JOINT MOTION-CORRECTION AND RECONSTRUCTION IN CRYO-EM TOMOGRAPHY

Yuanhao Wang, Ramzi Idoughi, and Wolfgang Heidrich

KAUST, Thuwal, KSA

ABSTRACT

Tilt-series cryo-electron tomography (cryoET) is an established imaging technique used in several scientific fields to determine samples' three-dimensional (3D) structures at nearatomic resolutions. However, the motion and misalignment that occur during the acquisition stage are major limiting factors to reaching smaller resolutions. Indeed, they introduce blur and artifacts, which deteriorate the reconstruction quality. In this paper, we propose a joint motion-correction and reconstruction framework to improve the quality of the output volume and, consequently, its resolution. Our framework first estimates the motion field in the sample in order to correct the captured data. Then an iterative plug-and-play prior approach is used to reconstruct the sample. The validation of our approach on real captured datasets shows a good quality reconstruction translated in a resolution improvement.

Index Terms— Cryo-Electron Tomography, Motion Compensation, Computational Imaging.

1. INTRODUCTION

Electron Microscopy (EM) Tomography is a widely used imaging modality in a variety of fields, from structural biology [1] to material science [2]. In the tilt-series EM technique, a 3D tomogram is reconstructed from a set of 2D captured projections of the sample. These projections are acquired from a limited range of tilt angles, usually less than 120° , resulting in a "missing wedge" of frequencies that limits the depth resolution. Moreover, the inevitable beam-induced motion and the sample deformation during the acquisition step introduce misalignment between the projections, which deteriorates the resolution of the reconstructed samples [3]. The motion correction is then required to improve the reconstruction quality.

Prior work: To tackle the alignment issue and to correct for the motion in the projections, the most common approach is to embed gold particles in the sample [4–9]. Then, the misalignment is estimated by tracking the fiducial markers in the different projections, which is commonly performed through a local similarity search between successive projections [10, 11]. However, this local approach does not guarantee global tracking consistency. Besides, the alignment will be of low quality in regions with no marker. After the track-

ing step, the sample' deformation can be generated from the marker's motion using different approaches, such as the use of a parametric 2D/3D quadratic polynomial surface model [9], or using a global thin-plate splines interpolation [12]. Finally, after applying the correction on captured projections, the tomographic reconstruction is usually performed using the Weighted Filtered Back-Projection method (WFBP) [13].



Fig. 1: General workflow for the reconstruction of a single highly resolved ribosome, from preprocessed tomograms.

Contributions: In this work, we introduce a new framework for tilt-series cryo-EM tomography. First, we use a marker-based technique to evaluate the misalignment in the sample. Then a plug-and-play optimization approach is used to jointly correct the motion in the sample and reconstruct the tomogram. The validation of our approach using real datasets shows better performance than commonly used reconstruction techniques. The main contributions of our work include: (1) a new framework for joint reconstruction and denoising of marker-based tilt-series EM tomography; (2) a new estimation method for the local motion from the embedded fiducial markers; (3) considerable improvement of the protein extraction, which results in a better resolution after the subtomogram averaging; and (4) evaluation and validation of our framework on two real datasets.

2. METHODOLOGY

We assume in the following that the input projections have been already preprocessed using existing software packages, such as IMOD [5, 8], or EMAN2 [7]. Thus, the 2D location of markers in the different projections is detected and used directly as input to our approach. In Figure 1, we illustrate our general framework for the reconstruction of a highly resolved molecule from the preprocessed input tilt-series images. The first step is to estimate the 3D position of the markers. By re-projecting these markers to each projection, we can estimate the motion of the sample at the markers' positions. Then, we apply a natural neighbor interpolation [14] to estimate the sample motion for all the pixels of the projections. Afterwards, we warp the input projections using the obtained motion field to correct the sample motion. The corrected projections are fed to our joint reconstruction algorithm to produce better quality reconstructed volumes. Finally, if applicable, the individual copies of the reconstruction accuracy.

2.1. Fiducial marker 3D position estimation

For a given fiducial marker j in the sample, the relationship between its 3D coordinates and the 2D coordinates of its projection onto the acquired image i is as follows:

$$\mathbf{w}_{\mathbf{i}}^{\mathbf{J}} = \mathbf{K}_{i}\mathbf{m}_{j} + \mathbf{d}_{i}^{j},\tag{1}$$

where $\mathbf{w}_i^j = (u_i^j, v_i^j)$ is the 2D position of the marker j projected in the captured image i. \mathbf{K}_i represents the projection matrix. $\mathbf{m}_j = (x_j, y_j, z_j)$ corresponds to the 3D coordinates of the j^{th} marker, and \mathbf{d}_i^j is a 2D offset induced by the motion of the j^{th} marker.

As mentioned above, the detection of the markers on the N_p captured projections is performed using existing software. We denote $\tilde{\mathbf{w}}_i^{\mathbf{j}} = (\tilde{u}_i^j, \tilde{v}_i^j)$ the measured 2D position of the j^{th} marker in the i^{th} projection. We formulate the estimation of the 3D position of the N_m detected markers as the following least-square problem:

$$E = \sum_{i=1}^{N_p} \sum_{j=1}^{N_m} \|\tilde{\mathbf{w}}_{\mathbf{i}}^{\mathbf{j}} - \left(\mathbf{K}_i \mathbf{m}_j + \mathbf{d}_i^j\right)\|^2.$$
(2)

To solve this optimization problem we follow a common strategy based on two step refinement. First, we set \mathbf{d}_i^j to an initial guess and solve the optimization problem for the different \mathbf{m}_i . Then, we compute the estimate of \mathbf{d}_i^j as follows:

$$\mathbf{d}_i^j = \bar{\mathbf{w}}_i^j - \mathbf{K}_i \bar{\mathbf{m}}_j, \tag{3}$$

where $\bar{\mathbf{m}}_j$ is the centroid of the fiducial j, and $\bar{\mathbf{w}}_i^j$ is the projection of the centroid j in the i^{th} image. These two steps are applied recursively until convergence.

2.2. Motion estimation

Next, we estimate the sample deformation field so we can compensate for it during the tomographic reconstruction. We only focus on 2D motion compensation for two main reasons. First, the deformation is mainly due to a beam-induced motion, which can be described accurately with a 2D model. Moreover, using a 3D model for the deformation requires a larger memory footprint and a higher computational time.

In the previous section, we estimated the 3D position of the markers (\mathbf{m}_j) and their corresponding 2D motion (\mathbf{d}_i^j) . The 2D deformation of all pixels in the different projections is obtained by interpolating the 2D motions of the markers \mathbf{d}_i^j . This interpolation is performed using the natural neighbor interpolation technique, presented in [14]. For a given pixel qin the *i*th projection, we interpolate its 2D motion $\mathbf{d}(q)$ from the markers' motion using the following weighted function:

$$\mathbf{d}(q) = \sum_{j=1}^{N_m} \omega_j(q) \, \mathbf{d}_i^j \tag{4}$$

where $\omega_j(q)$ are the weights, computed using the approach described in [14]. After computing these weights for each pixel in the N_p projections, we can obtain an estimation of the motion field as illustrated in Figure 1.

2.3. Geometry of the reconstructed volume

Before presenting our proposed reconstruction approach, we define in this section the boundaries of the reconstructed volume. In electron microscopy tomography, the sample is scanned using a parallel beam. Due to the rotation of the sample during the acquisition process, some regions will not be captured by the detector for all the angles. Thereafter our region of interest (ROI) should encompass only the region present in all the projections. Thus, we have to pre-compute the size of the geometry to minimize the computation load and preserve the ROI information. This pre-computation consists of assigning masks for all the projections and for the 3D volume in order to select only the ROI. These masks are defined in the following equation:

$$\mathbf{M}_{p}\mathbf{p} = \{\mathbf{K} \cdot \mathbf{f} (x, y, z) \mid (x, y, z) \in ROI\}$$
(5)
$$\mathbf{M}_{v}\mathbf{x} = \{\mathbf{K}^{-1} \cdot \mathbf{p} \mid \mathbf{p} \in \mathbf{M}_{v}\mathbf{p}\}$$

where \mathbf{M}_p and \mathbf{M}_v represent respectively the masks on projections and the 3D volume. **p** is a vector that combines all the projection pixels. **K** and \mathbf{K}^{-1} are the two matrix that model the Radon transform operator and its inverse for all used angles. **f** (x, y, z) is the density at the coordinates (x, y, z), and **x** is the 3D volume to be reconstructed.

2.4. 3D volume reconstruction

Next, we present an optimization-based framework to reconstruct the 3D volume according to the geometry defined in Section 2.3. This framework is designed to compensate for the estimated beam-induced motion, to reduce the artifacts caused by the missing wedge acquisition scenario and to reduce the noise level of the output volume. **Optimization framework:** The general optimization problem is formulated as:

$$(\mathbf{x}^*, \mathbf{z}^*) = \operatorname{argmin}_{\mathbf{x}, \mathbf{z}} \mathcal{L}_{data}(\mathbf{x}) + \lambda_1 \cdot \mathcal{L}_{TV}(\mathbf{x}) \quad (6)$$

subject to: $\mathbf{x} = \mathbf{z} = \mathcal{DN}(\mathbf{x})$ in ROI,

where \mathcal{L}_{data} and \mathcal{L}_{TV} are the data-fitting and the smoothness terms respectively. λ_1 is a smoothness weight, and z is a slack variable, corresponding to the denoised volume. By introducing a second slack variable $\mathbf{u} = \mathbf{x} - \mathbf{z}$, we can rewrite the optimization problem as follows:

$$(\mathbf{x}^*, \mathbf{z}^*) = \underset{\mathbf{x}, \mathbf{z}}{\operatorname{argmin}} \ \mathcal{L}_{data}(\mathbf{x}) + \lambda_1 \mathcal{L}_{TV}(\mathbf{x}) + \lambda_2 \|\mathbf{x} - \mathbf{z} + \mathbf{u}\|_2^2$$
(7)

where λ_2 is the denoising weight.

Data-fitting term: We define our data-fitting term as a modified version of the classical tomography least-square term. Specifically, we introduce a warping operator on the captured projections, to compensate for the beam-induced motion. In addition, we apply a pre-computed masks on those projections to take into account only the defined ROI. The proposed data-fitting term can be written as follows:

$$\mathcal{L}_{\text{data}}(\mathbf{x}) = \frac{1}{2} \|\mathbf{K}\mathbf{x} - \mathbf{M}_p \text{warp}(\mathbf{p}, -\mathbf{d})\|_2^2, \qquad (8)$$

where d is a vector regrouping the estimated motion for all pixels in the projections. warp $(\mathbf{p}, -\mathbf{d})$ is the warping operator, that warps the projections p by the deformation field $-\mathbf{d}$.

Smoothness prior: We introduce a total variation (TV) prior to our optimization loss function. The expected role of this term is to compensate for the ill-posedness of the limited-angle reconstruction problem and to reduce the streaking and ghost tail artifacts [15].

$$\mathcal{L}_{\mathrm{TV}}(\mathbf{x}) = \|\nabla_s \left(\mathbf{M}_v \mathbf{x}\right)\|_{\mathbf{H}_{\epsilon}} \tag{9}$$

In this term, ∇_s refers to the discrete spatial gradient operator, and \mathbf{H}_{ϵ} is the Huber penalty with the threshold equal to ϵ .

Denoising term: In order to improve the reconstruction accuracy, we also introduce a denoising step in our optimization framework. For the sake of flexibility, we implemented our framework following the plug-and-play prior scheme [16]. Thus, any denoiser can be easily incorporated in this step. In our implementation, we used the Topaz-Denoise algorithm [17], which is inspired by Noise2Noise approach [18]. This deep-learning approach does not require ground truth data for the learning, but instead, it uses a pair of noisy input data representing the same sample and having the same noise pattern. This pair of input is used for an inter-learning process. In Topaz-Denoise, the model is trained on 32 aligned cryoET tilt-series datasets. For each dataset, a pair of two

noisy reconstructions is computed from only odd/even projections. In our framework we apply the denoiser $\mathcal{DN}(\cdot)$ only on the ROI defined above.

Solver: To solve this optimization problem, we opt for an ADMM scheme [19] to split the optimization into three subproblems. The later are solved in an alternating and iterative fashion as detailed in Algorithm 1.

Algorithm	1	Proposed	Plug-and-Play	based	optimization				
framework for the volume reconstruction.									

Require: $(\mathbf{K}, \mathbf{d}, \mathbf{M}_p, \mathbf{M}_v, \lambda_1, \lambda_2, \epsilon, N_{iter})$

1: For $k = 0 : N_{iter} - 1$ do:

- 2: // Solve the sub-problem in x using the primal-dual Chambolle-Pock algorithm detailed in [20]
- 3: $\mathbf{x}^{k+1} \leftarrow \operatorname{argmin}_{\mathbf{x}} \ \mathcal{L}_{data}(\mathbf{x}) + \lambda_1 \mathcal{L}_{TV}(\mathbf{x}) + \lambda_2 \|\mathbf{x} \mathbf{z}^k + \mathbf{u}^k\|_2^2$
- 4: // Solve the sub-problem in z by applying a denoiser
- 5: $\mathbf{z}^{k+1} \leftarrow \operatorname{argmin}_{\mathbf{z}} \|\mathbf{x}^{k+1} \mathbf{z} + \mathbf{u}^k\|_2^2 = \mathcal{DN}(\mathbf{x}^{k+1} + \mathbf{u}^k)$
- 6: // update of the dual variable \mathbf{u}
- 7: $\mathbf{u}^{k+1} \leftarrow \mathbf{x}^{k+1} + \mathbf{u}^k \mathbf{z}^{k+1}$

9: return
$$\mathbf{x}^* = \mathbf{x}^{N_{iter}}$$

Subtomogram averaging: If the data contains multiple copies of the same molecule, a final optional step is to perform subtomogram averaging to improve the reconstruction quality. First, the copies are detected by cross-correlation calculation. Then, all the identified copies are aligned and averaged to obtain the final reconstruction of the molecule of interest. Note that for some samples several molecules are present. The subtomograms averaging should be done for each of these molecules. In our work we run this step using the Relion software and follow the standard protocols described in [21].

While we make no novel algorithmic contributions to the subtomogram averaging process itself, we show in the experiments that the number of detected molecules, and therefore the quality of the subtomogram averaging process, increases substantially when it is based on our reconstruction output instead of baseline results.

3. RESULTS AND DISCUSSION

Datasets: We evaluate our method on two publicly available datasets from the Electron Microscopy Pilot Image Archive (EMPIAR) [22]. The first dataset (EMPIAR 10045) [21] contains several purified *Saccharomyces cerevisiae* (yeast) 80S ribosomes. This data was scanned over an angular range of $[-60^{\circ}, 60^{\circ}]$, with a step size of 3°. The second sample (EMPIAR 10064) [23] contains purified mammalian 80S ribosomes, which were extracted from rabbit *Reticulocyte Lysate*. It was scanned over the same angular range, but

with an angular increment of 2° . For each dataset, several tilt-series were available; but we only used one series for our evaluations. These datasets were pre-processed using IMOD [8] to align projections, and to extract and localize markers.

Implementation parameters and baselines: After conducting a parameter search, we selected the following values: $\lambda_1 = 0.05$, $\epsilon = 0.04$, and $\lambda_2 = 0.003$ for all the experiments. We compare our reconstruction framework (**Ours**) to different baselines: (1) the **WFBP** method using the IMOD implementation, (2) **Motion-corrected WFBP** that we feed with corrected projections using our framework, (3) **SART+TV** [24], a model-based reconstruction with TV prior, (4) **Motion-corrected SART+TV** modelbased appraoch with motion-corrected input projections, (5) **Reference** a high-resolution reconstruction obtained through cryo-EM Single Particle Analysis (SPA). The last baseline involves several tilt-series (scans) for each sample.

Ablation study: The baselines are selected to evaluate separately each component of our framework. Comparing WFBP with SART+TV shows the importance of combining the TV prior to an iterative reconstruction. The motion-corrected versions of these approaches illustrate the impact of correcting the local motion in the projections. Finally, in our technique (**Ours**) all the modules are combined, including the denoiser.



Fig. 2: Slice visualization of the EMPIAR 10045 dataset (using a binning factor of 3) after reconstruction with **WFBP** method (left) and **Ours** (right).

The slice visualization in Figure 2 shows a clear improvement of our joint reconstruction method over the standard **WFBP**, especially in terms of contrast enhancement. Consequently, our method improves the accuracy of the protein extraction which performed prior to the subtomogram averaging step. For instance, 405, 450, and 458 proteins have been extracted from EMPIAR 10045 reconstructions using **WFBP**, **SART+TV**, and **Ours** respectively. This improvement is translated by a better reconstruction quality after the subtomogram averaging. Indeed, for the EMPIAR 10045, we run the **WFBP** and **Motion-corrected WFBP** methods using 2 different protein positions estimation strategies: ours and the one provided with the dataset (obtained after a **WFBP** reconstruction). Note that for all other cases we use the same protein positions obtained with our method to compute the subtomogram average.



Fig. 3: 3D visualization of the refined proteins after the subtomogram averaging.

The 3D visualization of the reconstructed proteins in Figure 3, and the computation of the Fourier Shell Correlation (FSC) to estimate the resolution of the reconstruction (see Table 1) both show the improvement introduced by each single module of our framework. Usually, after computing the FSC curve using existing software, the resolution of the reconstruction is the one obtained for a FSC = 0.143. The resolution obtained with FSC = 0.5 is also used in methods comparisons.

Table 1: Resolution (in Å) obtained with FSC 0.5 and 0.143.

	EMPIAR 10045		EMPIAR 10064	
	FSC 0.5	FSC 0.143	FSC 0.5	FSC 0.143
WFBP (Provided position)	24.218	15.507		
Motion-corrected WFBP	23.688	17.740		
(Provided position)				
WFBP	14.221	12.591	14.324	11.715
Motion-corrected WFBP	13.987	12.733	15.265	11.730
SART+TV	14.058	12.648	11.595	10.837
Motion-corrected	13.755	9.240	11.419	10.702
SART+TV				
Ours	13.548	9.178	11.396	10.593
Motion-corrected WFBP (Provided position) WFBP Motion-corrected WFBP SART+TV Motion-corrected SART+TV Ours	23.688 14.221 13.987 14.058 13.755 13.548	17.740 12.591 12.733 12.648 9.240 9.178	14.324 15.265 11.595 11.419 11.396	11.715 11.730 10.837 10.702 10.593

4. CONCLUSION

We proposed a new reconstruction framework to improve cryo-EM tomograms' quality and resolution. The main drawback of our approach is the dependence on the fiducial markers (accessibility, in-homogeneity, and artifacts). In the future we believe our approach can be adapted to marker-free tiltseries data by using sample features as virtual markers.

5. REFERENCES

- X-C. Bai, G. McMullan, and S.H.W. Scheres, "How cryo-EM is revolutionizing structural biology," *Trends Biochem. Sci.*, vol. 40, no. 1, pp. 49–57, 2015.
- [2] U. Valdrè, *Electron microscopy in material science*, Elsevier, 2012.
- [3] A.F. Brilot, J.Z. Chen, A. Cheng, J. Pan, S.C. Harrison, C.S. Potter, B. Carragher, R. Henderson, and N. Grigorieff, "Beam-induced motion of vitrified specimen on holey carbon film," *J. Struct. Biol.*, vol. 177, no. 3, pp. 630–637, 2012.
- [4] M.C. Lawrence, "Least-squares method of alignment using markers," in *Electron tomography*, pp. 197–204. 1992.
- [5] J.R. Kremer, D.N. Mastronarde, and J.R. McIntosh, "Computer visualization of three-dimensional image data using IMOD," *J. Struct. Biol.*, vol. 116, no. 1, pp. 71–76, 1996.
- [6] S.S. Brandt and U. Ziese, "Automatic TEM image alignment by trifocal geometry," *J. Microsc.*, vol. 222, no. 1, pp. 1–14, 2006.
- [7] G. Tang, L. Peng, P.R. Baldwin, D.S. Mann, W. Jiang, I. Rees, and S.J. Ludtke, "EMAN2: an extensible image processing suite for electron microscopy," *J. Struct. Biol.*, vol. 157, no. 1, pp. 38–46, 2007.
- [8] D.N. Mastronarde and S.R. Held, "Automated tilt series alignment and tomographic reconstruction in IMOD," J. Struct. Biol., vol. 197, no. 2, pp. 102–113, 2017.
- [9] J-J. Fernandez, S. Li, T.A.M. Bharat, and D.A. Agard, "Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion," *J. Struct. Biol.*, vol. 202, no. 3, pp. 200–209, 2018.
- [10] F. Amat, F. Moussavi, L.R. Comolli, G. Elidan, K.H. Downing, and M. Horowitz, "Markov random field based automatic image alignment for electron tomography," *J. Struct. Biol.*, vol. 161, no. 3, pp. 260–275, 2008.
- [11] M. Cao, A. Takaoka, H-B. Zhang, and R. Nishi, "An automatic method of detecting and tracking fiducial markers for alignment in electron tomography," *J. Electron Microsc.*, vol. 60, no. 1, pp. 39–46, 2010.
- [12] J-J. Fernandez, S. Li, and D.A. Agard, "Consideration of sample motion in cryo-tomography based on alignment residual interpolation," *J. Struct. Biol.*, vol. 205, no. 3, pp. 1–6, 2019.

- [13] M. Radermacher, "Weighted back-projection methods," in *Electron tomography*, pp. 245–273. 2007.
- [14] Q. Fan, A. Efrat, V. Koltun, S. Krishnan, and S. Venkatasubramanian, "Hardware-assisted natural neighbor interpolation.," in ALENEX/ANALCO, 2005, pp. 111–120.
- [15] B. Goris, W. Van den Broek, K.J. Batenburg, H.H. Mezerji, and S. Bals, "Electron tomography based on a total variation minimization reconstruction technique," *Ultramicroscopy*, vol. 113, pp. 120–130, 2012.
- [16] S.V. Venkatakrishnan, C.A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in 2013 IEEE GlobalSIP, 2013, pp. 945–948.
- [17] T. Bepler, K. Kelley, A.J. Noble, and B. Berger, "Topazdenoise: general deep denoising models for cryoEM and cryoET," *Nat. Methods*, vol. 11, no. 1, pp. 1–12, 2020.
- [18] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," in *Proc. ICML*, 2018, pp. 2965–2974.
- [19] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Now Publishers Inc, 2011.
- [20] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, 2011.
- [21] T.A.M. Bharat and S.H.W. Scheres, "Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION," *Nat. Protoc.*, vol. 11, no. 11, pp. 2054–2065, 2016.
- [22] A. Iudin, P.K. Korir, J. Salavert-Torres, G.J. Kleywegt, and A. Patwardhan, "EMPIAR: a public archive for raw electron microscopy image data," *Nat. Methods*, vol. 13, no. 5, pp. 387–388, 2016.
- [23] M. Khoshouei, S. Pfeffer, W. Baumeister, F. Förster, and R. Danev, "Subtomogram analysis using the Volta phase plate," *J. Struct. Biol.*, vol. 197, no. 2, pp. 94–101, 2017.
- [24] M. Zang, G.and Aly, R. Idoughi, P. Wonka, and W. Heidrich, "Super-resolution and sparse view CT reconstruction," in *Proc. ECCV*, 2018, pp. 137–153.