Challenge: Image Restoration/Superresolution for Single Particle Analysis

Yen-Chen Wu[1], Kuan-Hao Huang[2], Kai-Wen Liang[1], Shao-Hua Sun[1], Ming-Jen Yang[1], Po-Wen Hsiao[1], Guan-lin Chao[1], Ti-Fen Pan[1], Yi-Ching Chiu[1] Wei-Chih Tu[3], Shao-Yi Chien[3]

[1]Department of Electrical Engineering, National Taiwan University
[2]Department of Computer Science and Information Engineering, National Taiwan University
[3]Graduate Institute of Electronics Engineering, National Taiwan University

***Abstract*—All intrinsic deformation, microscope aberration, and sample preparation collude to severely limit the resolution of electron microscopy (EM) structures in the past. However, both the development of image processing techniques and the progress of the manufacturing technology of electron microscopes result in significant improvement in the resolution these years. In this paper, from the perspective of image processing, we tackle the low resolution problem of EM imaging results by exploiting the available information in Protein Data Bank (PDB) and ingeniously convert the arduous super-resolution problem into intuitive nearest neighbor (NN) problem. Moreover, a Beltrami flow-based approach is performed beforehand for image restoration and assessing the bounding of noisy protein structure. Experimental on different signal-to-noise ratio subsets of noisy structure verified the effectiveness of our proposed method.**

The upload username of our team is *ej0cl6*, and our submission ID are *1462*.

Keywords*—*Electron microscopy, image processing, ICASSP, signal processing cup

I. Introduction

In recent years, discovering the structures of macromolecular complexes, such as proteins, DNA, and carbohydrates, has become an important issue. The study of the macromolecular complex structures can lead to numerous applications in medicine. With the understanding of molecular structures of some biomedical complexes, we can further investigate specific biochemical functions which are carried out by these complexes. For instance, a new drug can be developed to stop some biochemical functions by blocking out certain proteins.

Recently, X-ray crystallography and Nuclear Magnetic Resonance (NMR) are the most common techniques to obtain high-resolution macromolecular complexes' structures [1]. Standard resolution for X-ray and NMR ranges between 1.5 and 3 Å. However, some macromolecular complexes are not amenable to X-ray and NMR, for not all macromolecular complexes can crystallize, which is needed by X-ray crystallography, and that NMR are not able to resolve large macromolecular complexes. Furthermore, these two methods require a relatively high macromolecular concentration, which is not available in all conditions. On the other hand, Electron Microscopy (EM) is free from the constraints mentioned above [2], being able to handle large complexes without the requirement of high specimen concentrations. Even though EM suffers from relatively low resolution (4∼20 Å) as a cost, this problem becomes less fatal compared to the immense benefit EM brings. Furthermore, EM gains higher resolution these years due to the advancement on manufacturing technology and thriving research studies on image processing methods [3], [4]. There is a trend that EM is becoming a more and more prevalent measure in the study of macromolecular complexes' structures.

In this paper, we propose an image processing framework to achieve higher resolution in structures produced by EM projection. The key observation to our method is that the high resolution information of manifold types of protein is avaliable in the protein data bank (PDB) [5]. With the aid of the clear protein structure selected from PDB, we can restore the degraded structure. A denoising procedure is applied beforehand, which helps reduce the impact of outliers and gives a precise bounding region of the target. The illustrator of our porposed framework is shown in Fig. 1

The remainder of this paper is organized as follows. First, we describe the Beltrami flow-based method for addressing the EM images restoration problem and present the exemplar-based strategy for resolving low-resolution EM images in Sec. II. In Sec. III, we present how we design the experiments to evaluate the performance of our proposed framework, and conclusions are drawn in Sec. IV.

II. Proposed Framework

A. TOMOBFLOW

The causes of resolution limitation in electronic microscopy (EM) include intrinsic deformation, microscope aberration, and sample preparation. The non-ideal effect of microscopic projection can be simulated by addition of gaussian noise, contrast transfer functions (CTFs) and random projections to the noise-free protein structures [6]. Hence, the state-of-the-art denoising techniques, such as bilateral filter [7], BM3D [8], K-SVD [9] which can only apply to additive gaussian noise are not sufficient to deal with unpredictability and irregularity of the noise pattern of EM images.

TOMOBFLOW [10] is a noise filtering approach based on Beltrami flow with the capability of preserving relevant

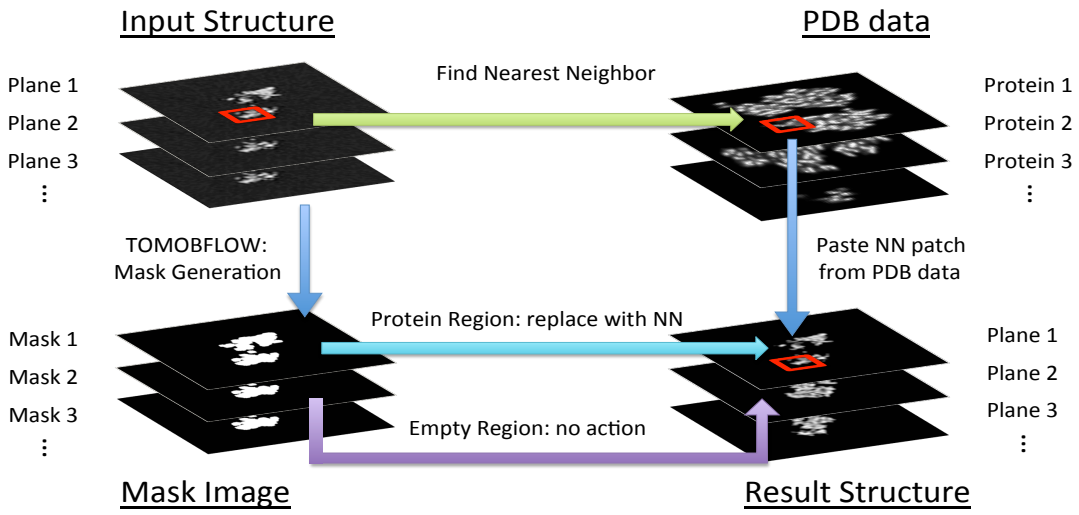

Fig. 1: Illustration of our proposed framework. Note that *Input Structure* represents the test or challenge data, *PDB data* represent the protein obtained from Protein Data bank, *Mask Image* are generated by TOMOBFLOW method and *Result Structure* is the noise-free and high resolution structure obtained by the framework.

biological information and thus eliminating the blurring of features in images. Beltrami flow is an edge-preserving approach which measures geodesic distance and gives the knowledge about underlying geometry between points. Therefore, TOMOBFLOW is able to minimize the area of the image and drive the flow to a minimal surface solution, and thus a good preprocessing algorithm in our framework in terms of complexity and robustness.

TOMOBFLOW can effectively eliminate noise based on the fact that the pattern of noise is rather unpredicable compared to that of proteins, with most of the protein structures being preserved. The result protein structures are therefore noise-free in the background with a potential loss of foreground details.

B. Nearest Neighbor

In order to further enhance the quality of imaging of the noisy structure, we introduce an exampler-based method due to the high degree of similarity within protein structures. It is applicable to exploit exterior protein images as prior information to enhance the resolution of the low-resolution input data. Therefore, several instances of high-resolution protein images from the PDB are selected as prior knowledge.

The data from the PDB are in the form of .pdb file, so we first have to convert the data into three-dimensional images using a file-converting function in the Xmipp program, which is a Microscopy Image Processing Package. Next, we treat the three-dimensional images as a series of two-dimensional planes, and divide the planes into a huge number of overlapping patches. All these patches form a database containing noise-free and high-resolution data, which we shall refer to as the prior database. Similarly, we can regard the three-dimensional input as a series of two dimensional planes formed by patches. For each patch in the input data, we search in the prior database for its "nearest neighbor," that is, the most similar patch to it. After this matching process, each patch in the input data will have a corresponding patch in the prior database. Owing to the fact that the data in the prior database are noise-free and at higher resolution than those in the input data, we can simply replace the patch in the input data with its corresponding patch, and derive images that look similar the original input but consist only of the patches from the prior database. The result can be acquired by assembling each enhanced resolution planes into three-dimensional volumes and converting the data into .mrc files.

Furthermore, based on the observation that in the data from the PDB, the voxels outside the protein structures have a gray-scale value of 0. Therefore, we adopt the *mask* produced by TOMOBFLOW, which is introduced in section II-A, to estimate the location of the protein in the noisy input data. A Boolean number is assigned to each voxel to indicate whether or not it contains protein information. In the nearest neighbor matching process, all the patches that contain no protein information are ignored and assigned 0. As a result, the noise in these regions is suppressed, and no further computation is needed. This adjustment not only enhances the quality of the result, but also significantly accelerates the image processing.

III. Experiments

A. Experimental Settings

The test and challenge data provided in SP Cup were four protein structures selected from Protein Data Bank. The four noise-free atomic structures were converted to voxel gray densities, randomly projected from 50,000 orientations, applied to artificial noise with 0.1, 0.2 and 0.4 Signal-to-Noise Ratio, and then superimposed by 34 different contrast transfer functions, as simulation of nonideal electron microscopic effect on protein structure images. In addition to low SNR, multiple highly-variant CTFs, the loss resolution also lies in low-pass filtering, the anisotropicness of transfer function correction, lack of projections and interpolation errors.

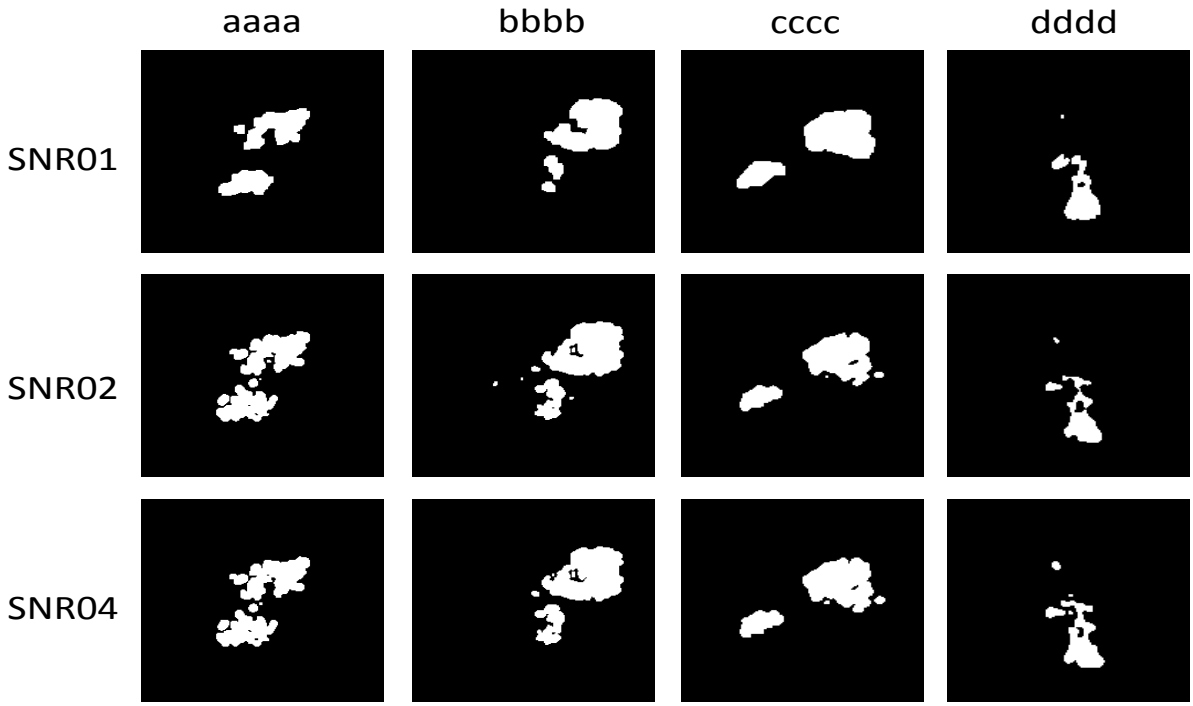

Fig. 2: The masks obtained from different SNR using TOMOBFLOW.

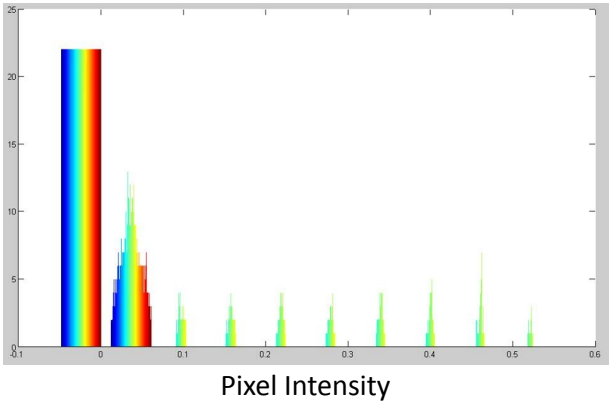

Fig. 3: The distribution of the intensity in noisy structure.

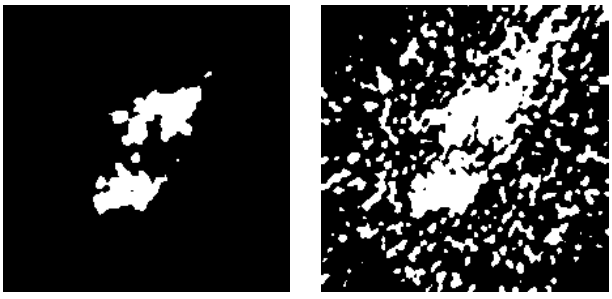

Fig. 4: The masks under (i) threshold = 0.07 and (ii) threshold = 0.01.

Iteration	150
γ in SNR 04	0.25
γ in SNR 02	0.25
γ in SNR 01	0.5
window size in open operation	3

TABLE I: Parameters used in TOMOBFLOW mask.

In light of the significantly low SNR in the provided data and unique atomic structures of proteins, the prior knowledge about protein structures plays an important role in determining the validity of the individual images from the noisy image datasets and serving as traing data in exampler-based super resolution. In the selection of the protein structures from PDB as prior information, we chose the three most complicated structures, Histidine Ammonia-Lyase from Pseudomonas Putida Inhibitaed with L-Cysteine (1GKM), Xylanase from Trichoderma Longibrachiatum (3AKT), and Uridine Phosphorylase from Shewanella oneidensis MR-1 (4HEN), among the 94 proteins which are at the same resolution of 1 Åas the noise-free structures.

B. Experimental Results

In this section, we present the experimental results from TOMOBFLOW and Nearest Neighbor approaches of our proposed method respectively, compare the performance of our method with 3D-Bilateral filter, and solelyTomoBFlow denoising, and summarize the results. The parameters used in TomoBFlow Mask are listed in table I.

Figure. 5 shows the masks obtained from different SNR

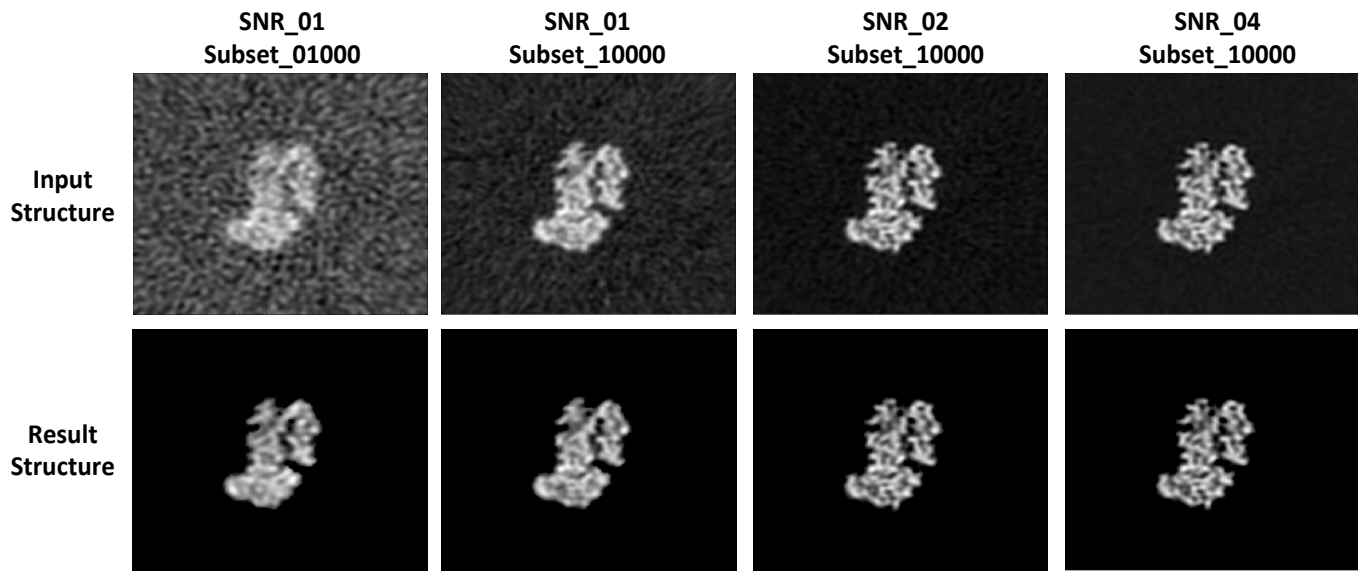

Fig. 5: Illustration of our result. The pictures in different columns are under different SNRs and projection numbers, and in the first rows represent the input structure and in the second rows represent our result structure.

	3D-Bilateral filter	TOMOBFLOW	Our Framework
SNR_04_10000	1.96	5.91	5.40
SNR_02_10000	5.39	9.70	10.25
SNR_01_10000	0.69	4.36	11.45
SNR_01_05000	N/A	4.24	9.74
SNR_01_01000	N/A	3.39	8.75
SNR_01_00500	N/A	2.84	8.33

TABLE II: Comparison between our framework, 3D bilateral filter and solely TOMOBFLOW method.

using TOMOBFLOW. Figure. 3 and Figure. 4 illustrate histogram by which we decide the parameters.This histogram shows the distribution of the intensity in mrc files. We assigned zero to the region where the intenity of the noise is lower than that of protein structure.

In Nearest Neighbor approach, we experiment with patch size = 2 and patch size = 3, and define the distance in Euclidean space. Figure. 2 compares the input noisy structures with the clear and sharpened results obtained by our method.

We use Fourier shell correlation (FSC) score rating from 3D Electron Microscopy Benchmark to quantify the effect of our method. This criterion has now become the standard quality measure on resolution. Table II demonstrates the comparison between our method and other two approaches. 3D-Bilateral filter stands for the classic baseline in noise removal, and TomoBFlow represents the state-of-art EM denoising techniques.

IV. Conclusion

In this paper, we propose a novel framework for image restoration and super-resolution of EM images. We advocated the TOMOBFLOW method for denoising and exampler-based strategy for resolving low-resolution EM images. With the preservation of relevant biological information, we significantly remove the unpredictable and irregular noise pattern by TOMOBFLOW approach. Then by exploiting the prior information obtained from Protein Data Bank (PDB), we convert the super-resolution problem into a fairly staightforward nearest-neighbor problem. The structures in the result images confirmed that our proposed framework quantitatively and qualitatively outperformed the existing denoising and super-resolution approaches.

References

[1] John-Marc Chandonia and Steven E Brenner, "The impact of structural genomics: expectations and outcomes," *Science*, 2006.

[2] Richard Henderson, "Realizing the potential of electron cryo-microscopy," *Quarterly reviews of biophysics*, 2004.

[3] Carlos Oscar Sorzano, JM de la Rosa Trevín, J Otón, JJ Vega, J Cuenca, A Zaldívar-Peraza, J Gómez-Blanco, J Vargas, A Quintana, Roberto Marabini, et al., "Semiautomatic, high-throughput, high-resolution protocol for three-dimensional reconstruction of single particles in electron microscopy," in *Nanoimaging*. Springer, 2013.

[4] Joachim Frank, Michael Radermacher, Pawel Penczek, Jun Zhu, Yanhong Li, Mahieddine Ladjadj, and Ardean Leith, "Spider and web: processing and visualization of images in 3d electron microscopy and related fields," *Journal of structural biology*, 1996.

[5] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne, "The protein data bank," *Nucleic Acids Res*, 2000.

[6] COS Sorzano, Roberto Marabini, Javier Velázquez-Muriel, José Román Bilbao-Castro, Sjors HW Scheres, José M Carazo, and Alberto Pascual-Montano, "Xmipp: a new generation of an open-source image processing package for electron microscopy," *Journal of structural biology*, 2004.

[7] Shachar Fleishman, Iddo Drori, and Daniel Cohen-Or, "Bilateral mesh denoising," in *ACM Transactions on Graphics (TOG)*. ACM, 2003.

[8] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, Karen Egiazarian, et al., "Bm3d image denoising with shape-adaptive principal component analysis," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[9] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, 2006.

[10] Jose-Jesus Fernandez, "Tomobflow: feature-preserving noise filtering for electron tomography," *BMC bioinformatics*, 2009.