Ultramicroscopy 157 (2015) 79–87

Contents lists available at ScienceDirect

Ultramicroscopy

journal homepage: www.elsevier.com/locate/ultramic

A fast iterative convolution weighting approach for gridding-based direct Fourier three-dimensional reconstruction with correction for the contrast transfer function

V. Abrishami [a], J.R. Bilbao-Castro [c], J. Vargas [a], R. Marabini [a,b], J.M. Carazo [a], C.O.S. Sorzano [a,d,*]

[a] Biocomputing Unit, Centro Nacional de Biotecnología-CSIC, C/Darwin 3, Cantoblanco, 28049 Madrid, Spain
[b] Escuela Politécnica Superior, Universidad Autónoma de Madrid, C/Francisco Tomás y Valiente, Cantoblanco, 28049 Madrid, Spain
[c] Unidad Asociada al CSIC, 'Supercomputación y Algoritmos', Universidad Almería & Málaga, Spain
[d] Bioengineering Lab, Escuela Politecnica Superior, University San Pablo CEU, Boadilla del Monte, 28668 Madrid, Spain

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form
18 May 2015
Accepted 24 May 2015
Available online 12 June 2015

Keywords:
Cryo-electron microscopy
Three-dimensional reconstruction
CTF correction
Single Particle Analysis

ABSTRACT

We describe a fast and accurate method for the reconstruction of macromolecular complexes from a set of projections. Direct Fourier inversion (in which the Fourier Slice Theorem plays a central role) is a solution for dealing with this inverse problem. Unfortunately, the set of projections provides a non-equidistantly sampled version of the macromolecule Fourier transform in the single particle field (and, therefore, a direct Fourier inversion) may not be an optimal solution. In this paper, we introduce a gridding-based direct Fourier method for the three-dimensional reconstruction approach that uses a weighting technique to compute a uniform sampled Fourier transform. Moreover, the contrast transfer function of the microscope, which is a limiting factor in pursuing a high resolution reconstruction, is corrected by the algorithm. Parallelization of this algorithm, both on threads and on multiple CPU's, makes the process of three-dimensional reconstruction even faster. The experimental results show that our proposed gridding-based direct Fourier reconstruction is slightly more accurate than similar existing methods and presents a lower computational complexity both in terms of time and memory, thereby allowing its use on larger volumes. The algorithm is fully implemented in the open-source Xmipp package and is downloadable from http://xmipp.cnb.csic.es.

1. Introduction

Single-Particle Analysis (SPA) is an Electron Microscopy (EM) method wherein the three-dimensional (3D) structure of a biological complex is determined from projections at random orientations of multiple instances of the specimen. Each projection is a two-dimensional (2D) projection of the 3D complex with a random spatial orientation that is additionally modulated by the Contrast Transfer Function (CTF) of the microscope. Upon determination of the orientation parameters, an inversion procedure yields a 3D volume that is compatible with the original projections. However, noisy imaging conditions, CTF effects, errors in orientation parameters, and a finite number of discrete projections not covering the whole spatial domain under study makes this inversion problem nontrivial [1]. Many approaches have been proposed to solve this ill-posed inversion, which can be categorized into three classes: algebraic, Weighted Back-Projection (WBP), and direct Fourier methods.

Algebraic methods treat this inversion problem as a system of linear equations where well-established algebra methods are employed to find the solution. In other words, the problem is formulated as $\bar{p} = W\bar{v}$, where the 3D object is decomposed into a finite set of basis functions whose coefficients are lexicographically stored in a vector $\bar{v}$, $\bar{p}$ is a vector with the values of all of the projections' pixels, and W encodes the weight of each of the basis functions onto each pixel. The Algebraic Reconstruction Technique (ART) [2] and Simultaneous Iterative Reconstruction Technique (SIRT) [3] are iterative approaches to solve this system of equations. The general idea behind these methods is to iteratively improve an initial volume by comparing each of the experimental projections with the projections from the current volume, thereby attempting to compensate for this difference. Although algebraic

---

[*] Corresponding author at: Biocomputing Unit, Centro Nacional de Biotecnología-CSIC, C/Darwin 3, Cantoblanco, 28049 Madrid, Spain. Fax: +34 91 585 4506.
E-mail address: coss@cnb.csic.es (C.O.S. Sorzano).

http://dx.doi.org/10.1016/j.ultramic.2015.05.018
0304-3991/© 2015 Elsevier B.V. All rights reserved.

methods have the potential to be applied to many different types of reconstructions [1] and to incorporate a variety of constraints [4], they suffer from a high computational complexity. In the SPA field, the introduction of blobs as spherically symmetric basis functions by [5] was one of the most efficient efforts towards a fast algebraic reconstruction, but it is still more computationally expensive compared to the two other aforementioned groups.

WBP [6,7] is the most intuitive method used to reconstruct a 3D object from its 2D projections based on the concept of the Fourier slice theorem. For a set of projections with known orientations, each projection is back-projected across the objective volume from its position in the projection space defined by its orientation. The superposition of all of these back-projections provides an estimation of the original 3D object. To account for the angular distribution of projections, these methods use a weighting function in Fourier space. Weighted back-projection methods are faster than algebraic methods, but they perform poorly in cases where large angular gaps exist, and they have been shown to underperform algebraic methods in a number of cases [5,8].

Based on the central slice theorem, direct Fourier reconstruction (DFR) methods try to obtain the 3D Fourier transform of an object directly from the 2D Fourier transform of its projections, so that an estimation of the original 3D object can be quickly obtained through an inverse 3D Fast Fourier Transform (FFT). In practice, the irregularity of the spatial distribution of the frequency of samples in the set of projections in experimental SPA studies makes the direct use of the inverse FFT unfeasible. Thus, an additional interpolation step is required to obtain the 3D Fourier transform of the object on a regular grid. The so-called gridding algorithm is an alternative method introduced by Penczek et al. [9] into the SPA field; this method was originally developed by Jackson et al. [10] to efficiently estimate the 3D Fourier transform in a regular grid of points using irregularly distributed samples in Fourier space. This algorithm uses an interpolation kernel; in our case this kernel is a modified Kaiser-Bessel (MKB) window function (also known as a blob). The gridding-based direct Fourier method can yield resolutions higher than the algebraic methods (and clearly weighted back-projection methods) in a fraction of their computing time.

Frequency samples from different projections mainly concentrate at the center of the 3D frequency domain, and their sparsity increases as we move away from the center. If no weighting scheme were employed, samples close to the 3D Fourier origin would be over-represented with respect to points away from the origin. In general, the sample values must be corrected by a weight function before inverting the Fourier transform of the volume. An accepted method in SPA to perform this weighting task is to use the volume of the Voronoi region [9] around each Fourier sample. This region is a polyhedron associated with each sample such that the distance between this sample and any point in the polyhedron is shorter than the distance from these points to any of the remaining samples. However, computing Voronoi cells is time consuming (particularly for SPA, where the algorithm has to handle millions of Fourier samples).

An alternate algorithm for obtaining the weighting function was proposed by Matej and Lewitt [11] for Positron Electron Tomography (PET). This approach seeks appropriate weights at each Fourier sample so that it participates with the right weight during the interpolation of regular points. The algorithm begins with the initial weights of the samples and uses convolution with a kernel to iteratively refine these weights. Their proposed method is practical for PET where the geometry of data is known and the number of sampling points is not large, but not for SPA where such conditions are not met.

In this paper, we introduce a gridding-based direct Fourier three-dimensional reconstruction in SPA following the method suggested by Matej and Lewitt [11] in PET. Our method follows the same iterative scheme for computing the weights, but we estimate the weights at each Fourier sample by evaluating a function instead of storing the collection of weights (which would become impractical in SPA). The proposed approach follows the general idea of Scheres [12], but differs in the way that the weights are calculated. Finally, our algorithm has an additional novel step to compensate for the trilinear interpolation of weights in Fourier space (described in Section 2.4), which improves the resolution of the final reconstruction. The CTF correction is applied during 3D reconstruction and is crucial for a high-resolution structure determination.

We compared the proposed reconstruction algorithm with the algorithms from the SPARX package [13] and the RELION package [12]. In the method by Zhang et al. [13], the projections are first padded to 2 times their original size (default value in the implementation), and then the Nearest Neighbor (NN) interpolation is used to calculate the target 3D Fourier volume. Finally, a weighting function using Bracewell's “local density” [14] is computed to correct the value for each voxel of the 3D Fourier volume.

The experiments showed that our approach is a superior method for 3D reconstruction in terms of accuracy, speed, and memory usage. The new algorithm is fully implemented in the open-source Xmipp package and is downloadable from http://xmipp.cnb.csic.es.

2. Material and methods

2.1. Preliminaries

The goal of the gridding-based direct Fourier method is to approximate frequency samples on a regular 3D Cartesian lattice $F_{3D}(\bar{R})$ from the measured samples of the 3D frequency domain $\hat{F}_{3D}(\bar{Q})$ as

$$F_{3D}(\bar{R}) = \int CTF^{-1}(\bar{Q})\hat{F}_{3D}(\bar{Q})K(\bar{R}-\bar{Q})d\bar{Q}, \tag{1}$$

where $\bar{R}$ is the frequency coordinate within the regular 3D grid and K is the kernel function by which the integration is accomplished. We recommend using a kernel function with some appealing attributes, such as finite size, bell-shaped decay, and differentiability at the borders. The modified Kaiser-Bessel (MKB or blob) is considered to be the best kernel for gridding interpolation by several authors [10,11,15,16]. Matej and Lewitt [16] generally assessed the optimal values for the parameters of the MKB to achieve a reconstruction with good quality. We use an MKB with the same parameter values suggested in their paper. CTF correction is incorporated during interpolation by dividing each irregular sample by $CTF^{-1}(\bar{Q})$, where $CTF(\bar{Q})$ is the value of the CTF at frequency $\bar{Q}$ (as a practical implementation issue, this division is performed as long as the CTF is above a given threshold).

Under experimental conditions, a limited number of projections from the specimen are available. Therefore, the discrete form of the integral in Eq. (1) should be considered because measurements are only available at a finite set of frequencies $\overline{Q} \in \{\bar{R}_i\}$. To obtain a discrete form of Eq. (1), the integral is substituted by a summation as shown

$$F_{3D}(\bar{R}) = \sum_i CTF^{-1}(\bar{R}_i)\hat{F}_{3D}(\bar{R}_i)K(\bar{R}-\bar{R}_i)w(\bar{R}_i), \tag{2}$$

where $w(\bar{R}_i)$ is the weighting factor for the i-th irregular sample. It is important to note that the weighting function is a substitution for $d\bar{Q}$ within Eq. (1). In fact, the value of each irregular sample should be corrected by the weighting factor and the related CTF