3D macromolecule structure reconstruction from electron micrograph by exploiting symmetry and sparsity

Min Woo Kim, Jiyoung Choi,and Jong Chul Ye

Bio Imaging & Signal Processing Lab., Dept. of Bio& Brain Engineering
Korea Advanced Institute of Science and Technology (KAIST)
373-1 Guseong-dong Yuseong-gu, Daejon 305-701, Korea

ABSTRACT

Single particle reconstruction is often employed for 3-D reconstruction of diverse macromolecules. However, the algorithm requires a good initial guess from *a priori* information to guarantee the convergence to the correct solution. This paper describes a novel *model free* 3-D reconstruction algorithm by employing the symmetry and sparsity of unknown structure. Especially, we develop an accurate and fully automatic iterative algorithm for 3D reconstruction of unknown helix structures. Because the macromolecule structure assumes only sparse supports in real space and the helical symmetry provides several symmetric views from a single micrograph, a reasonably quality 3-D reconstruction can be obtained from the limited views using the compressed sensing theory. Furthermore, the correct helix parameters usually provide the maximal variance of the reconstructed volume, facilitating the parameter estimation. Remarkably, the search space of helix parameter can be drastically reduced by exploiting the diffraction pattern. With the estimated helix parameter and additional 3-D registration, the multiple helix segments can be combined for the optimal quality reconstruction. Experimental results using synthetic and real helix data confirm that our algorithm provides superior reconstruction of 3-D helical structure.

Keywords: 3D macromolecule reconstruction, helical reconstruction, symmetry, sparsity, compressed sensing theory

1. INTRODUCTION

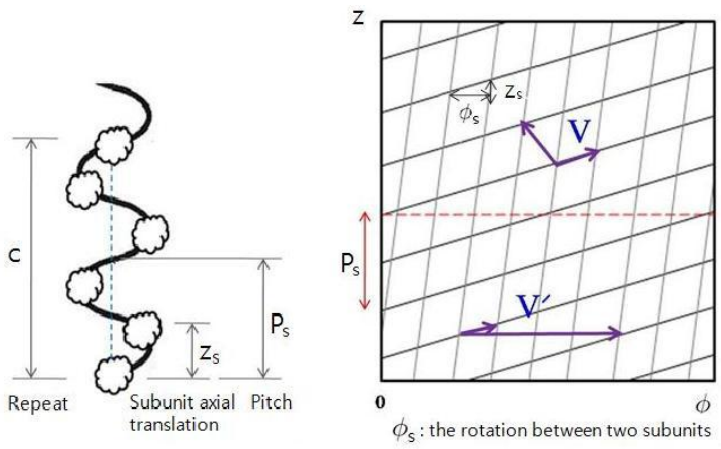
Many biological macromolecules are naturally assembled into helical polymers. This includes DNA, macrotubule, F-actin, filamentous viruses, cytoskeletal and muscle filaments, bacterial flagella, etc. The basic building block for these helical structures are the subunit that are spaced periodically building helical symmetric geometry. The most widely used reconstruction method for these helical structures is the so-called Fourier-Bessel transform approach that exploits the diffraction pattern in reciprocal space.[1] More specifically, the Fourier transform of a view of a helical filament consists of a discrete series of layer lines. The coordinate of the maximum intensity along the layer line conveys the useful information about the helical symmetry of the objects and the structure of its subunits. Consequently, a single view may be sufficient for 3-D reconstruction within a certain resolution limit.[2]

However, the Fourier-Bessel transform approach has its limitations when the underlying assumptions do not hold. For examples, flexible and disordered specimens,[3] and helices with variable pitch, seams,[4] or Bessel overlaps.[5] For these type of specimens, real space approaches based on single particle reconstruction techniques have been investigated as an alternative.[5] Specifically, short segment of filament are aligned with the reference projections and then back-projected to obtain 3-D reconstruction. In order to impose the helical symmetry, helix parameters such as subunit axial translation and azimuthal angle between subunits are estimated from the reconstructed 3-D volume rather than the diffraction pattern. Then, the entire processes are repeated until convergence. The main drawback of the single particle reconstruction is, however, that it does not guarantee the correct reconstruction unless a good initial estimate of unknown particle structure is available *a priori*.

The main contribution of this paper is a unified helical reconstruction theory that does not require any prior model, but encompasses the conventional Fourier-Bessel approach and the single particle reconstruction methods

Further author information: (Send correspondence to Prof. Jong Chul Ye.)
Jong Chul Ye: E-mail: jong.ye@kaist.ac.kr, Telephone: 82-42-869-4320

Computational Imaging VI, edited by Charles A. Bouman, Eric L. Miller, Ilya Pollak,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6814, 68140X, © 2008 SPIE-IS&T · 0277-786X/08/$18



(a) (b)

Figure 1. Parameters of a helical structure.

as special cases. The main idea of such breakthrough is to exploit the symmetry and the sparsity of the helical structure in real and reciprocal spaces. Recall that helical symmetry provides us specific number of symmetric views depending on the ratio between the repeat and the subunit axial translation. Therefore, as long as the helix parameters are correctly estimated, a reasonably good 3-D structure can be obtained even from a single projection. Furthermore, the reconstruction quality can be significantly improved by exploiting the sparsity of the helical structure in real space. Additional three dimensional registration between the reconstructed volumes may provide the sufficient number of views for the best resolution 3-D reconstruction.

The most important step in the proposed algorithm is estimating the helix parameters. As discussed before, analyzing the diffraction pattern in reciprocal space[1] is often not enough to completely estimate the helical parameters when the micrographs have high noise, or helices have bended structure or Bessel order overlap. To overcome the drawback, we employed the variance based parameter estimation technique called SPR (Single Projection Reconstruction)[6,7] to accurately estimate the unknown symmetry parameters. Generally, the variance of reconstructed volume with correct symmetry is higher than the others, because the reconstruction with the incorrect parameters become blurred and noisy. Therefore, the optimal helix parameters can be easily obtained by comparing variance of multiple reconstruction with all feasible helix parameters. However, the main technical hurdle for the parameter estimation using SPR is the huge computational overhead since the feasible parameter space is very large and often *uncountable*. This paper shows that exploiting the diffraction pattern reduces the computational burden drastically and the resultant search space is often *finite*.

2. HELICALLY SYMMETRIC DENSITY MODEL

Suppose that the molecules making up the microtubule are long slender objects that have a periodic electron density with the repeat c in the long direction as shown in Fig. 1(a). Because $\rho(r,\phi,z)$ is periodic in ϕ with period 2π and periodic in z with the repeat c, it follows that $\rho(r,\phi,z)$ can be expanded as Fourier series expansion with respect to ϕ and z:

$$\rho(r,\phi,z) \quad = \quad \sum_l \sum_n g_{n,l}(r) \exp\left[j2\pi\left(n\frac{\phi}{2\pi} - \frac{l}{c}z\right)\right] = \sum_l \sum_n g_{n,l}(r)\exp(-j2\pi\mathbf{k}\cdot\mathbf{t}) \tag{1}$$

where vectors $\mathbf{k} = [-n, l/c]^T$ and $\mathbf{t} = [\frac{\phi}{2\pi}, z]^T$ for $-\infty < l, n < \infty$. Note that we use $-l$ rather than l in Eq. (1) due to the notation of the classical Fourier-Bessel transform theory. In the simplified helix model shown in Fig. 1(a), the symmetry of a helical structure can be defined in terms of a number of helix parameters, such as the subunit axial translation (z_s), the pitch of the helix (P_s), the repeat of the helix (c) if the number of subunits in a pitch is not an integer.

More complete picture of helical symmetry can be described by the lattice model as shown in Fig. 1(b). The basis matrix $\mathbf{V}$ for the lattice $\mathbf{\Lambda}$ is not unique, and the upper-triangular form of the basis matrix is most useful:

$$\mathbf{V} = \begin{pmatrix} 1/Q & z_s/P_s \\ 0 & z_s \end{pmatrix} . \tag{2}$$

where Q is a non-zero integer. Under the lattice model, the helical symmetric density model is defined by

$$\rho(r, \phi, z) \quad = \quad \rho\left(r, \mathbf{t} + \mathbf{V}\mathbf{m}\right) \tag{3}$$

where $\mathbf{m}$ denotes any integer vector. The helical symmetry condition Eq. (3) then tells us

$$\mathbf{V}^T\mathbf{k} = \mathbf{p} \tag{4}$$

for any integer vectors $\mathbf{p} = [h, k]^T$. This implies

$$\frac{l}{c} = \frac{n}{P_s} + \frac{k}{z_s} \tag{5}$$

By multiplying with the repeat c, we arrive at the so called *selection rule*:[8]

$$l = \frac{n}{P} + \frac{k}{a} = tn + uk \tag{6}$$

where $P = P_s/c$ is the normalized pitch, $a = z_s/c$ is the normalized axial translation, and t denotes the number of turns of the basic helix per repeat, and u denotes the number of subunits per repeat, respectively.[8]

3. PROPOSED ALGORITHM

3.1 Reconstruction Flowchart

The overall reconstruction flowchart of our algorithm is illustrated in Fig. 2. The most important step of our model free reconstruction method is the estimation of optimal helix parameters by exploiting the variance of the single projection reconstruction and diffraction pattern in reciprocal space. More specifically, from a single or multiple micrographs, short segments are obtained by cutting helix structure. If a long and straight helix is available, the whole segment can be also used. Like the single particle approach, it is possible to use the flexible helix samples by cutting them in short segments. For each segment projection, three dimensional structure is obtained by exploiting the helix symmetry. Specifically, the number of identical views from the helical symmetry is given by the number of subunits per repeat $u = c/z_s$ in Eq. (6), where c is the repeat and z_s is the subunit axial translation. The main idea of the SPR (Single Projection Reconstruction)[6,7] is the observation that when an helix is reconstructed with incorrect helix parameters, the result model tends to be fuzzy and blurred. However, the detailed structures can be generated with the correct parameters. This implies that the optimal helix parameter can be easily obtained by comparing variance of SPRs with all available helix parameter pair (ϕ_s, z_s), where ϕ_s is the rotation between the two subunits and z_s is the axial translation between the two subunits, respectively. However, the multiple number of identical views from the helical symmetry is not enough to fill in the whole radon volume. In this case, the direct use of inverse radon transform incurs severe aliasing artifacts which hinders estimation of symmetric parameter from the SPRs. Therefore, we employ the compressed sensing theory[9] to reduce the artifacts and overcome the resolution loss. More detailed description about the compressed sensing will be explained in the next section.

One of the main technical difficulties in the variance based parameter estimation is its huge computational complexity since we need to calculate the three dimensional reconstruction for all feasible pairs of real numbers (ϕ_s, z_s). However, the computational burden can be dramatically reduced by exploiting the diffraction pattern. Recall that helix parameter pairs can be represented as following:

$$(\phi_s, z_s) = \left\{ \left(\frac{2\pi t}{u}, \frac{c}{u} \right) : c \in \mathbb{R}^+, t \in \mathbb{N}^+, u \in \mathbb{N}^+ \right\} \tag{7}$$

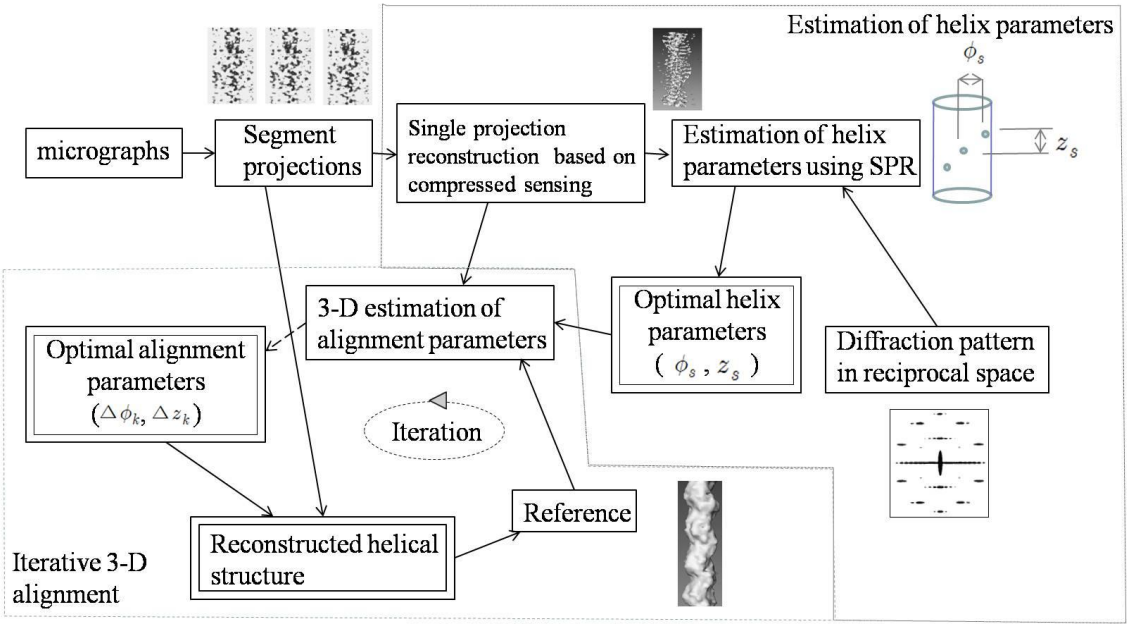


Figure 2. Flowchart of the proposed algorithm.

where $\mathbb{R}^+$ and $\mathbb{N}^+$ denote the space of positive real and integers; c,t and u denote the repeat, the number of turns of the basic helix per repeat, and the number of subunits per repeat, respectively, as described in Eq. (6). Unlike the conventional Fourier-Bessel transform approaches that estimate the (n, l) index in Eq. (6) from the diffraction patterns, our algorithm tries to estimate parameter triple (t, u, c). Interestingly, the search for the (t, u, c) triple turns out simpler than search for (n, l) pairs since the feasible (t, u, c) triples can be significantly reduced by exploiting diffraction patterns.

Fig. 3(a) shows the general diffraction pattern in reciprocal space. The repeat c and the subunit axial transition z_s can be calculated by the distance between the layer lines and the cross pattern, respectively. Then, Eq. (7) tell us that the only unknown helix parameter is t, the number of turns of the basic helix per repeat. Hence, the search space is one dimensional positive integer, which significantly reduces the computational burden for parameter estimation. In practise, due to the noise and limited resolution, the diffraction pattern often follows Figs. 3(b) or (c). More specifically, Fig. 3(b) is the diffraction pattern when the repeat c is bigger than the projection image field-of-view. In this case, the subunit axial transition z_s can be obtained from the diffraction pattern. Eq. (7) then tells us that the unknown helix parameters to be estimated are two positive integers (t, u). Similarly, Fig. 3(c) corresponds to the diffraction pattern when the subunit axial translation z_s is too small compared to the repeat. In this case, the unknown helix parameters are also the two positive integers (t, u). Therefore, the parameter search space is the two dimensional positive integers, which dramatically reduces the complexity of the parameter estimation procedure. Furthermore, by noting that the multiple segments from the same helix assumes the same helix parameter, the (t, u) parameter estimation can be made robust even under very low signal-to-noise ratio.

After obtaining the helix parameters, the next step is the alignment between the multiple segment to provide best three dimensional reconstruction as illustrated in Fig. 2. This is because that each single projection reconstruction quality is limited by the number of subunit per repeat; so aligning the multiple segment produces more views for better reconstruction. More specifically, let $\rho_k(r, \phi, z)$ denotes the k-th truncated helix segments. We can easily expect that each segment are rotated and translated from the reference helix. Let $(\Delta\phi_k, \Delta z_k)$ denote the unknown rotation and the translation parameters. Then, we have

$$\rho_k(r, \phi, z) = \rho_{Ref}(r, \phi + \Delta\phi_k, z + \Delta z_k) \quad , \quad k = 1, \cdots, K \ . \tag{8}$$

where ρ_{Ref} denotes the reference helix and K denotes the total number of helix segments. The estimation

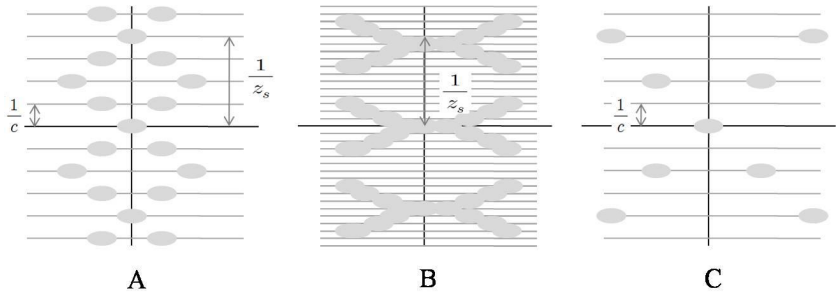


Figure 3. Diffraction pattern in reciprocal space in A : General case. B : The repeat is longer than the FOV. C : The axial transition is too short.

of the alignment parameter $(\Delta\phi_k, \Delta z_k)$ can be done in either two dimensional projection domain or three dimensional reconstruction domain. However, we found that the three dimensional alignment is more robust since the compressed sensing provides the reasonable three dimensional reconstruction and the three dimensional alignment minimizes the ambiguity. The choice of ρ_{Ref} can be updated iteratively for the best reconstruction performance. More specifically, at the first iteration, one single projection reconstruction is used as a reference to provide the first iteration reconstruction. Then, the reconstructed volume from the first iteration is used as a reference for the next iteration. This procedure continues until convergence.

3.2 Super-Resolution Reconstruction from Limited Views

Recently, there has been great interest in super-resolution reconstruction techniques from highly undersampled projection view data especially in the magnetic resonance imaging (MR) and x-ray computed tomography (x-ray CT) community.[10,11] Mathematically, the reconstruction problem from sparse k-space samples is an ill-posed inverse problem with infinitely many solutions; therefore, some kind of regularization or constraint is necessary to obtain a unique and stable reconstruction. Sparsity is a very promising constraint for reconstructing helical structures, because helical structures usually assume sparse support in real space, and the sparsity constraint does not usually sacrifice the image resolution. Assuming the image can be sparsified, the recent theory of so-called "compressed sensing"[9,12] in the signal processing community shows that perfect reconstruction is possible even from samples dramatically smaller than the Nyquist sampling limit. Even for the images not completely sparse, the compressed sensing theory tells us that significant features of images can be still obtained. Furthermore, the basis pursuit, matching pursuit methods, or the convex L_1 optimization method can be used to obtain sparse solutions.[9,12] Therefore, we employ the compressed sensing technique for the single projection reconstruction as well as multiple projection reconstruction.

4. EXPERIMENTAL RESULTS

4.1 Synthetic Microtubule

Before dealing with real data, we tested our algorithm using 3-D synthetic microtubules of rat kinesin motor obtained from European Bioinformatics Institute (EMBL-EBI). Synthetic micrograph images are made by projecting the synthetic microtubule along several rotation angles and axial transitions. One of the synthetic projections and the diffraction pattern are illustrated in Figs. 4(a) and (b), respectively. From the layer line pattern in Fig. 4(b), we can get the repeat c. Therefore, the unknown helix parameter to estimate is (t, u). Fig. 5 shows the variance of the single projection reconstruction with various pairs of (t, u) indices. We can confirm that at the optimal helix parameters (t, v), the variance was the largest. The estimation provides that the number of turns of the basic helix per repeat was one, and the number of subunit per repeat was 15. Figs. 6(a) and (b) denote the three dimensional structure of the original synthetic model and one of the single projection reconstruction, respectively. Thanks to the helical symmetry, the single projection reconstruction result still convey the most of the helix morphology. Fig. 6(c) illustrated the final reconstruction structure using our algorithm when the fifty short helix segments are available. We can clearly see that very accurate structures are obtained by our new method.

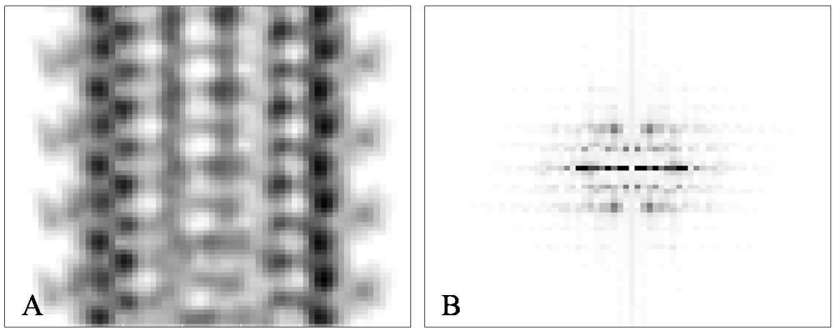


Figure 4. A : Synthetic microtubule projection image. B : The corresponding diffraction pattern.

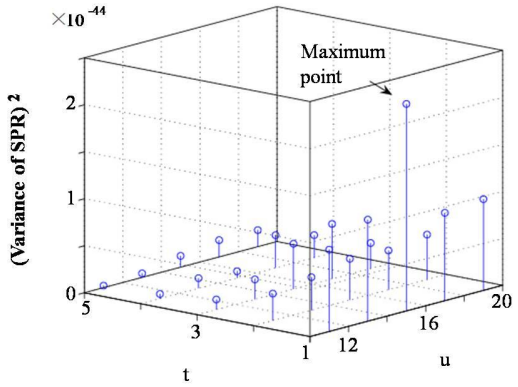


Figure 5. SPR variance along (t, u) where t is the number of turns in one repeat and u is the number of subunit per repeat.

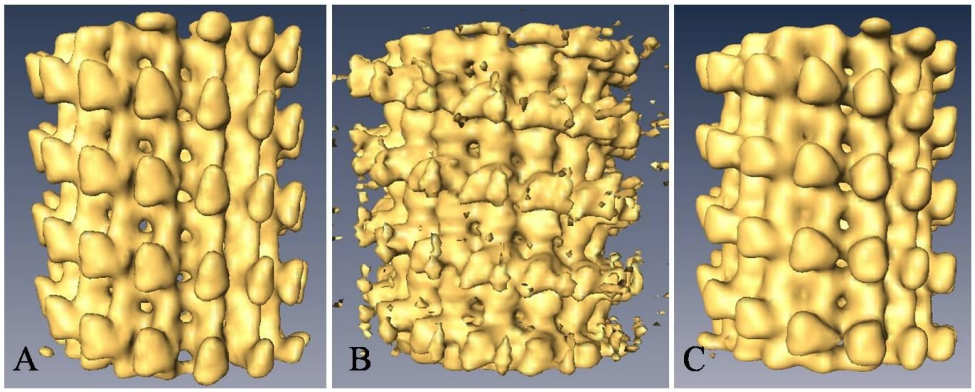


Figure 6. A : The synthetic microtubule structure. B : Single projection reconstruction. C : Final reconstruction structure using fifty short segments projection images.

4.2 Real F-actin data

We applied our algorithm to F-actin reconstruction from real data sample. Because F-actin[13]has different symmetric values depending on types (for example, muscle actin or yeast actin), it is suitable to test our algorithm. One of the micrographs and the average diffraction pattern are illustrated in Fig. 7(a) and (b), respectively. The image size is 100 by 100 pixels. As shown Fig. 7(b), the diffraction pattern is highly distorted due to noisy projection image. Ambiguity of layer line denotes that the repeat c is beyond the projection image FOV. However, the cross shape is repeated instead of layer line as described in Fig. 3(c). From the distance between the cross shapes, we can directly obtain z_s, the axial transition between subunits. As shown Eq. (7), we again need to find the best positive integer pairs (t, u) by comparing variance of the single projection

reconstruction. The estimate parameters was $(t, u) = (1, 13)$. After the parameter estimation, our algorithm apply the iterative three dimensional alignment to search the $(\Delta\phi_k, \Delta z_k)$ parameters using 400 real projections. Fig. 8 (a) and (b) show the three dimensional reconstruction from different start reference. We can show that the two structures are similar even with difference start references. Furthermore, the reconstructions coincide with the known three dimensional structures of F-actin in Fig. 8 (c).

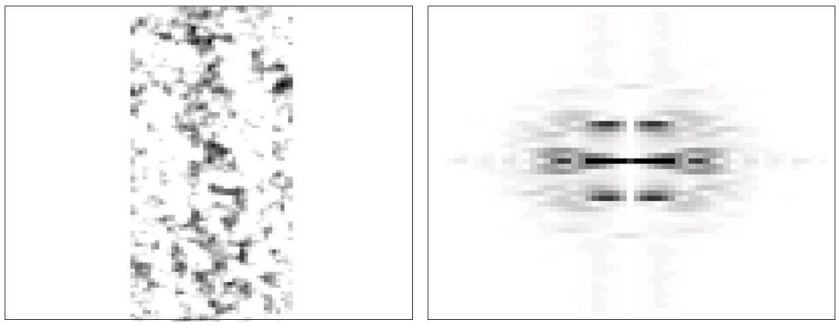


Figure 7. A : A real F-actin projection measurement. B : The average diffraction pattern.

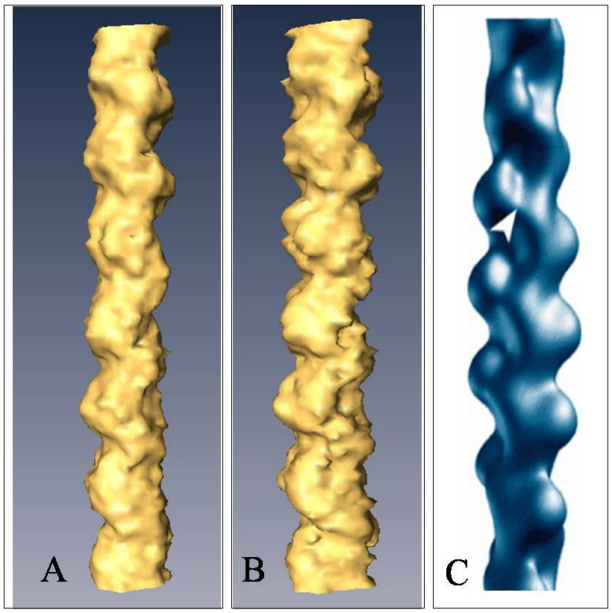


Figure 8. A and B : Reconstructed F-actin structure using our algorithm from two distinct start references. C : F-actin excerpted from (http://www.scripps.edu/ stoffler/publ/PDF/MRT99/MRTReview99.html)

5. CONCLUSION

This paper described a novel model free iterative reconstruction algorithm for helical structure by exploiting the helical symmetry and the sparsity of the sample in the real space. We showed that the helix parameters can be estimated by searching the maximum variance of the single projections for all feasible pairs of the helix parameters. Thanks to the diffraction analysis, we showed that the parameter search space can be reduced to a space of one or two dimensional positive integers, which drastically reduces the estimation complexity. We described the compressed sensing approach to improve the accuracy of the parameter estimation as well as the final reconstruction. An iterative 3-D alignment procedure was also described. Experiment results using synthetic and real helix data confirms that our algorithm provides superior reconstruction of 3D helical structure.

REFERENCES

1. D. J. DeRosier and A. Klug, "Reconstruction of three dimensional structures from electron micrographs," *Nature*, vol. 217, pp. 130–134, 1968.
2. D. J. Derosier and P. B. Moore, "Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry," *J. Mol. Biol.*, vol. 52, no. 2, pp. 355–369, 1970.
3. Andrew J. Pomfret, William J. Rice, and Divid L. Stokes, "Application of the iterative helical real-space reconstruction method to large membranous tubular crytals of p-type atpases," *Jour. of Structural Biology*, vol. 157, pp. 106–116, 2007.
4. F. Metoz and R. H. Wade, "Diffraction by helical structures with seams: Microtubules," *Jour. Structural Biology*, vol. 118, no. 2, pp. 128–139, 1997.
5. E. H. Egelman, "A robust algorithm for the reconstruction of helical filaments using single-particle methods," *Ultramicroscopy*, vol. 85, no. 4, pp. 225–234, 2000.
6. Francesca Cantele, Salvatore Lanzavecchia, and Pier Luigi Bellon, "The variance of icosahedral virus models is a key indicator in the structure determination: a model-free reconstruction of viruses, suitable for refractory particles," *Jour. of Structural Biology*, vol. 141, pp. 84–92, 2003.
7. Sjors H. W. Scheres, Roberto Marabini, and Salvatore Lanzavecchia, "Classification of single-projection reconstructions for cryo-electron microscopy data of icosahedral viruses," *Jour. of Structural Biology*, vol. 151, pp. 79–91, 2005.
8. Sarah J. Daniell, Eva Kocsis, Edward Morris, Stuart Knuttop, and Frank P, "Molecular microbiology," *Molecular Microbiology*, vol. 49, pp. 301–308, 2003.
9. D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
10. J. C. Ye, S. Tak, Y. Han, and H. W. Park, "Projection reconstruction MR imaging using FOCUSS," *Magn. Reson. Med.*, vol. 57, pp. 764–775, 2007.
11. M. Lusting, D. Donoho, and J. M. Pauly, "Sparse MRI : The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, 2007.
12. E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Info. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
13. Kiran Pant, Divid Chereau, and Victoria Hatch, "Cortactin binding to F-actin revealed by electron microscopy and 3d reconstruction," *Jour. of Structural Biology*, vol. 359, pp. 840–847, 2006.