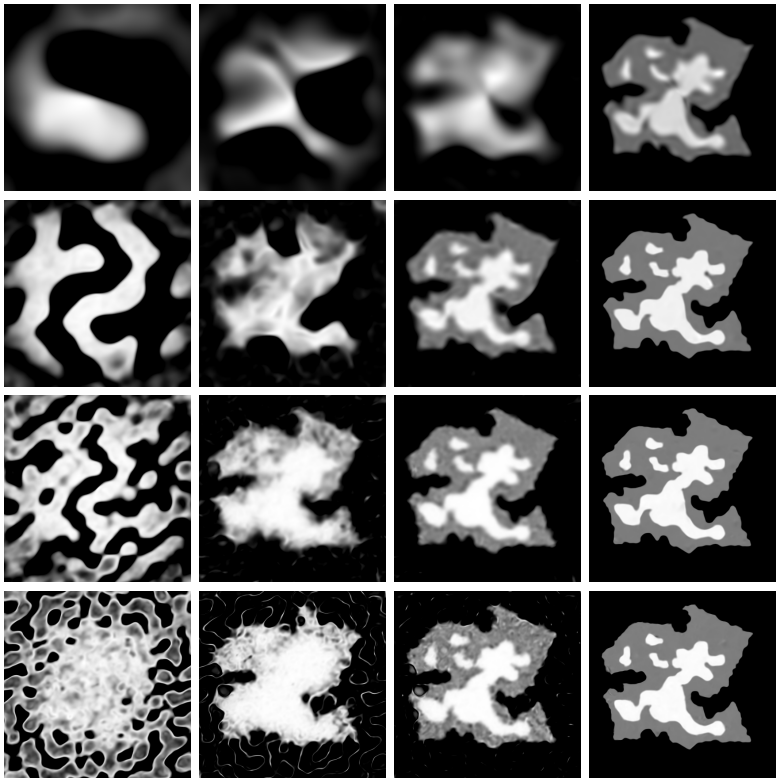


Figure 4.6: Intermediate results with varying the frequency parameter $\omega = 10$, 30, 50, and 70 (top-to-bottom).

multi-channel images from different X-ray energies. Total Nuclear Variation (TNV) (2.25) is one of the popular regularization methods for spectral CT. However, we observe that TNV can be vulnerable to noise, when the number of X-ray energies considered high (e.g. 12 energy bins). This may be because TNV often tries to align the image gradients with the noises over the multi-channel images, instead of aligning with object features.

In spectral CT, we want to reconstruct multi-channel images $\mathbf{u}_1, ..., \mathbf{u}_{n_e}$ representing energy-dependent attenuation coefficients from n_e projections, where n_e is the number of X-ray energies considered. As a way of regularization to correlate such multi-channel images, in Contribution D, we suggested to use a

vectorial total variation term based on the L_∞ norm, defined by

$$R_{\mathrm{VTV}}^{L_\infty}(\{\mathbf{u}_1, ..., \mathbf{u}_{n_e}\}) = \sum_{j=1}^{J} \left(\max_{1\leq i\leq n_e} |\partial_x u_{i,j}| + \max_{1\leq i\leq n_e} |\partial_y u_{i,j}| \right) \tag{4.2}$$

where J is the number of pixels in one channel image and $\partial_x u_{i,j}$ and $\partial_y u_{i,j}$ are the partial derivatives of the image $\mathbf{u}_i$ for the j-th pixel with respect to x and y directions, respectively. This term was suggested for color image denoising problems in [MS12], but has not been investigated for spectral CT data, to our knowledge. In this section, we call the reconstruction method using $R_{\mathrm{VTV}}^{L_\infty}$ as L_∞-VTV.

To analyze the effects of the different regularization terms, we compare several iterative methods based on Principal Component Analysis (PCA) [Jol86]. PCA, in a geometric viewpoint, attempts to find a low dimensional (affine) subspace to best fit a given set of points in a Euclidean space. The estimated low dimensional space can give insight on the original high dimensional points.

Here, PCA is used to visualize the patterns of reconstructed images by SIRT, TV, TNV and L_∞-VTV, where SIRT and TV reconstruct multi-channel images, independently. For this experiment, we use a real projection data (obtained from the 3D ImagingCenter at DTU) for a single-material object made of aluminum with 15 X-ray energy bands. We view one channel image as a point in a Euclidean space $\mathbb{R}^J$. Then, the reconstructed 15 multi-channel images are fitted to a low dimensional space with dimension 4 and the aim is to visualize the basis of this low dimensional space, known as *principal components*.

Fig. 4.7 shows several reconstructed images from different methods and the estimated principal components from the results. As SIRT has no regularization effects, the principal components show irregular non-smooth patterns. The principal components by TV have smooth background overall, but have some artifacts inside the object. The first principal components by TNV and L_∞-VTV look similar, but TNV yields other principal components with some artifacts. A possible explanation for this might be that TNV tries to align not only the boundaries, the but also the noises. On the other hand, the principal components by L_∞-VTV look smooth except for the boundary regions.

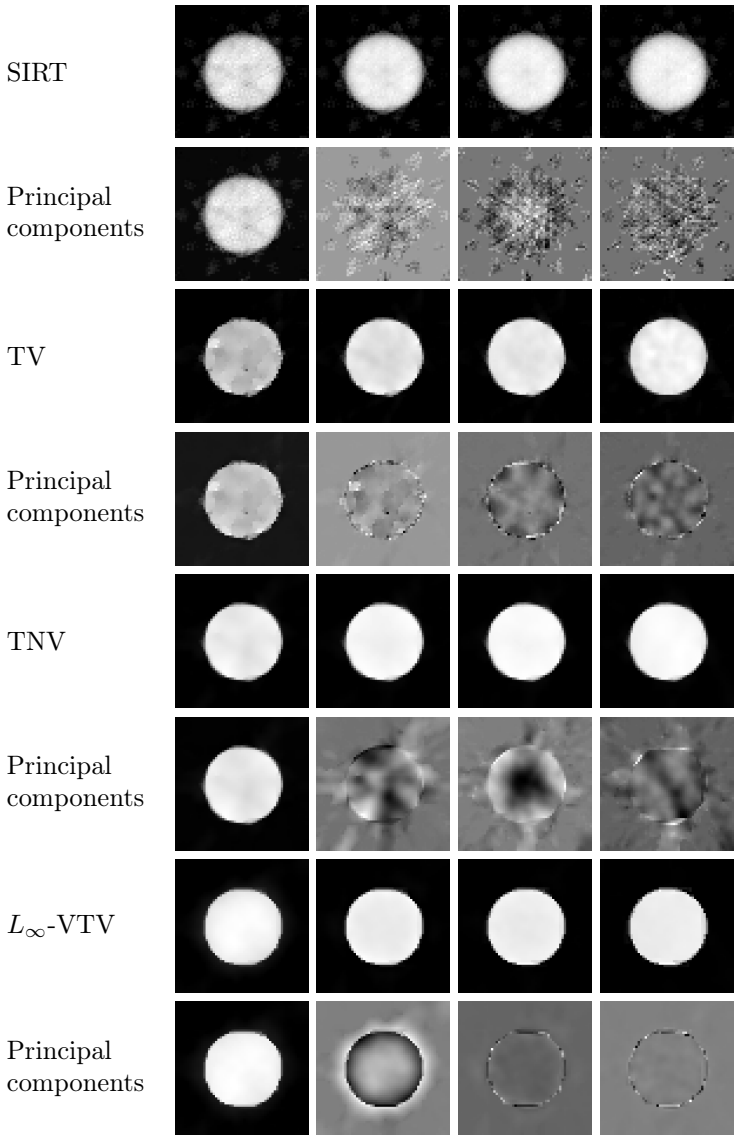


Figure 4.7: Reconstruction results and the principal components by different methods. Odd rows show four reconstruction images among 15 multi-channel images. Even rows show the 1st, 2nd, 3rd and 4th principal components which are estimated from the reconstructed images with 15 channels.

Chapter 5

Conclusion

The thesis has studied tomographic reconstruction methods and proposed several new methods which have been carried out during the Ph.D. study. The proposed methods have been shown to be applicable to X-ray CT, electron tomography or spectral CT.

The major work of the thesis is the direct tomographic segmentation methods from projections based on mesh deformation. Unlike the conventional pipeline of tomography, the proposed methods can estimate the segmentation directly from projections data for homogeneous objects. Contribution A addresses mesh deformation in 2D domain to align with object boundaries. One key advantage of this work is to support topological changes during deformation, so that the topology of the initialization mesh can be arbitrary.

Contribution B tackles 3D shape estimation with the assumption that the topology of objects is known and is kept fixed during deformation. Compared with other reconstruction algorithms, mesh-based methods require more computational costs, but can handle severely limited and noisy data. This makes them very suitable for problems where data is sparse, but we have a good prior knowledge about the objects to be reconstructed. We have found one such problem in electron tomography, where our 3D shape estimation has shown to be useful for reconstructing nano-particles. But there are other situations where mesh-based approaches may prove useful, for example in medical CT, where approximate

shape of organs is known, and low radiation dose limits data quality, or in dynamic tomography to estimate the dynamic behavior of an object. Furthermore, due to the differentiability, the proposed forward projection model can be easily incorporated into a deep learning framework.

We have also investigated another representation of objects using a coordinate-based neural network for tomographic reconstruction with a new spatial regularization term. Representing reconstructed object using coordinate-based neural network might offer benefits compared to image-based representations. For example, it may allow for capturing finer detail, not limited by the grid resolution. While our work has not been mature enough for a practical usage, it can add to our understanding of an exciting path.

Lastly, we have studied a regularization scheme for spectral CT reconstruction based on vectorial total variation in Contribution D. This regularization scheme shows the potential to be of practical use for security applications and may be valuable to any researcher using iterative methods for spectral CT data.

In summary, the contributions of the thesis are as follows:

- We have proposed a method for segmentation of homogeneous objects from projections based on mesh deformation with the support of multiple materials and topological changes during deformation in Contribution A.
- We have developed a differentiable forward projector for 3D triangular meshes representing homogeneous objects in Contribution B.
- We have proposed a direct 3D shape estimation method from projections and applied the proposed method to electron tomography data Contribution B.
- We have investigated a coordinate-baesd neural network for tomographic reconstruction and proposed a spatial regularization term in this framework in Contribution C.
- We have studied a regularization scheme for spectral CT reconstruction based on vectorial total variation based on the L_∞ norm in Contribution D.

I believe that the contributions of the thesis are of value to future research and practical applications of computed tomography.

Bibliography

[AK84] Anders H. Andersen and Avinash C. Kak. Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1), 1984.

[AKM11] Alireza Aghasi, Misha Kilmer, and Eric L. Miller. Parametric Level Set Methods for Inverse Problems. *SIAM Journal on Imaging Sciences*, 4(2), 2011.

[AY04] Christopher V. Alvino and Anthony J. Yezzi. Tomographic reconstruction of piecewise smooth images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2004.

[BC17] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer International Publishing, Cham, second edition, 2017.

[BS11] K. Joost Batenburg and Jan Sijbers. DART: A Practical Reconstruction Algorithm for Discrete Tomography. *IEEE Transactions on Image Processing*, 20(9), 2011.

[Buz08] Thorsten M. Buzug. *Computed Tomography from Photon Statistics to Modern Cone Beam CT*. Springer, 2008.

[BV18] Stephen P. Boyd and Lieven Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge University Press, Cambridge, UK ; New York, NY, 2018.

[Byr14] Charles L Byrne. *Iterative Optimization in Inverse Problems*. Monographs and Research Notes in Mathematics. CRC Press, Taylor & Francis Group, Boca Raton, 2014.

[CGL+19] Wenzheng Chen, Jun Gao, Huan Ling, Edward J. Smith, Jaakko Lehtinen, Alec Jacobson, and Sanja Fidler. Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer. In *Advances in Neural Information Processing Systems*, 2019.

[CKS95] Vincent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. In *International Conference on Computer Vision*, Cambridge, MA, USA, 1995. IEEE Comput. Soc. Press.

[CP11] Antonin Chambolle and Thomas Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1), 2011.

[CP16] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25, 2016.

[DDH18] Vedrana A. Dahl, Anders B. Dahl, and Per C. Hansen. Computing segmentations directly from x-ray projection data via parametric deformable curves. *Measurement Science and Technology*, 29(1), 2018.

[DMSC16] J. Duran, M. Moeller, C. Sbert, and D. Cremers. Collaborative Total Variation: A General Framework for Vectorial TV Models. *SIAM Journal on Imaging Sciences*, 9(1), 2016.

[EST20] Moshe Eliasof, Andrei Sharf, and Eran Treister. Multi-modal 3D shape reconstruction under calibration uncertainty using parametric level set methods. *SIAM Journal on Imaging Sciences*, 13(1), 2020.

[EW01] Vidya Elangovan and Ross T. Whitaker. From Sinograms to Surfaces: A Direct Approach to the Segmentation of Tomographic Data. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2001.

[Haw92] Peter W Hawkes. The electron microscope as a structure projector. In *Electron Tomography*. Springer, 1992.

[HSF12] Björn J Heismann, Bernhard T Schmidt, and Thomas Flohr. Spectral computed tomography. SPIE Bellingham, WA, 2012.

[Jol86] Ian T Jolliffe. Principal components in regression analysis. In *Principal Component Analysis*. Springer, 1986.

[KS88] Avinash C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. IEEEE Press, 1988.

[KUH18] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3D Mesh Renderer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[KvB18] Ajinkya Kadu, Tristan van Leeuwen, and K. Joost Batenburg. A parametric level-set method for partially discrete tomography. In *International Conference on Discrete Geometry for Computer Imagery*, 2018.

[KWT88] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), 1988.

[LCLL19] Shichen Liu, Weikai Chen, Tianye Li, and Hao Li. Soft Rasterizer: Differentiable Rendering for Unsupervised Single-View Mesh Reconstruction. In *International Conference on Computer Vision (ICCV)*, 2019.

[MB12] Marek Krzysztof Misztal and Jakob Andreas Bærentzen. Topology-adaptive interface tracking using the deformable simplicial complex. *ACM Transactions on Graphics*, 31(3), 2012.

[MS12] Takamichi Miyata and Yoshinori Sakai. Vectorized total variation defined by weighted L infinity norm for utilizing inter channel dependency. In *IEEE International Conference on Image Processing*, 2012.

[MST+20] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.

[NW99] Jorge Nocedal and Stephen J Wright. *Numerical Optimization*. Springer, New York, 1999.

[OS88] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, 79(1), 1988.

[PB14] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3), 2014.

[RBM03] Osborne Reynolds, Arthur William Brightmore, and William Henry Moorby. *Papers on Mechanical and Physical Subjects: The Sub-Mechanics of the Universe*, volume 3. The University Press, 1903.

[RLR15] David S Rigie and Patrick J La Rivière. Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization. *Physics in Medicine and Biology*, 60(5), 2015.

[SMB+20] Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit Neural Representations with Periodic Activation Functions. *arXiv:2006.09661*, 2020.

[Sun11] Ganesh Sundaramoorthi. Lecture Notes on Mathematical Models in Computer Vision. https://sites.google.com/site/ganeshsun/teaching/ee396, 2011.

[WE02] Ross T. Whitaker and Vidya Elangovan. A direct approach to estimating surfaces in tomographic data. *Medical Image Analysis*, 6(3), 2002.

[ZJL20] Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. Physics-Based Differentiable Rendering: A Comprehensive Introduction. https://shuangz.com/courses/pbdr-course-sg20/, 2020.

[ZY96] Song-Chun Zhu and Alan Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9), 1996.

Contribution A

DALM, Deformable Attenuation-Labeled Mesh for Tomographic Reconstruction and Segmentation

---

The following paper has been published in IEEE Transactions on Computational Imaging, vol. 7, pp. 151-163, 2021.

DALM, Deformable Attenuation-Labeled Mesh for Tomographic Reconstruction and Segmentation

Jakeoung Koo, Anders Bjorholm Dahl, and Vedrana Anderson Dahl

***Abstract*—Most X-ray tomographic reconstruction methods represent a solution as an image on a regular grid. Such representation may be inefficient for reconstructing homogeneous objects from noisy or incomplete projections. Here, we propose a mesh-based method for reconstruction and segmentation of homogeneous objects directly from sinogram data. The outcome of our proposed method consists of curves outlining the regions of constant attenuation, and this output is represented using a labeled irregular triangle mesh. We find the solution by deforming the mesh to minimize the residual given by the sinogram data. Our method supports multiple materials, and allows for topological changes during deformation. An integral part of our algorithm is an efficient forward projection of the labeled mesh onto the sinogram domain. We initialize our algorithm based on graph total variation, also here taking advantage of the mesh representation. Experimental results on simulated datasets show that our method gives a compact representation of the reconstruction and also accurate segmentation results for challenging data with e.g. large noise, a small number of angles or problems with limited angle. We also demonstrate the result on real fan-beam data. The proposed geometric solution shows a further step towards using alternative representations for tomographic reconstruction.**

***Index Terms*—Deformable models, tomographic reconstruction, tomographic segmentation.**

I. Introduction

In X-RAY CT, the aim of the tomographic reconstruction is typically to compute an image of a scanned object with image voxels representing attenuation coefficients arranged on a regular grid [1]. This representation is well-suited for a wide range of reconstruction problems. However, for homogeneous objects with simple geometry, grid-based representation may be very redundant.

In this work, we propose a reconstruction and segmentation method using an interface-based representation in the form of a labeled deformable mesh and we investigate the properties of our method. The proposed method is targeted at objects composed of homogeneous components. Our work extends the method by Dahl *et al.* [2] that employs a deformable closed curve to outline one object in the reconstruction. Here, we replace one single curve by an interface represented as edges in a labeled mesh, and we deform the interface based on Deformable Simplicial Complex (DSC) proposed by Misztal and Bærentzen [3]. This gives our method several advantages compared to [2]. First, we can reconstruct multiple objects thanks to topological adaptivity of DSC. Second, as DSC supports multi-label segmentation, our method also supports objects with different materials. Third, DSC may employ either fixed or adaptive mesh resolution, which gives an additional flexibility to our method. Fourth, while in [2] an initialization is done by circle, meshing of the reconstruction domain allows us to efficiently initialize our algorithm by finding a configuration close to the desired solution.

In Fig. 1 we show an example of a labeled mesh before and after the deformation. To obtain a segmentation, we compute the

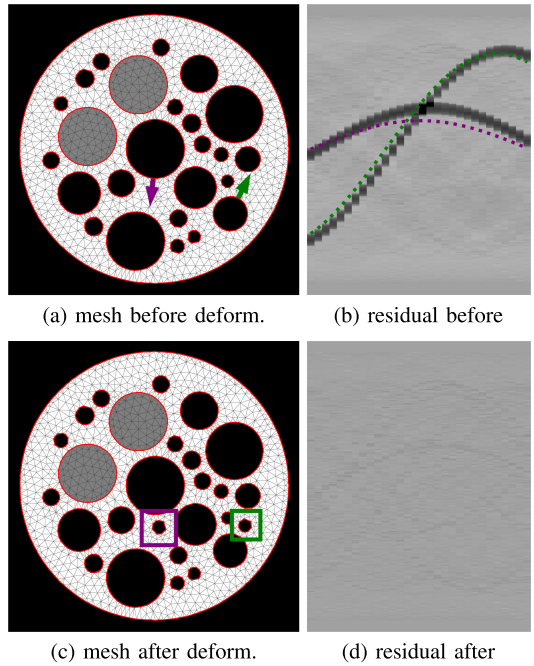


Fig. 1. Mesh deformation minimizes the residual. A mesh configuration (a) with triangle labels indicated by colors (black, gray and white) and interface drawn in red. This configuration yields a residual (b) where the dark regions denote large values. Two large displacements in the normal direction are shown as the arrows in (a) and their corresponding residuals are shown as dotted lines in (b). After deformation, we obtain the mesh with two new regions (c) shown in rectangles, with the residual (d).

Manuscript received July 2, 2020; revised October 31, 2020 and January 8, 2021; accepted January 9, 2021. Date of publication January 15, 2021; date of current version February 8, 2021. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement 765604. *(Corresponding author: Jakeoung Koo.)*

The authors are with the Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Lyngby, Denmark (e-mail: jakoo@dtu.dk; abda@dtu.dk; vand@dtu.dk).

Digital Object Identifier 10.1109/TCI.2021.3052034

displacements of the interface vertices to minimize the residual between data and the computed sinogram from the mesh. While moving those vertices, the mesh quality is maintained by splitting or merging nearby faces. These local mesh operations allow two new pores to be created from nearby pores.

A. Related Works

Two approaches to finding segmentations directly from sinogram data are related to our work: *discrete-valued tomography* and *deformable models*. We will review those two approaches in addition to *mesh-based reconstruction methods*.

Discrete-valued tomography methods assume that the reconstruction image has a small number of discrete attenuation coefficients. This acts as a prior to deal with situations where there are a small number of homogeneous objects from small angular range and noisy sinogram data. Batenburg and Sijbers in [4] proposed the DART method, which combines an existing iterative reconstruction method and a thresholding scheme. To improve DART's robustness to noise and to automatically compute the attenuation coefficients, extensions have been suggested in [5], [6], and [7].

The method in [8] splitted the variables representing attenuation coefficients into continuous and discrete. This splitting turns the problem into two simple problems solved by the conjugate gradient method and a submodular minimization solver. In [9] the authors suggested a joint reconstruction and segmentation method in a variational framework. Other techniques include graph cut [10], [11] and convex relaxation techniques [12]. These discrete-valued tomography methods relied on a regular image grid, while DALM does not restrict the solution on the image domain. Among these works, we choose one of the state-of-the-art works [7] to compare our segmentation results with.

Deformable models from sinogram data have been studied mostly based on level set methods [13]. Whitaker and Elangovan [14] formulated the Mumford-Shah piecewise constant model [15] in terms of level-set to minimize the reprojection error. This work was extended to a piecewise smooth model by Alvino and Yezzi [16]. Extending two-phase segmentation, a multi-phase piecewise constant model was studied in [17], [18]. The stated deformable models required dense and regular discretization. To reduce the unknown variables, a parametric level set method was proposed in [19] and used in [20]. Those methods required fewer parameters to represent a level set, which reduced the unknown variables and allows to use efficient second-order optimization methods.

Such level-set models rely on regular image grids and their forward projections use an image-based forward model, whereas the proposed method can directly project a labeled mesh into sinogram domain. Instead of level-set, our previous work [2] used an explicit representation, but limited to one simple closed object. We extend it to consider multiple materials and allow topological changes.

Several *mesh-based reconstruction methods* have been proposed. Such methods aim to find an adaptive representation of objects to match with the structure of objects and improve the reconstruction quality. In [21] Brankov *et al.* suggested a method to estimate an initial mesh configuration from the reconstruction image obtained by pixel-based filtered backprojection. This dependency on an analytical method could affect the mesh estimation step severely in the case of limited sinogram data. Cazasnoves *et al.* [22] employed a more sophisticated strategy to build an irregular mesh before reconstruction. They detected edges in sinogram data to extract structures in 2D and merge them to estimate the interface of objects in 3D. Around the interface, a more fine mesh was constructed and then the reconstruction was performed. Instead of estimating mesh configuration once, adaptive mesh refinement methods have been proposed. In [23], [24], a tetrahedron mesh was generated from a coarse regular grid and the Expectation Maximization (EM) method was used for reconstruction on the mesh. The mesh was finely refined around the nodes having large attenuation value variations. They repeated reconstruction and refinement in a coarse-to-fine way. On the other hand, the method in [25] began with a fine uniform grid and alternately use EM and coarsen the mesh, repeatedly. In [26] they iteratively refined the mesh based on the segmentation result by the level set method.

Unlike existing mesh-based reconstruction methods, our method aims to deform the interface between materials and uses the mesh as an auxiliary structure due to the assumption of a small number of materials.

B. Summary of Contributions and Outline

The main contribution of our work is to develop a mesh-based deformable method from sinogram data which supports multiple materials and topological changes during deformation (Sec. III). We develop an efficient forward projection algorithm mapping a labeled mesh into the sinogram domain. To avoid local optima, we employ a robust initialization scheme based on graph total variation on the mesh (Sec. IV). The numerical results are presented in noisy and incomplete synthetic data and in real fan-beam data (Sec. VI).

II. Problem Formuation

In this section, we formulate the tomographic segmentation problem to estimate the curves to segment the regions whose attenuation values are approximately homogeneous. We follow the similar formulation in [14], [16] which consider two-region segmentation, but we extend it to multi-region cases inspired by the region competition method [27]. Similar to such works, our formulation is based on variational methods where an energy over a continuous space is formulated and a low energy reflects a desired solution. Here, our aim is to find the minimizer. In the next section, we will propose a concrete algorithm to minimize the energy.

The sinogram data $p(\theta, s)$ is produced by projecting and rotating an unknown collection of objects onto the detector, and by θ we denote the projection angle, while s denotes the detector position.

Tomographic segmentation problem for homogeneous objects is to divide the reconstruction domain $\Omega \subset \mathbb{R}^2$ into M mutually disjoint regions $\{R_m\}$ such that $\Omega = \cup_{m=1}^{M} R_m$ and $R_i \cap R_j \neq \emptyset$ for $i \neq j$. Each region is associated to a homogeneous object with attenuation coefficient μ_m. We aim to find the curves $\{C_m\}$ to align with the boundaries of unknown objects such that $C_m = \partial R_m$ for each region R_m.

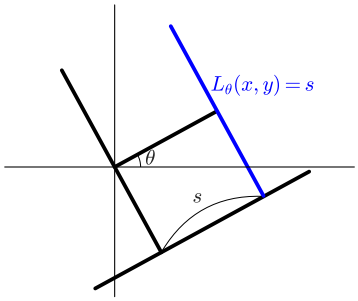


Fig. 2. Illustration of line equations for parallel beam. Given a position s along the detector, blue line represents the line equation $L_\theta(x,y)-s=0$.

From this configuration, we define $\mu_m\hat{p}_m$ as forward projection of a region R_m with the attenuation value μ_m where

$$\hat{p}_m(\theta,s)=\int_{R_m}\delta(L_\theta(x,y)-s)\,\mathrm{d}x\,\mathrm{d}y, \tag{1}$$

and δ is the delta function, s the detector position, θ the projection angle and $L_\theta-s=0$ a line ray equation. For the parallel beam, L is given by $L_\theta(x,y)=x\cos\theta+y\sin\theta$ (See Fig. 2). By combining all M materials, we define $\hat{p}$ as a linear combination of all the contributions

$$\hat{p}(\theta,s)=\sum_{m=1}^{M}\mu_m\hat{p}_m(\theta,s). \tag{2}$$

From the estimation $\hat{p}$, we aim to fit it to the sinogram data p, by minimizing the reprojection error $(p-\hat{p})^2$. This fitting term can lead to an undesirable solution if data is noisy or incomplete. To prevent it, we introduce a regularization term to make the curves smooth by penalizing their large lengths. With this smoothness term, we define the energy to minimize:

$$E(\{\mu_m,C_m\})=\frac{1}{2}\sum_{\theta,s}(p(\theta,s)-\hat{p}(\theta,s))^2+\lambda\sum_{m=1}^{M}\mathrm{Len}(C_m), \tag{3}$$

where λ is the weighting parameter for the regularization term and Len denotes the length of a closed curve, defined by $\mathrm{Len}(C_m)=\int_{C_m}\mathrm{d}r$, with arc length parameter r [28]. Choosing the optimal regularization parameter λ in (3) is not straightforward and other variational approaches [14], [16] have the same problem. We often have an idea of how object shapes look like before scanning. In this case, the prior knowledge on objects could help choose the parameter.

Given attenuation values $\{\mu_m\}$, we derive a curve evolution equation. Consider a curve C_m which encloses the region R_m with the attenuation μ_m and the adjacent region R_n with μ_n. Then, the curve evolution equation for C_m can be derived as (see Appendix)

$$C_m^{k+1}(r)=C_m^k(r)+\tau\left((\mu_m-\mu_n)\sum_{\theta}(p(\theta,\tilde{s})-\hat{p}(\theta,\tilde{s}))+\lambda\kappa_m(r)\right)N_m(r), \tag{4}$$

where k denotes the iteration number, τ the step size, N_m the outward normal vector, κ_m the curvature of the curve and $\tilde{s}=L_\theta(C_m(r))$ the detector positions onto which the point $C_m(r)$ are projected.

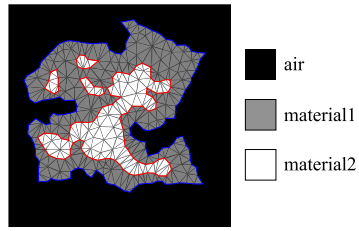


Fig. 3. Illustration of attenuation-labeled mesh. Each triangle is associated with one material. The blue edges indicate the interface between air and the material 1 and the red edges between the material 1 and the material 2. For the illustration, the triangles representing the background are omitted.

III. Deformable Attenuation-Labeled Mesh

In this section, we describe the proposed framework, deformable attenuation-labeled mesh (DALM), to optimize the energy in (3). We first propose our mesh-based representation, called ALM, and its forward projection to generate the estimated sinogram. From the estimation, we compute the attenuation values and deform ALM.

A. Attenuation-Labeled Mesh (ALM)

Consider a triangularization of a reconstruction domain $\Omega\subset\mathbb{R}^2$, where every triangle is labeled with a label m from $m\in\{1,\ldots,M\}$, and every label has an associated value μ_m. In terms of tomographic reconstruction, labels correspond to different materials, and label values correspond to material attenuations. While we can operate with an arbitrary number of materials, we assume that M is known as a prior and much smaller than the number of triangles.

The construction consisting of a triangle mesh, labels, and attenuation values is denoted by $\mathcal{X}$, and we call it *attenuation-labeled mesh (ALM)*. Any ALM configuration is fully defined by the list of mesh vertices, list of mesh triangles, list of triangle labels, and a list of label attenuations. An example of ALM is illustrated in Fig. 3.

Clearly, ALM gives a partitioning of the reconstruction domain Ω into different materials, i.e. regions R_m of attenuation μ_m, for $m=1,\ldots,M$. Furthermore, ALM gives an explicit representation of interface curves as a collection of *interface edges* – mesh edges whose two adjacent triangles are assigned two different labels. Furthermore, we define *interface vertices* as mesh vertices which are adjacent to at least one interface edge.

B. Forward Projection of ALM

We propose a forward projection algorithm to yield the estimated sinogram $\hat{p}$ from the interface edges of a configuration of ALM. Our forward model is based on the principle that a ray should hit a closed region an even number of times. Our forward algorithm has two advantages: First, we only need the information of the interface and its surrounding labels. Second, the forward projection can be done independently for each interface edge, which allows for efficient parallelization.

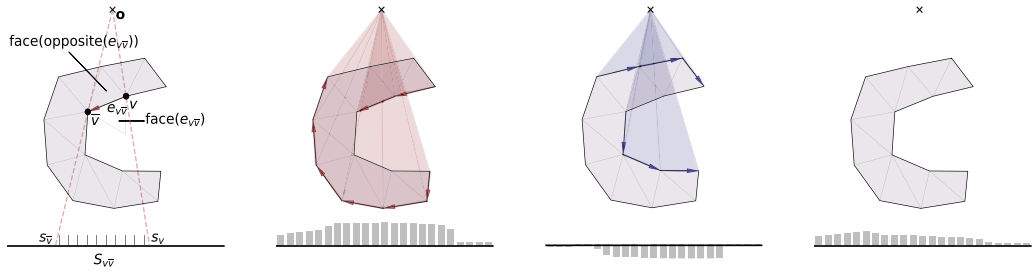


Fig. 4. Illustration of mesh representation and forward projection in fan-beam geometry for a projection angle. For the illustration purpose, we omit the triangles representing background and the interface edges are shown as darker than non-interface edges. (a) Half-edge based mesh is used to represent objects where each face is labeled with a material. The shaded region denotes a material and its outside denotes vacuum. The detector positions $s_v, s_{\bar{v}}$ are from the two vertices in a half-edge. Depending on the sign of $(s_v - s_{\bar{v}})$, we add (b) or subtract (c) the contribution of the segment between the source and the interface. After summing these values, we obtain (d) the final forward projection for a specific angle.

Algorithm 1: Forward Projection For One Angle.

Input: Attenuation-labeled mesh $\mathcal{X}$, a projection angle θ
Init: $\hat{p}_m(\theta, s) \leftarrow 0$ for every label m, detector pixel s
for all interface edge e_{ij} **do**
  $l \leftarrow$ label of triangle left of e_{ij}
  $k \leftarrow$ label of triangle left of e_{ji} (opposite triangle)
  $s_i \leftarrow$ projection of i onto the detector (continuous)
  $s_j \leftarrow$ projection of j onto the detector (continuous)
  for all detector pixel s between s_i and s_j **do**
    $o \leftarrow$ source position
    $x \leftarrow$ intersection of e_{ij} and ray $\overline{os}$
    $c \leftarrow$ length of $\overline{ox}$
    $\hat{p}_l(\theta, s) \leftarrow \hat{p}_l(\theta, s) + \text{sign}(s_i - s_j) \cdot c$
    $\hat{p}_k(\theta, s) \leftarrow \hat{p}_k(\theta, s) + \text{sign}(s_j - s_i) \cdot c$
  end for
end for
return $\hat{p}(\theta, s) \leftarrow \sum_{m=1}^{M} \mu_m \hat{p}_m(\theta, s)$

In Fig. 4, we provide an illustration of our forward algorithm for fan-beam geometry given one projection angle. We (b) add or (c) subtract the contribution between the source point and the interface, depending on the orientation of the half-edges. The sign of the contribution is determined by the sign of the difference of detector positions $s_v, s_{\bar{v}}$ corresponding to the initial and end vertices $v, \bar{v}$ in the edge, respectively. Then, we (d) sum those signed contributions. The whole procedure for one projection angle is summarized in Algorithm 1. After performing forward projection for all edges and angles, we obtain the final estimated sinogram.

C. Estimation of Attenuations

In the energy (3), we have two unknowns of attenuation values and interface curves. Our strategy is to optimize them separately. Here, we fix the interface curves and optimize M attenuation values only. The optimal condition with respect to μ_m reads

$$0 = \frac{\partial E}{\partial \mu_m} = -\sum_{\theta,s} (p(\theta, s) - \hat{p}(\theta, s))\hat{p}_m(\theta, s), \qquad (5)$$

which leads to a system of linear equations to solve:

$$\begin{pmatrix} \sum_{\theta,s} \hat{p}_1^2 & \cdots & \sum_{\theta,s} \hat{p}_1 \hat{p}_M \\ \vdots & \ddots & \\ \sum_{\theta,s} \hat{p}_M \hat{p}_1 & & \sum_{\theta,s} \hat{p}_M^2 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_M \end{pmatrix} = \begin{pmatrix} \sum_{\theta,s} \hat{p}_1 p \\ \vdots \\ \sum_{\theta,s} \hat{p}_M p \end{pmatrix}, \qquad (6)$$

where we omit (θ, s) in $\hat{p}_m$, p for simplicity of notation. The matrix on the left side is symmetric and positive semi-definite. In many cases, the background of a material is known as air or vacuum, so its attenuation coefficient can be set as zero. In this case, we only need to estimate $M - 1$ attenuation coefficients. If the number of unknown attenuation coefficients in (6) is less than 4, we can use the closed-form solution to compute the inverse of the matrix of size 2-by-2 or 3-by-3 in the left hand side. Otherwise, we can use Cholesky factorization [29] to solve (6).

D. Deformation of ALM

Given attenuation values, we now deform the interface in the direction of minimizing the energy (3). In ALM, each vertex v on the interface is associated with two materials: the inside μ_v and the outside $\tilde{\mu}_v$. For example, μ_v corresponds to the material for the triangle: `face(opposite(`$e_{v\bar{v}}$`))` in Fig. 4(a). For each interface vertex v, we compute the displacement from the current position, by the discretization of (4), as follows:

$$\tau \left((\mu_v - \tilde{\mu}_v) \sum_{\theta} (p(\theta, s_v) - \hat{p}(\theta, s_v)) + \lambda \kappa_v \right) N_v, \qquad (7)$$

where N_v denotes the outward unit normal vector for the vertex v and κ_v is the curvature for the vertex v which can be computed by the discrete version of Frenet formula [28]. The detector position s_v denotes the position determined by the vertex v and projection geometry, as shown in Fig. 4(a). Note that s_v needs to be interpolated, as detectors have finite measurements.

From the computed displacements, we aim to deform the interface. To allow for topological changes during evolution, we employ a mesh-based deformable model, Deformable Simplicial Complex (DSC) [3]. Here we briefly explain the principle of DSC.

DSC moves the interface vertices sequentially towards destinations, while improving the mesh quality. The improvement

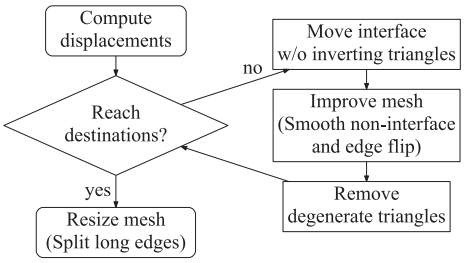


Fig. 5. A diagram of the deformation step in DSC. The loop is done for one vertex at a time. After all the vertices reach the destination, the mesh is resized.

step includes smoothing non-interface vertices, edge flipping and removing degenerate faces. The destination of each interface vertex is computed from (7) which is the displacement from the current position. When all the interface vertices reach their destinations, DSC refines the mesh further by splitting long edges, which are not on the interface, based on the parameter of average edge length l_1. This parameter allows some range of edge lengths around l_1. At this last step, DSC does not split the interface edges.

We provide a summary of DSC deformation in Fig. 5. We refer to the paper [3] for more detailed description of DSC.

IV. Initialization

Snake-based reconstruction suggested in [2] is initialized as a circle in the middle of the reconstruction domain. Such initialization would also work in our case, and topological adaptivity provided by DSC would split the curve if needed. Despite the topological adaptivity, the curve deformation might still stop in a local optimum, for example if a hole needs to be introduced in the material.

To avoid local optima, and to improve the efficiency of the reconstruction, it is desirable to start the mesh deformation from a good initialization. We present two initialization methods, the first based on filtered backprojection and the second based on a graph total variation. The first method is computationally efficient and aimed for complete and noise-free data. The second method is capable of handling incomplete or noisy data at an increased computational complexity.

Both initialization methods share a common pipeline. First, we construct a regular triangular mesh using a selected edge length l_0. Second, for each mesh triangle t we compute an attenuation coefficient ν_t. Third, we divide triangle attenuation coefficients into background and foreground objects by the k-means algorithm [30]. The difference between the two proposed initialization approaches lies in the second step, computing ν_t.

A. Filtered Backprojection on Mesh

Filtered backprojection method is one of the most popular methods in tomographic reconstruction [31]. This analytical method consists of two stages: filtering the sinogram and backprojection. Filtering sinogram does not depend on the representation of unknown objects, while the backprojection stage depends on the representation. In image reconstruction, for each pixel, the corresponding sinogram values are accumulated from all angles. Motivated by this way, we use a similar backprojection method on a triangular mesh.

For each mesh triangle t, we can compute its forward projection $\nu_t \hat{p}_t(\theta, s)$ with the attenuation ν_t by Algorithm 1. Then, the estimated sinogram corresponding to the whole mesh would be

$$\hat{p}_{\text{mesh}}(\theta, s) = \sum_{t=1}^{T} \nu_t \hat{p}_t(\theta, s)\,, \tag{8}$$

where T is the number of triangles in the mesh. This can be written as a matrix-vector product

$$\hat{p}_{\text{mesh}} = A\nu\,, \tag{9}$$

where $\hat{p}_{\text{mesh}}$ is unwrapped into a vector, A is the forward matrix whose t-th column is (unwrapped) $\hat{p}_t$ and ν is the unknown vector whose t-th element is ν_t.

From this setting, the backprojection operator is the adjoint operator of A, and the backprojected image is

$$\nu_{\text{B}} = A^T p. \tag{10}$$

Similar to filtered backprojection on images, we first filter the sinogram and backproject it onto the mesh as in (10).

B. Graph Total Variation

The filtered backprojection method is vulnerable to incomplete or noisy data. To deal with degraded data, we present another initialization method based on graph total variation [32]. In the mesh domain, we impose smoothness between the neighboring faces.

To formulate a minimization problem, we consider the absolute difference of attenuation coefficients of neighboring faces

$$E_{\text{smooth}} = \alpha \sum_{(t,\bar{t})\in\mathcal{E}} |\nu_t - \nu_{\bar{t}}|, \tag{11}$$

where $\mathcal{E} = \{(t, \bar{t}) : \text{trianges } t \text{ and } \bar{t} \text{ are adjacent}\}$ denotes a set of edges connecting two neighboring faces and α is a smoothness weight. This term can be written in a matrix multiplication form as

$$E_{\text{smooth}} = \|K\nu\|_1\,. \tag{12}$$

Here, K is an oriented edge-triangle incidence matrix with weight α. Each row of K represents one edge $(t, \bar{t})$, and has two non-zero elements: α in a position corresponding to triangle t and $-\alpha$ in a position corresponding to triangle $\bar{t}$.

Combining a data fidelity term of reprojection error and smoothness term (12), we aim at finding ν by solving the following minimization problem:

$$\nu_{\text{TV}} = \arg\min_{\nu} \frac{1}{2}\|A\nu - p\|_2^2 + \|K\nu\|_1. \tag{13}$$

Note that the weighting parameter α is incorporated into the matrix K.

Optimizing the convex energy (13) is not straightforward, as the energy has a non-smooth regularization term and two linear operators of A and K in the composite form. To deal with these difficulties, we employ the Hybrid Gradient Primal Dual (HGPD) algorithm [33] which can split the operators and lead to efficient iteration steps. To split the linear operators, we introduce two variables $z_1 = A\nu$, $z_2 = K\nu$ and the corresponding dual variables q_1, q_2. We also impose the positivity constraint

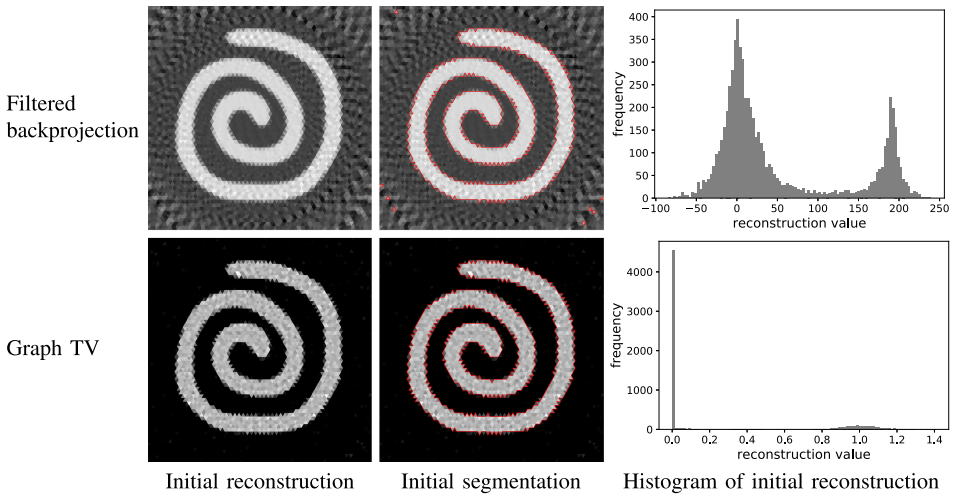


Fig. 6. An illustration of initialization schemes. In initial reconstructions, each triangle has its own attenuation value. In initial segmentation, the red line represents the initial guess of the object after applying k-means algorithm to initial reconstruction.

on our solution ν. To use HGPD, we formulate a saddle-point problem

$$\min_{\nu} \max_{q_1,q_2} \langle A\nu, q_1\rangle + \langle K\nu, q_2\rangle + \delta_+(\nu) - F^*(q_1, q_2), \quad (14)$$

where δ_+ is the indicator function of positive set and F^* is the conjugate function of $F(z_1, z_2) = \frac{1}{2}\|z_1 - p\|_2^2 + \|z_2\|_1$.

HGPD aims to find a saddle point solution of (14) by the minimization with respect to the primal variables and the maximization with respect to dual variables. The updates of primal and dual variables are provided in Algorithm 2. For the detailed derivation we refer to [34, Algorithm 4].

Algorithm 2: Primal Dual Updates For Solving (13).

Set the step sizes τ, σ as $1/(\|A\|_2 + \|K\|_2)$.
Initialize ν^0, q_1^0, q_2^0 as zero vectors.
for $k = 0, 1, 2, \ldots$

$$\nu^{k+1/2} := \nu^k - \tau(A^T q_1^k + K^T q_2^k) \quad (15)$$

$$\nu^{k+1} := \max(\nu^{k+1/2}, 0) \quad (16)$$

$$\bar{\nu} := 2\nu^{k+1} - \nu^k \quad (17)$$

$$q_1^{k+1} := \frac{q_1^k + \sigma(A\bar{\nu} - b)}{1 + \sigma} \quad (18)$$

$$q_2^{k+1} := \frac{q_2^k + \sigma K\bar{\nu}}{\max(1, |q_2^k + \sigma K\bar{\nu}|)} \quad (19)$$

Fig. 6 shows an illustration of two different methods. We use the same phantom and sinogram data with 30 angles. For filtered backprojection, we use Hann filter. The reconstruction from filtered backprojection has some unstable values around the boundary, which gives several outliers in the initial segmentation. On the other hand, graph total variation yields a better result with clean background due to the local smoothness and positivity constraint.

Algorithm 3: DALM.

Input: sinogram data p, number of materials M
Output: attenuation-labeled mesh $\mathcal{X}$
1: Construct an initial mesh with regular edge length l_0
2: Compute initial reconstruction ν by Algorithm 2
3: Construct an initial ALM $\mathcal{X}$ by applying k-means to ν
4: **while** stop condition not met**do**
5: Forward projection of ALM by Algorithm 1
6: Estimate attenuations by (6)
7: Estimate displacements by (4)
8: Deform the ALM $\mathcal{X}$ by DSC
9: **end while**

V. Combined Method

Combining the initialization scheme, we summarize our proposed method, DALM, in Algorithm 3. We estimate an initial configuration by reconstructing attenuations on regular mesh and clustering them by k-means method [30]. This clustering yields an initial attenuation-labeled mesh (ALM). We iteratively update attenuation coefficients and deform ALM. This procedure repeats until the magnitude of deformation is less than a threshold or the iteration number exceeds 500.

VI. Experimental Results

In this section, we perform experiments to demonstrate the robustness and the limitation of DALM. We illustrate the adaptivity of topological changes and investigate the initialization methods. We show the relative advantages over other reconstruction methods on various synthetic datasets. We also investigate the

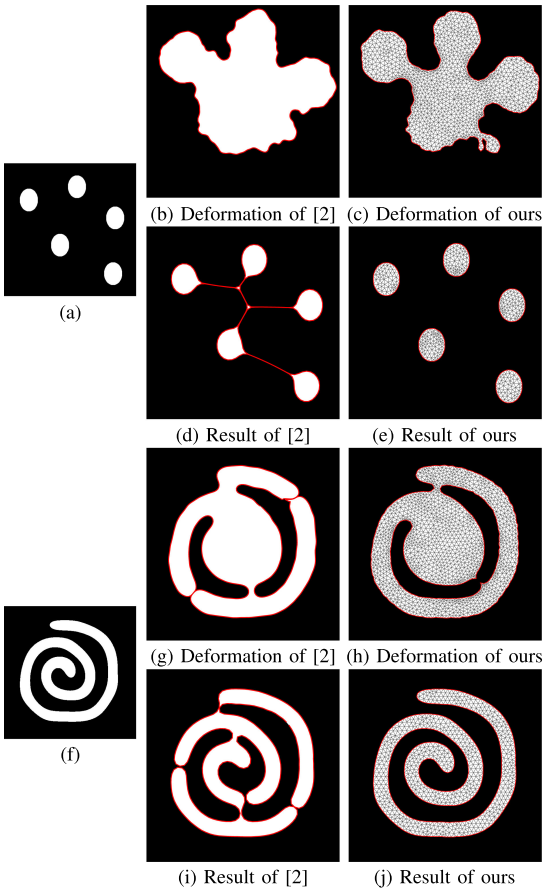


Fig. 7. DALM supports topological changes and captures complex shapes. The first column shows two phantom objects to be reconstructed. The second column shows the intermediate step and the final result of [2] from the circle initialization, where (b) and (d) are from the data generated by (a) and (g) and (i) are by (f). The last column shows the corresponding results of DALM from the same initialization.

numerical performance and the effects of algorithm parameters. Finally, we present results on real fan-beam data.

A. Support of Topological Changes During Deformation

Here, we illustrate one advantage of our method: topological adaptivity during deformation. Fig. 7 shows behaviour of our method compared to [2] on the two phantoms (a) and (f). The sinogram data are generated from the phantoms with parallel beam geometry and 30 angles. To better illustrate the curve evolution, we initialize both methods as circles. For the phantom (a), as shown in (b) and (d), a single curve used in [2] does not allow for change in topology during the curve evolution. The curves of the proposed method are shown in (c) and (e). The curves (interfaces) are red edges between faces labeled as foreground (white) and background (black). While moving interface vertices, the mesh quality is maintained by splitting or merging nearby faces. These local mesh operations allow for topological changes of objects during deformation. Support for topological adaptivity also helps reconstructing the complex shape (f). As shown in (g) and (i), the deformation of [2] stops in a local minimum, whereas our method successfully recovers the spiral shape (j).

B. Choice of Initialization and Parameters

As explained in Sec. IV, graph total variation can provide a more robust initialization than filtered backprojection. To deal with incomplete data, we choose the graph total variation method as the initialization throughout the next experiments.

Unless explicitly mentioned, we choose algorithm parameters as follows. As for initialization, we fix the initial edge length l_0 as 4, the smoothness parameter α as 8 and iterate the primal dual updates in (2) up to 200 times. As for the deformation of mesh, we fix the average edge length l_1 as 4, λ as 0.01 and set the maximum iteration number as 500. The step size τ for the curve evolution is set as 0.2. We use the public code for DSC provided the authors [3] and use the default parameters.

C. Robustness to Limited Data

In this set of experiment, we aim to show the relative advantage over other reconstruction methods on limited dataset.

We use the synthetic phantom images shown in the second column of Fig. 8 where Phantom 3 is generated from [35]. Each phantom has a fixed number of materials. Note that our forward projection requires a geometric representation of objects, while algebraic reconstruction methods need a discrete image. To compare fairly and avoid inverse crime, we generate sinogram data from images with 512-by-512 size by forward projection provided in the ASTRA toolbox [36] onto the detector with 256 pixels. Note that we do not employ our forward model to generate data, but employ an image-based forward model. We use a parallel beam geometry and choose the area projection model, which determines the weight in a pixel by the intersection area of the pixel and the ray whose width is the same as that of one detector pixel.

We compare DALM to other reconstruction methods including (image-based) Filtered Back Projection (FBP), Simultaneous Algebraic Reconstruction Technique (SART) [37] and Total Variation Regularized Discrete Algebraic Reconstruction Technique (TVR-DART) [7]. TVR-DART is one of the state-of-the-art discrete tomography methods where reconstruction images are assumed to have a fixed number of attenuation coefficients. This method is closely related to ours, as it gives an already segmented reconstruction. We carefully choose the regularization parameter of TVR-DART for fair comparison. Since the test data is generated from the image, these image-based methods have benefits in the reconstruction. We use the implementation of these methods based on ASTRA toolbox [36]. Other lines of works are learning-based methods using deep neural networks when training data is available. Although such approaches obtain impressive performance in some cases, our method is aimed at another set of problems where no training data is needed. Hence, we compare against the mentioned model-based methods, rather than data-driven methods.

Small number of projections: We use sinogram data with a small number of angles between 0 and 180 degrees. In Fig. 8,

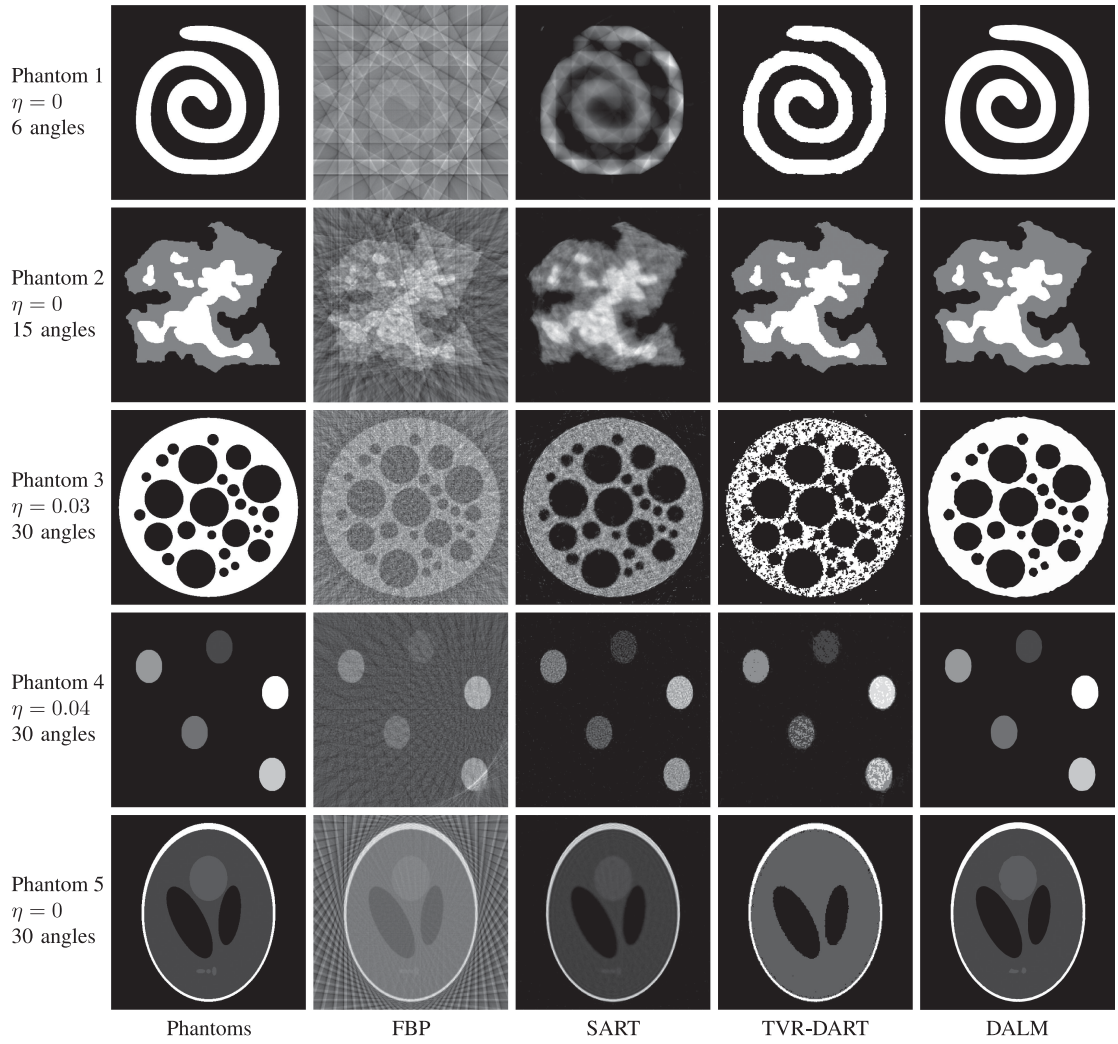


Fig. 8. Qualitative results on phantom data with small number of projections. Each row uses the different number of angles in sinogram data generated from the phantom images in the second column. The first column indicates the numbers of angles used and relative noise level η. The 3nd-5th columns show the reconstruction image by FBP, SART and TVR-DART, respectively. The last column shows our results based on geometric representation.

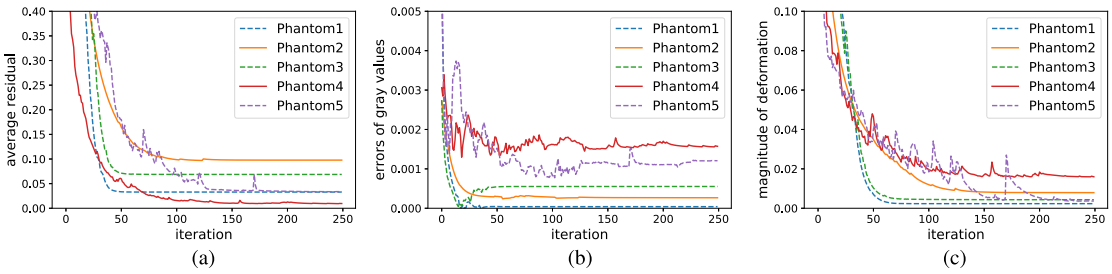


Fig. 9. Convergence behavior on phantom images in Fig. 8. (a) Convergence of residuals divided by the size of sinogram data. (b) Convergence of the gray value (attenuation coefficients) errors which are sum of absolute differences of true attenuation coefficients and estimations divided by the number of materials. (c) Convergence of the magnitude of deformation which is the mean value of displacements on the interface vertices.

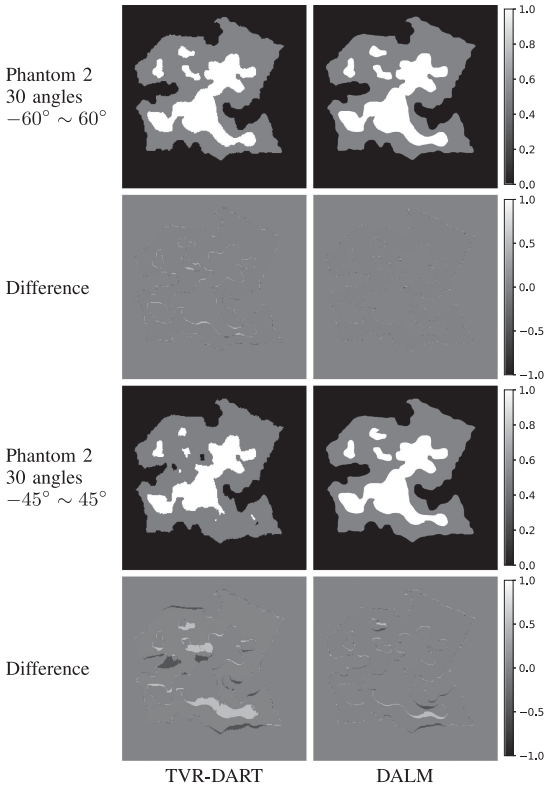


Fig. 10. Visual comparison of reconstruction results on limited angle data between TVR-DART and DALM. The odd rows show the reconstruction results and the even rows show the difference image to ground truth.

we present the qualitative results with different angles chosen to give reasonable results. In this figure, η denotes the relative Gaussian noise level. The details about imposing noise will be explained in VI-D. For Phantom 1 and Phantom 2, TVR-DART and our method are shown to give accurate reconstruction results even for a small number of angles. However, TVR-DART yields a degenerate reconstruction result for Phantom 3 due to high noisy levels, while DALM captures the overall shape of foams. TVR-DART also yields an inaccurate reconstruction in Phantom 4 which has 6 materials and in Phantom 5 which has 4 materials with small features. As indicated in its original paper, TVR-DART seems to be degenerate when the object have many materials. On the other hand, DALM overall captures the shapes in Phantom 4, but our method still misses one tiny feature at the right bottom part, as can be seen if we zoom in close enough. In Phantom 5, our method also has difficulty segmenting a fine detail on the bottom, but gives better result than TVR-DART.

Limited range of angles: In this experiment, we use Phantom 2, but generate sinogram with limited range of angles. In Fig. 10, we present the qualitative results from limited angle data. The first row shows the result from data with 30 limited angles between -60° to 60° and the second row shows the difference of the results to the ground truth. Both TVR-DART and our method yield good reconstruction shapes, but TVR-DART gives

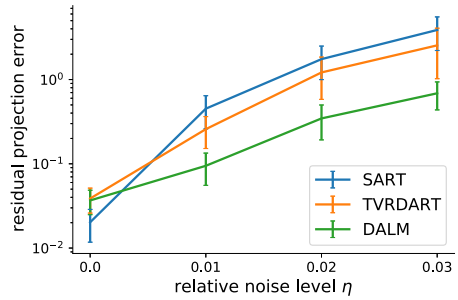


Fig. 11. Error measure with varying relative noise level for residual projection error between noise-free sinograms and estimations. The error bars show the mean and standard deviations of the errors with respect to 5 phantom data.

an inaccurate attenuation coefficient. The third row shows the visual results where a smaller range of angles (-45° to 45°) are used. The reconstruction by TVR-DART is inaccurate in shape and attenuation coefficient value. DALM gives a superior result to that of TVR-DART.

Convergence behavior: In Fig. 9, we investigate the convergence behavior of DALM for noise-free phantom data with uniformly-sampled 30 angles. In (a), we show the residual between the sinogram data and our estimated sinogram divided by the number of angles and detector pixels per iteration. The behaviors of phantoms 1, 2 and 3 show smooth curves, while the phantoms 4 and 5 have some peaks. This instability can arise when some regions are merged or splitted. In the end, the residual is shown to be stable. In (b) we plot the errors of attenuation coefficients per iteration which are measured as the sum of absolute differences between true attenuation coefficients and the estimates μ_m divided by the number of materials. As for Phantom 1 and 3, the gray value (attenuation coefficient) error is shown to increase around the 20th iteration. This happens because we estimate the optimal attenuation coefficients given the current mesh. Therefore, large changes in the mesh can affect the estimation of attenuation coefficients. In Phantom 4 and 5, the errors are shown to stablize after around 190 iterations. In (c) we show the magnitude of deformation per iteration which is calculated as the average of displacements on interface vertices. These displacements are calculated from the expression within the large parenthesis in (4). Because the step size is chosen around 0.2, the exact magnitude of deformation is smaller than those shown. Both errors of attenuation coefficients and magnitude of deformation begin to be stable in the end.

D. Robustness to Noise

In the following, we compare our method to other methods quantitatively with different noise levels. In this set of experiment, we use all the 5 phantoms and generate sinogram data with 30 projection angles. We impose Gaussian noise e on some sinogram data p such that a noisy sinogram is given as $\bar{p} = p + e$. The noise e is determined by varying the relative noise level $\eta = \|e\|_2/\|p\|_2$ from 0 to 0.03.

We employ a metric *residual projection error* [38] between the noise-free sinogram p and the estimated sinogram $\hat{p}$. This estimated sinogram $\hat{p}$ is obtained by forward projection of the

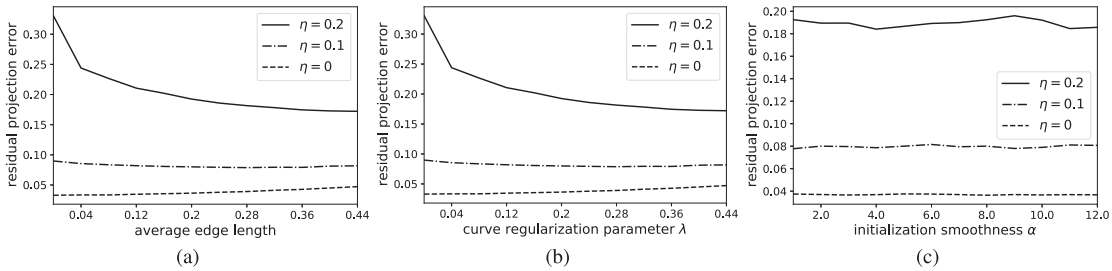


Fig. 12. Effect of algorithm parameters with different noise levels η in terms of residual projection error. Effect of (a) edge length parameter, (b) curve regularization parameter λ and (c) smoothness parameter α in the initialization. In (a) initial edge length l_0 and average edge length l_1 during deformation are set as the same values. Phantom 1 is used for this result.

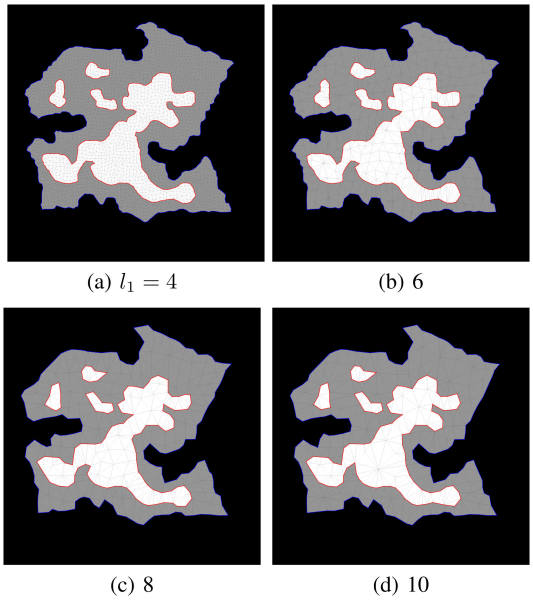


Fig. 13. Effect of average edge length l_1. Final mesh results of Phantom 3 are shown with varying average edge length, starting from the same initialization.

reconstruction. Then, the residual projection error is the mean squared error between between the noise-free sinogram p and the estimation $\hat{p}$.

In Fig. 11, we present residual projection errors with varying relative noise level η. When there is no noise, SART gives the best result as it aims to minimize the residual without regularization. However, as noise level increases, DALM is shown to yield better results than other methods.

In image-based reconstruction, it is common to evaluate the quality of reconstruction based on image quality metrics such as peak signal-to-noise ratio (PSNR) or structural similarity index (SSIM). To use such metrics, from our DALM solutions we produce images of size 256×256, the same size as reconstructions by image-based methods we compare against. As explained in Sec. VI-C, ground truth phantoms have the different image size of 512×512, so we downscale those phantoms by half, and then compute PSNR and SSIM between the downscaled phantoms

TABLE I
MEAN AND STANDARD DEVIATION OF PSNR AND SSIM WITH VARYING NOISE LEVELS η

η	PSNR			SSIM		
	SART	TVR-D	DALM	SART	TVR-D	DALM
0.00	24.5±4	24.2±4	24.8±4	0.88±0.08	0.96±0.02	0.97±0.01
0.01	23.8±4	22.7±3	24.7±3	0.84±0.10	0.93±0.05	0.97±0.01
0.02	22.4±4	19.4±8	24.1±3	0.76±0.12	0.82±0.15	0.96±0.02
0.03	20.8±5	19.0±7	23.1±4	0.70±0.13	0.81±0.16	0.90±0.11

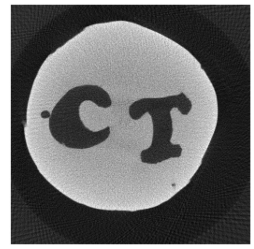


Fig. 14. Reference reconstruction image by filtered backprojection from the carved cheese data with 180 projection angles.

and the solutions. In Table I we provide the quantitative results by PSNR and SSIM with different noise levels. Our result is shown to give better results in all noise levels. In this result we use 5 phantoms and present the mean and standard deviation values.

The large standard deviation for the three methods might indicate that the difference between the means is not significant. Therefore, we performed a three-way ANOVA with no interaction between variables. The results for both PSNR and SSIM show that the phantom and the noise level have a very large influence on quality measure, and our method has a smaller, but also significant influence. In conclusion, ANOVA confirms that the quality means for the three methods are significantly different.

E. Effects of Algorithm Parameters

We investigate the main parameters of our method: the initial reconstruction regularization parameter α, the average edge length of the mesh l_1 and the curve regularization parameter λ. We use the same dataset and settings as in Section VI-C. Note that we have two parameters to control the fineness of mesh: the initial edge length l_0 and the average edge length l_1 during

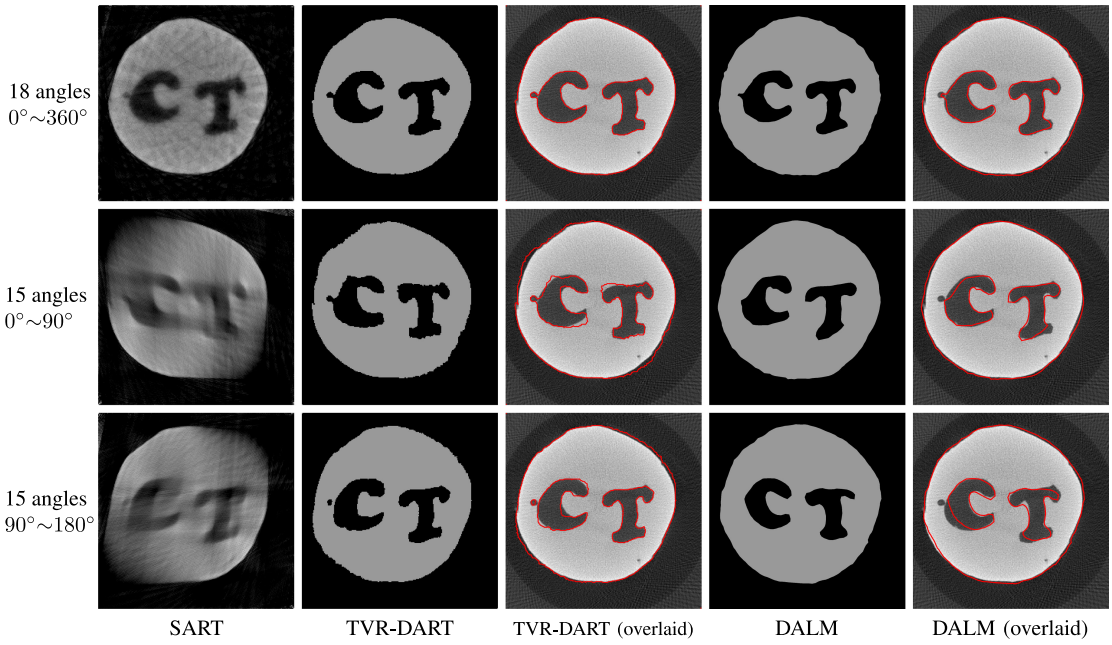


Fig. 15. Reconstruction and segmentation results on real fan-beam data.

TABLE II
COMPUTATIONAL COST AND COMPACTNESS OF MESH WITH WITH VARYING AVERAGE EDGE LENGTH l_1

avg. edge length (pixel)	4	6	8	10
# of vertices	6616	1278	815	723
# of edges	19565	3551	2162	1886
# of faces	12950	2274	1348	1164
# of interface vertices	685	469	264	248
initialization (sec)	12.7	12.7	12.7	12.7
compute $\hat{p}$, μ & disp.	10.3	8.5	5.1	4.6
deformation	77.1	41.0	24.8	20.4
total time	100.1	61.2	42.6	37.7

evolution. Throughout the experiment, we fix l_0 as 4 and this parameter affects the initial reconstruction quality and speed. It is desired to choose this parameter as small as possible to capture the details of objects. After initialization, we collapse non-interface edges for computational efficiency, but interface edges are unchanged. This way, we preserve the details on the interface, while non-interface regions have a more coarse mesh. Subsequently, the average edge length l_1 controls the fineness of the mesh during evolution.

In Fig. 13, segmentation results of Phantom 3 are provided with varying average edge length l_1. As l_1 increases, we have the compact representation of objects, while preserving the outlines of objects. In Table II, we provide the information of final mesh and computational cost with different average edge length. As expected, as average edge length is higher, we obtain more compact representation of objects and reduce the computational cost. To investigate the cost of each step, we divide the total time into three steps: first, initialization; second, computing forward projections, attenuation coefficients and displacements; third, the deformation step. Most of the computational cost is attributed to the deformation part. Based on this, we identify the mesh deformation model as the critical step. A more efficient deformation model could therefore be a candidate for improving the algorithm speed. For other image-based methods, the computational time is 13.6 seconds for SART and 42.4 seconds for TVR-DART. Both methods reconstruct 65 536 pixels. For the experiments, we use a laptop with 3.5 GHz processor and 16 GB memory.

The initial regularization parameter α controls the smoothness of the initial reconstruction result. We observe that our method is not sensitive to initialization in the simple datasets such as Phantom 1, 2, 3 and 4, but sensitive in Phantom 5. We leave as a future work to solve the apparent issue of initialization.

In Fig. 12, we provide the effects of the main parameters with different noise levels. For this experiment, we use Phantom 1 and show how residual reprojection errors are affected. In (a) we show the effect of fineness of mesh. We set both the initial edge length l_0 and the average edge length l_1 as the same values, so that the fineness of mesh is fixed from initialization to the end. As those values increase, overall, the residual projection error also increases. So there is a trade-off between accuracy and compactness of representation. In (b) it is shown that as the relative noise level increases, so does the optimal parameter of curve regularization λ. If the data is noisy or incomplete, we need stronger regularization. Note that, in addition, the deformation step has a regularization effect intrinsic to DSC. In (c) we see that the initialization smoothness parameter α does not affect the error, significantly.

F. Real Dataset

In the following, we perform an experiment on real fan-beam data provided by [39]. The authors of [39] carved out the letters

'CT' in a cheese. The carved cheese was scanned using micro-CT under the exposure time of 1000 ms, the acceleration voltage of 40 kV and the current of 1 mA in X-ray tube [39]. Two phases (materials) are present: cheese and air. The geometry of the scanning is flat fan-beam, and the number of detector pixels is 989. For this data we do not have the exact ground truth, but we consider the reconstruction by FBP from sinogram with 180 angles as the reference image shown in Fig. 14.

We compare our methods in challenging cases under the small number of projection angles or limited range of angles. In Fig. 15, we provide the reconstruction and segmentation results from SART, TVR-DART and our method, DALM. The first row shows the results when using sinogram from 18 projection angles. The second and third row present the results from limited angle data with 15 angles of 0–90 ° and 90–180 °, respectively. As TVR-DART and our method give the segmentation results, we provide the estimated boundaries overlaid on the reference image on the third and last column, respectively. DALM captures the shape of the object, even in the limited angle data. Overall, our method is shown to give slightly smaller letters due to regularization of curves.

VII. Conclusion and Discussion

We have presented DALM, a mesh-based method for 2D reconstruction and segmentation directly from sinogram data. In DALM, a labeled mesh is deformed to align interface edges with object boundaries. By using mesh deformation which supports topological changes and prevents self-intersections, DALM overcomes the drawbacks of existing explicit representation-based methods [2], [27], [40]. Moreover, our method easily supports multiple objects, while in popular level-set methods, supporting multiple regions is not straightforward and additional efforts are required [41]. Experimental results on synthetic data show that our method gives an accurate geometric solution with a compact representation of objects. We provide an efficient forward projection scheme, while the deformation step costs more in terms of computation time. We leave for future work to speed up the deformation step and extend to 3D reconstruction.

Appendix
Derivation of the Curve Evolution Equation (4)

To derive the evolution equation (4), we begin with the case of single material. Let C_m denote the curve to represent the boundary of foreground object with material μ_m and the air background. We introduce an artificial time parameter t and assume that the detector positions s are sampled dense enough. Following [16], the curve evolution equation for C_m is

$$\frac{\partial C_m(r)}{\partial t} = \left(\mu_m \sum_{\theta} (p(\theta, \tilde{s}) - \hat{p}(\theta, \tilde{s})) + \bar{\lambda} \kappa_m(r) \right) N_m(r), \tag{20}$$

where κ_m is the curvature, N_m is the outward normal vector, $\tilde{s} = L_\theta(C_m(r))$ with r the arc length parameter and $\bar{\lambda}$ is the regularization parameter.

We now extend to M multiple materials where the domain is divided by mutually disjoint regions $\{R_m\}$ and the curves $\{C_m\}$ represent their boundaries. Let the region R_m be adjacent to the region R_n with the boundary C_n and the attenuation μ_n. Then, for the points on $C_m \cap C_n \neq \emptyset$, the curve evolution equation for C_m is given [27]:

$$\begin{aligned} \frac{\partial C_m(r)}{\partial t} = & \left(\mu_m \sum_{\theta} (p(\theta, \tilde{s}) - \hat{p}(\theta, \tilde{s})) + \bar{\lambda} \kappa_m(r) \right) N_m(r) \\ & + \left(\mu_n \sum_{\theta} (p(\theta, \tilde{s}) - \hat{p}(\theta, \tilde{s})) + \bar{\lambda} \kappa_n(r) \right) N_n(r). \end{aligned} \tag{21}$$

For the points on $C_m \cap C_n$, the normal vectors and the curvatures have the opposite sign such that $N_m = -N_n$ and $\kappa_m = -\kappa_n$ [27]. Plugging these relations into (21), we have

$$\begin{aligned} \frac{\partial C_m(r)}{\partial t} = & \left((\mu_m - \mu_n) \sum_{\theta} (p(\theta, \tilde{s}) - \hat{p}(\theta, \tilde{s})) + 2\bar{\lambda} \kappa_m(r) \right) \\ & \times N_m(r). \end{aligned} \tag{22}$$

From this equation, we derive (4) with the fixed step size τ and $\lambda := 2\bar{\lambda}$.

Acknowledgment

The authors thank Patrick M. Jensen for the fruitful discussion and proofreading the paper.

References

[1] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2nd ed., *Ser. Advances in Pattern Recognit.*, Dordrecht, The Netherlands; New York, NY, USA: Springer, 2009.

[2] V. A. Dahl, A. B. Dahl, and P. C. Hansen, "Computing segmentations directly from x-ray projection data via parametric deformable curves," *Meas. Sci. Technol.*, vol. 29, no. 1, 2018, Art. no. 014003.

[3] M. K. Misztal and J. A. Bærentzen, "Topology-adaptive interface tracking using the deformable simplicial complex," *ACM Trans. Graph.*, vol. 31, no. 3, pp. 1–12, 2012.

[4] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, Sep. 2011.

[5] K. J. Batenburg, W. van Aarle, and J. Sijbers, "A semi-automatic algorithm for grey level estimation in tomography," *Pattern Recognit. Lett.*, vol. 32, no. 9, pp. 1395–1405, 2011.

[6] F. Bleichrodt, F. Tabak, and K. J. Batenburg, "SDART: An algorithm for discrete tomography from noisy projections," *Comput. Vis. Image Underst.*, vol. 129, pp. 63–74, 2014.

[7] X. Zhuge, W. J. Palenstijn, and K. J. Batenburg, "TVR-DART: A more robust algorithm for discrete tomography from limited projection data with automated gray value estimation," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 455–468, Jan. 2016.

[8] A. Tuysuzoglu, Y. Khoo, and W. C. Karl, "Variable splitting techniques for discrete tomography," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1764–1768.

[9] F. Lauze, Y. Quéau, and E. Plenge, "Simultaneous reconstruction and segmentation of CT scans with shadowed data," in *Scale Space and Variational Methods in Computer Vision*, F. Lauze, Y. Dong, and A. Dahl, Eds., vol. 10302. Cham, Switzerland: Springer, 2017.

[10] A. Tuysuzoglu, W. C. Karl, I. Stojanovic, D. Castañòn, and M. S. Ünlü, "Graph-cut based discrete-valued image reconstruction," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1614–1627, May 2015.

[11] J. H. Kappes, S. Petra, C. Schnörr, and M. Zisler, "TomoGC: Binary tomography by constrained graphcuts," in *German Conference on Pattern Recognition*, J. Gall, P. Gehler, and B. Leibe, Eds., Berlin, Germany: Springer, 2015, pp. 262–273.

[12] Z. Wei, B. Liu, B. Dong, and L. Wei, "A joint reconstruction and segmentation method for limited-angle X-ray tomography," *IEEE Access*, vol. 6, pp. 7780–7791, 2018.

[13] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Comput. Phys.*, vol. 79, no. 1, pp. 12–49, 1988.

[14] R. T. Whitaker and V. Elangovan, "A direct approach to estimating surfaces in tomographic data," *Med. Image Anal.*, vol. 6, no. 3, pp. 235–249, 2002.
[15] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.
[16] C. V. Alvino and A. J. Yezzi, "Tomographic reconstruction of piecewise smooth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004.
[17] S. Yoon, A. R. Pineda, and R. Fahrig, "Simultaneous segmentation and reconstruction: A level set method approach for limited view computed tomography," *Med. Phys.*, vol. 37, no. 5, pp. 2390–2340, 2010.
[18] R. Ramlau and W. Ring, "A mumford level-set approach for the inversion and segmentation of X-ray tomography data," *J. Comput. Phys.*, vol. 221, no. 2, pp. 137–166, 2007.
[19] A. Aghasi, M. Kilmer, and E. L. Miller, "Parametric level set methods for inverse problems," *SIAM J. Imag. Sci.*, vol. 4, no. 2, pp. 618–650, 2011.
[20] A. Kadu, T. van Leeuwen, and K. J. Batenburg, "A parametric level-set method for partially discrete tomography," in *Proc. Int. Conf. Discrete Geometry Comput. Imagery*, 2018.
[21] J. G. Brankov, Y. Yang, and M. N. Wernick, "Tomographic image reconstruction based on a content-adaptive mesh model," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 202–212, Feb. 2004.
[22] A. Cazasnoves, F. Buyens, and S. Sevestre-Ghalila, "Adapted sampling for 3D X-ray computed tomography," in *Proc. 13th Int. Meeting Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med.*, 2015.
[23] A. Sitek, R. Huesman, and G. Gullberg, "Tomographic reconstruction using an adaptive tetrahedral mesh defined by a point cloud," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1172–1179, Sep. 2006.
[24] N. F. Pereira and A. Sitek, "Evaluation of a 3D point cloud tetrahedral tomographic reconstruction method," *Phys. Med. Biol.*, vol. 55, no. 18, pp. 5341–5361, 2010.
[25] R. Boutchko, A. Sitek, and G. T. Gullberg, "Practical implementation of tetrahedral mesh reconstruction in emission tomography," *Phys. Med. Biol.*, vol. 58, no. 9, pp. 3001–3022, 2013.
[26] F. Buyens, M. Quinto, and D. Houzet, "Adaptive mesh reconstruction in X-ray tomography," in *Proc. MICCAI Workshop Mesh Process. Med. Image Anal.*, 2013.
[27] S. C. Zhu, T. S. Lee, and A. L. Yuille, "Region competition: Unifying snakes, region growing, energy/bayes/MDL for multi-band image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1995, pp. 416–423.
[28] B. O'Neill, *Elementary Differential Geometry*. Amsterdam; Boston: Elsevier Academic Press, 2006.
[29] G. Strang, *Linear Algebra and Its Applications*, 4th ed. Belmont, CA: Thomson, Brooks/Cole, 2006.
[30] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
[31] T. M. Buzug, *Computed Tomography From Photon Statistics to Modern Cone Beam CT*. Berlin, Germany: Springer, 2008.
[32] F. Mahmood, N. Shahid, U. Skoglund, and P. Vandergheynst, "Adaptive graph-based total variation for tomographic reconstructions," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 700–704, May 2018.
[33] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.
[34] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle-pock algorithm," *Phys. Med. Biol.*, vol. 57, no. 10, pp. 3065–3091, 2012.
[35] D. J. Ching and D. Gürsoy, "XDesign : An open-source software package for designing X-ray imaging phantoms and experiments," *J. Synchrotron Radiat.*, vol. 24, no. 2, pp. 537–544, 2017.
[36] W. van Aarle *et al.*, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Opt. Exp.*, vol. 24, no. 22, pp. 25129–25147, 2016.
[37] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrason. Imag.*, vol. 6, no. 1, pp. 81–94, 1984.
[38] T. Roelandts, K. J. Batenburg, A. J. den Dekker, and J. Sijbers, "The reconstructed residual error: A novel segmentation evaluation measure for reconstructed images in tomography," *Comput. Vis. Image Underst.*, vol. 126, pp. 28–37, 2014.
[39] T. A. Bubba *et al.*, "Tomographic X-ray data of carved cheese," *Med. Phys.*, 2017 *arXiv:170505732*.
[40] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.
[41] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the mumford and shah model," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 271–293, 2002.

Jakeoung Koo received the B.Sc. and M.Eng. degrees from Chung-Ang University, South Korea, in 2015 and 2017, respectively. He is currently working toward the Ph.D. degree with the Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU). His research interests include computational imaging and computer vision. His current work is on tomographic reconstruction where he investigates alternative representations such as meshes, instead of discrete images.

Anders Bjorholm Dahl is a Professor in 3D image analysis, and a Head of the Section for Image Analysis and Computer Graphics with DTU Compute, Technical University of Denmark (DTU). He is heading The Center for Quantification of Imaging Data from MAX IV, which focuses on research in quantitative analysis of 3D microscopy images. This is done through research collaboration with scientists using 3D imaging, development of software tools, and research in 3D image analysis tools.

Vedrana Andersen Dahl holds degrees in mathematics, multimedia technology, and the Ph.D. degree in geometry processing. She is an Associate Professor with the Department of Applied Mathematics and Computer Science, Technical University of Denmark. Her primary research interest focuses on using geometric models for the analysis of volumetric data. This includes volumetric segmentation, tomographic segmentation, and methods based on deformable meshes. She developed analysis tools with applications in material science, industrial inspection, and biomedicine.

Contribution B

Shape from Projections via Differentiable Forward Projector for Computed Tomography

The following paper has been accepted in Ultramicroscopy.
https://doi.org/10.1016/j.ultramic.2021.113239

Shape from Projections via Differentiable Forward Projector for Computed Tomography

Jakeoung Koo[a,*], Anders B. Dahl[a], J. Andreas Bærentzen[a], Qiongyang Chen[b], Sara Bals[b], Vedrana A. Dahl[a]

[a]*Technical University of Denmark, Anker Engelunds Vej 1, Kgs. Lyngby, 2800, Denmark*
[b]*University of Antwerp, Prinsstraat 13, Antwerpen, 2000, Belgium*

Abstract

In computed tomography, the reconstruction is typically obtained on a voxel grid. In this work, however, we propose a mesh-based reconstruction method. For tomographic problems, 3D meshes have mostly been studied to simulate data acquisition, but not for reconstruction, for which a 3D mesh means the inverse process of estimating shapes from projections. In this paper, we propose a differentiable forward model for 3D meshes that bridge the gap between the forward model for 3D surfaces and optimization. We view the forward projection as a rendering process, and make it differentiable by extending recent work in differentiable rendering. We use the proposed forward model to reconstruct 3D shapes directly from projections. Experimental results for single-object problems show that the proposed method outperforms traditional voxel-based methods on noisy simulated data. We also apply the proposed method on electron tomography images of nanoparticles to demonstrate the applicability of the method on real data.

Keywords: Computed Tomography, Electron Tomography, Tomographic Reconstruction, Mesh Deformation

1. Introduction

In computed tomography (CT), we aim at solving the inverse problem of computing the 3D structure (shape and attenuation) of an object from a set of projection images (Buzug, 2008) taken from different angles. Here, the geometry and the physics of the imaging system is known, which allows us to model the forward process, i.e. if we have a suggestion for the 3D structure of the imaged object, we can compute the projection images.

We need a data structure to represent the structure of the object that should be reconstructed. The most common data structure is a volumetric image, with voxel intensities representing local attenuation. This approach may be used for reconstructing any type of object. Since each voxel in the volume is a parameter that must be computed, we have a very large number of unknowns. This makes it difficult to reconstruct volumes in situations where we have projections from a limited angular view (e.g., in electron tomography) or noisy data, and it can be difficult to accurately compute the attenuations in all voxels. Therefore, we propose to use a mesh to represent the shape of the object. The mesh separates the object into parts with a constant attenuation.

In tomography, forward projection of 3D meshes has mostly been used for simulating tomographic data acquisition, i.e. modeling the forward projection. This includes modeling X-ray transmission imaging based on Monte-Carlo methods (Bonin et al., 2002; Freud et al., 2006) or ray tracing techniques (Freud et al., 2006; Marinovszki et al., 2018). Furthermore, Vidal et al. (2009); Sujar et al. (2017) took the advantage of the OpenGL library to simulate X-ray images in real time. However, none of the proposed methods are concerned with reconstruction, i.e. solving the inverse problem.

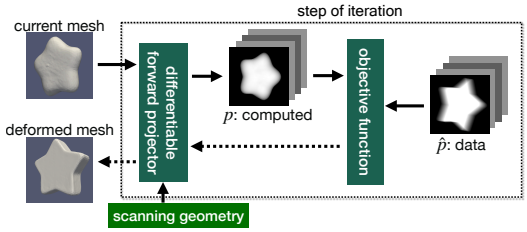


Figure 1: The proposed differentiable forward projector enables optimizing a 3D mesh from projections for tomographic reconstruction. Given a mesh and scanning geometry, the forward projector computes projections for the current mesh, which is updated by deforming vertices in the direction that minimizes our objective function.

For the mesh-based tomographic reconstruction that we propose, the reconstruction problem is two-fold. The mesh must be deformed to follow the boundaries of the depicted object, and in each part of the object, a single attenuation coefficient must be estimated. To forward project the mesh, we employ rendering techniques, which allow a very efficient projection of the 3D mesh to the detector plane. We extend the differentiable rasterizer recently proposed by Chen et al. (2019) to derive differentiable forward projection. This enables us to compute vertex displacements that deform the mesh based on the difference between the forward projection and projection data. In Fig. 1, we provide an overview of our shape estimation method. In Fig. 2, we illustrate how our work differs from existing image-based reconstruction methods (Andersen and Kak, 1984; Buzug, 2008).

Our model has been developed for problems like X-ray CT where the damping of the X-ray attenuation coefficient are mod-

*Corresponding author. E-mail: jakoo@dtu.dk
Accepted manuscript, Ultramicroscopy
DOI: https://doi.org/10.1016/j.ultramic.2021.113239

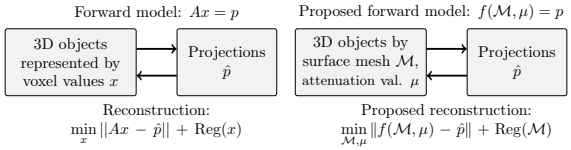


Figure 2: Algebraic reconstruction vs. our proposed approach. *Left:* Existing algebraic reconstruction methods use a linear system of equations to model the forward projections typically together with some regularization to constrain the solution. *Right:* We propose a differentiable forward model f which computes projections for an object represented by a triangular mesh and attenuation coefficients μ. This forward model is used to reconstruct the shape from projections data $\hat{p}$.

eled as linearly dependent on the path length through the sample (Buzug, 2008), which is also assumed in electron tomography up to a certain thickness (Midgley and Weyland, 2003) where coefficient is related to electron scattering.

In summary, our contribution is two-fold. We suggest a differentiable forward projector to generate projections from 3D meshes, and we propose a shape estimation method employing the differentiable forward projector. Our code is available online: https://github.com/jakeoung/ShapeFromProjections.

2. Related works

Differentiable rendering based on rasterization. In our work we deform a mesh by changing vertex positions using gradient descent. Therefore, we need the gradient of our objective function with respect to vertex coordinates, which is possible through differentiating the forward projection. As observed in (Vidal et al., 2009), the tomographic forward projection can be implemented by extending a rendering technique in computer graphics based on rasterization. Such rasterization-based rendering technique projects 3D models onto 2D image plane and involves a discrete step to choose the pixels covered by triangles of 3D models. This discrete step is not differentiable in conventional graphics pipelines. Making such rasterization-based rendering differentiable, called *differentiable rasterization*, has been studied by several works to connect rendering and optimization. As our work extends such differentiable rasterization techniques, we first review those works and explain the difference to our work.

The general framework OpenDR, that was proposed by Loper and Black (2014), approximated gradients of pixel values with respect to model parameters. Kato et al. (2018) suggested a heuristic forward and backward pass where blurring is used to avoid zero gradients. This approach has an inconsistency between the forward and backward pass, and to circumvent this inconsistency, Liu et al. (2019) proposed a method called SoftRas by relaxing discrete rasterization process into the aggregation of smooth probability functions. Unlike the conventional rasterization rendering, in SoftRas, each face in the mesh affects many pixels in the image plane, which is computationally costly and memory-demanding.

Chen et al. (2019) suggested an interpolation-based differentiable rasterizer called DIB-R. DIB-R reformulates the barycentric interpolation in the rasterization process to analytically derive the gradients. Our forward projector extends this reformulation when computing the thickness of an object for computed tomography, but differs in some aspects. The rendering techniques such as DIB-R can improve the performance by ignoring invisible faces, but since we model penetrating radiation, our forward projector needs to consider all the faces. DIB-R used the idea of SoftRas (Liu et al., 2019) for background pixels to propagate the gradients on those background, but our forward projector does not use it to reduce the computational cost.

Shape reconstruction from projections. Our proposed method is related to tomographic segmentation, where segments are directly computed from projections. This includes (Elangovan and Whitaker, 2001; Whitaker and Elangovan, 2002; Alvino and Yezzi, 2004) that are based on the Mumford-Shah model (Mumford and Shah, 1989) where boundaries are represented using level-sets (Osher and Fedkiw, 2004). Recently, the parametric level-set method (Aghasi et al., 2011) has been used for tomographic segmentation in (Kadu et al., 2018; Eliasof et al., 2020) where level-sets are represented as an aggregation of radial basis functions. Although the parametric level-set method has fewer unknown variables, its forward projection still depends on a regular grid. Gadelha et al. (2019) used a deep convolutional neural network for 2D tomographic reconstruction, where the forward projection is based on the transformation of a regular grid and resampling. On the other hand, the work (Dahl et al., 2018) based on snakes (Kass et al., 1988) avoids a dense grid – it represents curves explicitly and proposes a direct forward projection of the curves. However, this method is limited to a single 2D curve, while the proposed method supports 3D objects. Another difference is that (Dahl et al., 2018) evolves curves in the normal directions of curve points, while our deformation can displace the vertices in all directions.

In summary, existing shape estimation methods from projections are either based on regular grid or limited to single 2D curve. To our knowledge, our work is the first to propose the differentiable forward projector for 3D triangular mesh and use it for reconstructing shapes from projections.

3. Differentiable forward projector

In this section, we describe our main contribution of the differentiable forward projector. The goal is to forward project triangular meshes and make this process differentiable with respect to 3D vertex positions and attenuation coefficients. This differentiable forward projection will be used for optimizing the mesh shape described in Sec. 4. First we describe the case of a single object and then extend to composite objects.

3.1. Single object

Consider an object represented by a watertight triangular mesh. A watertight triangle mesh forms a closed surface that has a well-defined interior and exterior: any path from a point in the interior to a point in the exterior must cross the triangle mesh. We assume that the object is homogeneous, i.e. it has a certain attenuation coefficient μ associated with the volume

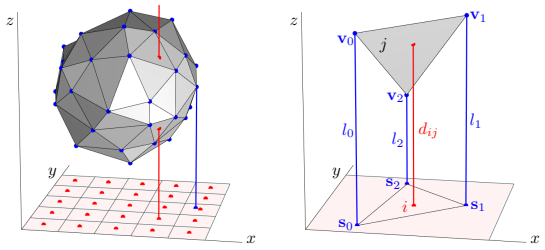


Figure 3: *Left:* The vertices of the triangle mesh (blue dots) are projected onto the detector. Each detector pixel (red dots) is associated with the projection ray which intersects mesh triangles. *Right:* One triangle j, here given by vertices $k = 0, 1, 2$, and one detector pixel i. Using barycentric coordinates, the distance d_{ij} may be expressed in terms of l_k.

inside the mesh. For now, we consider μ constant, we will later explain how it is computed in the next subsection. The mesh consists of K vertices, and we write $\mathbf{v}_k$ for the 3D coordinates of the vertex k.

To simplify the explanation, we start with a single projection and later expand to multiple projections from different angles. Let $\mathbf{P}$ and $\mathbf{R}$ be the position of the detector and a matrix that rotates from detector coordinates to the global frame, respectively. If we denote the position of vertex k in global coordinates by $\mathbf{V}_k$, the position in detector coordinates are $\mathbf{v}_k = \mathbf{R}^\intercal(\mathbf{V}_k - \mathbf{P})$. Note that in detector coordinates, the detector itself corresponds to the plane $z = 0$, its center is at the origin, and the positive z-axis points towards the object, see Fig. 3.

Expressed in detector coordinates, the distance of the vertex k from the detector is trivially $l_k = \mathbf{e}_3^\intercal \mathbf{v}_k$ while $\mathbf{s}_k = [\mathbf{e}_1\, \mathbf{e}_2]^\intercal \mathbf{v}_k$ are the coordinates of the projection of the vertex onto the detector. Here $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ are unit vectors in x, y and z direction, for example $\mathbf{e}_1 = [1\,0\,0]^\intercal$.

Projecting the object onto the detector pixel i we consider the projection ray associated with i (slightly sloppy, call it ray i), and its path length in the object. As explained in (Vidal et al., 2009), this can be broken into contribution of all intersections of the ray i with the mesh triangles

$$p_i = \mu \sum_{\substack{j \\ i \text{ intersects } j}} \operatorname{sign}(\mathbf{e}_3^\intercal \mathbf{n}_j) d_{ij} \tag{1}$$

where $\mathbf{n}_j$ is the normal of the triangle j (needed for determining the sign of the contribution), and d_{ij} is the distance of the intersection point to the detector. Here, we consider the sign value $\operatorname{sign}(\mathbf{e}_3^\intercal \mathbf{n}_j)$ as a constant attribute of each triangle.

Considering now a single triangle j we express d_{ij} using barycentric coordinates

$$d_{ij} = \sum_{\substack{k \\ k \text{ in } j}} w_{ij}^k l_k \ , \tag{2}$$

where k are the indices of the three vertices of the triangle j and w_{ij}^k are the corresponding three barycentric coordinates of pixel i with respect to the projection of triangle j onto the detector plane, see Fig. 3, *right*.

To make the forward projection differentiable, we derive

$$\frac{\partial p_i}{\partial \mathbf{v}_k} = \mu \sum_{\substack{j \\ i \text{ intersects } j}} \operatorname{sign}(\mathbf{e}_3^\intercal \mathbf{n}_j) \frac{\partial d_{ij}}{\partial \mathbf{v}_k} \ , \tag{3}$$

and

$$\frac{\partial d_{ij}}{\partial \mathbf{v}_k} = \sum_{\substack{k \\ k \text{ in } j}} \left(w_{ij}^k \frac{\partial l_k}{\partial \mathbf{v}_k} + \frac{\partial w_{ij}^k}{\partial \mathbf{v}_k} l_k \right) = \sum_{\substack{k \\ k \text{ in } j}} \left([0\,0\,w_{ij}^k] + \frac{\partial w_{ij}^k}{\partial \mathbf{v}_k} l_k \right). \tag{4}$$

For the last step, computation of $\partial w_{ij}^k / \partial \mathbf{v}_k$, we employ the idea from (Chen et al., 2019), which reformulates the barycentric form to express the coefficients w_{ij}^k in terms of 2D projected positions $\mathbf{s}_k$ and the position of detector pixel i.

3.2. Composite objects

The method described above generalizes to composite objects if certain conditions are met. Specifically, we require that we know beforehand the topology of the parts of the composites, and how parts are embedded within one another. Thus, for each interface triangle we will have a suggestion that what class of material is on either side, but we do not know the specific attenuation coefficient of the classes, since we solve for those.

In order to extend Eq. (1) to composites, we only have to observe that triangles may now be the interface between two materials and not just air and material. This can be handled simply by letting each triangle contribute twice

$$p_i = \sum_j \mu_j \operatorname{sign}(\mathbf{e}_3^\intercal \mathbf{n}_j) d_{ij} - \sum_j \overline{\mu}_j \operatorname{sign}(\mathbf{e}_3^\intercal \mathbf{n}_j) d_{ij} \tag{5}$$

$$= \sum_j (\mu_j - \overline{\mu}_j) \operatorname{sign}(\mathbf{e}_3^\intercal \mathbf{n}_j) d_{ij} \ , \tag{6}$$

where μ_j is the attenuation of the interior material and $\overline{\mu}_j$ of the exterior material according to normal orientation. Of course, either attenuation will be zero if the material on the corresponding side of the triangle is air.

The derivative of a pixel value with respect to the contributing attenuation μ_m for material m (by abuse of notation) is

$$\frac{\partial p_i}{\partial \mu_m} = \sum_j \pm \operatorname{sign}(\mathbf{e}_3^\intercal \mathbf{n}_j) d_{ij} \ , \tag{7}$$

where the $\pm$ is positive if the interior material of face j is labeled as m and negative if m is the exterior material. We also modify Eq. (3) by changing μ to $(\mu_j - \overline{\mu}_j)$.

We have derived the Jacobians in Eq. (3) and (7), which will be used to optimize an objective function E. That is, we can propagate the gradients from E

$$\frac{\partial E}{\partial \mathbf{v}_k} = \sum_i \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial \mathbf{v}_k}, \qquad \frac{\partial E}{\partial \mu_m} = \sum_i \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial \mu_m}, \tag{8}$$

where the summation is over all the detector pixels.

4. Shape from projections

In this section, we use the proposed forward projector to reconstruct shapes from projections. We assume that a template mesh with the correct topology is given. We aim to deform the template mesh and estimate the attenuation coefficients by minimizing the residual between data $\hat{p}$ and our estimation p.

Optimizing only the data fitting term can lead to degenerate meshes. To obtain high-quality meshes, we impose three regularization terms. The first term is the Laplacian term (Wang et al., 2018), which constraints the vertices to move similarly with their neighbors, defined by

$$E_{\text{lap}} = \sum_k \|\mathbf{V}_k - \frac{1}{|\mathcal{N}(k)|} \sum_{n \in \mathcal{N}(k)} \mathbf{V}_n\|^2, \tag{9}$$

where $\mathcal{N}(k)$ is the index set of neighboring vertices to k-th vertex. The second term is the edge length term to penalize long edges

$$E_{\text{edge}} = \sum_{(\mathbf{V}_k, \mathbf{V}_n) \in G} \|\mathbf{V}_k - \mathbf{V}_n\|^2, \tag{10}$$

where G denotes the set of edges. Lastly, we impose the flattening term (Kato et al., 2018; Liu et al., 2019)

$$E_{\text{flat}} = \sum_{e \in G} (1 - \cos\theta_e)^2, \tag{11}$$

where θ_e is the angle between the normal vectors of two faces sharing the edge e. Flattening term is needed to remove near-zero volume spikes. These thin artifacts have negligible contribution to the forward projection, and will be ignored by the data fitting term. As shown in Fig. 5, such artifacts can appear during the deformation, but disappear later.

With the data fidelity and regularization terms, the objective function to minimize is

$$E(\{\mathbf{V}_k\}, \{\mu_m\}) = \|p - \hat{p}\|_2^2 + \alpha E_{\text{lap}} + \beta E_{\text{edge}} + \gamma E_{\text{flat}} \tag{12}$$

where α, β, γ control the relative weights between the terms. Note that the size of projection data $\hat{p}$ is the number of detector pixels times the number of projection angles. We use automatic differentiation to minimize E. For large data, we can use stochastic gradient descent with mini batches in terms of projection angles. In this paper, however, we only consider full batch size of data.

5. Experiments and Results

In this section, we present the experimental results of the proposed method on synthetic data of single objects. We also show the results on real data of some nano particles from electron tomography, which has limited range of angles.

5.1. Shape reconstruction of a set of single objects

Datasets. This experiment is designed to test our shape estimation method on noisy simulated data. We use 5 watertight meshes (closed surfaces without any holes): star, spot, bunny, bob, kitten, shown in the first row of Fig. 6 and the attenuations of the objects are set to 1. Generating projections of those meshes using our forward model may resemble to the so-called inverse crime (Mueller and Siltanen, 2012). To avoid it, we employ the Blender software to make projection data based on

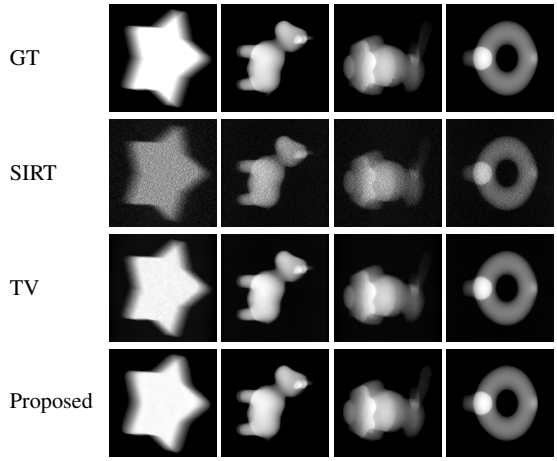


Figure 4: Qualitative results of estimated projections on noisy data with relative noise level 0.4. The first row shows the ground truth, *i.e.*, noise-free data. 2nd-5th row show the forward projection from the solutions by SIRT, TV and the proposed method, respectively.

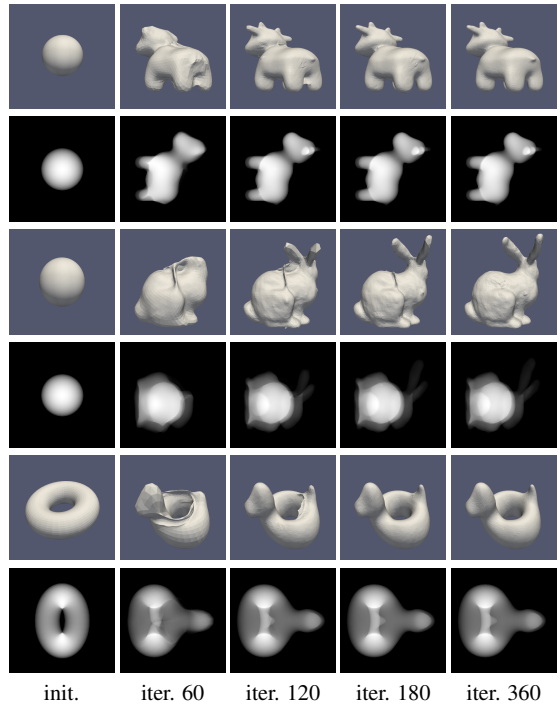


Figure 5: Deformation examples. The odd rows show the intermediate meshes during deformation and the even rows show the corresponding computed projections for one projection angle. We refine the mesh in finer resolution at iteration 60 and fix the mesh at iterations 120 and 180.

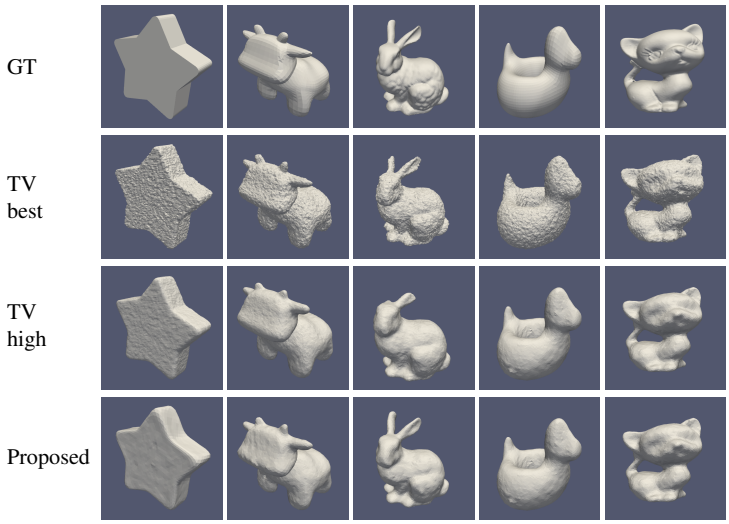


Figure 6: Qualitative results of extracted meshes on noisy data with relative noise level 0.4. The top row shows the ground truth meshes. Rows 2 and 3 show the extracted isosurface from the results of TV reconstruction with the optimal regularization parameter (best) and a high regularization parameter (high), respectively. The last row shows our results.

ray casting methods similar to (Marinovszki et al., 2018). We use 3D parallel projection geometry with 30 projection angles and the detector of size 192×192 pixels. Some projection images without noise are shown in the first row of Fig. 4 for one projection angle.

Evaluation metrics. We compare our result to two standard reconstruction methods: simultaneous iterative reconstruction technique (SIRT) (Andersen and Kak, 1984) and total variation (TV) based reconstruction (Chambolle and Pock, 2011; Sidky et al., 2012). These methods yield 3D images, whereas the proposed method produces surface mesh, making direct comparison of the main output challenging. For consistent comparison, we employ a residual-based metric: *residual projection error* (Roelandts et al., 2014), which measures the L^2 norm difference of data and computed projections. We impose the relative Gaussian noise on the original data and calculate the residual projection error between noise-free projections and the estimations of other methods and the proposed method. For SIRT and TV, the voxel size is set as $192 \times 192 \times 192$ and the algorithm parameters are chosen carefully.

Experimental details. Our implementation relies on PyTorch (Paszke et al., 2019) and uses Adam (Kingma and Ba, 2015) as an optimizer. The proposed forward projector is implemented as a module in PyTorch. As for the regularization parameters, we fix $\alpha = 10$, $\gamma = 0.01$ and iterate 500 times. The step size τ (learning rate in PyTorch) is set to 0.01 and reduced by half at 400 iteration. This reduction step is not really necessary but can yield a more stable result. We observe that 500 iterations are needed for capturing fine details of the complicated objects such as the *bunny* data. During the experiment, we only vary the edge length parameter β among the values of 1, 2, 4, 8, 16 and 32 and the optimal parameter would depend on the data. Finding the optimal regularization parameter is itself a research topic and not straightforward also in regularization-based image reconstruction methods such as TV. As for the initialization, the proposed method begins from an icosphere for genus-0 objects (*star, spot, bunny*), and from a torus for genus-1 objects (*bob, kitten*). Except for *star* data, we refine the mesh by (Huang et al., 2018) at iteration 60 and improve the mesh quality 3 times by a lightweight repair method (Attene, 2010) at iteration 60, 120, 180. These refine and repair steps help remove some artifacts and lead to fast convergence. As also observed in (Vidal et al., 2009), some artifact pixels can appear (e.g., when the ray hits an odd number of times). When a large deformation happens, we may observe some artifact pixels, which we exclude in the objective function. However, in the end, we only observe around 2 artifact pixels. In Fig. 5, we show the deformation of the estimated meshes with the corresponding computed projections.

Robustness to noise. The experiments show that the proposed method is robust to noise. Fig. 4 shows some computed projection images of the reconstruction results from noisy data with relative noise level 0.4 achieved using SIRT, TV reconstruction and the proposed method. Since SIRT has no regularization, it fits closely to the highly noisy data. The results of TV are relatively smooth, but sharp transitions appear blurred. On the other hand, the proposed method yields projections similar to noise-free data. Fig. 6 shows the final mesh results, where the proposed method yields qualitatively better results than TV – we omit highly noisy SIRT results. In Fig. 7, we provide the quantitative results of residual projection error with respect to relative noise levels. Without noise, SIRT gives the superior result as it fits to data without regularization. However, as noise

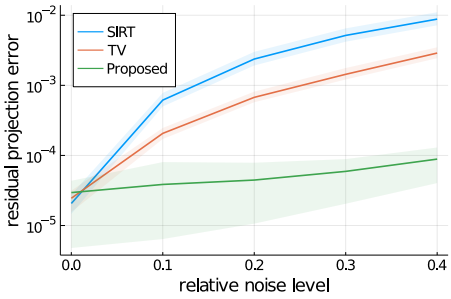


Figure 7: Quantitative results with varying relative noise level over 5 datasets. The *residual projection error* represents the error between the noise-free projections and the estimated projections. The error bars denote the average values with the maximum and minimum value and the y-axis is in logarithmic scale.

increases, the results of SIRT and TV are shown to be poorer.

Effect of parameters. Here, we investigate the effects of the parameters. In this experiment, we use data with a fixed relative noise level of 0.4 and use *residual projection error* to measure performance. Fig. 8 shows the effect of initializing the mesh by varying the radius of the icosphere used for initialization and then iterating 500 times. The final result is mostly not affected by the initial radius size thanks to the refinement steps during the optimization. We choose the *bunny* data model to generate the results shown in Fig. 8, because of its complicated shape.

In Fig. 9 we show the effects of the regularization parameters α, β and γ, while keeping other parameters fixed. Here we use all data models and the residual projection error for the *star* is large due to the coarse resolution of the mesh. We observe that if $\beta < 1.0$ or $\gamma > 0.1$, the final meshes might have some artifacts. As mentioned before, we use the default settings ($\alpha = 10$ and $\gamma = 0.01$), which in most experiments give a stable result. With these fixed values, we observe that it is enough to only vary the edge length parameter, depending on the desired degree of smoothness.

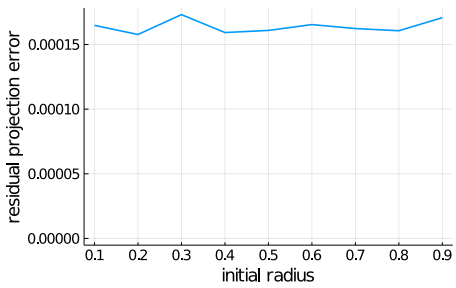
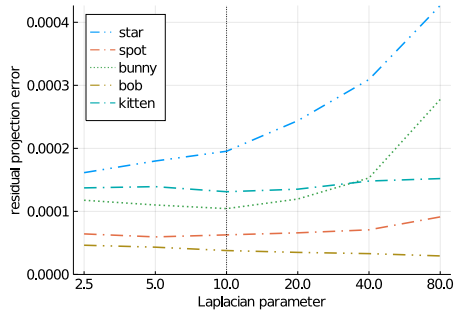
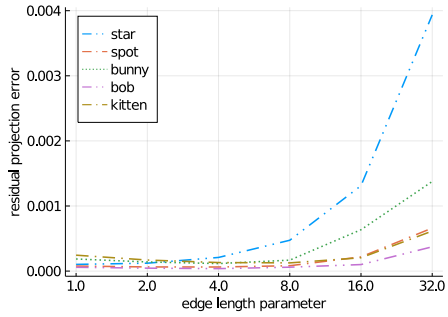


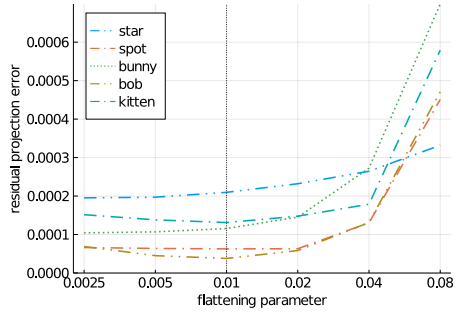
Figure 8: Effect of initializing the mesh with spheres of varying diameter. The x-axis represents the radius of the initialized sphere. The radius is determined relatively to an object space that is normalized to $(-1, 1)^3$. The experiments are based on the *bunny* data model.



(a) Effect of α



(b) Effect of β



(c) Effect of γ

Figure 9: Effect of the regularization parameters: (a) the Laplacian regularization parameter α given $\beta = 4$, $\gamma = 0.01$, (b) the edge length parameter β given $\alpha = 10$, $\gamma = 0.01$, and (c) the flattening parameter γ given $\alpha = 10$, $\beta = 4$. The vertical bars in (a) and (c) correspond to the values we keep fixed, when our method is compared with other methods.

Computational cost. Table 1 shows the size of the initial and final mesh and the corresponding running times. The mesh size is one of the major factors that contribute to computational time. For example, as we do not refine the mesh for *star* object, its final mesh size and its computational time is lower than others. We do the experiment on a Ubuntu server with 256GB RAM and Titan X GPU.

For a reference, the running times for the image-based methods are 38 seconds for SIRT and 152 seconds for TV which are implemented based on multithreading on 8 CPU cores, not GPU. For image-based methods, we use the fixed grid of size $192 \times 192 \times 192$ for all the data.

Table 1: Size of meshes and running times

Data	Star	Spot	Bunny	Bob	Kitten
Initial mesh	Icosphere (1280 faces)			Torus (3200 faces)	
Final mesh	no refine.	18680	17908	19308	20904
Run time	244.1 (sec.)	730.7	719.6	816.3	816.8

5.2. Application to electron tomography

The goal of this experiment is to estimate the shape of a bimetallic nanoparticle having a Au-core and a Ag-shell nano particles. We obtained 2 tilt series datasets of a nano triangular bipyramid and a nanocube using high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM). We hereby used a Thermo Fisher Tecnai Osiris electron microscope generated at 200kV. Each dataset contains 49 projection images acquired over ±72° with a tilt increment of 3º and a frame time of 4 seconds. This small range of projection angles is typical for electron tomography and makes the reconstruction challenging. Fig. 10 shows three images for each of the two datasets.

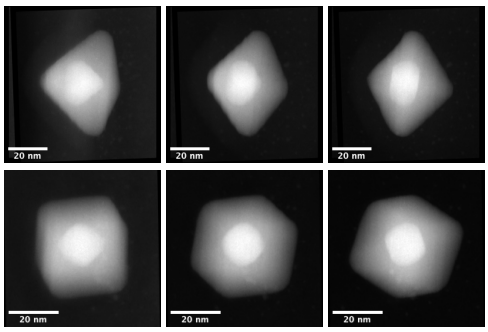


Figure 10: Projection images along 3 different tilt angles for a nano triangular bipyramid (top) and a nano cube (bottom).

To test the proposed method on the data, we initialize the meshes as two icospheres having total 2,560 faces. We use the collision detection method in (Lauterbach et al., 2009), to make sure that the core and the shell part do not collide. In this experiment, we set the number of iterations as 300, the step size τ as 0.005 and observe no collisions when a proper regularization weight is used. As the unknown shapes are relatively simple, we impose only the Laplacian regularization term, by setting $\alpha = 15$ for the nano bipyramid and $\alpha = 5$ for the nanocube. The nanocube data is not much affected by this parameter α, but for the nano bipyramid data, the parameter α should be greater than or equal to 10 to avoid the collision and obtain a high quality mesh when a small range of tilt angles is used. As for the initialization of two icospheres, the inner icosphere's size is set to half the size of the outer icosphere. This outer icosphere is initialized with a relative radius 0.4 (when the object space is normalized to $(-1, 1)^3$), but we obtain a similar result with the radius range of $0.2 \sim 0.8$ for the nano cube and $0.2 \sim 0.5$ for the nano triangular data.

Fig. 11 and Fig. 12 show the 3D reconstruction for the nano triangular bypyramid and the nano cube data, respectively, by SIRT, TV, and the proposed method. To evaluate the effect of tilt angle range, we test two cases: an angular range of ±72° and a subsampled tilt angles of ±18°. To compare our mesh result to the voxel-based methods, we extract meshes from the reconstruction images of SIRT and TV, by median filtering, thresholding, filling holes, and extracting the isosurfaces. For the highly-limited angle case with a tilt range of ±18°, we additionally apply Gaussian smoothing on the results of SIRT and TV before the thresholding step, to obtain the visually-appealing meshes.

For the tilt angle ranges of ±72°, the results of SIRT, TV, and the proposed method are in good comparison. However, when the range of tilt angles is reduced to ±18°, SIRT and TV yield a degenerated reconstruction, which also affects the quality of the final extracted surfaces. On the other hand, the proposed method is shown to be less affected by a tilt range. For the bypyramid in Fig. 11, the volume of the Au-core decreased for the limited angle case, due to the high regularization effect. In Table 2, we provide the estimated volumes of the Au-core and Ag-shell particles by the proposed method and the extracted isosurfaces from the images by SIRT and TV.

Table 2: Estimated volumes in the unit of 10^3 nm by SIRT reconstruction image followed by isosurface extraction, and the proposed method.

Tilt angles	Method	bipyramid		nano cube	
		core	shell	core	shell
-72° ~ 72°	SIRT+Iso.	3.6	33.2	2.4	29.8
	TV+Iso.	3.7	33.5	2.5	29.8
	Proposed	4.2	36.7	2.3	31.2
-18° ~ 18°	SIRT+Iso.	1.3	31.5	1.2	14.9
	TV+Iso.	3.5	25.0	1.3	11.2
	Proposed	2.7	37.7	2.5	29.8

6. Conclusion and Discussion

In this paper, we have developed a forward projector to map a triangular mesh onto the projections domain which is differentiable with respect to the mesh vertices and the attenuation values. Based on our differentiable forward model, we have suggested an optimization-based shape estimation method from projections. Our model should be chosen for reconstructing homogeneous objects with relatively simple geometry, but in situations where the data foundation is so noisy or limited that other methods will not allow a reconstruction. While a conventional approach most often consists of the two steps of image reconstruction followed by a segmentation to obtain a surface, the proposed method directly yields the surface, which among others allows imposing a shape prior directly on the reconstructed shapes. Our experiments on synthetic data show that the proposed method is robust to noise when reconstructing single objects. The experiments on the electron tomography data show how our method is robust even when the range of tilt angles is highly limited, compared with SIRT and TV. In this challenging case, the conventional image reconstructions by SIRT and

TV are highly degraded and these degenerate results propagate errors to the surface estimation from the reconstructed images. Our method can, however, capture the overall shapes well with the regularization term which can impose a smoothness prior directly on the object shapes.

To discuss the limitation of the proposed method, we divide our contributions into two parts: the differentiable forward model and the shape estimation method. The proposed forward model estimates the displacement direction of each vertex for minimizng the objective function. These displacements will only allow change in shape and not collisions of splitting of the mesh that would require topological changes. Handling such topological changes and collisions is itself an active research topic in mesh deformation in computer graphics or computer vision. Investigating how to handle topology changes for a more general mesh adaption and deformation is for future research.

The current differentiable forward model has only been tested on parallel beam geometry where rays are perpendicular to the detector plane. This can be extended to a more general setting such as the cone beam geometry (Buzug, 2008), but we have not investigated the effect of that, as electron tomography data are based on parallel beam geometry.

As a potential application, the proposed method can be useful for estimating the geometry precisely, e.g. for registering the reconstructed object to a 3D CAD model. Another future application can be dynamic tomography where we aim to analyze the objects which may change during scanning.

Acknowledgments

This work is funded by EU Horizon 2020 MSCA Innovative Training Network MUMMERING Grant Number 765604.

References

Aghasi, A., Kilmer, M., Miller, E.L., 2011. Parametric Level Set Methods for Inverse Problems. SIAM J. Imaging Sci. 4.

Alvino, C.V., Yezzi, A.J., 2004. Tomographic reconstruction of piecewise smooth images, in: IEEE Conference on Computer Vision and Pattern Recognition.

Andersen, A.H., Kak, A.C., 1984. Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. Ultrason. Imaging 6.

Attene, M., 2010. A lightweight approach to repairing digitized polygon meshes. Vis. Comput. 26.

Bonin, A., Chalmond, B., Lavayssière, B., 2002. Monte-Carlo simulation of industrial radiography images and experimental designs. NDT & E International 35.

Buzug, T.M., 2008. Computed Tomography from Photon Statistics to Modern Cone Beam CT. Springer.

Chambolle, A., Pock, T., 2011. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. J. Math. Imaging Vis. 40.

Chen, W., Gao, J., Ling, H., Smith, E.J., Lehtinen, J., Jacobson, A., Fidler, S., 2019. Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer, in: Neural Information Processing Systems.

Dahl, V.A., Dahl, A.B., Hansen, P.C., 2018. Computing segmentations directly from x-ray projection data via parametric deformable curves. Meas. Sci. Technol. 29.

Elangovan, V., Whitaker, R.T., 2001. From Sinograms to Surfaces: A Direct Approach to the Segmentation of Tomographic Data, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Eliasof, M., Sharf, A., Treister, E., 2020. Multi-modal 3D shape reconstruction under calibration uncertainty using parametric level set methods. SIAM J. Imaging Sci. 13.

Freud, N., Duvauchelle, P., Létang, J.M., Babot, D., 2006. Fast and robust ray casting algorithms for virtual X-ray imaging. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 248.

Gadelha, M., Wang, R., Maji, S., 2019. Shape Reconstruction Using Differentiable Projections and Deep Priors, in: International Conference on Computer Vision.

Huang, J., Su, H., Guibas, L., 2018. Robust Watertight Manifold Surface Generation Method for ShapeNet Models. ArXiv180201698 Cs .

Kadu, A., van Leeuwen, T., Batenburg, K.J., 2018. A parametric level-set method for partially discrete tomography, in: International Conference on Discrete Geometry for Computer Imagery.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. Int. J. Comput. Vis. 1.

Kato, H., Ushiku, Y., Harada, T., 2018. Neural 3D Mesh Renderer, in: IEEE Conference on Computer Vision and Pattern Recognition.

Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization, in: International Conference on Learning Representations.

Lauterbach, C., Garland, M., Sengupta, S., Luebke, D.P., Manocha, D., 2009. Fast BVH Construction on GPUs. Comput. Graph. Forum 28.

Liu, S., Chen, W., Li, T., Li, H., 2019. Soft Rasterizer: Differentiable Rendering for Unsupervised Single-View Mesh Reconstruction, in: International Conference on Computer Vision.

Loper, M.M., Black, M.J., 2014. OpenDR: An Approximate Differentiable Renderer, in: European Conference on Computer Vision.

Marinovszki, Á., Beenhouwer, J.D., Sijbers, J., 2018. An efficient CAD projector for X-ray projection based 3D inspection with the ASTRA Toolbox, in: Conference on Industrial Computed Tomography.

Midgley, P., Weyland, M., 2003. 3d electron microscopy in the physical sciences: the development of z-contrast and eftem tomography. Ultramicroscopy 96, 413–431.

Mueller, J.L., Siltanen, S., 2012. Linear and Nonlinear Inverse Problems with Practical Applications. volume 10. Society for Industrial and Applied Mathematics.

Mumford, D., Shah, J., 1989. Optimal approximations by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math. 42.

Osher, S., Fedkiw, R., 2004. Level set methods and dynamic implicit surfaces. Appl. Mech. Rev. 57.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library, in: Neural Information Processing Systems.

Roelandts, T., Batenburg, K.J., den Dekker, A.J., Sijbers, J., 2014. The reconstructed residual error: A novel segmentation evaluation measure for reconstructed images in tomography. Comput. Vis. Image Underst. 126.

Sidky, E.Y., Jørgensen, J.H., Pan, X., 2012. Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm. Phys. Med. Biol. 57.

Sujar, A., Meuleman, A., Villard, P.F., Garcia, M., Vidal, F., 2017. gVirtualXRay: Virtual x-ray imaging library on GPU, in: Computer Graphics and Visual Computing (CGVC).

Vidal, F.P., Garnier, M., Freud, N., Létang, J.M., John, N.W., 2009. Simulation of X-ray attenuation on the GPU, in: Proceedings of Theory and Practice of Computer Graphics, Eurographics Association, Cardiff, UK.

Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G., 2018. Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images, in: European Conference on Computer Vision.

Whitaker, R.T., Elangovan, V., 2002. A direct approach to estimating surfaces in tomographic data. Med. Image Anal. 6.

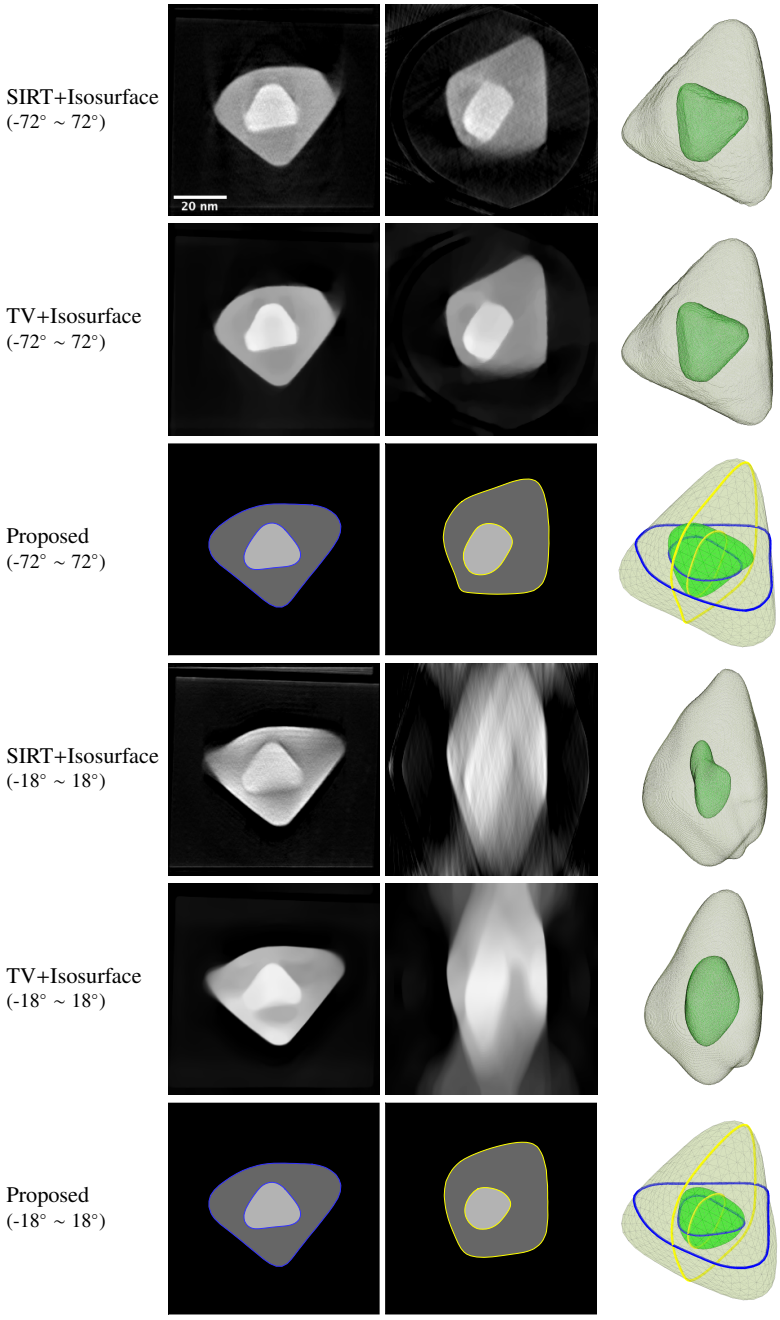


Figure 11: Reconstruction results on nano triangular bypyramid data. The first and the 4th rows show the reconstruction results by SIRT and from projections with tilt angles from -72° to 72° and -18° to 18°, respectively, where two central slices are visualized in the first and second column. The 2nd and the 5th rows show the reconstruction results by TV. The last column shows the extracted mesh from the reconstructed image by SIRT or TV with some post-processing procedures. The third and the last row show our direct shape estimation results, where two central slices are visualized in blue and yellow.

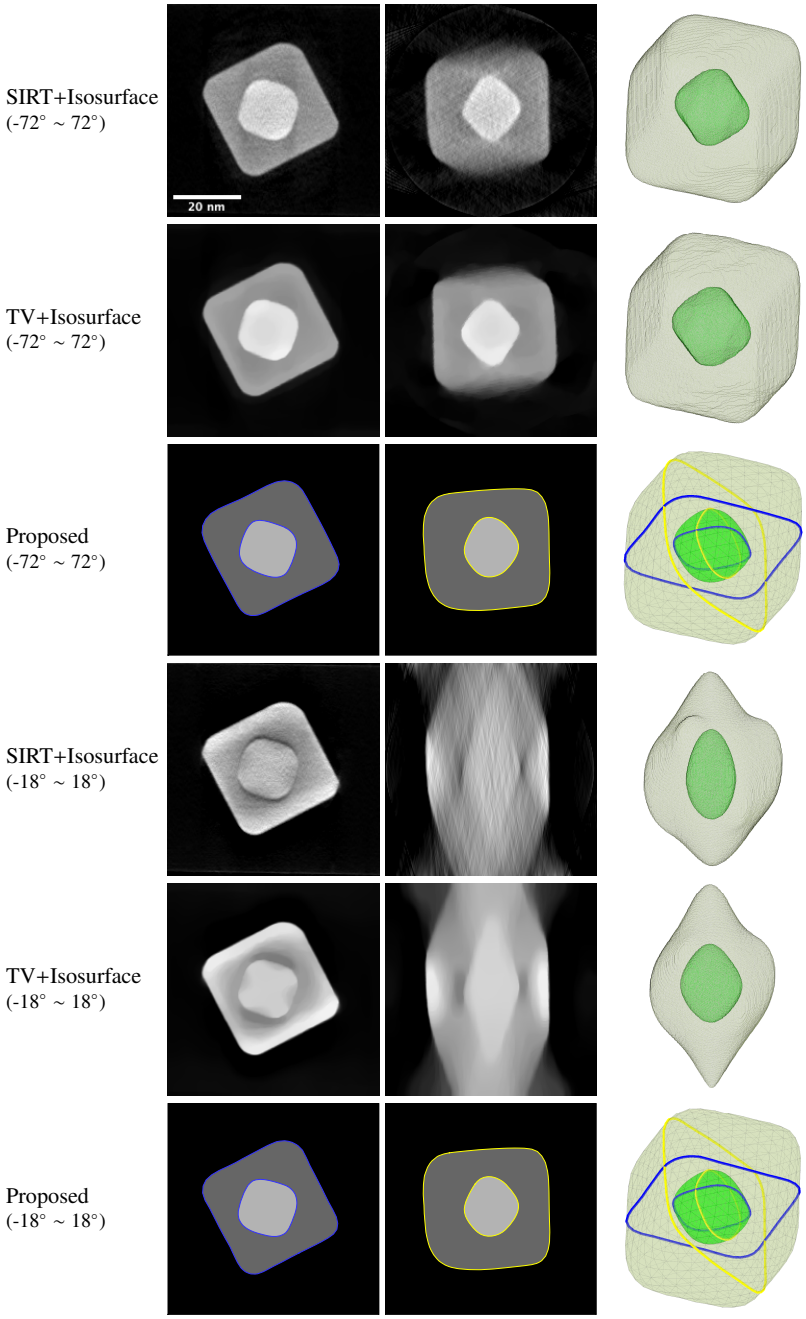


Figure 12: Reconstruction results on nano cube data. The first and the 4th rows show the reconstruction results by SIRT and from projections with tilt angles from -72° to 72° and -18° to 18°, respectively, where two central slices are visualized in the first and second column. The 2nd and the 5th rows show the reconstruction results by TV. The last column shows the extracted mesh from the reconstructed image by SIRT or TV with some post-processing procedures. The third and the last row show our direct shape estimation results, where two central slices are visualized in blue and yellow. The results on the second and the fourth row appear to be the same, but they are slightly different.

Contribution C

A Tomographic Reconstruction Method using Coordinate-based Neural Network with Spatial Regularization

---

The following paper was accepted to Northern Lights Deep Learning Workshop 2021 and presented as an oral presentation. This paper will appear at the Proceedings of Northern Lights Deep Learning Workshop 2021.

A Tomographic Reconstruction Method using Coordinate-based Neural Network with Spatial Regularization

Jakeoung Koo*[1], Elise O. Brenne[2], Anders B. Dahl[1], and Vedrana A. Dahl[1]

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark
[2]Department of Energy Conversion and Storage, Technical University of Denmark

Abstract

Tomographic reconstruction is concerned with computing the cross-sections of an object from a finite number of projections. Many conventional methods represent the cross-sections as images on a regular grid. In this paper, we study a recent coordinate-based neural network for tomographic reconstruction, where the network inputs a spatial coordinate and outputs the attenuation coefficient on the coordinate. This coordinate-based network allows the continuous representation of an object. Based on this network, we propose a spatial regularization term, to obtain a high-quality reconstruction. Experimental results on synthetic data show that the regularization term improves the reconstruction quality significantly, compared to the baseline. We also provide an ablation study for different architecture configurations and hyper-parameters.

1 Introduction

Computed tomography (CT) is a versatile imaging technique allowing the study of interior structures of objects and has many applications, ranging from clinical to industrial applications [5]. In CT, a procedure known as tomographic reconstruction aims to reconstruct the material properties from measurements, called projections, based on the interaction of objects and penetrating radiation such as X-ray. Specifically, the projections are obtained by line integrals of an attenuation coefficient function along straight lines and reconstructing an attenuation function is the goal of tomographic reconstruction.

Although the measurements are finite, we can consider different ways to represent the attenuation function. Among them, a discrete image on a regular grid has been popular in most of existing works including model-based [3, 14] or recent deep learning-based works [1, 8, 19]. In those works, the attenuation value within one pixel is typically assumed to be constant. From this image representation, conventional reconstruction works are based on solving a system of linear equations with some regularization terms.

Recently, there has been growing interest in coordinate-based neural networks to represent continuous functions [12, 16, 15]. Such networks input continuous spatial coordinate and output the signal on the coordinate. To represent high frequency features, a feature mapping of input coordinates was suggested in [12, 16] and another work called SIREN [15] used the sine function as the activation function with a specific initialization scheme.

In this paper, we study a coordinate-based neural network for tomographic reconstruction. We use the same architecture as in SIREN [15], to reconstruct a continuous attenuation function. We use the neural network, but our method does not require any training data. Our main contribution is to propose a regularization term to impose spatial smoothness. We also provide an ablation study with various configurations of the network and hyper-parameters.

Concurrent to our submission, there is a recent paper [16], which applied a coordinate-based network to

*Corresponding Author: jakoo@dtu.dk

different tasks including tomographic reconstruction. Our method differs in that we use another neural network architecture [15] and introduce a spatial regularization term. Our experimental results in Sec. 3 show that the regularization term improves the performance significantly.

1.1 Related works

Methods for tomographic reconstruction using deep neural networks can be classified into two categories: learning-based and learning-free approach. The learning-based approach utilizes the available data in a supervised or unsupervised manner. Zhu et al. [19] proposed a general method to train neural networks from pairs of projection data and ground truth reconstruction images. He et al. [8] considered two neural networks, where the first network simulate the filtered backprojection method [5] and the output of the first network is refined by the second convolutional neural network. The idea of filtered backprojection is also used in [10, 18]. Another approach is to combine conventional iterative methods and neural networks, called learned iterative methods [1, 2]. In this approach, convolutional layers replace some parts of the iterative methods and receive the forward projection operator and its adjoint as part of the inputs.

A learning-free approach does not need any training data, but still harnesses the power of neural networks. Gadelha et al. [7] extended Deep Image Prior (DIP) [17] to 2D tomographic reconstruction. DIP represents an image by the output of a learnable convolutional neural network on a fixed random input and have an implicit regularization effect. The disadvantage of [7] is that it requires early-stopping to prevent overfitting to noisy data.

As our method belongs to the learning-free approach, we compare our method to [7] shown in Sec. 3. The main difference of our work, compared to image-based methods, is that the outcome of our method is a continuous function, instead of discrete image, and the forward model does not depend on regular grid, which will be explained in the next section. Moreover, as we include the regularization term, our method does not need early-stopping.

2 Method

The aim of tomographic reconstruction is to reconstruct an attenuation coefficient function f from a finite number of projections. We first explain the representation of solution f based on a coordinate-based neural network and the forward model to connect the solution and measurements. Then, we introduce a regularization term and define our loss function.

2.1 Continuous representation of the attenuation function

To represent a continuous function, we employ a coordinate-based neural network, called SIREN [15]. As shown in Fig. 1, SIREN is a multi-layer perceptron whose input $\mathbf{x} = (x, y)$ is a spatial coordinate and the output $f_\Theta(\mathbf{x})$ represents the signal at that coordinate where Θ represents neural network parameters – the weight matrices and the bias vectors. In our case, the output corresponds to the attenuation coefficient. The key feature of SIREN is to use the sine function as the activation in the network and a principled initialization, which allows the representation of high frequency features.

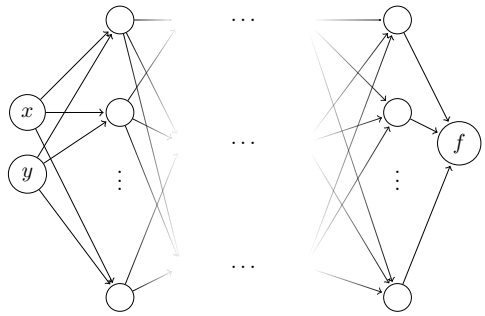


Figure 1: The coordinate-based neural network maps a spatial coordinate $\mathbf{x} = (x, y)$ to the attenuation coefficient $f_\Theta(\mathbf{x})$.

To be specific, each layer with the input $\mathbf{z}$ from the previous layer is constructed as

$$\sin\left(\omega(W\mathbf{z} + \mathbf{b})\right) \tag{1}$$

where W is a weight matrix and $\mathbf{b}$ is a bias vector in the layer and ω is a hyper-parameter to control the spatial frequency. We use Rectified Linear Unit (ReLU) as the last layer activation, to impose a non-negativity constraint.

2.2 Forward model

We model the process of obtaining tomographic measurements from the attenuation coefficient function $f : \mathbb{R}^2 \to \mathbb{R}$. We follow a common assumption in tomographic reconstruction based on Lambert-Beer's law [5], where radiation intensity is attenuated exponentially. After pre-processing, the projection data for a ray can be considered as the line integral of f along the path. The projection data consist of M measurements where M is the multiplication of the number of detector pixels and the number of projection angles. Each measurement can be assigned to a ray from the source of X-rays (or other waves) towards the corresponding detector pixel. Consider a ray i with an origin $\mathbf{o}$ and a normalized direction vector $\mathbf{d}$. Then, the projection p_i can be computed as the line integral of f along the ray:

$$p_i = \int_a^b f_\Theta(\mathbf{o} + t\,\mathbf{d})dt \tag{2}$$

where a and b denote the initial and the last value for t.

To numerically integrate Eq. (2), we employ the mid-point rule by dividing the limits of integration (a, b) into N subintervals:

$$p_i = \sum_{k=1}^{N} f_\Theta(\mathbf{x}_k)\,\delta \tag{3}$$

where $\delta = (b - a)/N$ and $\mathbf{x}_k$ is the midpoint of the intervals defined by

$$\mathbf{x}_k = a + \frac{\delta}{2} + (k - 1)\delta. \tag{4}$$

The forward projection in Eq. (3) is differentiable with respect to the neural network parameters and automatic differentiation tools can be used.

2.3 Regularization and loss function

To deal with noisy data and obtain a more accurate reconstruction, we introduce a regularization term to impose spatial smoothness, inspired by total variation [13]. The total variation term computes a spatial gradient from both the x- and y-direction. Instead of two directions, we impose smoothness on one direction along the ray, for computational efficiency. Specifically, we aim to minimize the variation on each point along the ray i:

$$R_i = \frac{1}{\sqrt{\delta}} \sum_{k=1}^{N-1} \sqrt{|f_\Theta(\mathbf{x}_{k+1}) - f_\Theta(\mathbf{x}_k)|} \tag{5}$$

where we omit the dependency of i in each $\mathbf{x}$, for notational simplicity. We use the square root function to make the regularization term more robust [4].

From the computed projection p in Eq. (3), we aim to fit it to the projection data $\hat{p}$ by minimizing the L_2 norm between them. With this data fitting term and the regularization term (5), we define the loss function to minimize with respect to neural network parameters Θ:

$$\mathcal{L}(\Theta) = \frac{1}{M} \sum_{i=1}^{M} \|p_i - \hat{p}_i\|_2^2 + \frac{\lambda}{M} \sum_{i=1}^{M} R_i, \tag{6}$$

where λ is a hyper-parameter to control the weighting between the data fidelity and regularization term.

To optimize Eq. (6), we use the mini-batch gradient descent method where at each iteration, we choose B random measurements among in total M available measurements.

3 Experimental Results

In this section, we compare our work to other image-based methods on simulated data qualitatively and quantitatively. We conduct an ablation study for different neural network settings and the regularization parameter. We also provide a result for real data.

3.1 Results on simulated data

We use phantom images with the size of 512 × 512 pixels shown in the first column of Fig. 2. These

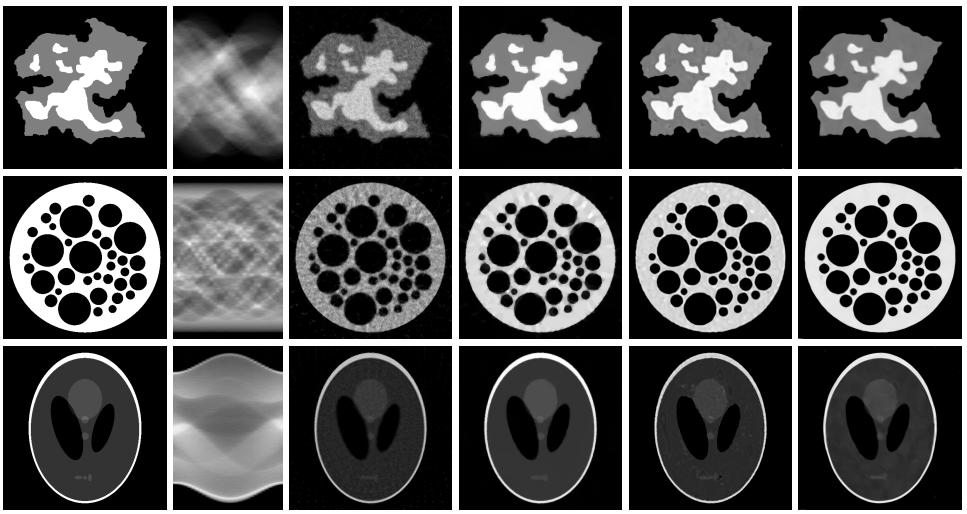


Figure 2: Qualitative results on noisy projection data. The first column shows the ground truth phantoms of Phantom 1, 2, 3 which are used to generate noisy projection data on the second column. The other columns show the reconstruction results.

Phantoms	PSNR					SSIM				
	SIRT	TV	DIP	Ours		SIRT	TV	DIP	Ours	
				$\lambda = 0$	$\lambda \neq 0$				$\lambda = 0$	$\lambda \neq 0$
Phantom 1	23.08	**28.12**	27.83	25.86	27.36	0.524	0.896	**0.941**	0.851	0.939
Phantom 2	16.52	19.70	21.20	20.88	**22.04**	0.468	0.764	0.881	0.820	**0.933**
Phantom 3	25.59	27.09	**27.44**	26.05	26.09	0.710	0.944	0.949	0.906	**0.958**

Table 1: Quantitative comparison to other methods on synthetic data from the phantom images in Fig. 2. We show two results for our method without regularization $\lambda = 0$ and with regularization $\lambda \neq 0$.

phantoms are used to generate projection data by an image-based linear forward model with parallel beam, shown in the second column. Each projection datum consists of 512 detector pixels and 30 projection angles. We impose Gaussian noise on the projection data with the relative noise level 0.02, which is chosen to reflect noise degree in real data.

Experimental details. Our method is implemented in the Julia language and based on a deep learning library called Flux [9]. Unless explicitly mentioned, we use a neural network with 3 hidden layers where each layer has size 128. That is, the weight matrix for each hidden layer has size 128×128, while the bias vector has 128 elements. We consider the reconstruction domain bounded into $(-1, 1) \times (-1, 1)$. In this bounded domain, the spatial frequency parameter $\omega = 30$ yields good performance overall. We initialize neural network parameters in

the same way as SIREN did [15] where the distributions of activations are preserved through the network. We use the Adam optimizer [11] with learning rate 0.0001, batch size $B = 512$ and iteration number 4,500 (150 epochs) as the solution changes very little after that. The batch size corresponds to the number of rays per iteration and the rays are sampled randomly. For the numerical integration (3), we divide the range of integration into $N = 256$ intervals.

Comparison to other image-based methods. As our method belongs to a learning-free approach, we compare our method to conventional model-based approaches including: SIRT [3] and a Total Variation (TV)-based method [14]. We also compare to a method [7] based on Deep Image Prior (DIP) [17]. As mentioned in Sec. 1.1, DIP needs an early stopping. For DIP, we iterate 2000 times and save the results every 50th iteration and among those candidate solutions, we choose the best result.

To compare our result to image-based methods, we produce images from our implicit solutions by evaluating the network at each pixel position. For quantitative comparison, we employ two image-quality metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). We choose the same configuration for 3 phantoms except for the regularization parameter λ, which is set 0.0005 for Phantom 1 and 2 and 0.0001 for Phantom 3.

Fig. 2 provides a visual comparison of the results by SIRT, TV, DIP and our method and Table 1 shows the corresponding quantitative results. Due to the noisy data, the result images have some high values, which make the images look darker than the ground truth images. Overall, our method gives competitive results. Fig. 3 shows some intermediate results of our method during optimization for Phantom 2.

3.2 Ablation study

In this ablation study, we use projection data without noise and use the same settings as before.

Effect of network size. We vary the layer size, but fix the number of hidden layers. As shown in Table 2, as the layer size increases, the number of network parameters increases approximately 4 times. For reference, for an image of size 512×512, image-based methods optimize 262,144 pixel values. When the layer size is 512 with a large number of parameters, the reconstruction quality is shown to degrade.

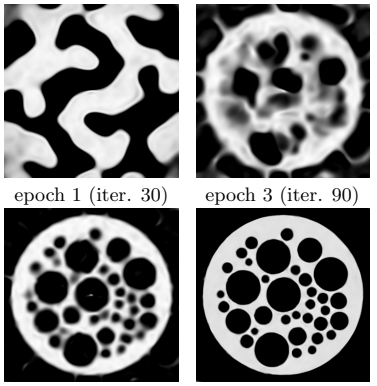


Figure 3: Intermediate results of our method during optimization.

layer size	# params.	PSNR	SSIM
32	3,296	19.1	0.827
64	12,736	21.2	0.907
128	50,048	23.4	0.953
256	198,400	**23.6**	**0.957**
512	790,016	21.5	0.930

Table 2: Effect of network size for the reconstruction of Phantom 2.

Effect of sampling points N. The number of sampling points per ray affects both reconstruction quality and the computational cost. Table 3 provides the effect of N. As expected, as N increases, the computation time also increases.

Effect of regularization hyper-parameter λ. Fig. 4 shows the effect of λ. Without regularization, some artifacts are observed and, with large regularization $\lambda = 0.0005$, our method could not capture some fine details. The value $\lambda = 0.0001$ is shown to yield the best result in terms of both visual quality and PSNR.

N	time per iter.	PSNR	SSIM
64	0.031	20.03	0.846
128	0.065	22.00	0.927
256	0.110	23.37	0.953
512	0.293	23.61	0.961
1024	1.115	**24.48**	**0.968**

Table 3: Effect of the number of sampling points for the reconstruction of Phantom 2. The computation time in GPU per iteration is shown in seconds.

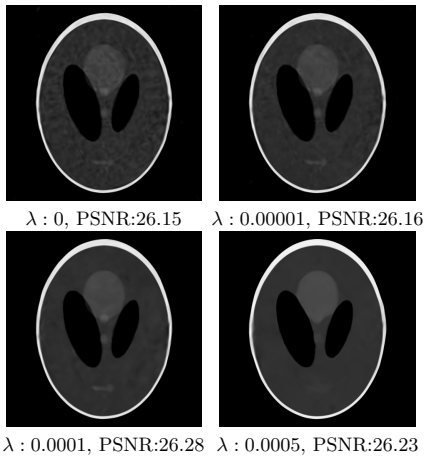


Figure 4: Effect of regularization parameter λ.

3.3 Results on real fan beam data

We test our method on cone-beam X-ray CT data from the SophiaBeads Dataset [6]. The scanned sample is a plastic tube filled with glass beads. Specifically, we use the provided central slice of the dataset "SophiaBeads_1024_averaged", choose 64 projections at angles in $[0°, 360°]$ and down-sample with a factor 2 so that each projection is 680 pixels wide. As a pre-processing, we correct the center-of-rotation by shifting the projections 12 pixels and remove a high-intensity edge artifact by subtracting 0.0015 from the projection. As we use the central slice of a cone-beam dataset, a fan-beam geometry is used for the reconstruction.

In Fig. 5, we provide the results of SIRT and our method. This demonstrates that our method can successfully be applied to a real data and that the performance is comparable to that of SIRT.

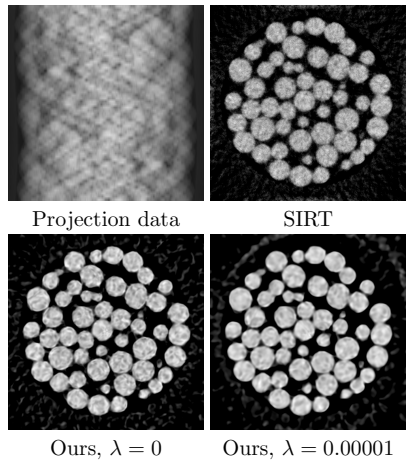


Figure 5: Reconstruction results for real fan beam data.

4 Conclusion and Discussion

We have proposed a spatial regularization term for tomographic reconstruction using a coordinate-based neural network. The experimental results show that the regularization term improves the reconstruction quality significantly, although choosing optimal regularization parameter is not trivial. For a practical use of our method, the major limitation is the high demand of memory requirement, which prevents us from choosing a large number of rays per iteration. This is an issue, especially in 3D reconstruction where projection data consist of a large number of measurements. We leave it as future work to improve the speed. Another future direction would be to include the encoding of projection data in the network, so that the network can infer to unseen projection data.

References

[1] J. Adler and O. Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12), 2017.

[2] J. Adler and O. Öktem. Learned Primal-dual Reconstruction. *IEEE Trans. Med. Imaging*, 37(6), 2018.

[3] A. H. Andersen and A. C. Kak. Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrason. Imaging*, 6(1), 1984.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, 2004.

[5] T. M. Buzug. *Computed Tomography from Photon Statistics to Modern Cone Beam CT*. Springer, 2008.

[6] S. B. Coban and S. A. McDonald. Sophiabeads dataset project, Mar. 2015.

[7] M. Gadelha, R. Wang, and S. Maji. Shape Reconstruction Using Differentiable Projections and Deep Priors. In *International Conference on Computer Vision (ICCV)*, 2019.

[8] J. He, Y. Wang, and J. Ma. Radon inversion via deep learning. *IEEE Transactions on Medical Imaging*, 39(6):2076–2087, 2020.

[9] M. Innes. Flux: Elegant machine learning with julia. *J. Open Source Softw.*, 2018.

[10] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Trans. Image Process.*, 26(9), 2017.

[11] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

[12] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.

[13] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. Nonlinear Phenom.*, 60(1-4), 1992.

[14] E. Y. Sidky, J. H. Jørgensen, and X. Pan. Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm. *Phys. Med. Biol.*, 57(10), 2012.

[15] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein. Implicit Neural Representations with Periodic Activation Functions. *arXiv:2006.09661*, 2020.

[16] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *arXiv:2006.10739*, 2020.

[17] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[18] T. Wurfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier. Deep Learning Computed Tomography: Learning Projection-Domain Weights From Image Domain in Limited Angle Problems. *IEEE Trans. Med. Imaging*, 37(6), 2018.

[19] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697), 2018.

Contribution D

Material clasificiation from sparse spectral X-ray CT using vectorial total variation based on L infinity norm

---

The following paper is in preparation. In this paper, my contribution is in the methodology part on the reconstruction.

Material classification from sparse spectral X-ray CT using vectorial total variation based on L infinity norm

Doniyor Jumanazarov[a,c,*], Jakeoung Koo[b], Henning F. Poulsen[a], Ulrik L. Olsen[a] and Mihai IOVEA[c]

[a]Technical University of Denmark, DTU Physics, 2800 Kgs. Lyngby, Denmark
[b]Technical University of Denmark, DTU Compute, 2800 Kgs. Lyngby, Denmark
[c]ACCENT PRO 2000 s.r.l. (AP2K), Nerva Traian 1, K6, Ap. 26, Bucharest, S3, ROMANIA 031041

ARTICLE INFO

Keywords:
Spectral X-ray CT
Electron density
Effective atomic number
Joint reconstruction
Security screening

ABSTRACT

It is common knowledge that the development of energy-resolving photon counting detectors (PCD) in spectral X-ray imaging opened the way to the rise of Spectral X-ray Computed Tomography (SCT), which simultaneously measures a material's linear attenuation coefficient (LAC) at multiple energies. In previous work, we introduced the method for system-independent material classification through X-ray attenuation decomposition (SIMCAD) from SCT, which estimates the material properties of electron density (ρ_e) and effective atomic number (Z_{eff}). In this work, using real experimental data we extend the SIMCAD method to the case where materials are identified from sparse-view (few) projections for enabling rapid scanning and from noisy projection data, which are of highly practical significance in security screening. Therefore, for reconstructing images from spectral sinogram we implement L_∞ vectorial total variation (L_∞-VTV) as a convex regularizer, which is for the first time implemented in SCT to our knowledge. This joint reconstruction algorithm employs the L_∞ norm to penalize deviations from the inter bin dependency, taking advantage of strong correlations among the energy bins. L_∞-VTV has been proven to be successful regularizer in image processing problems. We also investigate "total nuclear variation" (TNV), which is a state-of-the-art method for joint reconstruction and uses the nuclear norm to penalize the singular values of image matrices using the l^1 norm, so that gradient vectors of different energy bins are forced to point into the same direction. We use the correlation coefficient as a figure of merit (FOM) to measure the reconstructed image quality. We show that the joint reconstruction algorithms give the improved reconstructions compared to bin-by-bin reconstructions such as total variation (TV) and SIRT when scanning is performed with different detector noise level, and when significantly higher levels of noise are introduced to certain energy bins. We test also classification performance of the algorithms using as an another FOM. In 7-projections case, L_∞-VTV gives 8.6% and 75.0%, and 297.1% and 25.0% better accuracy for estimating ρ_e and Z_{eff} compared to TV and SIRT, respectively, while the improvements are 11.8% and 35.5%, and 308.8% and 3.2% for TNV, respectively.

1. Introduction

Many studies have been done in addressing the sparse data problems either in image processing or CT Systems. The total-variation (TV) introduced by Rudin, Osher, and Fatemi [1] has been effectively used for various image processing problems in noise suppression while enabling sharp discontinuities and preserving edges. Therefore, various TV based reconstruction algorithms have been implemented for sparse-view and limited angle CT [2, 3]. Sidky et al. (2006) [4] implemented a TV-regularized reconstruction in fan-beam CT for different insufficient data problems such as the few-view projections, limited-angle, and bad-bin cases. The results showed that the TV regularization outperformed the existing methods in the reconstructions in such cases suffering from incomplete projection data. Blomgren and Chan (1998) [5] introduced a first definition of vectorial TV (VTV) regularizer that extends the conventional scalar TV into the three-dimensional frame for image processing problems with considering the inter color channel dependency. The main question when formulating a VTV as a convex regularizer is how exactly to couple color channels. For an image denoising task, Miyata et al. (2012) [6] proposed using a L_∞ norm for defining VTV and showed that this norm can efficiently estimate the violation of inter-channel dependency and provide strong coupling among color channels. Consequently, it leads to better performance in image denoising compared to the commonly used norms of VTV.

Holt (2014) [7] investigated several Jacobian-based VTV types with color image denoising experiments that use nuclear norm in the Jacobian structure, which is the convex envelope of matrix rank [8]. They showed that minimizing Jacobian rank leads to promoting gradient vectors in different color channels to share a common direction, and using nuclear norm makes TNV to be stronger convex regularizer for encouraging shared directions compared to other convex Jacobian-based regularizers, including Total Frobenius Variation (TFV) and Total Spectral Variation (TSV). For image processing tasks, Duran et al. (2016) [9] implemented and analyzed several VTV regularizations using so-called collaborative norms that can penalize the three dimensions of the discrete gradient of the multi-channel image corresponding to pixels, spatial derivatives and color channels. They compared a performance of each norm by employing images with different noise level from various

*Corresponding author.
doniyor.jumanazarov@fysik.dtu.dk (D. Jumanazarov); jakoo@dtu.dk (J. Koo); hfpo@fysik.dtu.dk (H.F. Poulsen); ullu@dtu.dk (U.L. Olsen); office@accent.ro (M. IOVEA)
ORCID(s): 0000-0003-1583-7375 (M. IOVEA)

databases. They find that the joint regularizations based on the L_∞ norm and the Jacobian-based nuclear norm (TNV) give the best suppression of color artifacts.

Using simulated data, Rigie et al. (2015) [10] applied TNV to reconstructing tissue density maps from spectral CT, which simultaneously measures the energy dependence of a material's linear attenuation coefficient (LAC) using energy resolved photon counting detectors (PCD). They showed that TNV is better in suppressing noise and has less edge blurring compared to channel-by-channel TV reconstructions. Even though each energy bin of a spectral CT often has remarkably different noise levels, TNV showed robustness to undesired transfer of individual characteristics to other bins. With real data, Rigie et al. (2017) [11] implemented TNV and another VTV algorithm using the Frobenius norm as well as TV in dual-energy CT. They find that both VTV regularized couplings are better at suppressing noise compared the conventional TV while TNV was still the best regularizer among them. Zhong et al. (2018) [12] investigated to reconstruct energy-dispersive X-ray spectroscopy (EDS) tomographic images by combining with a more precise high-angle annular dark-field STEM (HAADF-STEM) tomography through TNV in case of a small number of tilts and a low number of photon counts. The experiments presented that TNV regularization is more precise in preserving features of reconstructed images compared to the simultaneous iterative reconstruction technique (SIRT) [13] and TV reconstruction methods in these difficult cases.

Spectral CT is just CT with multiple "color" channels, these advanced algorithms from image processing can be applied to spectral CT reconstruction. In this work, especially motivated by the work of Duran et al. [9] we implement a L_∞ norm-based VTV method in Spectral CT, which we borrow from Miyata et al. (2012) [6] in referring to it as L_∞-VTV. We also investigate TNV joint reconstruction algorithm using the experimental data. We compare their performances with TV and SIRT reconstructions. The term of a channel is referred to a bin.

In previous work we presented a novel method for material classification, named system-independent material classification through attenuation decomposition (SIMCAD) [14], which estimates the system-independent material properties (ρ_e, Z_{eff}), independent of the instrument or specifics of the scanner such as the X-ray spectrum, directly from energy resolved LACs in spectral CT. This method employs attenuation decomposition method introduced by Alvarez et al. [15] for the formulation of the method and adopts it for multi-energy case. The method can meet the requirements of rapid scanning for industrial applications such as check-in baggage control at airports [16]. In this work, stemming from practically importance we aim to extend the method into the case where materials are classified from sparse-view (few) projections again for increasing the classification speed motivated by security applications. We also test the method for noisy projection data. A broader range of materials is used for experimental validation. We focus on improving the reconstruction quality, which would lead to more accurate LAC extraction and as a result to targeted better material classification. To combat the heavy sparse-view artifacts, we leverage the structure coherence of a scanned object sharing the same structure information at various energies and jointly reconstruct the images simultaneously employing all the data instead of reconstructing each energy independently. We refer the reader to our previous work [14], especially to the subsections such as Experimental setup and data correction, Setting low- and high-energy thresholds and LAC extraction. We use the MultiX ME100 v2 CdTe detector ("Detection Technologies"), which is a good candidate to measure the energy-dependent LAC in high flux. The detector requires the correction of the detector's spectral response that helps to correct the LAC towards the reference values at the low energies, where the contrast between organics compounds is highest, and also at the high energies [17]. These deviations of the LAC from the reference occur due to the lower photon statistics and pulse pile-up, and they cannot be completely restored by the correction algorithm. We truncate these bins at the low and high energies setting the low- and high-energy thresholds. In previous work, the classification performance was tested for optimized bi-, 6, 15, 30, 45 and 90 energy bins. 15 energy bins approach spaced uniformly between low- (33.2 keV) and high- (132.4 keV) energy thresholds with equal width showed better results compared to 30, 45 and 90 energy bins. One reason may be that using energy bins with a width narrower than the energy resolution of the detector did not improve classification performance; Secondly the detector noise becomes greater as the width of each energy bin becomes narrower while the number of bins increases. Therefore, we use 15 energy bins approach for this study. Thus, SCT settings have a compromise between spectral resolution and noise level. The joint reconstruction of spectral sinogram is expected to compensate for this giving better quality compared to bin-by-bin reconstructions. We only do 2D reconstructions for slices, however the same methods can be applied to the third dimension.

The joint reconstruction algorithm has a positive weighting parameter that controls the balance between a good fit to spectral sinogram data and a smooth reconstruction. This parameter determines the strength of the regularization term. We use the correlation coefficient to estimate the reconstructed image quality against ground-truth images as a figure of merit (FOM). The convergence of each algorithm is ensured based on the sum of primal and dual residuals and the number of iterations is fixed when the correlation coefficient and its gradient remain stable. The weighting parameter is tuned for the best match on the reconstruction based on the correlation coefficients, which are assessed for each algorithm and different level of photon flux and data acquisition time. For each reconstruction method, the robustness to significantly higher noise in certain energy bins is tested, and we investigate a L-curve that is a plot and can be theoretically used for finding the optimal weighting parameter. Lastly, the material classification performance depending on the weighting parameter is tested for each reconstruction algorithm as an another FOM. We use 20 different materials for the calibra-

tion and 15 additional materials for the (ρ_e, Z_{eff}) calculation in the range of $6 \leq Z_{\text{eff}} \leq 15$. Moreover, the classification performance of L_∞-VTV in 7-projection is evaluated by artificially adding Gaussian noise to the experimental data after the spectral correction of detector response.

2. Method

2.1. Problem formulation

In spectral CT, the measurements with N energy bins consist of N sinograms $\mathbf{p}_1, ..., \mathbf{p}_N \in \mathbb{R}^M$ where M is the number of detector pixels times the number of projection angles. From the sinograms, we aim to reconstruct the corresponding N images $\mathbf{u}_1, ..., \mathbf{u}_N \in \mathbb{R}^J$ representing the linear attenuation coefficients for each energy bin E_i, where J is the number of pixels per each image. For notational simplicity, we define the stack of sinograms as a vector $\mathbf{p} \in \mathbb{R}^{NM}$ and the stack of N images as $\mathbf{u} \in \mathbb{R}^{NJ}$ and we denote by $u_{i,j}$ the attenuation coefficient for the energy bin E_i on the j-th pixel.

The forward-projection to map an image to sinogram domain can be modeled by a linear operator A such that $A\mathbf{u}_i$ estimates a sinogram for each energy bin E_i. By stacking the same operator for all energy bins, we define the linear operator $\mathbf{A}$ on the stack of images $\mathbf{u}$ such that $\mathbf{Au} = \mathbf{p}$. Then, a typical reconstruction approach is to minimize the reprojection error (difference between the computed and measured sinogram) in the L_2 sense as:

$$\mathcal{G}(\mathbf{Au}) = \frac{1}{2}\|\mathbf{Au} - \mathbf{p}\|_2^2. \tag{1}$$

This data fitting term is not enough, as data can be highly noisy or obtained by a small number of projection angles. To reconstruct high-quality images, we need a robust regulatization scheme, by imposing a prior knowledge on the solutions.

In the image reconstruction, a popular regularization scheme is Total Variation (TV) [18], which promotes sparse gradients of the images. One can apply (isotropic) total variation regularization on each energy channel E_i, independently, which can be written as

$$\mathcal{R}_{\text{TV}}(\mathbf{u}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\sqrt{\partial_x u_{i,j}^2 + \partial_y u_{i,j}^2}, \tag{2}$$

where $\partial_x u_{i,j}$ and $\partial_y u_{i,j}$ are the gradient of the image $\mathbf{u}_i$ on the pixel j with respect to x and y axis, respectively. This independent total variation regularization does not exploit any correlation between the images from different energy channels.

A more robust regularization can be achieved by correlating the gradients of the images. One way to correlate the gradients is Total Nuclear Variation (TNV) [9, 10], which is defined as the sum of nuclear norm of image gradients:

$$\mathcal{R}_{\text{TNV}}(\mathbf{u}) = \sum_{j=1}^{J}\left\|\begin{pmatrix}\partial_x u_{1,j} & \partial_x u_{2,j} & \cdots & \partial_x u_{n,j}\\ \partial_y u_{1,j} & \partial_y u_{2,j} & \cdots & \partial_y u_{n,j}\end{pmatrix}\right\|_*, \tag{3}$$

where the nuclear norm $\|\cdot\|_*$ of a matrix is the sum of its singular values. Note that TNV can be computed parallelly for each pixel j. TNV encourages the image gradients to align together over different energy channels and can be effective to correlate the images when the image gradients are reliable enough. However, TNV can be degenerate when there are some outliers in the gradients of the images.

To overcome the disadvantage of TNV, we suggest to use another vectorial total variation norm, L_∞-VTV, which is the sum of the maximum of gradients over the multi-channel images, defined as

$$\begin{aligned}\mathcal{R}_{\text{VTV}}^{L_\infty}(\mathbf{Du}) &= \|\mathbf{Du}\|_{\infty,1,1}\\ &:= \sum_{j=1}^{J}\left(\max_{1\leq i\leq N}|\partial_x u_{i,j}| + \max_{1\leq i\leq N}|\partial_y u_{i,j}|\right)\end{aligned} \tag{4}$$

where $\mathbf{D}$ is a linear operator for the discrete gradient operations such that $\mathbf{Du}$ is a 3D matrix. That is, $(\mathbf{Du})_{i,j,l}$ represents the gradient of the image $\mathbf{u}_i$ on the pixel j with respect to x axis when $l = 1$ or y axis when $l = 2$. This L_∞-VTV norm correlates the gradients strongly, while disallowing some outliers in the sense of the gradients magnitudes. This property will be illustrated in the subsection 3.4.

By combining the data fidelity term (Eq. (1)) and the regularization term L_∞-VTV (Eq. (4)), we formulate the optimization problem we want to solve:

$$\min_{\mathbf{u}\geq 0}\ \lambda\,\mathcal{G}(\mathbf{Au}) + \mathcal{R}_{\text{VTV}}^{L_\infty}(\mathbf{Du}), \tag{5}$$

where λ is the weighting parameter between the two terms. Here, we impose the non-negativity constraint on $\mathbf{u}$, as linear attenuation coefficients are desired to be non-negative.

2.2. Optimization

We describe the optimization of our objective function defined in Eq. (5). Although the objective function is convex, the challenge lies in having the composite operators $\mathbf{A}$ in the data fidelity term and $\mathbf{D}$ in the regularization term. To deal with such difficulty, we employ an efficient primal dual algorithm, called Hybrid Gradient Primal Dual method (HGPD) [19]. The basic idea of primal dual algorithms is to reformulate a minimiziaton problem as a minimization-maximization (min-max) problem and aim to find the saddle point. To derive the min-max problem, primal dual algorithms rely on Fenchel conjugate function [19] F^* of a convex function F, defined by,

$$F^*(\mathbf{y}) = \sup_{\mathbf{x}}\langle\mathbf{x},\mathbf{y}\rangle - F(\mathbf{x}). \tag{6}$$

To use the primal dual algorithm, we introduce two dual variables $\mathbf{q}$ and $\mathbf{r}$ corresponding to the operators $\mathbf{A}$ and $\mathbf{D}$, respectively. To rewrite the original minimization problem defined in Eq. (5) as a saddle point problem, we use the Fenchel conjugate function to obtain

$$\mathcal{G}(\mathbf{Au}) = \max_{\mathbf{q}}\langle\mathbf{Au},\mathbf{q}\rangle - \mathcal{G}^*(\mathbf{q}) \tag{7}$$

$$\mathcal{R}^{\mathrm{L}_\infty}_{\mathrm{VTV}}(\mathbf{Du}) = \max_{\mathbf{r}} \langle \mathbf{Du}, \mathbf{r} \rangle - \mathcal{R}^*(\mathbf{r}) \tag{8}$$

where $\mathcal{G}^*$ and $\mathcal{R}^*$ are the conjugate functions of

$$\mathcal{G}(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \mathbf{p}\|_2^2 , \tag{9}$$

$$\mathcal{R}(\mathbf{X}) = \|\mathbf{X}\|_{\infty,1,1} . \tag{10}$$

Based on Eq. (7) and Eq. (8), we now turn the minimization problem of Eq. (5) into the following saddle point problem

$$\min_{\mathbf{u}} \max_{\mathbf{q},\mathbf{r}} \langle \mathbf{Au}, \mathbf{q} \rangle + \langle \mathbf{Du}, \mathbf{r} \rangle - \mathcal{G}^*(\mathbf{q}) - \mathcal{R}^*(\mathbf{r}) + \delta_+(\mathbf{u}), \tag{11}$$

where δ_+ is the indicator function for the non-negative constraint, defined by,

$$\delta_+(x) = \begin{cases} x & \text{if } x \geq 0 , \\ \infty & \text{if } x < 0 . \end{cases} \tag{12}$$

HGPD provides an efficient way to solve the saddle point problem, by alternatively updating the primal and dual variables as follows:

$$\mathbf{u}^{k+1} = \mathrm{prox}_{\delta_+} \left(\mathbf{u}^k - \tau \left(\mathbf{A}^T \mathbf{q}^k + \mathbf{D}^T \mathbf{r}^k \right) \right) \tag{13}$$

$$\bar{\mathbf{u}} = 2\mathbf{u}^{k+1} - \mathbf{u}^k \tag{14}$$

$$\mathbf{q}^{k+1} = \mathrm{prox}_{\sigma_1 \mathcal{G}^*} \left(\mathbf{q}^k + \sigma_1 \mathbf{A}\bar{\mathbf{u}} \right) \tag{15}$$

$$\mathbf{r}^{k+1} = \mathrm{prox}_{\sigma_2 \mathcal{R}^*} \left(\mathbf{r}^k + \sigma_2 \mathbf{D}\bar{\mathbf{u}} \right) \tag{16}$$

where k is the iteration number, τ is the step size for the primal variable, σ_1, σ_2 are the step sizes for the dual variables and prox is the proximal operator [20] defined by

$$\mathrm{prox}_{\sigma f}(\mathbf{x}) = \arg\min_{\mathbf{y}} \frac{1}{2\sigma} \|\mathbf{x} - \mathbf{y}\|_2^2 + f(\mathbf{x}). \tag{17}$$

In Algorithm 1, we provide the whole procedure of optimization, where $\|\mathbf{A}\|_2$ denotes the operator norm of $\mathbf{A}$, which is the largest eigenvalue of $\mathbf{A}$. To guarantee the convergence, we carefully choose the step sizes, following [19], based on the operator norms, which can be computed by the power method [21]. The concrete solutions to Eq. (13) and Eq. (15) are provided in Eq. (21) and Eq. (23), respectively. To evaluate Eq. (16), we use the fact that the proximal operator to L_∞ norm is the projection operator to L_1 unit ball, denoted by $\Pi_{\|\cdot\|\leq 1}$ [20]. Then, by the abuse of notation, we can write the solution to Eq. (16) as follows:

$$\left(\mathrm{prox}_{\sigma_2 \mathcal{R}^*}(\mathbf{V}) \right)_{i,j,l} = \mathrm{sgn}\left(\mathbf{V}_{i,j,l} \right) \Pi_{\|\cdot\|\leq 1} \left(\sigma_2 \|\mathbf{V}_{:,j,l}\| \right) \tag{18}$$

where sgn is the sign function, i denotes the index for the energy bin, j for the pixel index and $l \in \{1, 2\}$ denotes the index for the gradient with respect to x or y axis. The notation $\mathbf{V}_{:,j,l}$ represents a vector consisting of the elements with respect to the energy bins, given j and l. We refer to [9] for a detailed derivation.

To check the convergence, we introduce two residuals $\mathcal{P}$ and $\mathcal{D}$ for the primal variables and the dual variables, respectively, as follows:

$$\mathcal{P}^{k+1} = \|(\mathbf{u}^k - \mathbf{u}^{k+1})/\tau - \mathbf{A}^T(\mathbf{q}^k - \mathbf{q}^{k+1}) - \mathbf{D}^T(\mathbf{r}^k - \mathbf{r}^{k+1})\|_1, \tag{19}$$

$$\mathcal{D}^{k+1} = \|(\mathbf{q}^k - \mathbf{q}^{k+1})/\sigma_1 - \mathbf{A}(\mathbf{u}^k - \mathbf{u}^{k+1})\|_1 + \|(\mathbf{r}^k - \mathbf{r}^{k+1})/\sigma_2 - \mathbf{D}(\mathbf{u}^k - \mathbf{u}^{k+1})\|_1. \tag{20}$$

These primal and dual residuals measure the changes of the solutions per iteration for primal and dual variables, respectively. Such residuals are expected to decrease along the iterations. The behaviors of these residuals and the stopping criteria will be discussed in the subsection 3.1.

Algorithm 1 Primal dual updates for solving Eq. (11)

Set the step sizes:

$$\tau = \frac{1}{\|\mathbf{A}\|_2 + \|\mathbf{D}\|_2}, \ \sigma_1 = \frac{1}{\|\mathbf{A}\|_2}, \ \sigma_2 = \frac{1}{\|\mathbf{D}\|_2}$$

Initialize $\mathbf{u}^0$, $\mathbf{q}_1^0$, $\mathbf{q}_2^0$ as zero vectors.
for $k = 0, 1, 2, \ldots$

$$\mathbf{u}^{k+1} = \max\left(\mathbf{u}^k - \tau(\mathbf{A}^T\mathbf{q}^k + \mathbf{D}^T\mathbf{r}^k), 0 \right) \tag{21}$$

$$\bar{\mathbf{u}} = 2\mathbf{u}^{k+1} - \mathbf{u}^k \tag{22}$$

$$\mathbf{q}^{k+1} = \frac{\mathbf{q}_1^k + \sigma_1(\mathbf{A}\bar{\mathbf{u}} - \mathbf{p})}{1 + \sigma_1} \tag{23}$$

$$\mathbf{r}^{k+1} = \mathbf{r}^k + \sigma_2(\mathbf{D}\bar{\mathbf{u}} - \mathbf{g}) \tag{24}$$

3. Results and discussions

3.1. Stopping criteria for the iteration number

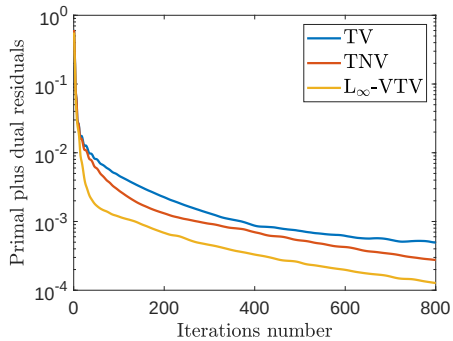


Figure 1: The sum of the primal and dual residuals with respect to the number of iterations for TV, TNV and L_∞-VTV obtained with 36 projections. In this subsection, we use the experimental sample shown in Fig. 2 scanned with 8 s of the total integration time per projection. The weighting parameter λ for each reconstruction defined in Eq. (5) is taken from Fig. 4.

Fig. 1 shows the sum of primal and dual residuals for reconstruction algorithms that decreases as the number of iterations increases for each algorithm, as expected. The convergence of each algorithm is therefore ensured. L_∞-VTV

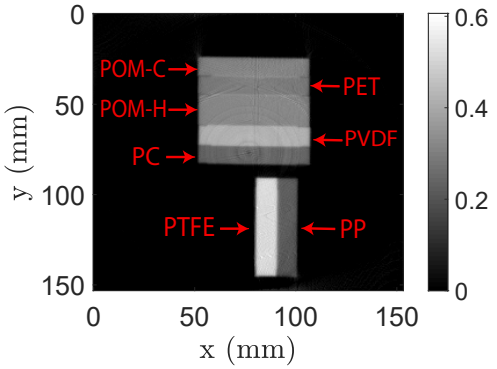


Figure 2: A display of the sample including the seven plastics indicated by red arrows. This is reconstructed with SIRT using 360 projections at 42 keV. The color bar shows linear attenuation coefficients (LACs) with the unit of cm^{-1}.

Table 1
The number of iterations fixed for each algorithm and projection number.

Projection number	SIRT	TV	TNV	L_∞-VTV
7	80	550	750	550
12	80	550	600	500
36	100	400	500	400
360	150	400	400	400

converges noticeably faster than TNV and TV whereas TV shows a slowest convergence rate.

We estimate the reconstructed image quality through the correlation coefficient, r, which measures how the reconstructed image $\mathbf{u}$ is linearly connected to the ground-truth (true) image $\mathbf{v}$, and it is expressed as

$$r = \frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})^2 \sum_i (v_i - \bar{v})^2}} \tag{25}$$

where $\bar{u}$ and $\bar{v}$ are the mean values, and u_i and v_i are the i^{th} pixel values of $\mathbf{u}$ and $\mathbf{v}$, respectively.

We use correlation coefficients for stopping criteria to define whether or not the algorithms will converge to the ground-truth image. In this work, the ground-truth images are created by measuring the mean LAC based on the normal distribution [14] through SIRT reconstruction and 360 projections. Fig. 3 presents how the correlation coefficients change with the number of iterations for each algorithm, and for 7 and 36 projections. We stop SIRT reconstruction algorithm when the correlation coefficient reaches the maximal value. For other algorithms, we truncate the iteration numbers at the points when the correlation coefficient and its gradient remain stable. Table 1 shows the fixed number of iterations for each reconstruction method and projection number obtained with the stopping criteria.

3.2. Investigating the reconstruction quality as a function of the weighting parameter λ

We performed the reconstructions for a set of weighting parameters λ for each method. A too small value of λ may result in an over-regularized image with blurred edges, while a too large value may give insufficient regularization effects, as notable in Eq. (5). We then computed the correlation coefficients with reference to the ground-truth images for the different number of projections, which are visible in Fig. 4. For 7 and 12 projections, the maximal correlation coefficient values for TV and SIRT reconstructions decrease noticeably more compared to the spectral reconstructions, TNV and L_∞-VTV . The maximal correlation coefficients for L_∞-VTV reconstruction are clearly larger than the coefficients for TNV and TV reconstructions, while SIRT reconstruction has the lowest correlation coefficients. Fig. 5 compares the TV, the TNV and the L_∞-VTV reconstructions corresponding to the maximal correlation coefficients presented in Fig. 4 for 7, 12 and 36 projections. The TNV and the L_∞-VTV algorithms appear to have clearly less structural artifacts than the TV reconstruction that suffers from over-smoothing and more distortion in shape, which are more visible for 7 and 12 projections. L_∞-VTV, in turn, has more accurate structure with better edges compared to TNV.

3.3. Robustness to different noise levels in all the energy bins

MultiX ME100 photon counting X-ray detector can have integration times from 0.5 ms to 100 ms (in 10 μs increments). In this subsection, we scanned the sample shown in Fig. 6 with total integration times of 2 ms, 10 ms and 50 ms per projection using 36 projections for testing robustness of each algorithm to such decreasing photons statistics of each energy bin. Fig. 7 shows the correlation coefficients evaluating reconstruction qualities for each method. The maximal correlation coefficients for TNV and L_∞-VTV reconstructions are larger than the coefficient for TV reconstruction, which are more noticeable in 2 ms and 10 ms cases. As an integration time decreases, λ values corresponding to maximal correlation coefficients also decrease for each algorithm, in other words the higher noise in scanning the more regularization is needed, as expected. Fig. 8 compares the reconstructions of TV, TNV and L_∞-VTV corresponding to such maximal correlation coefficients. 2 ms integration time shows that the TV reconstruction suffers noticeably from blurring and over-smoothing in a such low integration time whereas the TNV and the L_∞-VTV reconstructions appear to be clearer and better at preserving the edges. The images produced by the L_∞-VTV, in turn, have noticeably sharper edges and less structural artifacts compared to TNV, which are more visible in 2 ms and 10 ms levels. Thus, the joint reconstruction algorithms show more robustness to such higher noise levels compared to the TV. Robustness to lower integration time may be important for industrial applications, which can enable faster material classification through faster data acquisition.

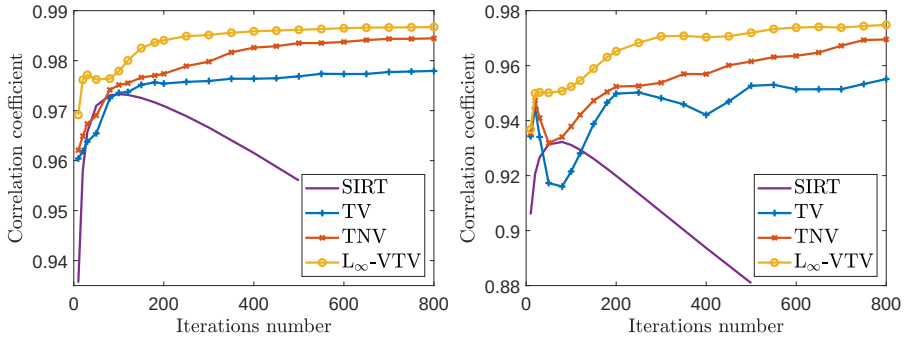


Figure 3: Correlation coefficients as a function of iteration number for 36 (left) and 7 projections (right). The weighting parameters λ used for the reconstructions are obtained based on Fig. 4. The correlation coefficients represent the mean values calculated from the coefficients for 15 energy bins used.

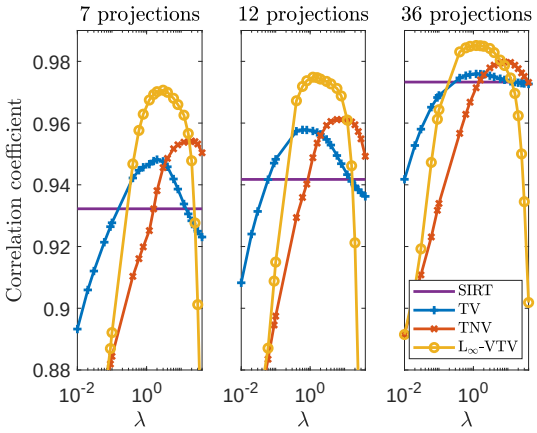


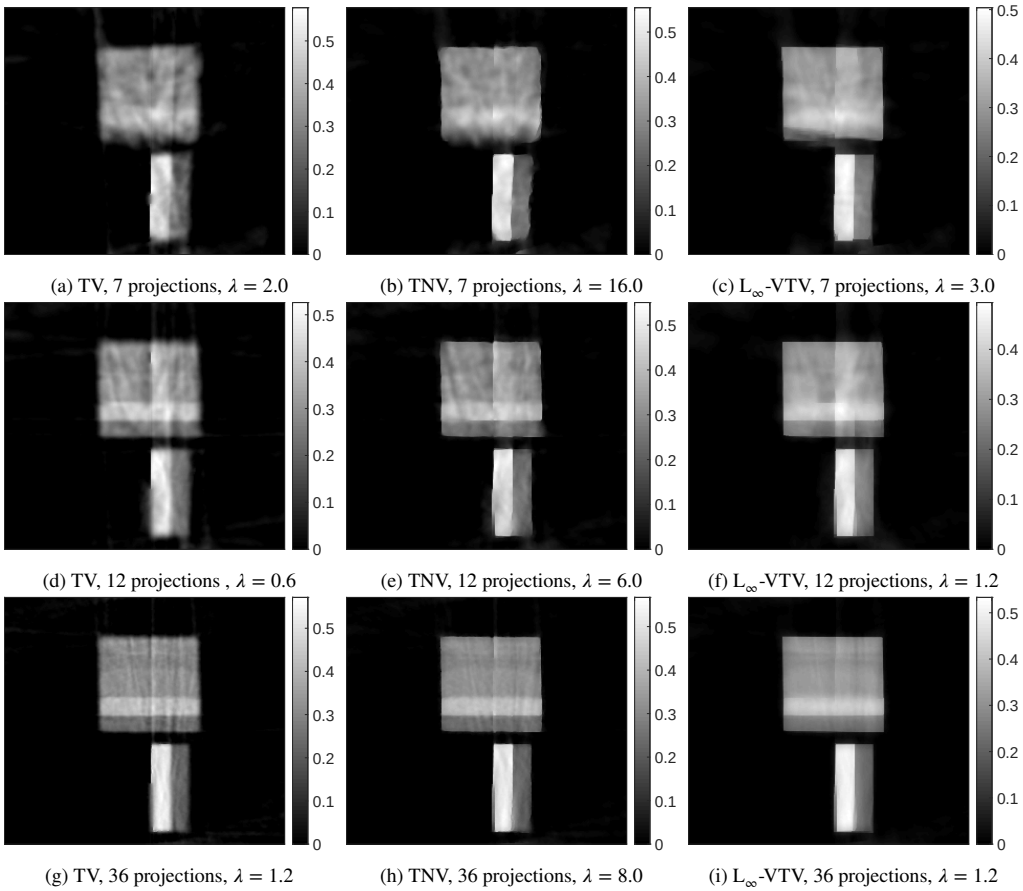
Figure 4: The correlation coefficients as a function of weighting parameter λ for 7, 12 and 36 projections. In this subsection, we use the plastics sample shown in Fig. 2 scanned with 8 s of the total integration time per projection. The correlation coefficients represent the mean values calculated from the coefficients for 15 energy bins used. Note the logarithmic scale in the x−axis.

3.4. Robustness to significantly higher noise levels in certain energy bins

Some spectral energy bins may suffer from sudden noise during scanning. Significantly higher noise in certain energy bins can be caused by the metal objects in the sample, which reduce the photon counts leading to photon starvation and create metal artifacts in the reconstructions. Metal artifacts influence primarily the lower energy bins and may be more severe resulting in less reliable characterization of the material properties. To estimate the robustness of the algorithms to such artifacts, we added Gaussian noise to the sinograms with a standard deviation (σ) of 0.5, 1.0 and 1.5. The noise is introduced to two bins out of 15 energy bins: 48.7 keV and 101.6 keV. We estimate the reconstruction qualities of the algorithms for different σ. Fig. 9 shows correlation coefficients depending on energies for each σ value. The mean λ values optimized for $\sigma = 1.5$ and calculated from the coefficients for 15 energy bins used are applied to all the noise levels for reconstructions with each algorithm. It can be noted from the graphs that for TV there are no interactions between energy bins as expected whereas TNV and L_∞-VTV show high correlations between channels and have noticeably smaller decreases in correlation coefficients for the affected bins. Thus, the joint reconstruction algorithms compensate for significantly higher noise in certain energy bins by effectively using the unaffected energy bins, for which coupling leads to slight degradation in the reconstruction quality. The figures also show that L_∞-VTV outperforms TNV in compensating such noises, and improves sharply the reconstruction quality in the lower energy bins, which can be appropriate property to reduce the metal artifacts.

3.5. The L-curve

The L-curve is a plot of the norm of a regularized solution as a function of the norm of the respective residual vector. This tool can be used for showing the trade-off between the weight of a regularized solution and its fit to the tomographic data, as the weighting parameter λ changes. Differently from correlation coefficients, using the L-curve one can rely only on the experimental sample to establish the optimal λ without the ground truth image. For TV, TNV and L_∞-VTV, the weights of the regularized solutions are measured by the 2-norm of the image gradients ($\| \mathbf{u}_\lambda \|_{\mathrm{TV}}$, Eq. (2)), the nuclear norm of image gradients ($\| \mathbf{u}_\lambda \|_{\mathrm{TNV}}$, Eq. (3)) and the 2-norm of the maximum of gradients ($\| \mathbf{u}_\lambda \|_{\mathrm{L}_\infty-\mathrm{VTV}}$, Eq. (4)), respectively. The fit is defined by the 2-norm $\| \mathbf{A}\mathbf{u}_\lambda - \mathbf{p} \|_2$ of the residual vector (Eq. (1)) for each method. Thus, one may use L-curve for selecting an optimal weighting parameter for the given data. L-curve may be represented on a log-log scale, such as for Tikhonov regularization [22]. In our work, we employ a linear-linear representation of the L-curve as in [23], which fits better to our reconstruction methods. Fig. 10 shows the correlation coefficient and the L-curve for each algorithm. Theoreti-



(a) TV, 7 projections, $\lambda = 2.0$ (b) TNV, 7 projections, $\lambda = 16.0$ (c) L_∞-VTV, 7 projections, $\lambda = 3.0$

(d) TV, 12 projections , $\lambda = 0.6$ (e) TNV, 12 projections, $\lambda = 6.0$ (f) L_∞-VTV, 12 projections, $\lambda = 1.2$

(g) TV, 36 projections, $\lambda = 1.2$ (h) TNV, 36 projections, $\lambda = 8.0$ (i) L_∞-VTV, 36 projections, $\lambda = 1.2$

Figure 5: TV, TNV and L_∞-VTV reconstructions of the plastics sample (Fig. 2) for 7, 12 and 36 projections at 42 keV. The weighting parameters λ used correspond to maximal correlation coefficients shown in Fig. 4.

cally, the appropriate λ values lie on the corner of the curve and the values on the flat and vertical parts lead to the solutions (reconstructions) dominated by over-regularization and under-smoothing (the data errors), respectively. Note that we use the real experimental data and sparse-view projections that may lead to some limitations of the L-curve for the methods. Hanke [24] showed that the smoother the reconstruction the lower accuracy of the estimated λ. Such limitation is more visible for TV for which the L-curve criterion can produce a weighting parameter that leads to slight over-regularization. The reason for this may be due to that TV tends to be over-regularized for sparse-view and high noise cases as discussed in the subsections 3.2, 3.3. The λ values corresponding to maximum correlation coefficients for TNV and L_∞-VTV are in good correspondence with the λ values located on the corner of the L-curve, as shown in Fig. 10. L_∞-VTV, in turn, demonstrates clearly accurate estimation of the optimal λ based on the L-curve criterion.

3.6. Classification performance as a function of λ

Linear attenuation coefficient (LAC) is proportional to a material's electron density [26]. The electron density is the number of electrons per unit volume ($\mathrm{electron - mole/cm^3}$). For a compound or mixture that has the composition of N total different elements, the electron density can be calculated as

$$\rho_e = \frac{\sum_{i=1}^{N} \alpha_i Z_i}{\sum_{i=1}^{N} \alpha_i A_i} \rho. \tag{26}$$

where ρ is mass density ($\mathrm{g/cm^3}$), A_i and Z_i are atomic mass and atomic number for each element, i, respectively, α_i is the number of atoms that have atomic number Z_i. For compounds, the atomic number is referred to as effective atomic number, Z_{eff} and a now classical parameterization proposed by [27, 28] is defined as

$$Z_{\mathrm{eff}} = \sqrt[l]{\sum_{i=1}^{N} r_i Z_i^l}, \tag{27}$$

Table 2
The list of the materials scanned and processed for the calibration step [14], and their physical properties ρ_e and Z_{eff}. The mass density, ρ for the plastic materials was measured with uncertainties of $\pm$ 0.15%. The mass densities for the rest of materials represent the theoretical values found in PubChem data, [25].

Material	Chemical Formula	Width×length/ Diameter (mm)	ρ (g/cm^3)	ρ_e (e$^-$mol/cm^3)	Z_{eff}
Graphite	C	12.7	1.8	0.899	6
PC	$(CO_3C_{13}H_8)_n$	8.2×53.5	1.18	0.610	6.82
PMMA	$(C_5O_2H_8)_n$	40×42	1.18	0.636	7.02
POM-C	$(CH_2O)_n$	9×53.5	1.41	0.753	7.40
PTFE	$(C_2F_4)_n$	9×53.3	2.16	1.035	8.70
N,N-Dimethylhydrazine	$C_2H_8N_2$	67	0.791	0.447	6.44
Ethylenediamine	$C_2H_8N_2$	67	0.90	0.509	6.44
Acetone 2	C_3H_6O	54	0.785	0.432	6.90
Nitrobenzene	$C_6H_5NO_2$	49	1.20	0.624	7.00
Ethanol 96%	C_2H_6O (96%)	67×67	0.798	0.450	7.06
Methanol	CH_3OH	20	0.792	0.446	7.29
Hydrazine solution	H_4N_2 (35%)	54	1.0	0.561	7.43
Nitromethane	CH_3NO_2	20	1.14	0.597	7.50
Water	H_2O	20	0.997	0.554	7.78
Water 3	H_2O	12.7	0.997	0.554	7.78
Hyd. Peroxide 2	H_2O_2 (50%)	73×74	1.22	0.661	7.83
Magnesium 2	Mg	18	1.74	0.859	12
Aluminum 2	Al	25	2.70	1.3	13
Aluminum 3	Al	20×20	2.70	1.3	13
Silicon	Si	25	2.33	1.161	14

Table 3
The list of all the materials scanned and processed for the (ρ_e, Z_{eff}) calculation step [14], and their physical properties ρ_e and Z_{eff}. The mass density, ρ for the plastic materials was measured with uncertainties of $\pm$ 0.15%. The mass densities for the rest of materials represent the theoretical values found in PubChem data, [25].

Material	Chemical Formula	Width×length/ Diameter (mm)	ρ (g/cm^3)	ρ_e (e$^-$mol/cm^3)	Z_{eff}
PET	$(C_{10}H_8O_4)_n$	9×53.5	1.39	0.721	7.09
POM-H	$(CH_2O)_n$	15.5×53.3	1.43	0.763	7.40
PVDF	$(C_2H_2F_2)_n$	9×53.5	1.79	0.896	8.40
PTFE 2	$(C_2F_4)_n$	12.7	2.2	1.056	8.70
2-Butanone	C_4H_8O	83	0.805	0.447	6.76
Acetone	C_3H_6O	20	0.785	0.432	6.90
Methanol 2	CH_3OH	81	0.792	0.446	7.29
Ethanol 40%	C_2H_6O (40%)	67×67	0.947	0.532	7.63
Water 2	H_2O	51×51	0.997	0.554	7.78
Nitric acid	HNO_3 (65%)	83	1.39	0.714	7.80
Hyd. Peroxide	H_2O_2 (50 %)	20	1.22	0.661	7.83
Magnesium	Mg	12.7	1.74	0.859	12
Aluminum	Al	25	2.70	1.3	13
Silicon powder	Si	48	0.65	0.324	14
Silicon 2	Si	12.7	2.33	1.161	14

where r_i is "relative electron fraction" contribution of an element, i, which is determined as

$$r_i = \frac{\alpha_i Z_i}{\sum_{j=1}^{N} \alpha_j Z_j},$$

In previous work [14], the exponent l was investigated to tune this value on the best match for the materials, source spectrum and system features. The optimal value was l = 8.0 that showed the lowest percent relative deviations in classification performance, therefore we choose this value for the calculations of reference Z_{eff} values in this work. Based on [15], one can define a linear attenuation coefficient (LAC) of

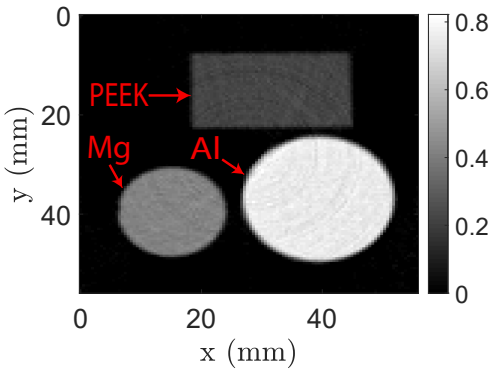


Figure 6: A display of sample consisting of aluminum, magnesium and PEEK plastic indicated by red arrows, reconstructed with SIRT, 360 projections and 100 ms of total integration time per projection at 61.9 keV. The color bar shows linear attenuation coefficients (LACs) with the unit of cm^{-1}.

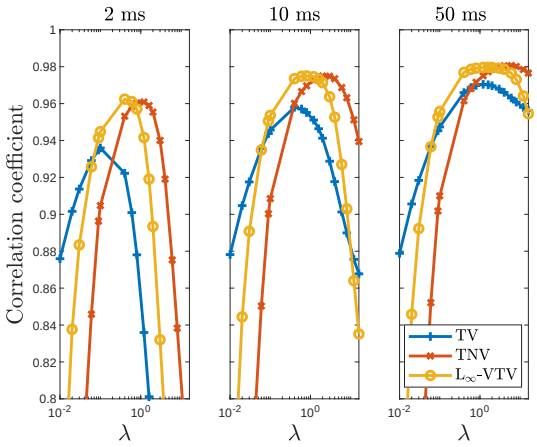


Figure 7: Correlation coefficient as a function of weighting parameters λ for the sample (Fig. 6) obtained with total integration times of 2 ms, 10 ms and 50 ms per projection for each method. The number of projections is 36. The correlation coefficients represent the mean values calculated from the coefficients for 15 energy bins used. Note the logarithmic scale in the x−axis.

a material, m, for each energy bin as

$$\mu_m(E_k) = \rho_{e,m}\left(Z_{\mathrm{eff},m}^{n-1}p(E_k) + c(E_k)\right), \tag{28}$$

where $m = 1, 2, ..., M$, $k = 1, 2, ..., K$ with M being the total number of materials and K being the number of energy bins used. Note that we use 15 energy bins approach for this study based on previous work [14]. $p(E_k)$ and $c(E_k)$ are photoelectric absorption and Compton scattering basis functions for energy bin E_k, respectively. The basis functions and parameter n being same for all materials are calibrated by using the reference ρ_e and Z_{eff} values and measured LACs of the scanned materials listed in Table 2 for the best classification performance. After the calibration step, the classification performance is estimated calculating (ρ_e, Z_{eff}) values of the additionally scanned materials listed in Table 3 by using the calibrated values and the measured LACs. We refer the reader to our previous work [14] for the derivation of calibration of basis functions and n and (ρ_e, Z_{eff}) calculation steps. The reference ρ_e and Z_{eff} values of the materials listed in both tables were calculated by using Eq. (26) and Eq. (27) and the exponent $l = 8.0$, respectively. Sample dimensions are also given in the tables through width×length or diameter for rectangular and circular samples, respectively. The classification of plastics is an efficient gauge for a system's performance to detect explosive materials in luggage. A number of explosives have similar chemical compositions, e.g. POM has been regarded as an explosive simulant [29]. The materials used in this study reflect the same Z_{eff} variation as different organic tissues [30] and none possess K-edge absorption within the detector's energy range.

The accuracy of material classification was measured as the percent relative deviation from the reference values for ρ_e and Z_{eff} as:

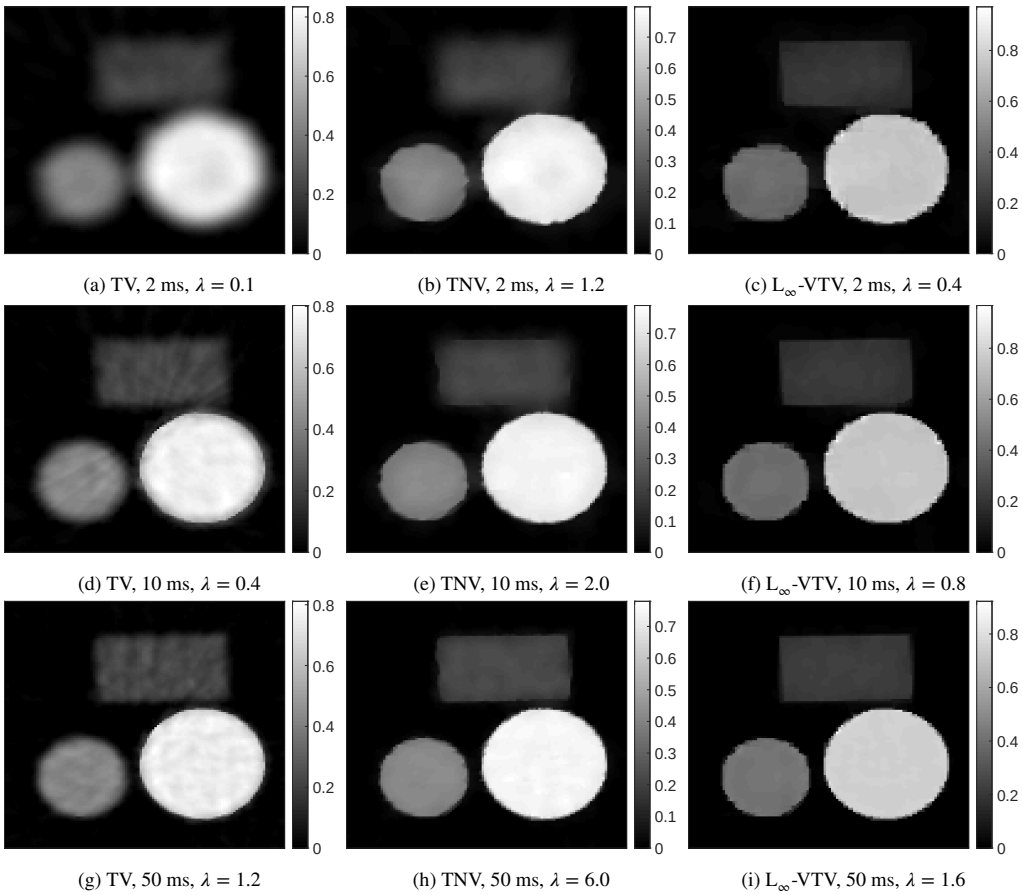
$$\Delta Z_{\mathrm{eff}}^{\mathrm{rel}} = 100\% \cdot \frac{Z_{\mathrm{eff}}^{\mathrm{est}} - Z_{\mathrm{eff}}^{\mathrm{ref}}}{Z_{\mathrm{eff}}^{\mathrm{ref}}}, \tag{29}$$

$$\Delta \rho_e^{\mathrm{rel}} = 100\% \cdot \frac{\rho_e^{\mathrm{est}} - \rho_e^{\mathrm{ref}}}{\rho_e^{\mathrm{ref}}} \tag{30}$$

where superscripts est and ref refer to the estimated and reference values, respectively.

Fig. 11 highlights the percent relative deviations for (ρ_e, Z_{eff}) depending on weighting parameters λ for sparse-view and noisy projections. For 12 projections, the algorithms have almost identical relative deviations. For 7 projections, TV gives the minimal overall deviations of 3.8% for ρ_e, and 4.2% for Z_{eff} (at $\lambda = 16.0$), whereas TNV and L_∞-VTV yield deviations of 3.4% and 3.1% (at $\lambda = 150.0$), and 3.5% and 2.4% (at $\lambda = 30.0$), respectively. In 7 projections case, Gaussian noise is introduced to the experimental data after the spectral correction [17] with a standard deviation of $\sigma = 0.05$, which may be realistic for some scanning applications due to characteristics of X-ray sources. The noise is added to all of the materials used for the calibration and calculation, and to all of the 15 energy bins. As shown in the figure, in the noisy case of 7 projections the deviations for TV reconstructions increased to 5.0% for both ρ_e and Z_{eff}, whereas TNV and L_∞-VTV have the increases to 4.0% and 3.9%, and 3.6% and 3.2% at the same corresponding λ values, respectively. Thus, spectral reconstruction algorithms show better classification performance in such sparse-view and noisy cases, and L_∞-VTV, in turn, outperforms TNV. It can be noted from the figure that in the noisy case of 7 projections L_∞-VTV uses the lower λ value that gives the lowest deviations compared to the case without the noise, i.e. the algorithm uses stronger regularization term to overcome the noise, as expected.

Fig. 12 shows the relative deviations for the different materials in 7 projections case. Compared to the other al-



(a) TV, 2 ms, $\lambda = 0.1$ (b) TNV, 2 ms, $\lambda = 1.2$ (c) L_∞-VTV, 2 ms, $\lambda = 0.4$

(d) TV, 10 ms, $\lambda = 0.4$ (e) TNV, 10 ms, $\lambda = 2.0$ (f) L_∞-VTV, 10 ms, $\lambda = 0.8$

(g) TV, 50 ms, $\lambda = 1.2$ (h) TNV, 50 ms, $\lambda = 6.0$ (i) L_∞-VTV, 50 ms, $\lambda = 1.6$

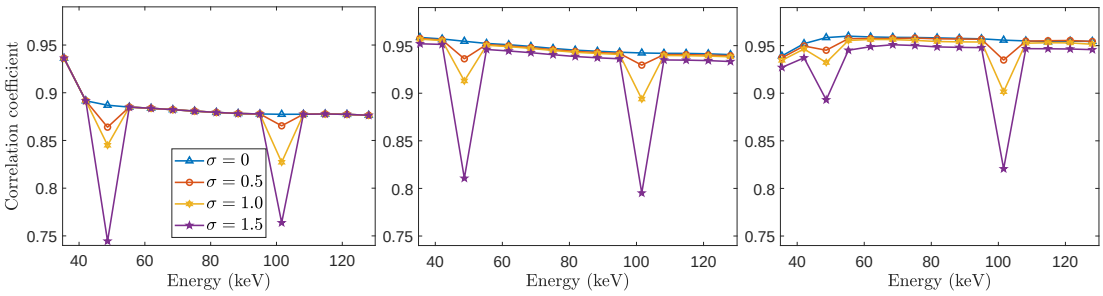
Figure 8: TV, TNV and L_∞-VTV reconstructions of the sample (Fig. 6) obtained with total integration times of 2 ms, 10 ms and 50 ms per projection and with 36 projections at 61.9 keV. The weighting parameters λ used for each method correspond to maximal correlation coefficients presented in Fig. 7.

gorithms, L_∞-VTV gives the lower relative deviations for POM-H that can be an explosive simulant. Fig. 13 shows the classification performance of L_∞-VTV for different materials depending on weighting parameter λ for 7 projections. It can be noted from the figure that the λ values between 20.0 and 40.0 for L_∞-VTV can provide the optimal overall performance for material classifications with this Z_{eff} range of materials used.

We found in our implementation that for material classification L_∞-VTV involves 10.6% and 52.3% shorter computation time compared to TNV that is a state-of-the-art method for joint reconstruction and TV, respectively.

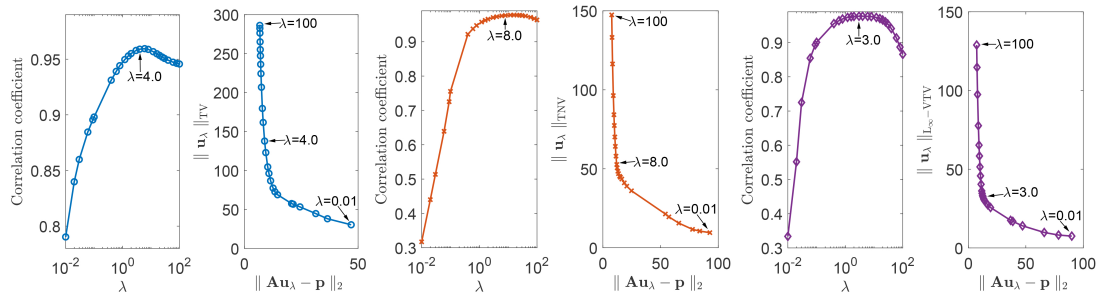
4. Conclusion

In this paper, we presented the joint reconstruction of vectorial total variation defined by the L infinity norm (L_∞-VTV) for material classification using the SIMCAD method from sparse spectral X-ray CT. We compared the obtained results with a state-of-the-art joint reconstruction of total nuclear variation (TNV) and channel-by-channel reconstructions of TV and SIRT. The optimization of L_∞-VTV algorithm has been displayed. We showed how we truncate the iteration number based on the correlation coefficients for each algorithm. When image reconstructions are performed from sparse-view projections and with low levels of photon statistics due to decreased integration times of the detector, the joint reconstruction algorithms showed clearly better performance in image reconstruction compared channel-by-channel reconstructions. L_∞-VTV, in turn, outperformed TNV. In study of robustness to significantly higher noise levels in certain energy bins, we found that the joint reconstruction algorithms effectively distribute such noises in affected bins to unaffected bins and achieve significantly higher reconstruction performance for all the energy bins compared to TV. L_∞-VTV, in turn, was found to be better in compensating for such noise levels than TNV, especially at lower energies. This property can be very useful for handling metal



(a) TV, $\lambda = 0.06$ (b) TNV, $\lambda = 0.8$ (c) L_∞-VTV, $\lambda = 0.4$

Figure 9: Correlation coefficient as a function of photon energy when different levels of Gaussian noise with standard deviation, σ, are added to certain energies of 48.7 keV and 101.6 keV for each method. The mean λ values optimized for $\sigma = 1.5$ case are used to produce correlation coefficients for all the noise levels for each method. In this subsection, the sample shown in Fig. 6 was scanned with 12 projections and 100 ms of the total integration time per projection.



(a) TV (b) TNV (c) L_∞-VTV

Figure 10: The correlation coefficients depending on λ and L-curve for each method. In this subsection, the sample shown in Fig. 6 was scanned with 12 projections and 100 ms of the total integration time per projection. Note the logarithmic scale in the x−axis for correlation coefficients.

artifacts, which affect mainly the low-energy bins. For each algorithm, we studied the practical use of L-curve method for finding the optimal weighting parameter based on the scanned sample itself while computing a correlation coefficient requires a ground truth image. For L_∞-VTV, the optimal weighting parameter found based on correlation coefficient was accurately fitted with the weighting parameter that is considered as the optimal value for L-curve located on its corner. The joint reconstruction methods showed noticeably better classification performance compared to channel-by-channel reconstruction methods for such sparse-view projections, also in the case when noise is added to all the materials used. L_∞-VTV, in turn, is more accurate in (ρ_e, Z_{eff}) estimation compared to TNV. We found that L_∞-VTV requires significantly less evaluation time for material classification compared to TNV and TV.

CRediT authorship contribution statement

Doniyor Jumanazarov: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Jakeoung Koo:** Methodology, Software, Validation, Formal analysis, Writing - original draft. **Henning F. Poulsen:** Writing - review and editing, Supervision. **Ulrik L. Olsen:** Validation, Resources, Data curation, Writing - review and editing, Supervision . **Mihai IOVEA:** Conceptualization, Validation, Writing - review and editing, Supervision, Funding acquisition.

Acknowledgements

This project has received funding from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 765604 as part of the MUltiscale, Multimodal and Multidimensional imaging for EngineeRING project (MUMMERING Innovative Training Network, www.mummering.eu).

The authors want to acknowledge also the 3D Imaging Center at DTU, where the experiments have been conducted.

References

[1] L. I. Rudin, S. Osher, E. Fatemi, 1-s2.0-016727899290242F-main.pdf, Physica D 60 (1992) 259–268.

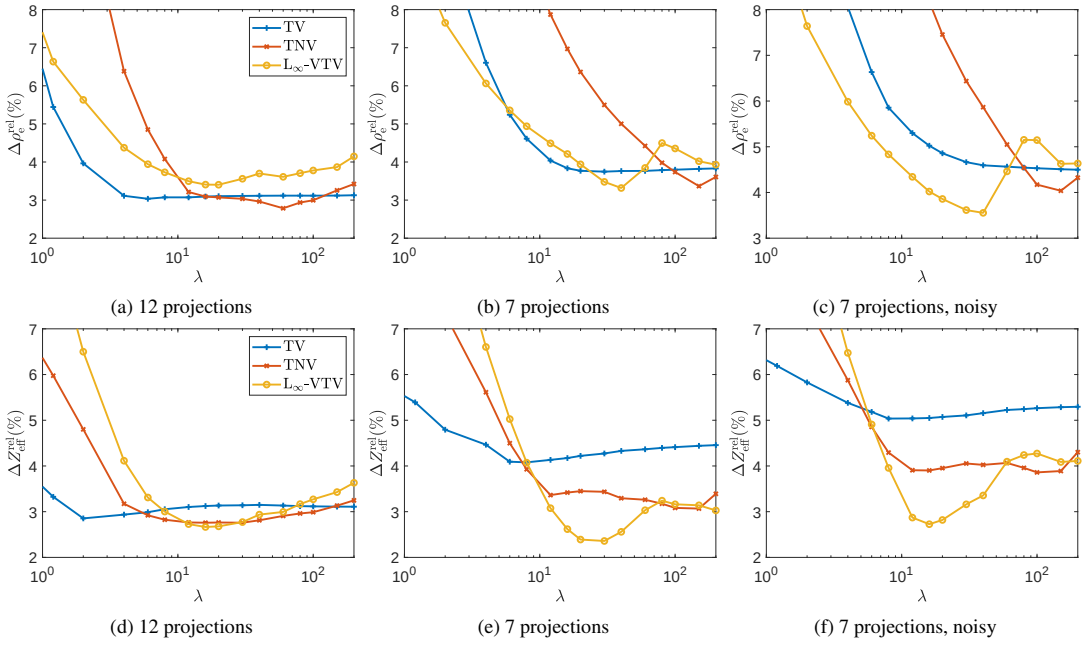


Figure 11: The dependence of the percent relative deviations for (ρ_e, Z_{eff}) on weighting parameters λ for 12 projections, 7 projections and 7 projections in the case of added Gaussian noise. The first and second rows are respectively the percent relative deviations for ρ_e and Z_{eff}. The deviations represent the mean values calculated from absolute values of relative deviations for each material listed in Table 3. SIRT reconstruction algorithm that is not shown in the figure gives relatively higher deviations of 8.4% for ρ_e, and 2.4% for Z_{eff} for 12 projections, and 13.9% and 3.0% for 7 projections, respectively. The experimental samples used in this subsection are listed in Table 2 and Table 3, and scanned with 8 s of the total integration time per projection. Note the logarithmic scale in the x−axis.

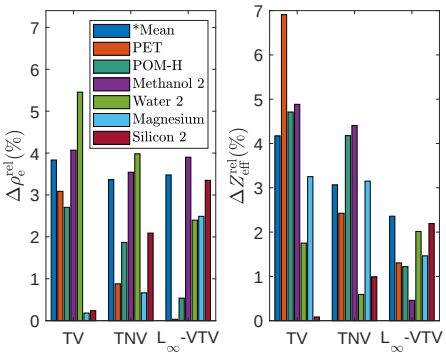


Figure 12: The percent relative deviations for ρ_e (left) and Z_{eff} (right) for the different materials in 7 projections case. The mean values are calculated from absolute values of relative deviations for each material listed in Table 3. The deviations representing the absolute values for each algorithm correspond to λ values that provide the lowest overall deviations presented in Fig. 11.

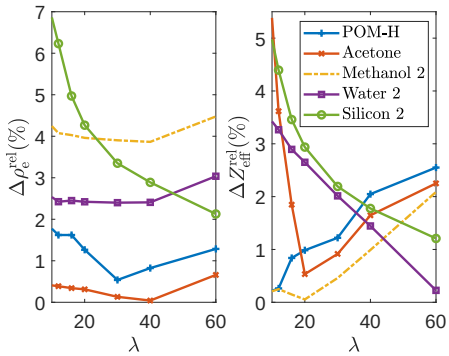


Figure 13: The dependence of percent relative deviations for ρ_e (left) and Z_{eff} (right) on weighting parameter λ for the different materials obtained with L_∞-VTV in 7 projections case. The deviations for each material represent the absolute values. The materials correspond to the Table 3 used for the calculation.

[2] J. Song, Q. H. Liu, G. A. Johnson, C. T. Badea, Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT, Medical Physics 34 (2007) 4476–4483.

[3] J. Bian, J. Wang, X. Han, E. Y. Sidky, L. Shao, X. Pan, Optimization-based image reconstruction from sparse-view data in offset-detector CBCT, Physics in Medicine and Biology 58 (2013) 205–230.

[4] E. Y. Sidky, C. M. Kao, X. Pan, Accurate image reconstruction from

few-views and limited-angle data in divergent-beam CT, Journal of X-Ray Science and Technology 14 (2006) 119–139.

[5] P. Blomgren, T. F. Chan, Color TV: Total variation methods for restoration of vector-valued images, IEEE Transactions on Image Processing 7 (1998) 304–309.

[6] T. Miyata, Y. Sakai, Vectorized total variation defined by weighted L infinity norm for utilizing inter channel dependency, Proceedings - International Conference on Image Processing, ICIP (2012) 3057–3060.

[7] K. M. Holt, Total Nuclear Variation and Jacobian Extensions of Total Variation for Vector Fields, IEEE Transactions on Image Processing 23 (2014) 3975–3989.

[8] B. Recht, M. Fazel, P. A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, 45th Annual Allerton Conference on Communication, Control, and Computing 2007 1 (2007) 42–48.

[9] J. Duran, M. Moeller, C. Sbert, D. Cremers, On the Implementation of Collaborative TV Regularization: Application to Cartoon+Texture Decomposition, Image Processing On Line 5 (2016) 27–74.

[10] D. S. Rigie, P. J. La Rivière, Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization, Physics in Medicine and Biology 60 (2015) 1741–1762.

[11] D. S. Rigie, A. A. Sanchez, P. J. La Riviére, Assessment of vectorial total variation penalties on realistic dual-energy CT data, Physics in Medicine and Biology 62 (2017) 3284–3298.

[12] Z. Zhong, W. J. Palenstijn, J. Adler, K. J. Batenburg, EDS tomographic reconstruction regularized by total nuclear variation joined with HAADF-STEM tomography, Ultramicroscopy 191 (2018) 34–43.

[13] J. Gregor, T. Benson, Computational analysis and improvement of SIRT, IEEE Transactions on Medical Imaging 27 (2008) 918–924.

[14] D. Jumanazarov, J. Koo, M. Busi, H. F. Poulsen, U. L. Olsen, M. Iovea, System-independent material classification through X-ray attenuation decomposition from spectral X-ray CT, NDT and E International 116 (2020) 102336.

[15] R. E. Alvarez, A. Macovski, Energy-selective reconstructions in X-ray computerised tomography, Physics in Medicine and Biology 21 (1976) 733–744.

[16] M. Kheirabadi, A. Bjorholm Dahl, U. Lund Olsen, W. Mustafa, M. Lyksborg, Multispectral x-ray CT: multivariate statistical analysis for efficient reconstruction, in: Developments in X-Ray Tomography XI, volume 1039113, International Society for Optics and Photonics, 2017, p. 38. doi:10.1117/12.2273338.

[17] E. S. Dreier, J. Kehres, M. Khalil, M. Busi, Y. Gu, R. Feidenhans, U. L. Olsen, Spectral correction algorithm for multispectral CdTe x-ray detectors, Optical Engineering 57 (2018) 16.

[18] L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. Nonlinear Phenom. 60 (1992).

[19] A. Chambolle, T. Pock, A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging, J. Math. Imaging Vis. 40 (2011).

[20] H. H. Bauschke, P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics, Springer International Publishing, Cham, 2017. doi:10.1007/978-3-319-48311-5.

[21] L. N. Trefethen, D. Bau, Numerical Linear Algebra, Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[22] P. C. Hansen, The L-Curve and its Use in the Numerical Treatment of Inverse Problems, in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering 4 (2000) 119–142.

[23] X. Yang, R. Hofmann, R. Dapp, T. van de Kamp, T. d. S. Rolo, X. Xiao, J. Moosmann, J. Kashef, R. Stotzka, TV-based conjugate gradient method and discrete L-curve for few-view CT reconstruction of X-ray in vivo data, Optics Express 23 (2015) 5368.

[24] M. Hanke, Limitations of the L-curve method in ill-posed problems, BIT Numerical Mathematics 36 (1996) 287–301.

[25] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, PubChem 2019 update: Improved access to chemical data, 2019. URL: https://pubchem.ncbi.nlm.nih.gov. doi:10.1093/nar/gky1033.

[26] A. H. Compton, S. K. Allison, et al., X-rays in theory and experiment, New York, NY, USA: Van Nostrand (1935).

[27] Mayneord, WV, The significance of the roentgen, Acta of the International Union Against Cancer 2 (1937) 271.

[28] F. W. SPIERS, Effective atomic number and energy absorption in tissues., The British journal of radiology 19 (1946) 52–63.

[29] J. Rinkel, G. Beldjoudi, V. Rebuffel, C. Boudou, P. Ouvrier-buffet, G. Gonon, L. Verger, A. Brambilla, Experimental Evaluation of Material Identification Methods With CdTe X-ray Spectrometric Detector, IEEE Transactions on Nuclear Science 58 (2011) 2371–2377.

[30] M. M. Goodsitt, E. G. Christodoulou, S. C. Larson, Accuracies of the synthesized monochromatic CT numbers and effective atomic numbers obtained with a rapid kVp switching dual energy CT scanner, Medical Physics 38 (2011) 2222–2232.