Particle Picking in Cryo-EM

Khursheed Ali
IIT-Bombay

Guided By

Prof. Ajit Rajwade

April 25, 2018

Contents

Chapter 1

Background

1.1 Tomographic Reconstruction

1.1.1 Introduction

Objective of tomographic reconstruction is to obtain 3D representation of the internal structure of 3d object. For example in computed tomography (CT) scan of human head (figure 1.1), in this we are able see inside the human head which is structure of brain.

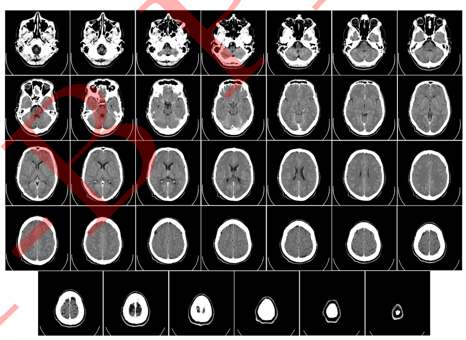

Figure 1.1: Computer tomography of human brain, slice from 3D volume from base of the skull to top. Source: wiki

But before for creating 3d internal structure of the object, their are in between many steps which has to be performed. Just to make it simple, we will take the example of CT machine, In CT machine, X-rays are used because they have ability to penetrate the most of the objects. Using X-rays, many 2D projections are taken for the 3D object (in case of 2D object, 1D projection are taken) at different angles. These projections are also called as *tomographic projections*. These tomographic projection are defined as *Radon transform*. Then using all these 2D tomographic projections 3D object internal structure is created.

1.1.2 Tomographic Projection

For simple understanding in this chapter, we will take 2D object and it's 1D projection (figure 1.2). Because it is much easier to understand. But this concept can easily be extended for 3D objects and it's 2D projections. For taking projection, a parallel beam of X-rays are fired on the object at some fixed angle θ_k. Degree of absorption of X-rays by the object are the recorded on to detector which is called as projection at θ_k. Now will formulate mathematical equation for the projection. Let 2D object be defined as $f(x, y)$. Then Projection at angle θ_k is given by

$$R(f) = g(\rho_j, \theta_k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x \cos\theta_k + y \sin\theta_k - \rho_j)dxdy \tag{1.1}$$

$R(f)$ is called as *Radon transform* of function f. $g(\rho_j, \theta_k)$ is read as projection of $f(x, y)$ at angle θ_k and ρ_j distance from origin.

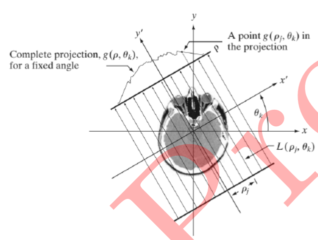

Figure 1.2: Parallel Beam (X-ray) Projection.
Source:Rafael C. Gonzalez , Richard E. Woods, Digital Image Processing

1.1.3 Fourier Slice Theorem

Fourier Slice Theorem gives us relationship between Fourier transform and Radon transform. Radon transform of $f(x, y)$ is by eq. (1.1). The 1D Fourier transform $g(\rho, \theta)$ w.r.t ρ with fixed θ is given by eq (1.2).

$$G(\mu, \theta) = \int_{-\infty}^{\infty} g(\rho, \theta)e^{-j2\pi\mu\rho}d\rho \tag{1.2}$$

Now expanding the $g(\rho_j, \theta_k)$ and simplifying the equation

$$G(\mu, \theta) = \int_{-\infty}^{\infty} g(\rho, \theta)e^{-j2\pi\mu\rho}d\rho \tag{1.3}$$

$$G(\mu, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x \cos\theta + y \sin\theta - \rho)e^{-j2\pi\mu\rho}dxdyd\rho \tag{1.4}$$

$$G(\mu,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\Big[\int_{-\infty}^{\infty}\delta(x\cos\theta + y\sin\theta - \rho)e^{-j2\pi\mu\rho}d\rho\Big]dxdy \quad (1.5)$$

$$G(\mu,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi\mu(x\cos\theta + y\sin\theta)}dxdy \quad (1.6)$$

$$G(\mu,\theta) = \big[F(u,v)\big]_{u=\mu\cos\theta, v=\mu\sin\theta} \quad (1.7)$$

Here, $F(u,v)$ represent the 2D discrete Fourier transform of function $f(x,y)$. By eq (1.7) we can say that, Fourier transform of projection of 2D object along some direction θ i.e $G(\mu,\theta)$ is equal to a slice of 2D Fourier transform of object along same direction θ in frequency plane, passing through origin. This is called as *Fourier Slice Theorem* or *Projection Theorem*. Similar theorem is also their for 3D object

1.1.4 Filtered Back Projection (FBP)

From section (1.1.2) and (1.1.3) we know how projections (tomographic projections) are taken and it's relation with Fourier transform of object. By exploiting these facts, we to reconstruct the internal structure of the object. Let suppose object is 2D denoted as $f(x,y)$ and it's 2D F.T (Fourier transform) is represented as $F(u,v)$.

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j2\pi(xu+yv))}dudv \quad (1.8)$$

Let u=$\mu\cos\theta$ v=$\mu\sin\theta$,

$$f(x,y) = \int_{0}^{2\pi}\int_{0}^{\infty} F(\mu\cos\theta, \mu\sin\theta)e^{j2\pi(x\mu\cos\theta + y\mu\sin\theta))}\mu d\mu d\theta \quad (1.9)$$

By using Fourier slice theorem eq (1.7) we get,

$$f(x,y) = \int_{0}^{2\pi}\int_{0}^{\infty} G(\mu,\theta)e^{j2\pi(x\mu\cos\theta + y\mu\sin\theta))}\mu d\mu d\theta \quad (1.10)$$

After simplification we get,

$$f(x,y) = \int_{0}^{\pi}\int_{0}^{\infty} G(\mu,\theta)e^{j2\pi(x\cos\theta + y\sin\theta))}|\mu| d\mu d\theta \quad (1.11)$$

So eq (1.11) says that, if we are having lot many projection at different angle then we will able to recover $f(x,y)$ i.e object internal structure from them. But one problem is their i.e in eq (1.11) $|\mu|$ (ramp filter) is a *unbounded* function, therefore it is not integrable. So to solve this problem we uses Ram-Lak filter

$(rect(\mu D))$ or Ram-Lak hamming filter. Updated FBP equation is given by eq (1.12) using Ram-Lak filter

$$f(x, y) = \int_0^{\pi} \int_0^{\infty} |\mu| rect(\mu D) G(\mu, \theta) e^{j2\pi(x \cos\theta + y \sin\theta))} d\mu d\theta \tag{1.12}$$

1.1.5 Reconstruction using Compressed Sensing

In this section, we will see another method for reconstruction using the set of projections taken at different angle. This method uses the concept of Compressed sensing (CS). CS works here because number of angles of projection are limited due to various reason like cost, energy and health considerations. This type of of problem is also called as *angle starved* problem. We know that images are sparse or compressible in standard basis like DCT. So CS based optimization equation for tomographic reconstruction is give by eq (1.13)

$$E(\boldsymbol{\beta}) = ||\boldsymbol{y} - \boldsymbol{RU\beta}||^2 + \lambda||\boldsymbol{\beta}||_1 \tag{1.13}$$

Here, $\boldsymbol{y}$ vector is created concatenating 1D projections as various angle. $\boldsymbol{U}$ is the basis matrix in which image/object f(x,y) is sparse or compressible. $\boldsymbol{\beta}$ is f(x,y) representation in $\boldsymbol{U}$ basis i.e $f(x, y)$ satisfy eq (1.14). $\boldsymbol{R}$ is the Radon matrix (it can also be Radon operator)

$$vec(f)) = \boldsymbol{U\beta} \tag{1.14}$$

We can solve eq (1.13) by using many optimization algorithm such as ISTA.

1.1.6 Application

Tomographic reconstruction has many application mostly is health care such as CT scan, industrial application such as fault detection in machine, observation of plant roots, remote sensing i.e observing underground objects or phenomena and most importantly study of protein, bacteria, ribosome, cells and virus by biologist for creating medicine or vaccination or for any other purpose.

1.2 Tomography under unknown angle

In previous section (1.1), we seen reconstruction from projections where angle at which each projections are known. But if only projections are known but not angles then previously seen method for reconstruction will not be applicable.
There are various application where angles are not known such as patient moving during CT scanning (this mostly happens if patient is baby), machine fault, moving insect tomography, Cryo-electron tomography. In all these case, we are only having projections but not angles. Under these scenarios, this are called as *tomography under unknown angle.*

In this section, we see how to reconstruct internal structure of object back from projection when angles are not know i.e only with the help of tomographic projections.

1.2.1 Moment Based Reconstruction

Again, for simple understanding we will take 2D object and it's 1D projection. But this logic can easily be extend for 3D object and it's 2D projection.
Let say f(x,y) be the object then moment of order (p,q) for f(x,y) is given by eq (1.15).

$$\boldsymbol{M}_{p,q} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)x^p y^q dxdy \tag{1.15}$$

Let $P_\theta(s)$ be the projection at angle θ then moment of order (n) is given by eq (1.16)

$$\boldsymbol{M}_\theta^{(n)} = \int_{-\infty}^{\infty} P_\theta(s)s^n ds \tag{1.16}$$

where,

$$P_\theta(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - s)dxdy \tag{1.17}$$

Therefore by substituting $P_\theta(s)$ in eq(1.16) by eq(1.17) and after simplification we get,

$$\boldsymbol{M}_\theta^{(n)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)(x\cos\theta + y\sin\theta)^n dxdy \tag{1.18}$$

So, the Helgasson Ludwig Consistency Conditions (HLCC) [2] gives us the relation between the image moment and its projection moment and which is given by eq(1.19).

$$\boldsymbol{M}_\theta^{(n)} = \Sigma_{l=0}^{n} f(x,y)\boldsymbol{C}_n^{n-l}(\cos\theta)^{n-l}(\sin\theta)^l \boldsymbol{M}_{n-l,l} \tag{1.19}$$

Here we know the projections of the f(x,y), so we also knows it moment i.e $\boldsymbol{M}_\theta^{(n)}$. By eq(1.19) we can find out unknown angles by using iteratively solving for angles [1]. Once we found the angles for these projection then we can apply any of the earlier mentioned method for reconstruction.

Advantage of this method are simple to implement. makes direct use of consistency conditions and work well for small number of angles. But major drawback of this method is that it is highly sensitive to noise.

1.2.2 Order Based Reconstruction

This method very much similar to machine learning based method. In this, all acquired projections are sorted i.e in order of increasing angles. Only issue is that this method required large number of projection and it assumes the all unknown projection angles are independently sampled from uniform distribution (it can be from any known distribution but then accordingly below mentioned algorithm will change). Assuming distribution to be uniform then this problem reduce to matching problem where we have to match each projection to one of the angles sampled evenly from unit circle.

It's a iterative algorithm where we choose one initial projection say P_0 and one angle. Without lose of generality we map that projection to with angle say θ_0. Then we find closest projection to P_0 and then that projection is mapped to next angle θ_1. This process is repeated till full mapping is completed.

Once we found the angles for all projection then we can apply any of the earlier mentioned method for reconstruction.

Chapter 2

Introduction to Cryo-EM

2.1 Motivation

2.2 Basic Pipeline

2.2.1 Particle Picking

2.2.2 CTF correction

2.2.3 Clustering

2.2.4 Angle Assignment

2.2.5 3D Reconstruction

2.3 Challenges

2.3.1 Noise

2.3.2 Unknown angles

2.3.3 Heterogeneity

2.3.4 Manual Particle Picking

Chapter 3

Experiments

3.1 Dataset

Bibliography

[1] Eeshan Malhotra and Ajit Rajwade. Tomographic reconstruction from projections with unknown view angles exploiting moment-based relationships. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 1759–1763. IEEE, 2016.

[2] Frank Natterer. *The mathematics of computerized tomography*. Siam, 2001.