Journal of Structural Biology xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Technical Note

Optimal and fast rotational alignment of volumes with missing data in Fourier space ☆

Maxim Shatsky [a,*], Pablo Arbelaez [b], Robert M. Glaeser [c], Steven E. Brenner [a,d]

[a] Physical Biosciences Division, Lawrence Berkeley National Laboratory, CA 94720, USA
[b] Electrical Engineering and Computer Science Division, University of California, Berkeley, CA 94720, USA
[c] Life Sciences Division, Lawrence Berkeley National Laboratory, CA 94720, USA
[d] Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 20 May 2013
Received in revised form 12 August 2013
Accepted 13 August 2013
Available online xxxx

Keywords:
Missing data in Fourier space
Rotational alignment of tomograms
Spherical harmonics
Constrained cross-correlation

ABSTRACT

Electron tomography of intact cells has the potential to reveal the entire cellular content at a resolution corresponding to individual macromolecular complexes. Characterization of macromolecular complexes in tomograms is nevertheless an extremely challenging task due to the high level of noise, and due to the limited tilt angle that results in missing data in Fourier space. By identifying particles of the same type and averaging their 3D volumes, it is possible to obtain a structure at a more useful resolution for biological interpretation. Currently, classification and averaging of sub-tomograms is limited by the speed of computational methods that optimize alignment between two sub-tomographic volumes. The alignment optimization is hampered by the fact that the missing data in Fourier space has to be taken into account during the rotational search. A similar problem appears in single particle electron microscopy where the random conical tilt procedure may require averaging of volumes with a missing cone in Fourier space. We present a fast implementation of a method guaranteed to find an optimal rotational alignment that maximizes the constrained cross-correlation function (cCCF) computed over the actual overlap of data in Fourier space.

1. Introduction

The problem of incomplete data appears in several areas of electron microscopy (EM) applied to biological samples. For example, EM tomographic reconstruction is achieved by combining two-dimensional projections obtained at different tilt angles. Current experimental methodology limits the tilt angle of a specimen to about ±70 degrees. This results in reconstructed three-dimensional volumes that lack significant information in the corresponding Fourier space. This missing structural volume is called the *missing wedge* and corresponds to at least 20 percent of the structural data. If similar particles are recorded at different orientations, then averaging these sub-tomogram volumes can restore the missing information and at the same time improve the overall resolution. Sub-tomogram averaging has been successfully applied in the past (Frangakis et al., 2002; Bartesaghi and Subramaniam, 2009; Briggs et al., 2009; Winkler et al., 2009; Stölken et al., 2011).

Volume averaging is not limited to EM tomography. In single particle cryo-EM an initial model of negatively stained particles is usually obtained using the Random Conical Tilt approach. For the Random Conical Tilt approach, two images are collected, for example, at zero and 45–60 degrees. Particles with similar 2-D projections are identified in the untilted image and their corresponding tilted counterparts are used to partially fill the Fourier space. The volumetric space that cannot be filled with this information results in the shape of a *missing cone*. Here, as well, the missing data can be identified by combining two or more volumes that have complementary information (Penczek et al., 1994; Scheres et al., 2009).

Volume averaging requires solving the problem of rotational and translational alignment since the relative orientation of the volumes is usually not known. Volume alignment involves optimization of some similarity measure under six degrees of freedom – three degrees for rotation and three degrees for translation. The cross-correlation function (CCF) is widely used to measure the similarity of images and volumetric data. This function is especially attractive because its optimal value can be efficiently computed by applying the Fast Fourier Transform (FFT) (Cooley and Tukey, 1965). Therefore, finding an optimal translation between two

* Corresponding author. Address: University of California, 461 Koshland Hall, Berkeley, CA 94720-3102, USA.
E-mail address: max.shatsky@gmail.com (M. Shatsky).

1047-8477/$ - see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.08.006

volumes of dimension $n \times n \times n$ can be done efficiently in $O(n^3 \log(n))$ time. However, optimizing cross-correlation in rotational space has proved to be more challenging. An exhaustive search can take $O(n^6)$ time if sampling every $360/n$ degrees. Improved algorithms came when it was realized that the FFT could successfully be used in the space of spherical harmonics, and that the volume's real space representation can be very accurately approximated in the space of spherical harmonics. The first proposed method resolved two degrees of freedom by applying the FFT on spherical harmonics in 2-D, while the third rotational dimension was left for the exhaustive sampling (Crowther, 1972). Two decades later, a solution for 3-D FFT on spherical harmonics was proposed with time complexity of $O(n^4)$ (Kovacs and Wriggers, 2002). Currently, this is asymptotically the fastest known algorithm to optimally solve the rotational CCF in 3-D.

The missing wedge and the missing cone in Fourier space result in elongated density in the real space. This in turn creates a bias during the rotational alignment in which volumes tend to be aligned along the axis of elongation. To avoid such bias, several solutions have been proposed (Schmid and Booth, 2008; Bartesaghi et al., 2008; Förster et al., 2008). These are based on constraining the CCF to the overlapping data in Fourier space between rotationally aligned volumes. In terms of computational complexity these methods can be classified into two categories: (1) exhaustive, very slow, but accurate, and (2) faster heuristics which do not guarantee optimal orientation. The first category comprises approaches that apply an exhaustive rotational search in which a variant of the constrained CCF is calculated for each sampled rotation. These methods are very slow and supercomputers are required to process such large numbers of alignments (Frangakis et al., 2002; Förster et al., 2008; Scheres et al., 2009; Amat et al., 2010; Stölken et al., 2011). The second type, faster heuristics, consists of a method that applies the FFT and spherical harmonics to compute the constrained CCF faster, but does not guarantee to achieve an optimal solution (Bartesaghi et al., 2008).

Here we present a method that is guaranteed to find an optimal rotational alignment that maximizes the constrained cross-correlation function computed over the actual overlap of data in Fourier

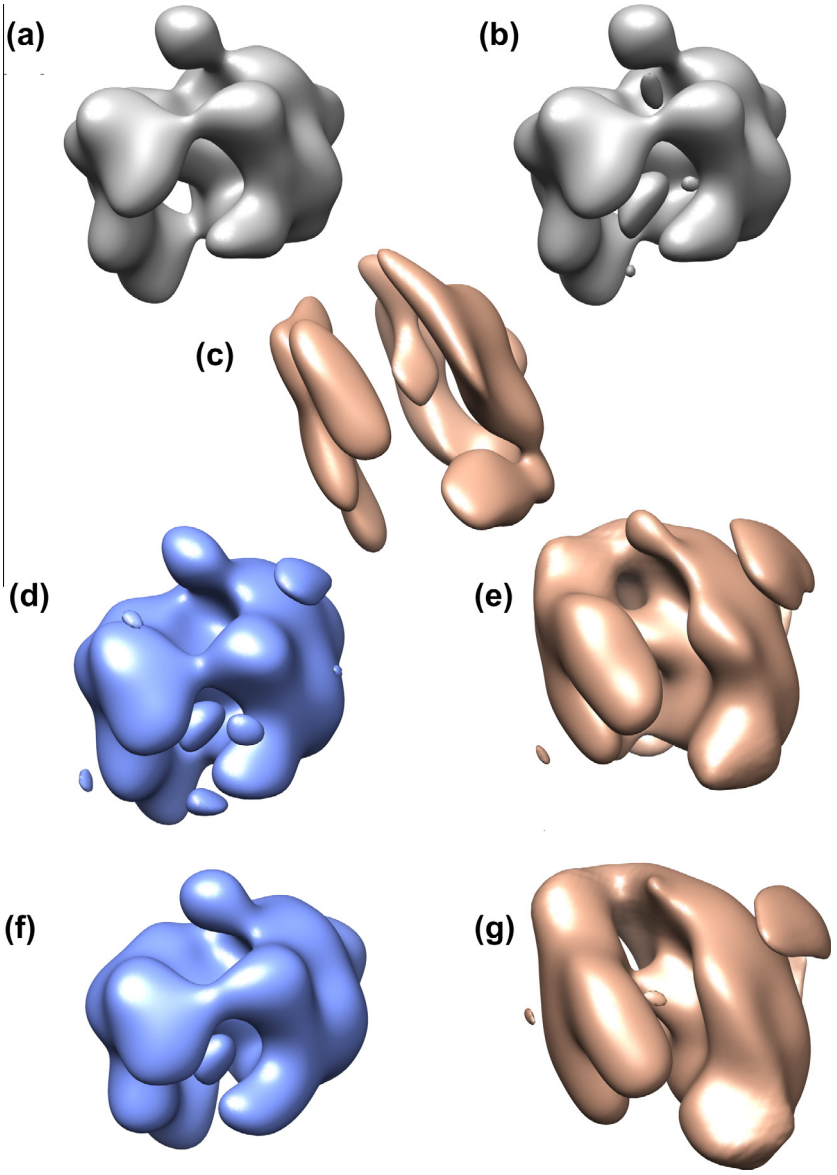

Fig. 1. (a) 70S ribosome filtered to 50 Å. (b) The average volume obtained from applying the CCF function on 100 randomly oriented volumes with complete information in Fourier space. (c) Fourier space wedge mask is applied to the volume from (a). (d) Result of averaging 100 randomly oriented, wedge mask filtered volumes aligned with the cCCF. (e) Same as (d) but the CCF is used for the alignment. Notice elongation bias. (f) Volume refinement. Model from (d) is used to align 100 volumes using the CCF method. (g) Model from (e) is used to align 100 volumes using the CCF method.

Please cite this article in press as: Shatsky, M., et al. Optimal and fast rotational alignment of volumes with missing data in Fourier space. J. Struct. Biol. (2013), http://dx.doi.org/10.1016/j.jsb.2013.08.006

space. The method's key idea is based on the observation that correlation between two functions in real space is equivalent to correlation in Fourier space. The constrained cross-correlation function (cCCF) is defined in Fourier space as the CCF function normalized by the size of the overlap of the two volumes' actual experimentally obtained data. The overlap function itself can be represented as a correlation in real space of the masks that define the location of experimental data in Fourier space. Consequently, the cCCF function is represented as the product of two real-valued correlation functions, each of which can be efficiently computed using the algorithm by Kovacs and Wriggers (2002) (see Supplementary information for more details). Recently, an independently realized Matlab implementation of this approach was presented (Xu et al., 2012) which takes minutes to rotationally align two volumes. Here we report an implementation orders of magnitude faster, for which we modified the source code in **C** of Kovacs and Wriggers (2002).

2. Validation and conclusions

We validated our implementation of the cCCF procedure on a synthetic data set generated from a density map of the 70S ribosome taken from the EBI Database EMDB (Tagari et al., 2002), entry EMD-1003, resolution 11.3 Å (Rawat et al., 2003). First, the density map was filtered to 50 Å (Fig. 1(a)). One hundred random rotations were applied to the filtered volume. Then, each orientation was transformed to Fourier space and a wedge mask that corresponded to ±40 degree tilt was applied. The volumes were then transformed back to real space. This filtered set of one hundred volumes was used to test the alignment. The alignment was performed by using the first volume (Fig. 1(c)) as a model and superimposing the rest of the volumes onto it. The average of these 99 alignments was computed by the weighted wedge mask overlap as described in the Suppl. Inf. The average of cCCF alignment is shown in Fig. 1(d). Notice that the cCCF average is very close to the original undistorted volume, while average of CCF alignment, Fig. 1(e), displays a heavy wedge bias. The averages from (d) and (e) were then used as the models for the second round of alignment. While the cCCF average achieves a small improvement, Fig. 1(f), the CCF average continues to be elongated in the direction of the missing wedge, Fig. 1(g).

The method presented here is general for all types of missing areas in Fourier space, and is not limited to a missing wedge or cone. Its fast running time, about a second for volumes of size $135 \times 135 \times 135$, allows thousands of alignments to be performed in one day on a standard desktop computer. The code can be downloaded at: compbio.berkeley.edu/people/maxshats/volalign/.

Acknowledgments

We thank Julio Kovacs and Pablo Chacon for providing source code of the frm3d program. This work conducted by ENIGMA – Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, was supported by the Office of Science, Office of Biological and Environmental Research, of the US Department of Energy under Contract No. DE-AC02-05CH11231.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2013.08.006.

References

Amat, F., Comolli, L.R., Moussavi, F., Smit, J., Downing, K.H., Horowitz, M., 2010. Subtomogram alignment by adaptive Fourier coefficient thresholding. Journal of Structural Biology 171 (3), 332–344.

Bartesaghi, A., Subramaniam, S., 2009. Membrane protein structure determination using cryo-electron tomography and 3D image averaging. Current Opinion in Structural Biology 19 (4), 402–407.

Bartesaghi, A., Sprechmann, P., Liu, J., Randall, G., Sapiro, G., Subramaniam, S., 2008. Classification and 3D averaging with missing wedge correction in biological electron tomography. Journal of Structural Biology 162 (3), 436–450.

Briggs, J.a.G., Riches, J.D., Glass, B., Bartonova, V., Zanetti, G., Kräusslich, H.-G., 2009. Structure and assembly of immature HIV. Proceedings of the National Academy of Sciences 106 (27), 11090–11095.

Cooley, J.W., Tukey, J.W., 1965. An algorithm for the machine calculation of complex Fourier series. Mathematics of Computation 19 (90), 297–301.

Crowther, R.A., 1972. The fast rotation function. In: Rossmann, M.G. (Ed.), The Molecular Replacement Method. Gordon and Breach, New York, pp. 173–178.

Förster, F., Pruggnaller, S., Seybert, A., Frangakis, A.S., 2008. Classification of cryo-electron sub-tomograms using constrained correlation. Journal of Structural Biology 161 (3), 276–286.

Frangakis, A.S., Böhm, J., Förster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R., Baumeister, W., 2002. Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. Proceedings of the National Academy of Sciences 99 (22), 14153–14158.

Kovacs, J.A., Wriggers, W., 2002. Fast rotational matching. Acta Crystallographica Section D Biological Crystallography 58 (8), 1282–1286.

Penczek, P.A., Grassucci, R.A., Frank, J., 1994. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy 53 (3), 251–270.

Rawat, U.B.S., Zavialov, A.V., Sengupta, J., Valle, M., Grassucci, R.A., Linde, J., Vestergaard, B., Ehrenberg, M., Frank, J., 2003. A cryo-electron microscopic study of ribosome-bound termination factor RF2. Nature 421 (6918), 87–90.

Scheres, S.H.W., Melero, R., Valle, M., Carazo, J.-M., 2009. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. Structure 17 (12), 1563–1572.

Schmid, M.F., Booth, C.R., 2008. Methods for aligning and for averaging 3D volumes with missing data. Journal of Structural Biology 161 (3), 243–248.

Stölken, M., Beck, F., Haller, T., Hegerl, R., Gutsche, I., Carazo, J.-M., Baumeister, W., Scheres, S.H.W., Nickell, S., 2011. Maximum likelihood based classification of electron tomographic data. Journal of Structural Biology 173, 77–85.

Tagari, M., Newman, R., Chagoyen, M., Carazo, J.-M., Henrick, K., 2002. New electron microscopy database and deposition system. Trends in Biochemical Sciences 27 (11), 589.

Winkler, H., Zhu, P., Liu, J., Ye, F., Roux, K.H., Taylor, K.A., 2009. Tomographic subvolume alignment and subvolume classification applied to myosin V and SIV envelope spikes. Journal of Structural Biology 165 (2), 64–77.

Xu, M., Beck, M., Alber, F., 2012. High-throughput subtomogram alignment and classification by Fourier space constrained fast volumetric matching. Journal of Structural Biology 178 (2), 152–164.