

rAMI – Rapid Alignment with Moment of Inertia for Cryo-EM Image Processing

Szu-Chi Chung¹, Shao-Hsuan Wang², Cheng-Yu Hung¹, Wei-Hau Chang³ and I-Ping Tu¹

¹Institute of Statistical Science, Academia Sinica, United States, ²Institute of Statistics, National Central University, United States, ³Institute of Chemistry, Academia Sinica, United States

Cryogenic Electron microscopy (cryo-EM) is one of the most important instruments for determining the structures of protein complexes in near-atomic resolution, evidenced by the fact that the first 3D structure of spike protein on COVID-19 virus was solved using cryo-EM [1]. Data processing is a crucial step in solving a cryo-EM 3D protein-complex structure and usually takes several days to emerge a reasonable 3D structure. There is a cry for accelerating data processing in cryo-EM community and here we address the alignment step. Alignment is critical in the whole workflow to boost the signal-to-noise ratio (SNR) by grouping similar particles together. Current approaches compute the cross-correlation between each image and the current model at all possible orientations to search the best parameter. Even though this costly procedure has benefited significantly from GPU parallelization [2] or the carefully designed heuristic method [3] and the branch-and-bound algorithm that exclude regions of the search space [4], it still remains as a computational bottleneck in the entire workflow due to the fast-growing rate of automatic data acquisition.

Moment of Inertia (MoI) as a 2 by 2 matrix I containing the central moments with order two, whose first eigenvector corresponds to the object's orientation, has been a popular tool for image alignment [5]. However, the low SNR nature of cryo-EM images has prevented the direct application of MoI. Here, we proposed an algorithm called Rapid Alignment with Moment of Inertia (rAMI), as illustrated in Figure 1 (a) and we show that it can be widely applicable to the current alignment steps in cryo-EM. Our method starts with denoising the particle images with a novel dimension reduction method called 2SDR [6]. The denoised particles are then

used to compute the matrix
$$I = \frac{1}{\mu_{00}} \begin{pmatrix} \mu_{02} & \mu_{11} \\ \mu_{11} & \mu_{20} \end{pmatrix}$$
 for each image, where $\mu_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q X(x, y)$ is the 2D Cartesian central moment (X

is an image represented as a $M \times N$ matrix, $X(x, y)$ is the pixel value at coordinate (x, y) and \bar{x}, \bar{y} is the mean position of coordinates). Then, its first eigenvector is recorded as the major axis for each image. The major axis is then aligned with the x-axis in 2D plane for coarse alignment. Finally, it will check whether flipping along y-axis or rotation by 180 degrees for each image sequentially can improve the alignment results. As for those particles whose shapes are close to spherical, we propose a composite algorithm which use rAMI to produce an initial value for a state-of-the-art reference-free alignment algorithm (RFA) implemented in EMAN2 package [8]. To quantify the alignment performance, we use the Frobenius norm of the averaged image of the whole dataset (that contains N images X_1, X_2, \dots, X_N and each one is represented as a 2D matrix)

defined as $\left\| \left(\sum_{i=1}^N X_i \right) / N \right\|_F$ which has been previously used in [7].

We find that the application of rAMI can significantly reduce the computation time while maintaining comparable performance, especially when the target resembles elliptical structures such as Betagal. We test rAMI on a synthetic Betagal dataset generated from RELION tutorial [3] (5000 particles with box size equal to 130×130 pixel and $\text{SNR}=0.1$). We compare our algorithm with RFA and the results are shown in Figure 1 (b-d). As shown from the figure, each optimizing step (i.e., denoise, flip, rotation by 180 degrees) can substantially improve the Frobenius norm and the performance of rAMI is comparable to the RFA while we can save the computation time by one order of magnitude. We further test on the experimental 70S ribosome

dataset (5000 particles with box size equal to 130×130 pixel) [9] and 80S ribosome dataset (105,247 particles with box size equal to 90×90 pixel) [10] whose particle shapes are closer to a sphere. The results of the composite algorithm are comparable to RFA, while the computation time is saved by 40 percent, due to that the initial by rAMI can save the iteration number of RFA.

The results are encouraging and we are optimistic that other tasks like 2D or 3D classification in cryo-EM that require alignment can also benefit from rAMI. This new approach to align cryo-EM images without searching a large amount of orientation space may solve the scalability problem due to the growing size of dataset.

*The authors acknowledge funding from Academia Sinica.

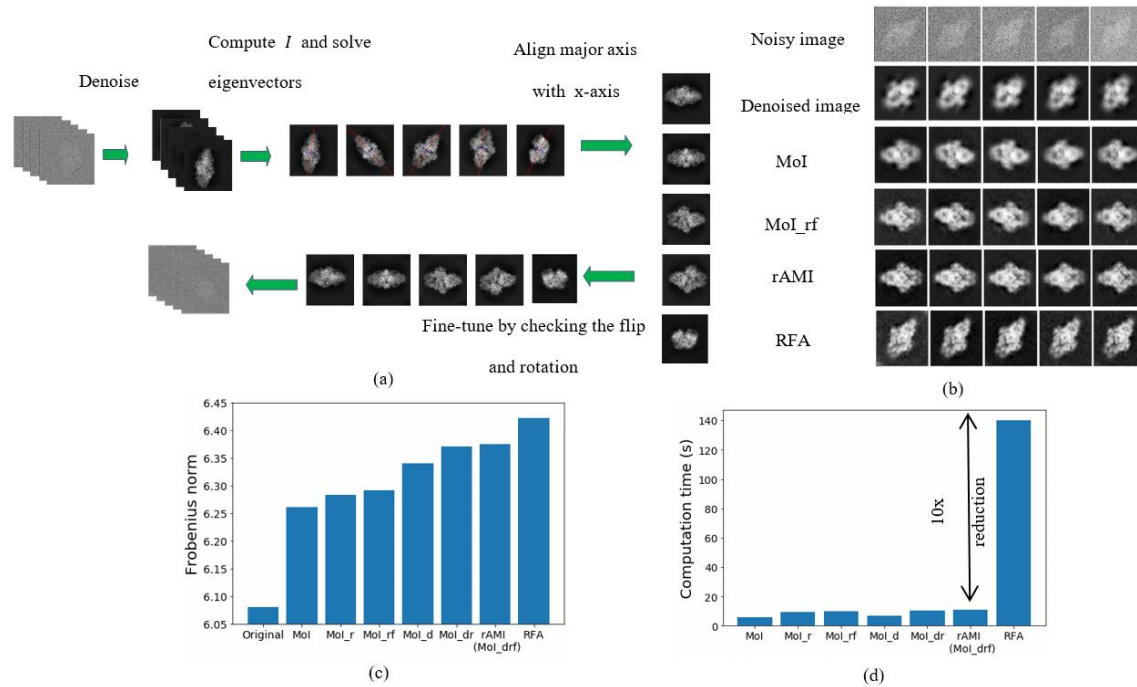


Figure 1. (a) Visual illustration of our proposed rAMI algorithm. The red line and blue line in each image represent the major axis and minor axis extracted from MoI. Note that we use clean particles in the figure for better visualization purposes, but the actual algorithm is performed on the noisy images. (b) Visual comparison between different alignment steps on synthetic Betagal dataset. Note that all the results are further denoised by 2SDR for better visualization. Methods include alignment by MoI, MoI with refinement (MoI_rf, rotation and flip), MoI with denoised particle (MoI_d), MoI with denoised particle and refinement (rAMI) and the state-of-the-art alignment algorithm RFA. (c) Frobenius norm after different steps on synthetic Betagal dataset (Higher is better). (d) Computation time of different steps on synthetic Betagal dataset (Lower is better).

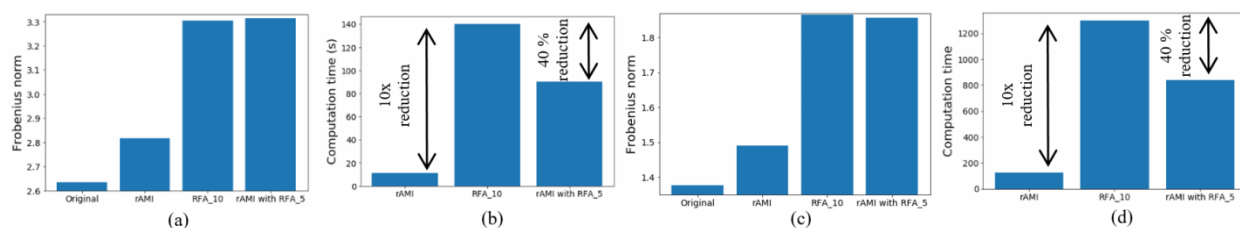


Figure 2.(a) Frobenius norm of different methods on experimental 70S ribosome dataset (Higher is better). The state-of-the-art alignment algorithm is denoted as RFA_x where x represents the number of iterations. (b)

Computation time of different methods on experimental 70S ribosome dataset (Lower is better). (c) Same as (a) but on experimental 80S ribosome dataset. (d) Same as (b) but on experimental 80S ribosome dataset.

References

- [1] Wrapp, Daniel, et al. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation." *Science* 367.6483 (2020): 1260-1263.
- [2] Dari Kimanius, Björn O. Forsberg, Sjors H.W. Scheres, and Erik Lindahl. "Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2". *eLife*, (2016). 3.
- [3] Zivanov, Jasenko, et al. "New tools for automated high-resolution cryo-EM structure determination in RELION-3." *eLife* 7 (2018): e42166.
- [4] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. "CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination". *Nature Methods*, (2017). 2, 3.
- [5] Flusser, Jan, Tomas Suk, and Barbara Zitová. "2D and 3D image analysis by moments". John Wiley & Sons, (2016).
- [6] S.C. Chung, S.H. Wang, P.Y. Niu, S.Y. Huang, W.H. Chang, and I.P. Tu, "Two-stage dimension reduction for noisy high-dimensional images and application to cryo-em," *Annals of Mathematical Sciences and Applications*, vol. 5, no. 2, pp. 283–316, (2020).
- [7] Penczek, P., Radermacher, M. & Frank, J. "Three-dimensional reconstruction of single particles embedded in ice". *Ultramicroscopy* 40, 33–53 (1992).
- [8] Guang Tang, Liwei Peng, Philip R Baldwin, Deepinder S Mann, Wen Jiang, Ian Rees, and Steven J Ludtke, "Eman2: an extensible image processing suite for electron microscopy," *Journal of Structural Biology*, vol. 157, no. 1, pp. 38–46, (2007).
- [9] Herman, Gabor T., and Mirosław Kalinowski. "Classification of heterogeneous electron microscopic projections into homogeneous subsets." *Ultramicroscopy* 108.4, 327-338 (2008).
- [10] Wong, W. et al. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the antiprotozoan drug emetine. *eLife* 3, e03080 (2014).