IMAGE RESTORATION IN SETS OF NOISY ELECTRON MICROGRAPHS

Steven J. Ludtke, Wah Chiu

National Center for Macromolecular Imaging, Baylor College of Medicine, Houston, TX 77030

ABSTRACT

The imaging process in a transmission electron microscope (TEM) produces a number of artifacts including the contrast transfer function (CTF) and envelope functions. In addition, when electron cryomicroscopy is performed for purposes of single particle reconstruction, signal to noise ratios are very low, generally peaking between 0.1 and 1.0, and averaging 0.01 to 0.1. We apply a semi-empirical model of the microscope artifacts to optimally correct data used for 3D single particle reconstructions.

1. INTRODUCTION

Electron cryo-microscopy applied to single particle processing is a methodology for reconstructing the 3D structure of biological molecules and macromolecular assemblies at resolutions currently ranging between 0.7 and 2 nm. In this technique ostensibly identical molecules/assemblies are embedded in a thin layer of vitreous ice and imaged in the microscope. The individual randomly oriented particles are then located in the resulting micrographs and through a complicated process are merged to produce a 3D structure of the particle in question. Such biological molecules are fragile, and easily damaged in the high energy (100 – 400 keV) electron beam. So, particularly for high resolution work, the signal to noise ratio in the micrographs is dose-limited. Typical spectral signal to noise ratios for moderate sized particles (10-30 nm) at intermediate resolutions (1-2 nm) peak between 0.1 and 1.0, and average only 0.01 to 0.1.

Additionally, the TEM produces a number of artifacts in the recorded images known as the CTF and envelope functions. The well-known image contrast theory [1] provides a good approximation to the true CTF. While theoretical expressions exist for several of the resolution-limiting envelope functions [2] they can be grouped together and approximated as a single Gaussian [3, 4].

The problem we seek to address is optimal deconvolution and filtration when data is being averaged together from several micrographs with different CTFs and envelope functions. This problem has been addressed in different contexts several times in the past [5, 6]. Our approach makes optimal use of the available data and provides accurate envelope function correction/filtration for 2D averaging. Figure 1 demonstrates these effects and the appearance of the raw data.

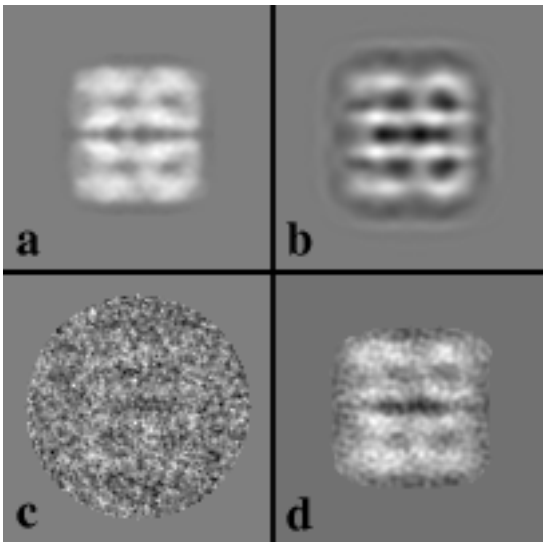

Figure 1 – a) computed projection of a GroEL particle [3]. b) the same particle with CTF and envelope functions applied. c) a typical single particle image from a real micrograph aligned to a. d) average of 48 particles with corrections applied as derived below.

2. MICROSCOPE ARTIFACTS

We begin by defining the CTF and how it affects the data. The Fourier transform of the measured microscope image can be described as:

$$\overline{M}(s,\theta) = C(s)E(s)\overline{F}(s,\theta) + \overline{N}(s,\theta), \qquad (1)$$

where $\overline{M}(s,\theta)$ is the Fourier transform of the measured image data expressed in polar coordinates. The overbar marks complex valued functions. *C(s)* and *E(s)* represent the CTF and envelope functions of the microscope, respectively. They are real-valued functions of *s* (spatial frequency) only. This is our first assumption; the image is stigmatic and drift-free eliminating any azimuthal dependence. Extension of this derivation to astigmatic and/or drifting images is straightforward, but makes the derivation considerably more complicated. $\overline{F}(s,\theta)$ represents the Fourier transform of a true particle

projection. In this case, this is the unknown we wish to regenerate from a set of measurements, $\overline{M}_n(s,\theta)$. $\overline{N}(s,\theta)$ is a complex vector representing background noise present in the image. As noise, it is completely random and uncorrelated with $\overline{F}(s,\theta)$. The noise distribution in real space is approximately Gaussian.

Image contrast theory provides a reasonably accurate theoretical model of *C(s):*

$$C(s) = -k(\sqrt{1-Q^2}\,\sin(\gamma) + Q\cos(\gamma)) \qquad (2)$$

$$\gamma = -2\pi(\frac{C_s\lambda^3 s^4}{4} - \frac{\Delta Z\lambda s^2}{2})$$

k is a trivial scaling factor relating the arbitrary scale of $\overline{F}(s,\theta)$ to the specific electron dose, film speed, etc. C_S is the spherical aberration of the objective lens (constant for most microscopes), λ is the electron wavelength, and ΔZ is defocus. Image contrast theory produces two terms, one for phase contrast and one for amplitude contrast. Q is the fractional amplitude contrast in this model.

The envelope function, *E(s)* can be complicated, and contains contributions from a number of effects, such as spatial and temporal coherence, specimen motion, etc. [2] Functional forms exist for many of these effects, but other possible effects like specimen charging and complex beam-induced movement are not well characterized. In addition, accurately determining all of the parameters for such a model is virtually impossible since many of them are nearly singular. It has been shown that in most cases a simple Gaussian with a single factor, B, adequately describes the cumulative envelope function:

$$E(s) = e^{-Bs^2} \qquad (3)$$

It is, of course, impossible to obtain an expression for $\overline{N}(s,\theta)$, since the phases, and individual amplitudes are completely random. The best we can do is to characterize the spectral profile of the intensity of the background. Generally the background intensity is rotationally symmetric, so we characterize the noise as a function of spatial frequency only. The sources of noise in the microscope and post-processing are difficult to explicitly enumerate, so the expression for background noise is purely empirical:

$$N^2(s) = n_1 e^{n_2 s + n_3 s^2 + n_4\sqrt{s}} \qquad (4)$$

This functional form has been found to fit the data from several microscopes sufficiently well to provide accurate corrections. As will be seen below, minor inaccuracies in background fitting have a minimal effect on the corrected images.

3. PERFORMING 2D IMAGE RESTORATION

As seen in eqn. 2, the functional form of the CTF is oscillatory with a varying period, with the position of the first zero-crossing varying with microscope defocus. This provides an interesting deconvolution problem. Q is generally in the range 0.05 to 0.15, meaning *C(s)* is largely sinusoidal and is very small at low spatial frequencies. For this reason, electron micrographs are typically collected somewhat out of focus to improve low resolution contrast.

While the oscillations of the CTF make deconvolution more difficult, we can also take advantage of this fact. Since we know the signal component is exactly zero at several spatial frequencies, we can determine the precise value of the background at these points. Focal series demonstrate that we can assume the background varies smoothly between these points, which allows complete separation of the signal from the background in the image power spectrum. Not only does this provide an accurate estimate of the spectral signal to noise ratio, but it also provides the opportunity to accurately determine CTF and envelope function parameters.

The data we wish to process consists of i micrographs, each containing a large number of particles in random orientations. In this derivation, we consider the process of combining n particles, all in the same orientation, selected from the set of micrographs. This set of n may contain several particles from each micrograph. Our goal is to produce a particle average that matches the corresponding projection of the true 3D structure as closely as possible.

In considering how to optimize the deconvolution/filtration process, we begin with our model of the raw data (eqn 1):

$$\overline{M}_{\Psi,n}(s,\theta) = C_n(s)E_n(s)\overline{F}_\Psi(s,\theta) + \overline{N}_n(s,\theta). \qquad (5)$$

$\overline{M}_{\Psi,n}(s,\theta)$ represents any particle, n, in approximately the same 3D orientation, Ψ, from any of the i micrographs. $\overline{F}_\Psi(s,\theta)$ represents the true 2D structure factor (Fourier transform) of the particle in this orientation. Note that, as the true answer, $\overline{F}_\Psi(s,\theta)$ is the same for all n particles. Our problem is to recover $\overline{F}_\Psi(s,\theta)$ as accurately as possible given only $\overline{M}_n(s,\theta)$. Clearly we will need to characterize $C_n(s)$, $E_n(s)$ and the spectral profile of $\overline{N}_n(s,\theta)$ to achieve this result. We currently utilize an eight parameter model, including k, Q, ΔZ, B and $n_{1\text{-}4}$. All other parameters in the above equations are known microscope constants. To determine the needed parameters, we consider the one dimensional power spectrum of all of the particles in a single micrograph.

This assumes that none of the micrograph parameters vary substantially within a micrograph:

$$M_i^2(s) = \frac{1}{2\pi} \sum_l \int_\theta \left| \overline{M}_l(s,\theta) \right|^2 \quad (6)$$

The l subscript is a sum of all particles within a micrograph regardless of orientation and the i subscript discriminates between different micrographs. $M_i^2(s)$ approximates the one dimensional power spectrum of the particles in a micrograph. There is an arbitrary scaling factor associated with the final solution, which allows us to drop the normalization constant at the front of this expression. We thus obtain

$M_i^2(s) = C_i^2(s)E_i^2(s)F^2(s) + N_i^2(s)$, where

$$F^2(s) = \int_\Psi \int_\theta \left| \overline{F}_\Psi(s,\theta) \right|^2 \text{ and } N^2(s) = \int_\theta \left| \overline{N}(s,\theta) \right|^2 \quad (7)$$

This expression does not contain a vector cross term between F and N since N is incoherent with respect to F. This makes the cross term the average of a random vector, which goes to zero with sufficient averaging. Techniques for obtaining the eight parameter values using eqn. 6 are described elsewhere ([3, 4, 7]). This expression is used solely for determining the micrograph parameters, and F in this case is an average over all 3D particle orientations.

Assuming we have obtained accurate parameter values for *C(s), E(s)* and *N(s)* for each micrograph, we need to determine how to best reconstruct $\overline{F}_\Psi(s,\theta)$ using particles in the same orientation from several micrographs. We will look for the best solution in a least-squares sense. In the following derivation, we determine $\overline{A}_\Psi(s,\theta)$, the final average of particles in orientation Ψ, which should approach $\overline{F}_\Psi(s,\theta)$. To perform a 3D reconstruction, this process is repeated for all orientations. $\overline{A}_\Psi(s,\theta)$ will be represented as a weighted average in Fourier space of the individual particle images. That is:

$$\overline{A}_\Psi(s,\theta) = \sum_n k_n(s) \overline{M}_{\Psi,n}(s,\theta) \quad (8)$$

where $k_n(s)$ are the particle weights. We wish to obtain an $\overline{A}_\Psi(s,\theta)$ with the highest possible signal to noise ratio. The absolute signal to noise ratio of an individual particle (using the micrograph average parameters for each particle) is:

$$\frac{F_\Psi^2(s)C_n^2(s)E_n^2(s)}{N_n^2(s)} \quad (9)$$

The absolute signal to noise ratio of $\overline{A}_\Psi(s,\theta)$ is then simply:

$$R_A(s) = F_\Psi^2(s) \frac{(\sum_n k_n(s)C_n(s)E_n(s))^2}{\sum_n k_n^2(s)N_n^2(s)} \quad (10)$$

We wish to maximize R_A, giving n constraints for all s:

$$\frac{\partial R_A(s)}{\partial k_n(s)} = 0 \quad (11)$$

In addition, we must provide a normalization constraint to define the overall filtering of the data. We have assumed that the data in each image, $\overline{M}_{\Psi,n}(s,\theta)$, is normalized such that *F(s)* has the same amplitude in each image. For the moment, we wish to maintain this normalization:

$$\sum_n k_n(s)C_n(s)E_n(s) = 1 \quad (12)$$

Solving eqn 11,12 simultaneously gives:

$$k_n(s) = \frac{\frac{F_\Psi^2(s)C_n(s)E_n(s)}{N_n^2(s)}}{\sum_j \frac{F_\Psi^2(s)C_j^2(s)E_j^2(s)}{N_j^2(s)}} \quad (13)$$

This can be rewritten a bit more transparently as:

$$k_n(s) = \frac{1}{C_n(s)E_n(s)} \frac{R_n(s)}{\sum_j R_j(s)}, \text{ where} \quad (14)$$

$$R_n(s) = \frac{C_n^2(s)E_n^2(s)}{N_n^2(s)},$$

which is simply the relative spectral signal to noise ratio of the image. That is, the individual micrographs are weighted by their relative signal to noise ratios at each spatial frequency. The division by *C(s)E(s)* performs the actual deconvolution. The weighting term prevents infinities at the zeros of the CTF except in cases where there is a coincident zero in all data sets. It is therefore critical that several micrographs with different defocuses are used. Note also that $F_\Psi^2(s)$, the 1D structure factor of the projection, is not required to determine $k_n(s)$.

The normalization imposed in eqn. 12 produces an $\overline{A}_\Psi(s,\theta)$ with a signal component approaching $\overline{F}_\Psi(s,\theta)$, but it ignores the presence of noise, which will be substantially upweighted. For interpretation of the results we wish $\overline{A}_\Psi(s,\theta)$ to match $\overline{F}_\Psi(s,\theta)$ as closely as possible in the least-squares sense in real space. That is, we wish to adjust the normalization such that:

$$\sum_{x,y} (a_\Psi(x,y) - f_\Psi(x,y))^2 \quad (15)$$

is minimized, where $a_\Psi(x,y)$ and $f_\Psi(x,y)$ are the inverse Fourier transforms of $\overline{A}_\Psi(s,\theta)$ and $\overline{F}_\Psi(s,\theta)$ respectively.

The solution to this well-known problem is simply to include a Wiener filter [8]:

$$\frac{1}{1+\frac{1}{SNR(s)}} \tag{16}$$

where *SNR(s)* is the spectral signal to noise ratio of $\overline{A}_\Psi(s,\theta)$. Obtaining a reliable *SNR(s)* is the main difficulty involved in applying Wiener filters in most applications. However, in this case, an accurate estimate of the absolute *SNR(s)* is already available from the derivation above. Eqn. 10 reduces to:

$$R_A(s) = F_\Psi{}^2(s)\sum_n \frac{C_n{}^2(s)E_n{}^2(s)}{N_n{}^2(s)}. \tag{17}$$

$F_\Psi(s)$, or an adequate approximation to it can be obtained using a variety of methods, including x-ray solution scattering experiments or combination of low resolution EM data with representative PDB data.

This final filtration process is completely separable from the deconvolution/weighting process, and can be incorporated into a single overall correction equation:

$$\overline{A}(s,\theta) = \frac{1}{1+\frac{1}{R_A(s)}}\cdot\sum_n k_n(s)\overline{M}_n(s,\theta), \text{ where}$$

$$k_n(s) = \frac{1}{C_n(s)E_n(s)}\cdot\frac{R_n(s)}{\sum_j R_j(s)} \tag{18}$$

This provides an optimal solution for the CTF correction problem in a least-squares sense. It can be fairly trivially extended to include astigmatism and drift. However, presence of these artifacts may also lead to errors in 2D alignment and nonisotropic resolutions, so such data should be avoided whenever possible.

4. CONCLUSIONS

We have presented a mathematical model for optimal correction of TEM artifacts when performing 2D averaging. This methodology is incorporated into the EMAN [3, 7] single particle reconstruction software suite, which provides nearly automated correction of these artifacts in the context of a 3D single particle reconstruction.

If particles treated with this methodology in the context of producing a 3D model, the additional averaging should be taken into account in the Wiener filter, otherwise the 3D model will be moderately overfiltered. The overfiltration is mitigated by the fact that the most averaging occurs at low spatial frequencies where the signal to noise ratio is already high, and hence the Wiener filter is effectively 1.0. Relatively little averaging occurs at higher spatial frequencies when using typical reconstruction parameters.

This research is supported by NCRR (RR02250) and the Agouron Foundation.

5. REFERENCES

[1] H. P. Erickson and A. Klug, "The Fourier transform of an electron micrograph: effects of defocussing and aberrations, and implications for the use of underfocus contrast enhancement," *Phil. Trans. Roy. Soc. Lond. B*, vol. 261, pp. 105-18, 1970.

[2] K. J. Hanszen, "The optical transfer theory of the electron microscope: fundamental principles and applications," in *Advances in optical and electron microscopy*, vol. 4, R. Barer and V. E. Cosslett, Eds. N.Y.: Academic Press, 1971, pp. 1-84.

[3] S. J. Ludtke, J. Jakana, J. L. Song, D. T. Chuang, and W. Chiu, "A 11.5 Å single particle reconstruction of GroEL using EMAN," *J Mol Biol*, vol. 314, pp. 253-62, 2001.

[4] A. Saad, S. J. Ludtke, J. Jakana, F. J. Rixon, H. Tsuruta, and W. Chiu, "Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination," *J Struct Biol*, vol. 133, pp. 32-42, 2001.

[5] J. Zhu, P. A. Penczek, R. Schroder, and J. Frank, "Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome," *J Struct Biol*, vol. 118, pp. 197-219, 1997.

[6] J. F. Conway and A. C. Steven, "Methods for reconstructing density maps of "single" particles from cryoelectron micrographs to subnanometer resolution," *J Struct Biol*, vol. 128, pp. 106-18, 1999.

[7] S. J. Ludtke, P. R. Baldwin, and W. Chiu, "EMAN: semiautomated software for high-resolution single-particle reconstructions," *J Struct Biol*, vol. 128, pp. 82-97, 1999.

[8] W. H. Press, W. T. Vetterling, S. A. Teukolsky, and B. P. Flannery, in *Numerical Recipes in C*, 2nd ed: Press Syndicate of the University of Cambridge, 1992, pp. 547-549.