Journal of Structural Biology 178 (2012) 152–164

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

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

High-throughput subtomogram alignment and classification by Fourier space constrained fast volumetric matching

Min Xu [a], Martin Beck [b], Frank Alber [a,*]

[a] Program in Molecular and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA
[b] Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany

ARTICLE INFO

Article history:
Available online 7 March 2012

Keywords:
Cryo-electron tomography
Subtomogram alignment and classification
Fast rotational matching

ABSTRACT

Cryo-electron tomography allows the visualization of macromolecular complexes in their cellular environments in close-to-live conditions. The nominal resolution of subtomograms can be significantly increased when individual subtomograms of the same kind are aligned and averaged. A vital step for such a procedure are algorithms that speedup subtomogram alignment and improve its accuracy to allow reference-free subtomogram classifications. Such methods will facilitate automation of tomography analysis and overall high throughput in the data processing. Building on previous work, here we propose a fast rotational alignment method that uses the Fourier equivalent form of a popular constrained correlation measure that considers missing wedge corrections and density variances in the subtomograms. The fast rotational search is based on 3D volumetric matching, which improves the rotational alignment accuracy in particular for highly distorted subtomograms with low SNR and tilt angle ranges in comparison to fast rotational matching of projected 2D spherical images. We further integrate our fast rotational alignment method in a reference-free iterative subtomogram classification scheme, and propose a local feature enhancement strategy in the classification process. As a proof of principle, we can demonstrate that the automatic method can successfully classify a large number of experimental subtomograms without the need of a reference structure.

1. Introduction

Cryo-electron tomography (cryoET) enables the visualization of a cell's interior under close to live conditions (Beck et al., 2009, 2011; Nickell et al., 2006). The 3D tomogram is reconstructed from a set of two-dimensional micrographs, which are collected by tilting the sample around a single rotational axis. The reconstructed tomograms have typically resolutions that are sufficient to detect individual macromolecular complexes in their cellular context (Nicastro et al., 2006; Kühner et al., 2009; Medalia et al., 2002; Komeili et al., 2006). Once detected, subtomograms of the same complexes can be aligned and averaged to achieve a higher nominal resolution and signal-to-noise ratio (*SNR*) for their 3D density maps (Frank, 2006).

Subtomogram alignment and subsequent classification is an integral part of this strategy. The alignment relies on the search for the best rigid transformation of one subtomogram with respect to the second so that the similarity measure between them is maximized. However, several factors make the alignment of subtomograms challenging. Tomograms contain relatively high noise levels (Frangakis and Förster, 2004) and are typically of low non-isotropic resolution ($\geqslant$4 nm after averaging (Lucic et al., 2005; Förster et al., 2005; Briggs et al., 2009)). Moreover, tomograms are subject to distortions. One source of distortions is the variation of the Contrast Transfer Function (CTF) within and between individual micrographs used in the 3D density reconstruction (Förster et al., 2008). More critical are orientation specific distortions as a result of the so-called missing wedge effect, which is a consequence of the limited data collection due to the limited tilt ranges when collecting individual micrographs (with a maximal tilt range from −70 to +70 degrees). As a result, in Fourier space structure factors are missing in a characteristic wedge shaped region. This missing data leads to anisotropic resolution and different kinds of artifacts that depend on the structure of the object and its orientation with respect to the direction of the tilt-axis.

Several subtomogram alignment approaches take into account the missing wedge effects by using a constrained similarity measure (e.g., Förster et al., 2008; Bartesaghi et al., 2008; Amat et al., 2010; Volkmann, 2010). For instance, Förster et al. (2008) introduced a correlation metric that constrains the similarity measure only to the structure factors (i.e., Fourier coefficients) common to both pairs of subtomograms. Because sample thickness can be variable in a tomogram, this method also corrects for the local

* Corresponding author. Address: 1050 Childs Way, RRI 413E, Los Angeles, CA 90089, USA. Fax: +1 213 740 2437.
E-mail address: alber@usc.edu (F. Alber).

1047-8477/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2012.02.014

contrast difference in individual subtomograms by normalizing the similarity measure with respect to the mean and variance of the intensity distributions in both subtomograms (Förster et al., 2008).

In another method the influence of noise has been reduced in the alignment process by considering only a small percentage of high magnitude Fourier coefficients when computing the cross-correlation based similarity metric and therefore this method excludes those coefficients that are expected to be dominated by noise (Amat et al., 2010).

Most subtomogram alignment methods use an exhaustive scanning over all rotations of one subtomogram relative to the second to identify the orientation that maximizes the similarity metric (e.g., Förster et al., 2008; Amat et al., 2010; Volkmann, 2010). The scanning is performed at fixed angle intervals and at each sampled rotation a fast translational search is performed using the Fast Fourier Transform (FFT) (Frank, 2006). For highly accurate alignments, relatively small sampling intervals are needed, which significantly increases the cost of the rotational search. Exhaustive rotational scanning is computationally intensive, which limits its applicability when large data sets need to be aligned for subtomogram classification. This problem becomes even more relevant with increasing cryoET resolutions (Murata et al., 2010) and resulting larger subtomogram volumes. Therefore development of new algorithms that improve speed and accuracy of subtomogram alignment is a vital step in automation of tomography analysis and overall high throughput in data processing and reconstruction.

To enhance computational efficiency, Bartesaghi et al. (2008) use a fast rotational matching in which the best rotational transformation is computed rapidly using the convolution theorem within the Spherical Harmonics framework. Each subtomogram is transformed into a two-dimensional spherical image by integrating the magnitude of Fourier coefficients positioned along rays through the Fourier space origin. However, such a projection of the Fourier coefficient magnitudes onto the points of a unit sphere may increase ambiguities in the alignment. Moreover, the applied alignment dissimilarity score is not normalized with respect to the mean and variance of two subtomograms, therefore does not consider the non-uniformness of the tomogram.

In this paper, we formulate a Fourier-space equivalent similarity measure for the normalized constrained correlation introduced by Förster et al. (2008) and combine it with a fast rotational alignment based on 3D volumetric rotational matching. We adapt a 3D volumetric rotational matching method that was previously used for fitting atomic structures into density maps (Kovacs and Wriggers, 2002; Garzon et al., 2007) and extend its applicability to subtomogram alignments by including missing wedge effects in the matching score. This method significantly increases the alignment speed in comparison to the standard exhaustive rotational scanning approach (Förster et al., 2008). It also improves the alignment accuracy for subtomograms with low SNR and small tilt angle ranges in comparison to the fast rotational the rotational matching of 2D spherical images (Bartesaghi et al., 2008). Moreover, by using a Fourier equivalent score of the normalized constrained correlation our method also corrects for the non-evenness of contrast in tomograms.

As test case for our fast alignment method we perform reference-free subtomogram classifications. Reference-free classification is fundamental for providing an unbiased structural categorization of the macromolecular complexes in subtomograms because the initial classification is directly derived from the input data. Several types of reference-free subtomogram classification approaches exist, including methods based on maximum likelihood approaches (e.g., Scheres et al., 2009), methods using rotation invariant subtomogram features (e.g., Xu et al., 2009, 2011), and finally methods that rely on iterative successive alignment and classification steps (i.e., the alignment-through-classification approach) (e.g., Bartesaghi et al., 2008; Winkler, 2007; Winkler et al., 2009; Hrabe et al., 2012).

Our alignment method is sufficiently fast for carrying out all pair-wise alignments even for a relatively large number of subtomograms (e.g., a few thousands). Therefore we are able to embed the method into the alignment-through-classification framework, which is widely popular in 2D single particle averaging. Furthermore, we propose an additional generic and automatic local feature enhancement step to the framework. This step emphasizes the most discriminative local features between subtomograms, which improves the clustering performance. Moreover, we also integrate an automatic optimal cluster selection into the framework. We can demonstrate that the automated framework can successfully classify experimental subtomograms even for highly similar but distinct complexes.

2. Materials and methods

2.1. Fourier space equivalence of constrained correlation

Two subtomograms f and g are defined as two integrable functions $f, g : \mathbb{R}^3 \to \mathbb{R}$. To calculate the similarity between two subtomograms, Förster et al. (2008) proposed a constrained correlation with missing wedge correction. It is based on a transform of the subtomograms that eliminates the Fourier coefficients in the missing wedge region. This goal is achieved by introducing a binary missing wedge mask function as $\mathcal{M} : \mathbb{R}^3 \to \{0, 1\}$. The missing wedge mask function $\mathcal{M}$ defines valid and missing Fourier coefficients in Fourier space. For example, in single tilt electron tomography with tilt angle range $\pm\theta$, the missing wedge mask function can be defined as $\mathcal{M}(\boldsymbol{\xi}) := \mathbb{I}_{(|\xi_3| \leqslant |\xi_1| \tan(\theta))}(\boldsymbol{\xi})$, where $\mathbb{I}$ is the indicator function. Given $\mathcal{M}$, the real space subtomogram then excludes any coefficients located inside of any of the two missing wedge regions and is defined as

$$f^* := \Re\{\mathcal{F}^{-1}[(\mathcal{F}f)\,\Omega]\} \tag{1}$$

where $\Re$ denotes the real part of a complex function; Λ_R is the rotation operator such that $(\Lambda_R e)(\mathbf{x}) := e[R^{-1}(\mathbf{x})]$ for any function $e : \mathbb{R}^3 \to \mathbb{C}$; $\mathcal{F}$ is the Fourier transform operator; and $\Omega := \mathcal{M}(\Lambda_R \mathcal{M})$ ensures that only structure factors are considered that are defined in both subtomograms, therefore excluding any structure factors that are located in the missing wedge region of any of the two subtomograms. Correspondingly, the real space function of the second subtomogram is defined as

$$g^* := \Re\{\mathcal{F}^{-1}[(\mathcal{F}\tau_{\mathbf{a}}\Lambda_R g)\,\Omega]\} \tag{2}$$

where $\tau_{\mathbf{a}}$ is defined as the translation operator so that $(\tau_{\mathbf{a}} g)(\mathbf{x}) := g(\mathbf{x} - \mathbf{a})$ with $\mathbf{a} \in \mathbb{R}^3$. Then the normalized constrained correlation value between two subtomograms f and g that also considers the missing-wedge corrections can be calculated as (Förster et al., 2008)

$$c = \frac{\int [f^* - \mu(f^*)]\,[g^* - \mu(g^*)]}{\sqrt{\int [f^* - \mu(f^*)]^2}\,\sqrt{\int [g^* - \mu(g^*)]^2}} \tag{3}$$

where μ is the mean operator that returns the mean value of a function h, and is defined as $\mu h = \frac{\int h(\mathbf{x})}{\mathcal{S}h}$. $\mathcal{S}$ is the size operator, and correspondingly $\mathcal{S}h$ returns the size of the subtomogram h (i.e., the total number of voxels in the subtomogram). Because this similarity measure c is normalized by the mean and variance of intensity values of both subtomograms it considers the local contrast differences for individual regions in a tomogram.

To allow a fast rotational alignment of the local constrained similarity measure, we formulate the constrained correlation