COMPUTING THE 3-D STRUCTURE OF VIRUSES FROM ELECTRON MICROSCOPE IMAGES

*Junghoon Lee, Peter C. Doerschuk**

Purdue University
School of Electrical and Computer Engineering
West Lafayette, IN

Jinghua Tang, John E. Johnson

The Scripps Research Institute
Department of Molecular Biology
La Jolla, CA

ABSTRACT

3-D reconstruction of the electron scattering intensity of a virus from cryo electron microscopy is essentially a 3-D tomography problem in which the orientation of the 2-D projections is unknown. Many biological problems concern mixtures of different types of virus particles or mixtures of different maturation states of the same type of virus particle. For a variety of reasons, especially low SNR, it can be very challenging to label the type or state shown in an individual image. Algorithms capable of computing multiple reconstructions, one for each type or state, based on images which are not labeled according to type or state, are described and demonstrated on experimental images.

1. INTRODUCTION

We describe the basic approach and recent computational results concerning the computation of the 3-D scattering density of virus particles from 2-D cryo electron microscope (cryo-EM) images. The basic approach is documented in detail in Ref. [1] and initial numerical results on experimental images are documented in detail in Ref. [2]. The contribution of this paper is additional numerical results demonstrating the computation of multiple 3-D reconstructions from unlabeled mixtures of particle images.

Each cryo-EM image is a 2-D projection of the 3-D scattering density modified by the so-called contrast transfer function (CTF) of the microscope. Two central problems in cryo-EM imaging are the unknown projection orientation and the sensitivity of the specimen to the electron beam. Because the orientation of the 3-D specimen on the stage of the microscope is not known, it follows that the image is related to an unknown-orientation 2-D projection of the 3-D specimen. This would not be a problem if the microscopist could rotate the 3-D specimen and take a series of images with known relative orientation, which is essentially what is done in medical imaging. However, this approach is not possible because the 3-D specimen is rapidly damaged by the electron beam (and also due to technical problems with the range of achievable rotation). Therefore, taking multiple images of one 3-D specimen in different orientations is replaced by taking one image (or a very few images called a tilt series) of each of many identical 3-D specimens where each specimen is in a different random unknown orientation. The collection of images from random unknown orientations is a less informative data set exactly because of the unknown orientations. In addition, because the specimen is sensitive to damage by the electron beam, it is desirable to keep the beam current low which leads to very noisy images with as few as 5 or 6 electrons contributing to a particular pixel though more commonly higher beam currents are used which lead to roughly 10^2 electrons contributing to a typical pixel. Finally, the CTF of the microscope is never known exactly but is of great importance when computing high spatial resolution structures because in the spatial frequency domain it has multiple zeros and sign reversals at higher frequencies.

Reconstruction methods for cryo-EM images have been extensively studied. From the structural biology point of view, recent special issues of the *Journal of Structural Biology* include Refs. [3, 4, 5] and recent reviews include Refs. [6, 7]. In the engineering literature, new results and an extensive literature review on estimation of projection orientations are contained in Ref. [8, 9, 10]. In particular, Refs. [9, 10] describe new methods of projection orientation estimation based on moments and new results on uniqueness of reconstruction. High performance computing implementations are described in Refs. [11, 12, 13]. Once the projection orientations are determined, the reconstruction problem is large (since it is in 3-D) but tractable. Two important and highly-developed approaches for determining the projection orientations are methods based on the symmetry of the particle (so-called "common lines" methods) and methods based on correlation with a model. Broad classes of virus particles have 3-D scattering densities with rotational symmetries. Therefore the 3-D Fourier transforms of the

*This work was supported by NIH (1R01EB000432-01), NSF (MCB-9873139, EIA-0112672, CCR-0098156, and EIA-0130538) and ARO DAAD19-99-1-0015.

scattering densities also have rotational symmetries. Due to the 3-D projection slice theorem, the 2-D Fourier transform of a projection is a 2-D slice through the 3-D Fourier transform of the scattering density where the slice includes the origin. The rotational symmetries of the 3-D Fourier transform of the scattering density appear in the 2-D Fourier transform of the projection as radially directed lines along which the variation of the function is identical. By locating the number and position of such lines it is possible to determine the projection orientation. A major problem is the noise in the images. Sophisticated methods are used to address this issue such as using pattern recognition ideas to collect sets of images with similar projection orientations followed by averaging of the images to increase signal to noise ratio and then estimation of the projection orientation in the averaged image. A second important approach to determining projection orientation is based on models. In many biological investigations there is some idea of how the virus under investigation is structured, for instance, it may be related to a second virus with known structure. Therefore it is often possible to propose a 3-D structure, typically at low spatial resolution, which can be used as a template in a 3-D correlation calculation to determine which projection orientation for the template best matches the observed image. Both common-lines and model-based correlation procedures are then embedded in larger iterative algorithms in which projection orientation estimation alternates with reconstruction and the resolution of the orientation estimates and reconstructions is slowly increased from iteration to iteration.

We have developed an integrated statistical approach to both reconstruction and evaluation of alternative experimental designs that addresses the issues raised in the first paragraph. This approach is based on models of the entire image formation process:

The model of the particle emphasizes the symmetry and support in 3-D of the scattering density of the particle. While the scattering density is always positive, we do not emphasize that constraint because what is actually imaged is the difference between the scattering intensity of the particle and the scattering intensity of the vitreous ice in which the particle is embedded. A simple *a priori* statistical model for the particle is included which can also be interpreted as a regularization term.

The model for the specimen is that the specimen is composed of mixtures of different types of particles where the number of types is known and all particles of a particular type are identical. For details of the experiment please see Refs. [6, 7]. We consider two different statistical descriptions of the uncertainty in particle type. The result of the reconstruction is a separate model for each type of particle.

A key feature of the specimen is the random orientation of the particles in the frozen specimen. We consider an extensive set of models of orientation: No preferred orientation, a precise preferred orientation, and a preferred orientation with wobble. For particles with a preferred orientation, we consider two types of preference.

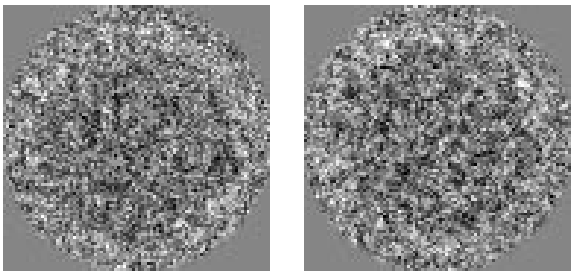


Fig. 1. Examples of the boxed Flockhouse virus [24] images. All panels use the same intensity scale.

The model for the image formation process is standard. It includes an unknown origin location in the images and a CTF that is not unity.

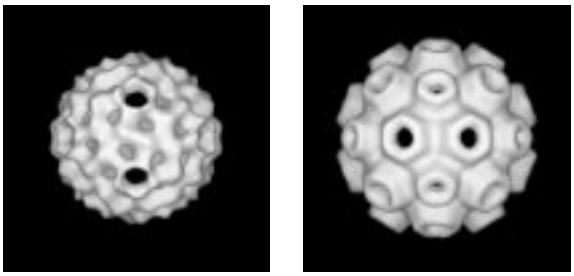
The image recording process is modeled as the addition of additive white Gaussian noise to the true image. At least for the corruption due to electron counting statistics, a Poisson model is more accurate. However, there are both algorithmic and modeling advantages to a Gaussian model. Such approaches are also discussed in, for example, Refs [14, 15].

In addition to the previously described features, we consider tilt series containing an arbitrary number of images with arbitrary relative projection orientations.

Our approach is to formulate a statistical description of the data that includes measurement uncertainty, orientation uncertainty, and type-of-particle uncertainty. We then solve a maximum likelihood estimation problem for the structure of each type of particle. In this paper we solve the likelihood equations using the expectation-maximization (EM) algorithm [16, p. 459] but other methods could also be considered. From the point of view of reconstruction procedures, the novel aspects of this work are the detailed statistical models of orientation, the attempt to deal simultaneously with orientation and reconstruction, and the ability to work with images that come from a mixture of particles of different types. While it is not the subject of this paper [1], we have used the same models to develop a methodology for evaluating alternative experimental protocols based on Cramer-Rao bounds [17, 18, 19].

2. IMAGE PROCESSING AND NUMERICAL RESULTS

The contribution of this paper is that we have developed new image pre-processing tools that have enabled these algo-



(a) (b)

Fig. 2. 3-D reconstructions of Cowpea Chlorotic Mottle Virus from Step 4 of Ref. [2]. (a) Reconstruction of the $T = 1$ (dimer) structure from 488 mixture images with prior probability 0.41. (b) Reconstruction of the $T = 3$ structure from 488 mixture images with prior probability 0.59.

rithms to successfully simultaneously compute multiple 3-D reconstructions from unlabeled mixtures of images. The images, e.g., Figure 1, have low SNR, the particle is not in the center of the image, and in some cases there is additional material in the image. The image pre-processing primarily deals with the second and third images.

In theory, non-centered particles are not a problem because the origin location can be treated as a nuisance parameter in the software of Ref. [2]. However, treating them as nuisance parameters leads to increased computational burden. Furthermore, in the existing software, the probability density function for the nuisance parameters is the same for all images. Therefore, no advantage can be gotten from the fact that a particular image is known to be well centered. In a qualitative sense, pre-processing to center the particle in the image is equivalent to changing from treating the origin location as a nuisance parameter to treating the origin location as a parameter to be estimated. However, it is a much simplier algorithm than treating the origin location as a parameter to be estimated in a, for example, maximum likelihood approach. We center a particle in an image by minimizing the l_2 correlation of the image with an annular template whose diameters are determined by the radii used to define the radial basis functions in the orthonormal expansion of the electron scattering intensity.

The algorithms of Ref. [1, 2] process the Fourier transform of the image not the image itself. We have approximated the Fourier transform by the DFT which we compute by the FFT. This approach implies that if there is a non-virus object within the image then the Fourier transform is inaccurately computed. For this reason, once the particle has been centered in the image, the image is masked to remove anything outside of the particle. This masking operation would be difficult to perform if the origin location was treated as a nuisance parameter.

Various variants of Cowpea Chlorotic Mottle Virus [20] can assembly in to a variety of types of particles, e.g., quasi $T = 3$ symmetry particles, which are the native state, and quasi $T = 1$ dimer symmetry particles. These particles differ in average radius. In work in submission, Dr. Jinghua Tang *et al.* have collected images of mixtures of the different types of particles and, by hand labeling of each image, determined multiple 3-D reconstructions using the Spider/Web system [21].

By extending the centering operation to include correlation with different diameter templates, we have automatically extracted subsets of the images based on radius and have created two-component mixtures. Surface renderings of some of the resulting reconstructions are shown in Figure 2. The visual correspondence with some of the submitted work of Dr. Jinghua Tang *et al.* is excellent. Quantitative correlation, e.g., the Fourier Shell Correlation function [22, Eq. 2] [23, Eq. 17] [7, p. 879] is underway.

3. REFERENCES

[1] Peter C. Doerschuk and John E. Johnson, "*Ab initio* reconstruction and experimental design for cryo electron microscopy," *IEEE Trans. Info. Theory*, vol. 46, no. 5, pp. 1714–1729, Aug. 2000.

[2] Zhye Yin, Yili Zheng, Peter C. Doerschuk, Padmaja Natarajan, and John E. Johnson, "A statistical approach to computer processing of cryo electron microscope images: Virion classification and 3-D reconstruction," *J. Struct. Biol.*, vol. 144, pp. 24–50, 2003.

[3] Bridget Carragher and Pawel Penczek, "Special issue: Analytical methods and software tools for macromolecular microscopy," *J. Struct. Biol.*, vol. 133, no. 1/2, pp. 1–252, Oct. 2003.

[4] Wah Chiu and Steven Ludtke, "Special issue: Single-particle analysis," *J. Struct. Biol.*, vol. 133, no. 2/3, pp. 89–266, Feb/Mar 2001.

[5] B. Carragher and P. R. Smith, "Special issue: Advances in computational image processing for microscopy," *J. Struct. Biol.*, vol. 116, no. 1, pp. 1–249, Jan. 1996.

[6] Joachim Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, San Diego, 1996.

[7] T. S. Baker, N. H. Olson, and S. D. Fuller, "Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," *Microbiology and Molecular Biology Reviews*, vol. 63, no. 4, pp. 862–922, Dec. 1999.

[8] Peter D. Lauren and N. Nandhakumar, "Estimating the viewing parameters of random, noisy projections of asymmetric objects for tomographic reconstruction," *IEEE Trans. PAMI*, vol. 19, no. 5, pp. 417–430, May 1997.

[9] Samit Basu and Yoram Bresler, "Uniqueness of tomography with unknown view angles," *IEEE Trans. Image Proc.*, vol. 9, no. 6, pp. 1094–1106, June 2000.

[10] Samit Basu and Yoram Bresler, "Feasibility of tomography with unknown view angles," *IEEE Trans. Image Proc.*, vol. 9, no. 6, pp. 1107–1122, June 2000.

[11] Ioana M. Martin and Dan C. Marinescu, "Concurrent computation and data visualization for spherical-virus structure determination," *IEEE Comput. Sci. Eng.*, pp. 40–52, October–December 1998.

[12] Christopher J. Lanczycki, Calvin A. Johnson, Benes L. Trus, James F. Conway, Alasdair C. Steven, and Robert L. Martino, "Parallel computing strategies for determining viral capsid structure by cryoelectron microscopy," *IEEE Comput. Sci. Eng.*, pp. 76–91, April–June 1998.

[13] Z. H. Zhou, W. Chiu, K. Haskell, H. Spears, Jr., J. Jakana, F. J. Rixon, and L. R. Scott, "Refinement of Herpesvirus B-capsid structure on parallel supercomputers," *Biophys. J.*, vol. 74, pp. 576–588, Jan. 1998.

[14] Ken D. Sauer and Charles A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[15] Jeffrey A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, no. 2, pp. 290–300, June 1994.

[16] E. L. Lehmann and George Casella, *Theory of Point Estimation*, Springer-Verlag, New York, 2nd edition, 1998.

[17] John. D. Gorman and Alfred. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Info. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.

[18] Thomas L. Marzetta, "A simple derivation of the constrained multiple parameter Cramer-Rao bound," *IEEE Trans. Sig. Proc.*, vol. 41, no. 6, pp. 2247–2249, June 1993.

[19] Harry L. van Trees, *Detection, Estimation, and Modulation Theory: Part I*, John Wiley and Sons, Inc., New York, 1968.

[20] Jeffrey A. Speir, Sanjeev Munshi, Guoji Wang, Timothy S. Baker, and John E. Johnson, "Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by x-ray crystallography and cryoelectron microscopy," *Structure*, vol. 3, pp. 63–78, 15 January 1995.

[21] Joachim Frank, Michael Radermacher, Pawel Penczek, Jun Zhu, Yanhong Li, Maheiddine Ladjadj, and Ardean Leith, "SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields," *J. Struct. Biol.*, vol. 116, pp. 190–199, 1996, `http://www.wadsworth.org/resnres/frank.htm`

[22] Marin van Heel, "Similarity measures between images," *Ultramicroscopy*, vol. 21, pp. 95–100, 1987.

[23] George Harauz and Marin van Heel, "Exact filters for general geometry three dimensional reconstruction," *Optik*, vol. 73, no. 4, pp. 146–156, 1986.

[24] Andrew J. Fisher and John E. Johnson, "Ordered duplex RNA controls capsid architecture in an icosahedral animal virus," *Nature*, vol. 361, pp. 176–179, 14 January 1993.