

Bioimage informatics

# Robust and ultrafast fiducial marker correspondence in electron tomography by a two-stage algorithm considering local constraints

Renmin Han <sup>1,2</sup>, Guojun Li <sup>1,\*</sup> and Xin Gao <sup>2,\*</sup>

<sup>1</sup>Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao 266237, China and <sup>2</sup>King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia

\*To whom correspondence should be addressed.

Associate Editor: Jinbo Xu

Received on July 2, 2020; revised on December 8, 2020; editorial decision on December 22, 2020; accepted on December 28, 2020

## Abstract

**Motivation:** Electron tomography (ET) has become an indispensable tool for structural biology studies. In ET, the tilt series alignment and the projection parameter calibration are the key steps toward high-resolution ultrastructure analysis. Usually, fiducial markers are embedded in the sample to aid the alignment. Despite the advances in developing algorithms to find correspondence of fiducial markers from different tilted micrographs, the error rate of the existing methods is still high such that manual correction has to be conducted. In addition, existing algorithms do not work well when the number of fiducial markers is high.

**Results:** In this article, we try to completely solve the fiducial marker correspondence problem. We propose to divide the workflow of fiducial marker correspondence into two stages: (i) initial transformation determination, and (ii) local correspondence refinement. In the first stage, we model the transform estimation as a correspondence pair inquiry and verification problem. The local geometric constraints and invariant features are used to reduce the complexity of the problem. In the second stage, we encode the geometric distribution of the fiducial markers by a weighted Gaussian mixture model and introduce drift parameters to correct the effects of beam-induced motion and sample deformation. Comprehensive experiments on real-world datasets demonstrate the robustness, efficiency and effectiveness of the proposed algorithm. Especially, the proposed two-stage algorithm is able to produce an accurate tracking within an average of  $\leq 100$  ms per image, even for micrographs with hundreds of fiducial markers, which makes the real-time ET data processing possible.

**Availability and implementation:** The code is available at <https://github.com/ichrm/auto-tilt-pair>. Additionally, the detailed original figures demonstrated in the experiments can be accessed at <https://rb.gy/6adt4>.

**Contact:** [xin.gao@kaust.edu.sa](mailto:xin.gao@kaust.edu.sa) or [guojunsdu@gmail.com](mailto:guojunsdu@gmail.com)

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

## 1 Introduction

Electron tomography (ET) is a powerful and indispensable tool to solve the three-dimensional (3D) ultrastructure (Frank, 2006), by reconstructing from a series of micrographs (tilt series) taken with different tilt angles. The development of ET has bridged the resolution gap between cellular imaging and high-resolution ultrastructure analysis (Rigort et al., 2010; Wan and Briggs, 2016). Especially, the recent application of ET in subtomogram averaging has advanced the limits of *in situ* ultrastructure analysis, which may lead to the next revolution in structural biology (Himes and Zhang, 2018).

The reconstruction of a high-quality ultrastructure relies on the consistency between the 3D projection model and the real-world projections. Before reconstruction, an accurate tilt series alignment is required. Usually, gold beads are used as fiducial markers to assist the alignment (Brandt and Ziese, 2006; Castaño-Díez et al., 2007; Kremer et al., 1996; Lawrence, 1992). So far, fiducial marker-based alignment is still the most widely used alignment method in high-resolution ET (Mastrorarde and Held, 2017).

Finding the correspondence of the fiducial markers on different tilted micrographs is the first and most important step in tilt series alignment. After many years of effort, several automated methods have been proposed: RAPTOR (Amat et al., 2008) uses the Markov

random field (MRF) to encode the positions of fiducial markers and utilizes the Loopy belief propagation theory to establish the correspondence. The defect of this method is the high computational cost in building the probabilistic model and the high failure rate. Later, Han et al. (2015) utilizes a random sampling method to determine the correspondence. Though the run time has been significantly reduced compared with RAPTOR, the method still faces an increasing computational cost when the number of fiducial markers increases. IMOD makes an automatization for its alignment workflow by first detecting the fiducial markers from the low tilt micrograph and then propagating the correspondence with a nearest neighbor search (Mastrorade and Held, 2017). Nevertheless, a pre-alignment of the tilt series is necessary and the method is not applicable when a significant transformation happens. With the proof of the error bound in a fiducial marker tracking model, a Gaussian mixture model (GMM) based fast-tracking method is proposed (Han et al., 2018). In the fast-tracking model, the correspondence is determined by minimizing the Bayesian probability in fiducial marker assignment. However, the method is not robust to the undesired outliers and relatively large transformation. Therefore, a robust, accurate and automatic fiducial marker correspondence is still one of the scientific challenges in the field (Jensen, 2019).

Recently, non-linear alignment and reconstruction have been further proposed in ET (Fernandez et al., 2018; Lawrence et al., 2006). According to the most recent research, sample deformation and beam-induced motion have been underestimated (Fernandez et al., 2018; Zheng et al., 2017). Fernandez et al. (2019)'s work has clearly demonstrated the warping during sample tilt and how the deformation correction benefits the reconstruction. However, there are only very few works that take the effect of non-uniform distortion into consideration in fiducial marker tracking.

In this article, we propose to divide the determination of fiducial marker correspondence into two stages: (i) initial transformation determination and (ii) local correspondence refinement. A novel two-stage algorithm by considering the local geometric constraint is proposed to ensure the efficient fiducial marker tracking while keeping robustness and accuracy. In the first stage, we model the initial transformation estimation as a correspondence pair inquiry and verification problem. The local invariant features are used for the fast comparison of the geometric similarity between different fiducial marker positions on different micrographs. Within each similarity inquiry, we limit the range of local feature extraction and propose a multiple-check technique of the extracted features for fast consistency verification. In the second stage, we encode the geometric distribution of the fiducial markers by a weighted Gaussian mixture model and solve the correspondence of distributions by an expectation-maximization algorithm, where the parameters about non-uniform drift are introduced to correct the potential distortion. The obvious outliers are pre-excluded based on the solved initial transformation.

With the two-stage algorithm, our aim is to completely solve the fiducial marker correspondence and tracking problem in ET. The utilization of local geometric constraints further reduces the computational complexity and the introduction of drift parameters ensures the accuracy in fiducial marker correspondence determination. Comprehensive experiments on real-world datasets demonstrate the robustness, efficiency and effectiveness of the proposed algorithm. Especially, the proposed algorithm is able to produce an accurate tracking within an average of  $\leq 100$  ms per image, even for micrographs with hundreds of fiducial markers, which is more than  $40\times$  faster than the state-of-the-art methods, while achieving  $\sim 99\%$  detection accuracy.

## 2 Materials and methods

### 2.1 Problem formulation

First, we would like to redefine the problem under the terminology of point set registration:

Denoting the positions of fiducial markers from a micrograph as the fixed 'scene' point set  $\mathcal{X} = \{x_n\}_{n=1,\dots,N}$  and the positions of

fiducial markers from another micrograph as the moving 'model' point set  $\mathcal{Y} = \{y_m\}_{m=1,\dots,M}$ , the problem of the correspondence determination is to find a transform  $\mathcal{T}(\cdot)$  so that there is a subset of  $\mathcal{T}(\mathcal{Y})$  with the maximum cardinality that aligns the points from a subset of the fixed 'scene' set  $\mathcal{X}$  under a selected measure of distance or similarity.

Figure 1 gives a schematic illustration of the determination of fiducial marker correspondence between two tilted micrographs.

### 2.2 Fast estimation of initial transformation

The fiducial marker correspondence upon micrographs with different tilt angles approximately follows an affine relationship (Han et al., 2018). For a single point  $p = [x, y]^T$ , the affine transformation is defined as

$$\mathcal{T}(p'; A, t) = Ap + t, \quad (1)$$

where  $A$  is a  $2 \times 2$  affine matrix and  $t$  is a  $2 \times 1$  translation vector. Without considering the sample deformation, an affine transformation is able to describe the correspondence of the fiducial markers extracted from two different tilted micrographs (Fig. 1E). Given a 'scene' point set with  $N$  points and a 'model' point set with  $M$  points, the time complexity is  $O(M^3N^3)$  for a baseline method (Three pairs of corresponding points are needed for the estimation of an affine transformation with six free parameters. A simple algorithm to find three corresponding points is: (i) choose three points from the 'scene' and 'model' point sets respectively; (ii) make an enumeration of the combination of these three points, estimate a possible transformation and finding if it could produce reasonable mapping; (iii) repeat the previous steps until a transformation that maximizes the congruent subset is found. It can be found that this procedure has two loops for 3-point combination, which requires  $A_N^3 A_M^3$  operations in average, resulting in an  $O(M^3N^3)$  complexity.). However, such complexity is prohibitively high considering the practically large number of fiducial markers and the number of micrographs.

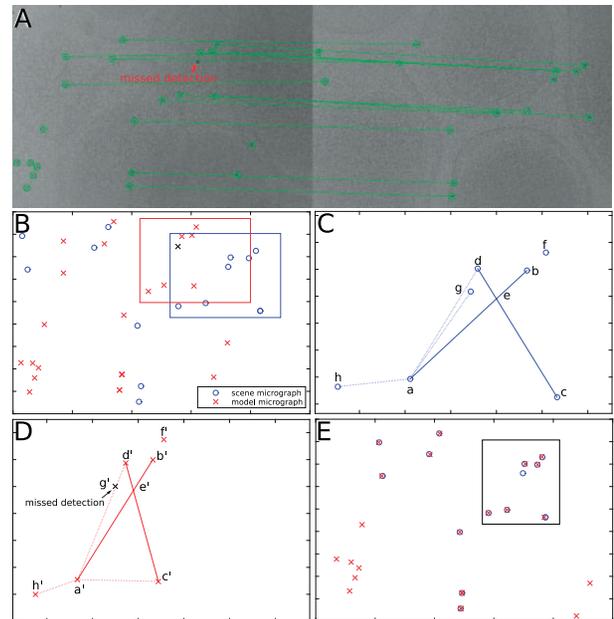


Fig. 1. (A) The detected fiducial markers and correspondences between two micrographs, where the left is with a high tilt angle and the right is with a low tilt angle. (B) Superimposition of fiducial marker positions extracted from the micrographs. (C) The geometric constraint within the blue rectangle. (D) The geometric constraint within the red rectangle. (E) Superimposition of fiducial marker positions after an affine transformation has been applied to the 'model' points

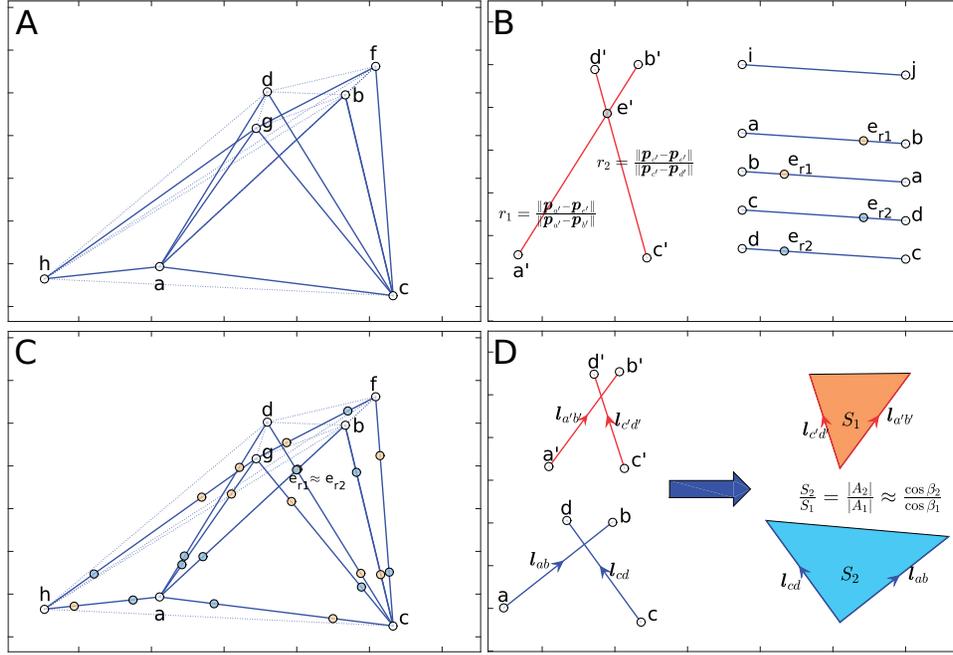


Fig. 2. Fast inquiry and verification for a certain 4-point invariant feature. Given the minimum and maximum limitation of 4-point feature's diagonal, a pruned virtual topology could be built on the 'scene' point set and the inquiry cost will be reduced when a 4-point feature of the 'model' point set comes. (A) A pruned virtual topology generated from the point set shown in Figure 1C, with the setting of minimal diagonal length  $l_{min} = l_{gb}$  and maximal length  $l_{max} = l_{bg}$ . (B) A 4-point feature chosen from the point set shown in Figure 1D, with the corresponding invariant ratio  $r_1$  and  $r_2$ . (C) The inquiry of a consistent 4-point feature on the 'scene' point set. For each pair of points  $\{i, j\}$ , two kinds of possible intersection ( $e_{r_1}$  and  $e_{r_2}$ ) are calculated. A conflict of  $e_{r_1}$  and  $e_{r_2}$  indicates a possible candidate of the corresponding 4-point feature. (D) The area ratio constraint underlies the two corresponding 4-point features

### 2.2.1 Local constrained 4-point invariant feature

Here, we propose a novel strategy, which combines the 4-point affine invariant feature (Aiger et al., 2008; Koenderink and van Doorn, 1991) with local constraints and invariant area ratio, to reduce the parameter searching space.

**4-point affine invariant feature:** For a point set  $\mathcal{P} = \{p_a, p_b, p_c, p_d\}$  in which line  $l_{ab} = p_a + \alpha(p_b - p_a)$  intersect with line  $l_{cd} = p_c + \beta(p_d - p_c)$  at point  $p_e$ , the ratios  $r_1 = \frac{\|p_e - p_c\|}{\|p_e - p_b\|}$  and  $r_2 = \frac{\|p_e - p_d\|}{\|p_e - p_a\|}$  are preserved under any affine transformation. The 4-point set  $\mathcal{P}$  along with the ratios  $r_1$  and  $r_2$  compose the 4-point invariant feature (A brief proof is provided in Supplementary S1.1).

Figure 1C and D demonstrate an example of the 4-point invariant feature, where Figure 1C shows the fiducial markers of the low tilt micrograph within the blue rectangle of Figure 1B, and D shows the correspondences of the high tilt micrograph within the red rectangle. By considering the intersections, point set  $\{a, b, c, d\}$  forms a 4-point invariant feature in Figure 1C and point set  $\{a', b', c', d'\}$  forms the congruent invariant feature in Figure 1D. These two 4-point subsets could be fitted into each other within a suitable transformation (Fig. 1E).

**Local geometric constraint and numerical stability:** The fast inquiry of a 4-point invariant feature is still a problem for large point sets (Aiger et al., 2008). Various local geometric constraints have been proposed to reduce the complexity, within which the nearest neighbor constraint is the most popular one (Amat et al., 2008; Hauer et al., 2013; Vilas et al., 2016). Though the nearest neighbor constraint is able to prune the topology and reduce the complexity, the nearest neighbor itself, however, can be easily corrupted by outliers and distance changes. Figure 1C and D shows such an example, in which a 3-neighbors local linear embedding (LLE) system is built for point  $a$ . Unfortunately, affected by the missed detection of  $g'$ , we get points  $g, d$  and  $b$  as the 3-neighbors for point  $a$  in Figure 1C but points  $d', b'$  and  $c'$  as the 3-neighbors for its correspondence in Figure 1D.

Here, we define lines  $l_{ab}$  and  $l_{cd}$  as the *diagonal* of an invariant feature  $\mathcal{P}$ , and propose to extract the 4-point invariant feature

within the length of diagonal constraints instead of the nearest neighbor geometry:

1. Build a nearest neighbor search tree (Bentley, 1975) to get the set of the distance  $\{l_k\}$  between each point in a point set  $\mathcal{X}$  with its nearest neighbor.
2. Get the average  $l_{avg}$  and standard deviation  $l_{sdv}$  of  $\{l_k\}$ .
3. Define a diagonal length for a 4-point invariant feature to be no more than  $l_{max} = 3\sqrt{2}l_{avg} + l_{sdv}$ .
4. Get all the possible point pairs within the point set  $\mathcal{X}$ , and only select the point pairs with length no more than  $l_{max}$  to produce the 4-point invariant feature.

By doing this, we have limited the extraction of 4-point invariant feature within a local constrained area (if the distribution of fiducial markers is approximately even, we will get all the 4-point invariant features within about 10-neighbors).

We further exclude the 4-point invariant feature with too small diagonal lengths: if the localization error for a fiducial marker is 5 pixels, a 4-point invariant feature with a diagonal length of 20 pixels will suffer from  $5/20 = 25\%$  inaccuracy in the invariant ratio, while a 4-point invariant feature with a diagonal length of 200 pixels will only suffer from  $5/200 = 2.5\%$  inaccuracy. The minimum diagonal length  $l_{min}$  is related to the numerical stability of a given coordinate system, which depends on the nanoscale of fiducial markers and the pixel size of the micrograph. Here, we call this local geometry constraint the *min&max rule* for the 4-point invariant feature.

Figure 2A shows an example of the 'min&max rule', where we set  $l_{min} = l_{gb}$  and  $l_{max} = l_{bg}$  for the convenience of demonstration. Here, for the points selected in Figure 1C, we have calculated all the possible point pairs  $\{l_{p_i p_j} | p_i, p_j \in \mathcal{X}\}$ . Then, the point pairs with  $\|l_{p_i p_j}\| > l_{max}$  or  $\|l_{p_i p_j}\| < l_{min}$  are discarded. Consequently, a pruned virtual topology is built based on the remaining point pairs. As shown in Figure 2A, about half of the point pairs that violate the min&max rule have been pruned from the topology, being plotted in dash lines. Furthermore, because of the min&max rule, there are

few connections between the points depicted in Figure 2A and the other points outside the selected region, which limits the combination of the geometric features within a local area.

### 2.2.2 Fast inquiry and verification of invariant feature

We could extract a ‘query’ subset from the ‘model’ point set  $\mathcal{Y}$  and find its congruent subset in the ‘scene’ point set  $\mathcal{X}$ . If we obtain an affine transformation from the inquiry that covers enough points after being applied to the ‘model’ point set, we may have found the correct initial transformation.

**Partial inquiry for a certain feature:** For a 4-point invariant feature  $\mathcal{P} \in \mathcal{Y}$  with invariant ratios  $r_1$  and  $r_2$ , we try to check its correspondence in  $\mathcal{X}$  based on the supposed position of its intersection  $e$ .

If a point pair  $l_{ij}$  corresponds to one of  $\mathcal{P}$ ’s diagonals, it has

$$\begin{cases} \mathbf{p}_{e_{r_1(ij)}} = \mathbf{p}_i + r_1(\mathbf{p}_j - \mathbf{p}_i) \text{ or } \mathbf{p}_{e_{r_1(ij)}} = \mathbf{p}_j + r_1(\mathbf{p}_i - \mathbf{p}_j), \\ \mathbf{p}_{e_{r_2(ij)}} = \mathbf{p}_i + r_2(\mathbf{p}_j - \mathbf{p}_i) \text{ or } \mathbf{p}_{e_{r_2(ij)}} = \mathbf{p}_j + r_2(\mathbf{p}_i - \mathbf{p}_j), \end{cases} \quad (2)$$

where  $e_{r_1(ij)}, e_{r_1(ji)}$  represent the two types of intersection  $e_{r_1}$  defined by  $r_1$ , and  $e_{r_2(ij)}, e_{r_2(ji)}$  represent the two types of intersection  $e_{r_2}$  defined by  $r_2$  (an illustration is shown in Fig. 2B). Consequently, to find  $\mathcal{P}$ ’s corresponding 4-point feature in  $\mathcal{X}$ , we may build a nearest search tree of  $\{e_{r_2}\}$  and search  $e_{r_1}$  on the nearest search tree.

Figure 2C shows the detailed positions of the candidates  $e_{r_1}$  and  $e_{r_2}$  on the virtual topology, and how a candidate  $e_{r_1}$  conflicts with a candidate  $e_{r_2}$  (to make the figure clear, only the first type of  $e_{r_1}$  and  $e_{r_2}$  is depicted). Once a conflict of the query is detected, the corresponding subset will be selected as a candidate correspondence to the inquiry subset.

It can be found that the partial inquiry procedure will take  $O(N^2)$  time complexity even without the constraint of local geometry, which is much faster than the  $O(N^3)$  complexity method used by the 3-point combination in a baseline algorithm.

**Fast consistency verification:** Given a query subset as the 4-point invariant feature, multiple candidate correspondences may exist in the ‘scene’ point set. To retrieve the invariant feature that exactly corresponds to the query, the affine transformation should be estimated and tested on the entire point set, which is still time-consuming. Here, we introduce another affine constraint, the constraint of the area ratio, to simplify the verification:

Lemma Given a micrograph with tilt angle  $\beta_1$  and another one with tilt angle  $\beta_2$ , the area of the corresponding plane shapes on these two micrographs have an approximate ratio of  $\cos \beta_1 / \cos \beta_2$ .

The interested readers may refer to Supplementary S1.2 for the proof of the lemma. Figure 2D shows how the area ratio constraint is applied to the 4-point invariant feature. Here, it should be noted that the area ratio constraint is rotation and translation invariant. A relaxation parameter  $\xi$  is further defined. Given a candidate correspondence  $\mathcal{P}'$  of the 4-point invariant subset  $\mathcal{P}$ ,  $\mathcal{P}'$  is accepted if and only if  $(1 - \xi) \frac{\cos \beta_1}{\cos \beta_2} \leq \frac{S_{\mathcal{P}'}}{S_{\mathcal{P}}} \leq (1 + \xi) \frac{\cos \beta_1}{\cos \beta_2}$ . **Time complexity of the algorithm:** Algorithm 1 summarizes the proposed fast inquiry and verification algorithm, where the input is a 4-point invariant feature  $\mathcal{P}$  extracted from the moving ‘model’ point set  $\mathcal{Y}$ , the pruned point pairs  $\ell$  extracted from the moving ‘scene’ point set  $\mathcal{X}$ , the value of the two tilt angles  $\beta_1, \beta_2$  and two parameters  $\delta, \xi$  used to define the inquiry accuracy.  $\mathcal{E}_{r_1}$  and  $\mathcal{E}_{r_2}$  are the point sets generated from  $\ell$  with ratios  $r_1$  and  $r_2$ , and  $\Theta$  is the candidate 4-point subset that correspond to  $\mathcal{P}$ .  $\mathcal{KD}(\cdot)$  is the operation to build a k-d tree served for the nearest neighbor search (Bentley, 1975) and  $dist(\cdot)$  is the operation to calculate the Euclidean distance between two points.

Because point pairs  $\ell$  have been pruned by the min&max rule, the cardinality  $K$  of  $\ell$  should be  $\ll N^2$  ( $N$  is the cardinality of  $\mathcal{X}$ ). Then, the calculation of possible intersections has  $O(K)$  time complexity, building the nearest search tree has  $O(K \log(K))$  time complexity, and the query of the 4-point invariant subsets that correspond to  $\mathcal{P}$  takes  $O(K \log(K))$  time complexity ( $O(\log(K))$  for each query and  $O(K)$  for the number of tries). Therefore, the total complexity of the algorithm is  $O(K + K \cdot \log(K))$ , which is far smaller than  $O(N^2 \log(N))$ .

---

#### Algorithm 1: Fast inquiry of a 4-point invariant feature

---

**Input:** 4-point subset  $\mathcal{P} \in \mathcal{Y}$ , point pairs  $\mathcal{L} = \{\mathbf{l}_{p_i p_j} | p_i, p_j \in \mathcal{X}\}$

```

1 Procedure FastInquiry( $\mathcal{P}, \mathcal{L}, \beta_1, \beta_2, \delta, \xi$ )
2    $\Theta \leftarrow \emptyset, r_\beta \leftarrow \frac{\cos \beta_1}{\cos \beta_2}$ ;
3   Get the ratio  $r_1$  and  $r_2$  from  $\mathcal{P}$ ;
4   foreach  $\mathbf{l}_{p_i p_j} \in \mathcal{L}$  do
5     Calculate the possible intersection  $e_{r_1(ij)}, e_{r_1(ji)}, e_{r_2(ij)}$  and  $e_{r_2(ji)}$ 
     by Equation 2 with ratio  $r_1$  and  $r_2$ ;
6   end
7   Let  $\mathcal{E}_{r_1} \leftarrow \{e_{r_1(ij)}\} \cup \{e_{r_1(ji)}\}$  and  $\mathcal{E}_{r_2} \leftarrow \{e_{r_2(ij)}\} \cup \{e_{r_2(ji)}\}$ ;
8   Create search tree  $\mathcal{KD}(\mathcal{E}_{r_2})$ ;
9   foreach  $e_{r_1(ij)} \in \mathcal{E}_{r_1}$  do
10    Inquire  $e_{r_1(ij)}$  in  $\mathcal{KD}(\mathcal{E}_{r_2})$ , get point  $e_{r_2(kl)}$  nearest to  $e_{r_1(ij)}$ ;
11    if  $dist(e_{r_1(ij)}, e_{r_2(kl)}) < \delta$  then
12       $\mathcal{P}_X \leftarrow \{p_i, p_j, p_l, p_k\}$ ;
13      if  $(1 - \xi) \cdot r_\beta \leq \frac{S_{\mathcal{P}'}}{S_{\mathcal{P}}} \leq (1 + \xi) \cdot r_\beta$  then
14         $\Theta \leftarrow \Theta \cup \mathcal{P}_X$ ;
15      end
16    end
17  end
18  Return the candidate congruent subsets  $\Theta$ .
```

---

### 2.2.3 Robust estimation of the affine transformation

Considering the effect of noise and the missed detection of fiducial markers, multiple trials of congruent subset inquiry is necessary. Here, we adopt a random sample consensus (RANSAC) procedure (Fischler and Bolles, 1981) to robustly estimate the affine transformation between the two point sets.

Based on the min&max rule, the sets of range limited point pairs  $\ell_X$  and  $\ell_Y$  are generated from  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. Then, the following operations are carried out:

1. Compose a random 4-point set  $\mathcal{P}$  by the point pairs within  $\mathcal{Y}$ , and get its candidate congruent subsets  $\Theta$  in  $\mathcal{X}$  by the fast inquiry algorithm.
2. For each congruent subset  $\mathcal{P}_X$  within  $\Theta$ , calculate the possible transform  $\mathcal{T}(\cdot; \mathbf{A}, \mathbf{t})$  by the least square estimation. Apply the transform to  $\mathcal{Y}$  and count the number of points in  $\mathcal{T}(\mathcal{Y})$  that are close enough (congruent) to the points in  $\mathcal{X}$ .
3. If the number of congruent points between  $\mathcal{T}(\mathcal{Y})$  and  $\mathcal{X}$  is larger than the current maximal record  $c_{max}$ , save the current value as  $c_{max}$  and update the termination condition  $I$ .
4. Repeat Steps 1 ~4 until termination.

Algorithm 2 elaborates the details of the algorithm, where the algorithm accepts the point sets  $\mathcal{X}, \mathcal{Y}$  and the related tilt angles  $\beta_1, \beta_2$  as input, the distance threshold  $d$ , the area ratio relaxation  $\xi$  as parameters, and outputs the estimated transformation  $\mathcal{T}(\cdot; \mathbf{A}, \mathbf{t})$ . The distance threshold  $d$  can be set according to the value of fiducial marker diameter and the relaxation  $\xi$  can be estimated based on the experiments of mechanical instability. Generally,  $d$  is set to  $\lambda \cdot D$ , where  $D$  is the fiducial marker diameter and  $0.5 < \lambda < 0.8$ . Because the fiducial markers are sparsely distributed on the specimen, the value of  $\lambda$  is not essential to the system. In the algorithm, we denote the operation to count the corresponding points between  $\mathcal{X}$  and  $\mathcal{Y}$  under distance threshold  $d$  as  $(\mathcal{X}, \mathcal{Y}, d)$ . Because the transformation of  $\mathcal{Y}$  needs a sequence of matrix multiplication, a further optimization is to randomize the verification of congruent points, by fast verifying a constant number of random points first and then the whole dataset. The maximum iteration is initialized and updated according to  $I = \log(1 - p_s) / \log(1 - p_g^k)$ , where  $p_s$  is the required success probability,  $p_g$  is the percentage of traceable fiducial markers that appear in both  $\mathcal{X}$  and  $\mathcal{Y}$ , and  $k$  is set to 4 as the inquiry of 4-point invariant subset.

**Algorithm 2:** Randomized estimation of the transform

---

**Input:** Point sets  $\mathcal{X}$  and  $\mathcal{Y}$ , distance threshold  $d$

**1 Procedure** RandomizedEstimation( $\mathcal{X}, \mathcal{Y}, \beta_1, \beta_2, d, \xi$ )

2 Get the  $l_{min}$  and  $l_{max}$  as described in Section 2.2.1;

3  $\mathcal{L}_X \leftarrow \{l_{p_i, p_j} | p_i, p_j \in \mathcal{X} \wedge l_{min} < \|l_{p_i, p_j}\| \leq l_{max}\};$

4  $\mathcal{L}_Y \leftarrow \{l_{p'_i, p'_j} | p'_i, p'_j \in \mathcal{Y} \wedge l_{min} < \|l_{p'_i, p'_j}\| \leq l_{max}\};$

5 Calculate the maximum iteration  $I$ ;

6 Let  $i \leftarrow 0, c_{max} \leftarrow 0$ ;

7 **while**  $i < I$  **do**

8 Randomly get two intersected line  $l_{p'_i, p'_j}, l_{p'_k, p'_l} \in \mathcal{L}_Y$ , so that

9  $\mathcal{P} \leftarrow \{p'_i, p'_j, p'_k, p'_l\};$

10  $\Theta \leftarrow \text{FastInquiry}(\mathcal{P}, \mathcal{L}_X, \beta_1, \beta_2, \frac{d}{2}, \xi);$

11 **foreach**  $\mathcal{P}_X \in \Theta$  **do**

12 Estimate  $\mathcal{T}(\cdot; A, t)$  from  $\mathcal{P}$  to  $\mathcal{P}_X$ ;

13 Apply  $\mathcal{T}(\cdot; A, t)$  to  $\mathcal{Y}$ ;

14  $c \leftarrow \text{Congr}(\mathcal{T}(\mathcal{Y}; A, t), \mathcal{X}, d)$ ;

15 **if**  $c > c_{max}$  **then**

16  $c_{max} \leftarrow c$ ;

17 Refresh the maximum iteration  $I$ ;

18 **end**

19 **end**

20  $i \leftarrow i + 1$ ;

21 **end**

**Return** the estimated affine transformation  $\mathcal{T}(\cdot; A, t)$ .

---

### 2.3 Local correspondence refinement

By considering the local geometric constraint, the optimized randomized algorithm is able to produce a fast and robust estimation of the initial transformation. However, the distortion caused by sample deformation or beam-induced motion is non-negligible (Fernandez *et al.*, 2019; Lawrence *et al.*, 2006; Zheng *et al.*, 2017), which may corrupt the affine relationship of two micrographs within a local area and lead to spurious correspondence.

Here, we first correct the large deviation and outliers in  $\mathcal{Y}$  by the affine transformation  $\mathcal{T}(\cdot; A, t)$ , and then feed the transformed ‘model’ point set and ‘scene’ point set to the algorithm for the second stage (for the concision of text, we still denote the corrected point set  $\mathcal{T}(\mathcal{Y}; A, t)$  by  $\mathcal{Y}$  in the following discussion.).

#### 2.3.1 GMM interpolation for the scattered points

For the efficiency and accuracy, the data interpolation and transformation refinement are carried out by the non-rigid coherent point drift, based on the Gaussian mixture model (GMM) (Jian and Vemuri, 2011; Myronenko and Song, 2010).

Given the fixed ‘scene’ point set  $\mathcal{X} = \{x_n\}_{n=1, \dots, N}$  and the moving ‘model’ point set  $\mathcal{Y} = \{y_m\}_{m=1, \dots, M}$ , the probability that a point  $x \in \mathcal{X}$  is corresponding to a point  $y \in \mathcal{Y}$  can be described by an isotropic Gaussian function:

$$p(x|y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right), \quad (3)$$

where  $\sigma$  is a parameter to describe the instability of the system. Similarly, the probability that the point  $x$  belongs to the point set  $\mathcal{Y}$  can be defined as  $p(x|\mathcal{Y}) = \sum_{m=1}^M P(m)p(x|y_m)$ , where  $P(m)$  is the prior of  $x$  under the condition of the  $m$ th point  $y_m$ .

Considering a drift transform  $\mathcal{T}(\cdot; v)$  applied to the model for distortion correction, and assuming the point  $x$  either belongs to outliers with  $w$  probability or sampled from the point set  $\mathcal{Y}$  with a uniform distribution, the GMM probability density function can be defined as:

$$\begin{aligned} p(x|v_m) &= w \frac{1}{N} + (1-w) \sum_{m=1}^M \frac{1}{M} p(x|y_m), \\ &= w \frac{1}{N} + (1-w) \sum_{m=1}^M \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{\|x - \mathcal{T}(y_m; v_m)\|^2}{2\sigma^2}}, \end{aligned} \quad (4)$$

where  $v_m$  is the drift corresponding to the  $m$ th point  $y_m$ . Our aim is to find such a transformation  $\mathcal{T}(\cdot; v)$  and parameter  $\sigma$  applied to all the points  $y \in \mathcal{Y}$  so that the negative log-likelihood from  $\mathcal{Y}$  to  $\mathcal{X}$  is minimized:

$$E(\mathcal{T}(\cdot), \sigma^2) = - \sum_{n=1}^N \log \sum_{m=1}^{M+1} P(m)p(x|y_m), \quad (5)$$

where  $P(m)$  is the reweighted i.i.d prior and  $p(x|y_{M+1}) = \frac{w}{N}$  presents the probability of outliers.

#### 2.3.2 Transform parameter optimization

A negative log-likelihood objective function could be effectively solved by the expectation-maximization (E-M) optimization, where the algorithm iterates between the E-step and the M-step until convergence. With Jensen’s inequality (Jensen, 1906; Redner and Walker, 1984), the negative log-likelihood defined in Eq. 5 is upper bounded by the following function in each iteration:

$$\mathcal{Q} = - \sum_{n=1}^N \sum_{m=1}^{M+1} p(m|x_n) \log (P^{new}(m)p^{new}(x_n|y_m)), \quad (6)$$

where  $p(m|x_n) = P(m)p(x_n|y_m)/p(x_n)$  is the probability that  $y_m$  corresponds to  $x_n$ , the ‘old’ superscript indicates that a parameter is guessed in the E-step and the ‘new’ superscript indicates that a parameter is optimized from the negative log-likelihood function in the M-step.

**E-step:** By ignoring the constants independent of  $v$  and  $\sigma$ , we rewrite Eq. 6 as:

$$\mathcal{Q}(v, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M p(m|x_n) \|x_n - \mathcal{T}(y_m; v_m)\|^2 + N_p \log \sigma^2, \quad (7)$$

where  $N_p = \sum_{n=1}^N \sum_{m=1}^M p(m|x_n) \leq N$  (with  $N = N_p$  only if  $w = 0$ ), and  $p(m|x_n)$  denotes the posterior probabilities of GMM components calculated using the previous parameter values:

$$p(m|x_n) = \frac{\exp\left(-\frac{1}{2} \left\| \frac{x_n - \mathcal{T}(y_m; v_m)}{\sigma} \right\|^2\right)}{\sum_{k=1}^M \exp\left(-\frac{1}{2} \left\| \frac{x_n - \mathcal{T}(y_k; v_m)}{\sigma} \right\|^2\right) + 2\pi\sigma^2 \frac{w}{1-w} \frac{M}{N}}. \quad (8)$$

**M-step:** Here we model the transform  $\mathcal{T}(\cdot; v)$  to correct the local distortion but not global affine transformation. For the convenience of further discussion, here we introduce the following notations:

1.  $X_{N \times 2} = (x_1 \cdots x_N)^T$ —matrix presentation of the point set  $\mathcal{X}$ ;
2.  $Y_{M \times 2} = (y_1 \cdots y_M)^T$ —matrix presentation of the point set  $\mathcal{Y}$ ;
3.  $\mathbf{1}$ —the column vector of all ones;
4.  $d(a)$ —the diagonal matrix formed from vector  $a$ ;
5.  $P$ —the matrix that is composed by  $p_{mn} = p(m|x_n)$ .

Based on the Tikhonov regularization framework (Chen and Haykin, 2002), a non-rigid parameterization for  $\mathcal{T}(\cdot; v)$  is adapted to minimize the objective function  $\mathcal{Q}$ :

$$\mathcal{T}(Y; v) = Y + v(Y), \quad (9)$$

where  $v(\cdot)$  is the expected drift. A regularization term  $\frac{\lambda}{2} \phi(v)$  is added to Equation 7 to enforce the smoothness and compensate for the drift, and  $\lambda$  is a trade-off parameter.

If we model the regularization function  $\phi(v) = \|Lv\|^2$  within a Kernel Hilbert Space (RKHS) (Chen and Haykin, 2002; Myronenko and Song, 2010), the negative log-likelihood function in Equation 7 will be defined as:

$$Q(v, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M p(m|x_n) \|x_n - (y_m + v(y_m))\|^2 + N_p \log \sigma^2 + \frac{\lambda}{2} \|Lv\|^2. \quad (10)$$

To minimize Equation 10, we should find a function  $v(\cdot)$  for all the elements  $y_m$  of  $\mathbf{Y}$  under the Euler-Lagrange differential equation:

$$\frac{1}{\sigma^2 \lambda} \sum_{n=1}^N \sum_{m=1}^M p(m|x_n) (x_n - (y_m + v(y_m))) \Delta y_m = \hat{L}Lv(\Delta y_m), \quad (11)$$

where  $\hat{L}$  is the adjoint operator to  $L$ . By rewriting the equation and using an integral of a Green's function  $g(x, y) = e^{-\frac{\lambda}{2} \|x-y\|^2}$  instead of the self-adjoint operator, the solution  $v(\cdot)$  of such a partial differential equation has the form of

$$\begin{aligned} v(z) &= \frac{1}{\sigma^2 \lambda} \sum_{n=1}^N \sum_{m=1}^M p(m|x_n) (x_n - (y_m + v(y_m))) g(z, y_m) \\ &= \sum_{m=1}^M w_m g(z, y_m), \end{aligned} \quad (12)$$

where  $w_m = \frac{1}{\sigma^2 \lambda} \sum_{n=1}^N p(m|x_n) (x_n - (y_m + v(y_m)))$ . To further get the values of  $v(\cdot)$ , we could solve each  $w_m$  first by evaluating Eq.12 at  $y_m$ :

$$(G + \lambda \sigma^2 d(P1)^{-1})W = d(P1)^{-1}PX - Y, \quad (13)$$

where  $\mathbf{W} = (w_1, \dots, w_M)^T$ ,  $\mathbf{G}$  is an  $M \times M$  kernel matrix with elements  $g_{ij} = g(y_i, y_j)$ , and  $d(\cdot)^{-1}$  is the inverse diagonal operation.

Consequently, the transform  $T = T(Y; \mathbf{W}) = Y + \mathbf{G}\mathbf{W}$ . By substituting  $T$  back into  $Q$  and solving the partial derivation,  $\sigma^2$  is updated according to the result of  $\frac{\partial Q}{\partial \sigma^2}$  as

$$\sigma^2 = \frac{1}{2N_p} \left( \text{tr}(\mathbf{X}^T d(\mathbf{P}^T \mathbf{1}) \mathbf{X}) - 2\text{tr}((\mathbf{P}\mathbf{X})^T \mathbf{T}) + \text{tr}(\mathbf{T}^T d(\mathbf{P}\mathbf{1}) \mathbf{T}) \right). \quad (14)$$

The overall iterative optimization for the GMM-based drift correction is summarized in Algorithm 3. After the update of the positions by  $T(Y, \mathbf{W})$ , we could recalculate the correspondence between point sets  $\mathcal{Y}$  and  $\mathcal{X}$  under a given distance threshold  $d$ . With the initial transformation inherited from the first stage, Algorithm 3 is able to quickly converge. At the same time, because the initial transformation is very close to the global optimum, Algorithm 3 will converge to the global optimum in a high probability.

---

**Algorithm 3:** Correspondence refinement by local drift correction

---

**Input:** Vectorized point set  $\mathbf{X}$  and  $\mathbf{Y}$

- 1 **Procedure** LocalRefinement( $\mathbf{X}, \mathbf{Y}, \beta, \lambda$ )
- 2 Initialize  $\mathbf{G}$  with each  $g_{ij} \leftarrow \exp(-\frac{1}{2\beta^2} \|y_i - y_j\|^2)$ ;
- 3  $\sigma^2 \leftarrow \frac{1}{2MN} \sum_{m,n=1}^{MN} \|x_n - y_m\|^2$ ,  $\mathbf{W} \leftarrow 0$ ;
- 4 **repeat**
- 5     **foreach**  $p_{mn} \in P$  **do**
- 6          $p_{mn} \leftarrow \frac{\exp(-\frac{1}{2} \|\frac{x_n - G(m, \cdot) \mathbf{W}}{\sigma}\|^2)}{\sum_{k=1}^M \exp(-\frac{1}{2} \|\frac{x_n - G(m, \cdot) \mathbf{W}}{\sigma}\|^2) + 2\pi\sigma^2 \frac{w}{1-w} \frac{M}{N}}$ ;
- 7     **end**
- 8     Update objective function  $Q$ ;
- 9     Solve  $\mathbf{W}$  in  $(\mathbf{G} + \lambda\sigma^2 d(\mathbf{P}\mathbf{1})^{-1})\mathbf{W} = d(\mathbf{P}\mathbf{1})^{-1} \mathbf{P}\mathbf{X} - \mathbf{Y}$ ;
- 10      $N_p \leftarrow \mathbf{1}^T \mathbf{P}\mathbf{1}$ ,  $\mathbf{T} \leftarrow \mathbf{Y} + \mathbf{G}\mathbf{W}$ ;
- 11      $\sigma^2 \leftarrow \frac{1}{2N_p} (\text{tr}(\mathbf{X}^T d(\mathbf{P}^T \mathbf{1}) \mathbf{X}) - 2\text{tr}((\mathbf{P}\mathbf{X})^T \mathbf{T}) + \text{tr}(\mathbf{T}^T d(\mathbf{P}\mathbf{1}) \mathbf{T}))$ ;
- 12 **until** the convergence of  $Q$ ;
- 13 **Return** The corrected position  $\mathcal{T}(\mathbf{Y}, \mathbf{W}) = \mathbf{Y} + \mathbf{G}\mathbf{W}$ .

---

### 3 Experiments and results

#### 3.1 Datasets

Six real-world datasets are used to evaluate the proposed method. The first dataset is a tilt series that has been used in the previous studies, provided by the Institute of Biophysics, Chinese Academy of Sciences (Han et al., 2018). The remaining five datasets are downloaded from the Caltech ETDB (Ortega et al., 2019).

The first dataset, Hemocyanin, is a tilt series of vitrified key-hole limpet hemocyanin solution (Fig. 3A). It is a cryo-ET dataset with about 100 ~150 fiducial markers embedded in. The tilt series were collected by FEI Titan Krios (300kV) with a Gatan US4000 camera. The total dose used during data collection was around 8000 e/nm<sup>2</sup>. There are 95 images with the tilt ranging from -70° to 70° at 1° ~ 2° intervals (2K × 2K pixels with 0.4 nm/px).

The second and third datasets, Vibrio1-2 (Vibrio1 is downloaded from <https://bit.ly/35HJ0WS>; Vibrio2 is downloaded from <https://bit.ly/3kGvl8l>), are two cryo-ET datasets of isolated Vibrio cholerae cells (Fig. 3B and C). Vibrio1 and Vibrio2 have about 150 ~200 and 200 ~250 fiducial markers embedded in the specimens, respectively. Both of the tilt series were collected by FEI Tecnai Polara (F30) (300kV) with a Gatan K2 camera, operated at 145eV/Å<sup>2</sup> dosage. There are 121 images with the tilt ranging from -60° to 60° at 1° interval (4K × 4K pixels with 0.4 nm/px).

The fourth to sixth datasets, Nitrosop1-3 (Nitrosop1 is downloaded from <https://bit.ly/3pCgsHL>; Nitrosop2 is downloaded from <https://bit.ly/36L2zyq>; Nitrosop3 is downloaded from <https://bit.ly/3kF1Qub>), are three cryo-ET datasets of isolated Nitrosopumilus maritimus cells (Fig. 3D-F). Nitrosop1 has about 250 ~300 fiducial markers embedded in the specimen, while both Nitrosop2 and Nitrosop3 have about 400 ~500 fiducial markers. The Nitrosop1-3 were collected by FEI Tecnai Polara (F30) (300kV) with a Gatan K2 camera. The Nitrosop1 was operated at 150eV/Å<sup>2</sup> dosage with -10 μm defocus; the Nitrosop2 and Nitrosop3 were operated at 180eV/Å<sup>2</sup> dosage. There are 121 images for Nitrosop1 with the tilt ranging from -60° to 60° at 1° interval (4K × 4K pixels with 0.64 nm/px), and 111 images for both Nitrosop2 and Nitrosop3 with the tilt angles ranging from -55° to 55° at 1° interval (4K × 4K pixels with 0.49 nm/px).

As our focus is on the determination of fiducial marker correspondence, all the fiducial markers on the micrographs have been detected in advance. Here, we used the sampling and classification algorithm proposed in *markerauto* (Han et al., 2015) to automatically and exhaustively detect the fiducial markers. Nevertheless, other sophisticated techniques can also be used to provide precise fiducial marker positions.

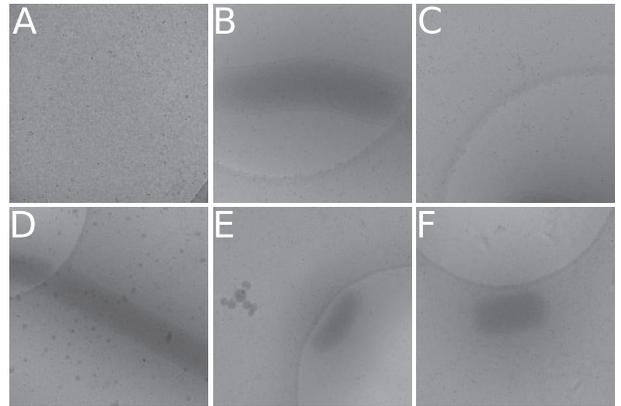


Fig. 3. Illustration of the test datasets. (A) Hemocyanin, (B) Vibrio1, (C) Vibrio2, (D) Nitrosop1, (E) Nitrosop2 and (F) Nitrosop3. Limited by the space, only the small thumbnails of the 0° micrographs are shown here. Please refer to Supplementary S2.1 for detailed information

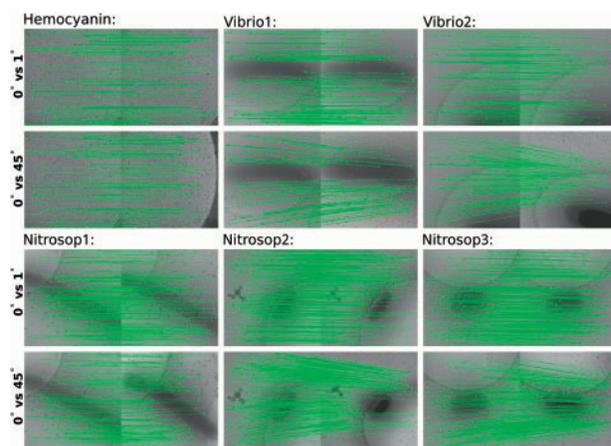


Fig. 4. A snapshot of the fiducial marker correspondence determined by the two-stage algorithm on micrographs with different tilt angle intervals (for the demonstrated micrograph pairs, the left is with  $0^\circ$  tilt angle and the right is with  $1^\circ$  or  $45^\circ$  tilt angle, respectively). Please refer to [Supplementary S2.2](#) for detailed information

## 3.2 Results

### 3.2.1 Robustness of the algorithm under various conditions

The robustness is very important for a fully automatic fiducial marker correspondence method. Here, we first test the robustness of the proposed method on the datasets with different micrograph pairs.

For each dataset, the tracking of fiducial markers on the micrographs with tilt angle intervals increasing from  $1^\circ$  to  $45^\circ$  is carried out by the two-stage algorithm. [Figure 4](#) demonstrates a snapshot of the tracking results, in which the experimental results on micrographs with tilt angles of  $0^\circ$  and  $1^\circ$ ,  $0^\circ$  and  $45^\circ$  are selected for illustration. It should be noted that, from the dataset Hemocyanin to Nitrosop3, the number of fiducial markers has increased from one hundred to more than five hundred. Nevertheless, for all the datasets with various tilt angle intervals, our algorithm successfully finds the correct correspondence.

Particularly, the proposed method could handle the conditions with numerous fiducial markers and a mass of outliers well. [Figure 5](#) illustrates the superimposed fiducial marker positions of the Nitrosop2 and Nitrosop3 datasets that are extracted from the micrographs with  $0^\circ$  and  $45^\circ$  tilt angles (labeled by blue ‘circle’ and red ‘dot’, respectively). The two-stage algorithm accepts the raw fiducial marker positions shown in [Figure 5A and C](#) as input and outputs the transformed fiducial marker positions as shown in [Figure 5B and D](#). There are 394 and 366 detected fiducial markers on the  $0^\circ$  tilted micrographs of the Nitrosop2 and Nitrosop3 datasets, and 513 and 452 detected fiducial markers on the  $45^\circ$  tilted micrographs of the Nitrosop2 and Nitrosop3 datasets, respectively. As shown in [Figure 5A and C](#), the fiducial markers from the  $0^\circ$  and  $45^\circ$  tilted micrographs are difficult to be directly corresponded together. Especially, for the micrographs with high tilt angles, the blurring and blocking of fiducial markers cause a mass of missed detection. In [Figure 5B and D](#), we use the green ellipses to indicate the missed detection of fiducial markers caused by the increasing dark shadows and use the red ellipses to indicate the introducing of outliers by the extension of the field of view. In total, there are more than one hundred fiducial markers appearing as outliers and hampering the estimation of transformation. However, after the execution of the algorithm, the proposed method has generated high-quality fiducial marker correspondence.

The proposed method is then compared with the state-of-the-art methods, including the probabilistic graphical model for robust point set registration ([Qu et al., 2017](#)), the GMM-based fast fiducial marker tracking model ([Han et al., 2018](#)) and the affine transform-based naive sampling method ([Han et al., 2015](#)), which are referred to as ‘VBPSM model’, ‘GMM model’ and ‘naive sampling’, respectively [Because RAPTOR ([Amat et al., 2008](#)) needs a pre-alignment of the tilt series and cannot be applied on micrographs with large tilt intervals (for example, intervals  $> 15^\circ$ ), we compare with the

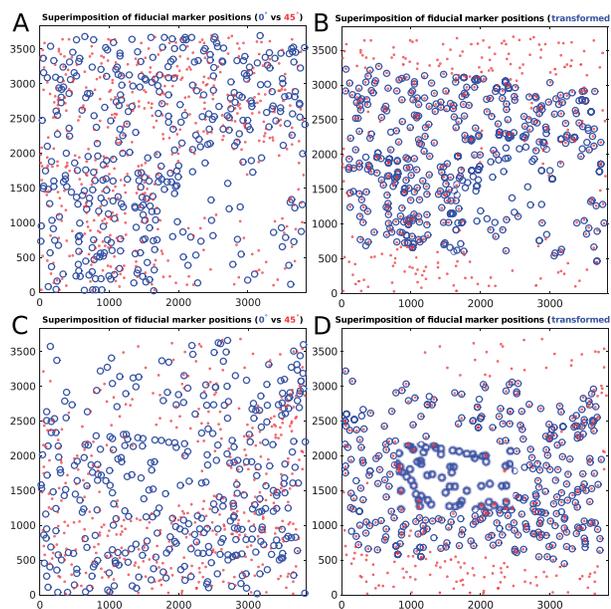


Fig. 5. Superimposition of fiducial marker positions from the  $0^\circ$  and  $45^\circ$  tilted micrograph before (the left) and after (the right) applying transformation produced by the two-stage algorithm. (A) and (B) fiducial marker positions extracted from the Nitrosop2 dataset. (C) and (D) fiducial marker positions extracted from the Nitrosop3 dataset

VBPSM model instead of RAPTOR, which uses a similar probabilistic graphical model as RAPTOR does. IMOD’s another fiducial marker correspondence solution, beadtrack script, is based on nearest neighbor search, which cannot solve the problem with large deviation or tilt angles. Thus, it is also not compared here.]. [Figure 6](#) demonstrates the comparative results of these methods on the tilted micrographs with  $0^\circ$  and  $45^\circ$  tilt angles. To finish the task, the two-stage algorithm, the GMM model and the VBPSM model cost hundreds of milliseconds to several minutes for each dataset, whereas the naive sampling method costs tens of minutes.

As shown in the 1st column of [Figure 6](#), the distributions of fiducial marker positions on the  $0^\circ$  and  $45^\circ$  tilted micrographs cannot be easily aligned. The Hemocyanin dataset has a relatively small number of fiducial markers and fewer outliers. Consequently, the correspondences are successfully determined by all the methods, though the VBPSM model has failed in several local areas. With the increase of the fiducial marker number and outliers, the VBPSM model and GMM model performed poorly on the Vibrio1 and Nitrosop1 datasets. Though the methods tried to maximize the correlation of the underlying geometric information depicted by the fiducial markers, the numerous outliers resulted in a trap of local optimum. On the contrary, the naive sampling and the two-stage algorithm produced reasonable results. However, as pointed out by the arrows in [Figure 6](#), the results of naive sampling suffer from a severe non-uniform deviation, whereas the two-stage algorithm produces a tight fitting.

### 3.2.2 Efficiency and effectiveness of the algorithm

When the tilt angle and the field of the view increase, an affine transformation is not able to accurately describe the potential correspondence related to two tilted micrographs. With the introducing of local drift refinement, the two-stage algorithm can easily solve the problem.

[Figure 7](#) shows the non-uniform drift of the fiducial markers solved from [Figure 5A and B](#), in which the drift is estimated based on the initial affine transformation solved in the first stage. There are two drift directions shown in [Figure 7](#), the drift upward (marked by blue) on the left and the drift downward (marked by red) on the right, which may be caused by a considerable global deformation of the sample. Furthermore, by analyzing the drift direction pointed

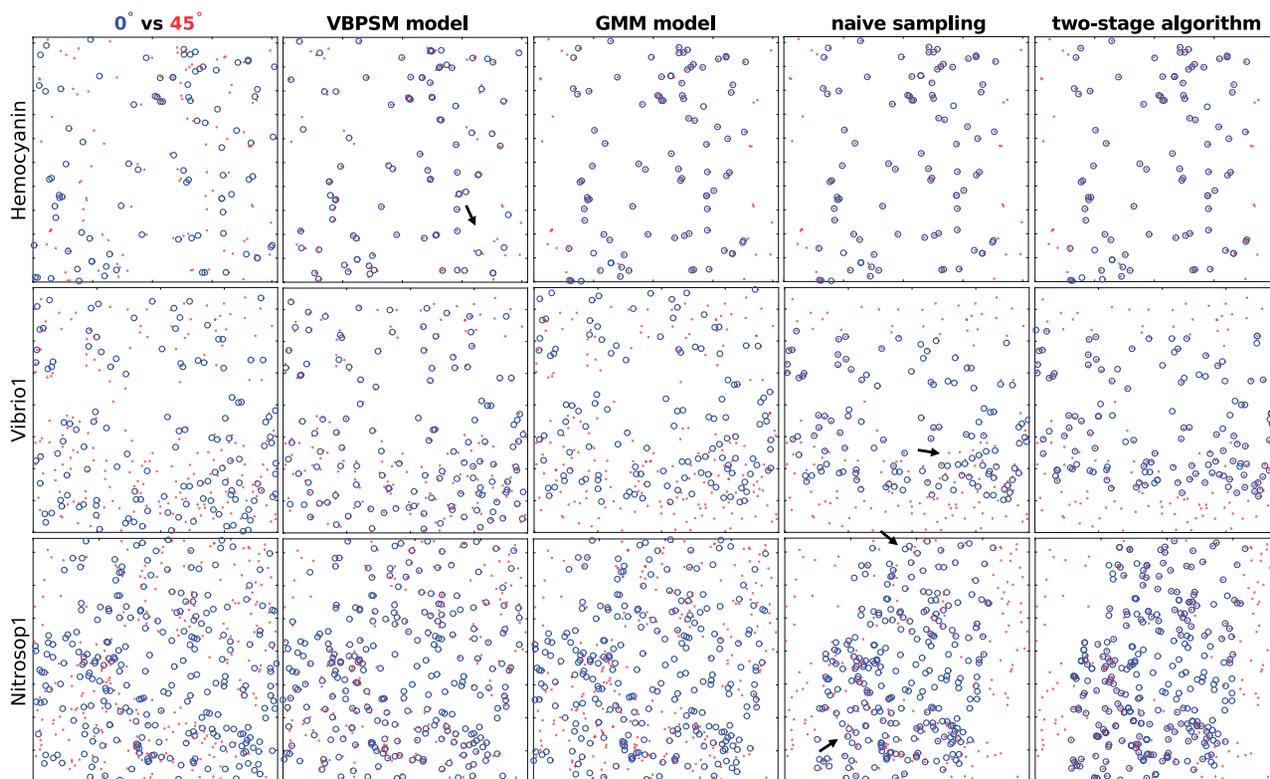


Fig. 6. Performance comparison between different fiducial marker tracking methods. The 1st column illustrates the superimposition of the raw fiducial marker positions, where the blue ‘circle’ and red ‘dot’ denote the fiducial markers extracted from the  $0^\circ$  and  $45^\circ$  tilted micrographs, respectively. The 2nd, 3rd, 4th and 5th columns illustrate the transformed fiducial marker positions of the  $0^\circ$  tilted micrograph solved by the VBPSM model, the GMM model, naive sampling and the two-stage algorithm, respectively. Due to space limitation, only the experimental results of the Hemocyanin, Vibrio1 and Nitrosop1 datasets are shown here. Please refer to [Supplementary Figure S14](#) for the results of the other datasets

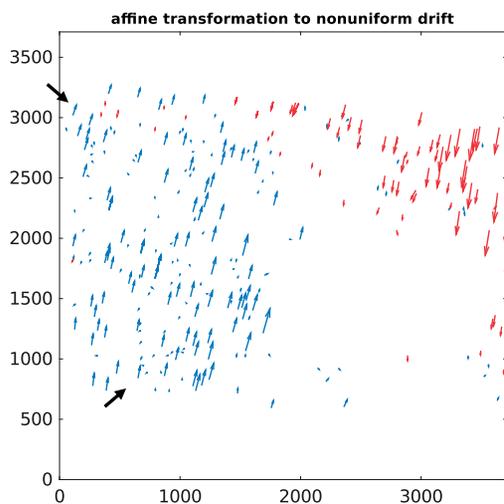


Fig. 7. Demonstration of the non-uniform drift of the fiducial markers. The presented drift vectors are estimated from the  $0^\circ$  and  $45^\circ$  tilted micrographs of Nitrosop2 (Fig. 5A and B). Here, the drifts with different directions are marked by different colors. Please refer to [Supplementary Figure S15](#) for the other datasets

out by the black arrows in [Figure 7](#), an additional local non-uniform deformation can be noticed. The fiducial markers embedded nearby or on the close depths may have a similar drift, as shown in the right top of [Figure 7](#). Generally, most of the fiducial markers are laid on the surface of the specimen, thus, the deviations of the fiducial markers on two micrographs will grow in a gradient way. If there are two neighbor fiducial markers with very close  $x$ - $y$  distance but

with quite different  $z$  distance (different heights) in the space, they may result in a very different drift magnitude, even though both markers appear next to each other in the projection image, which could be observed in [Figure 7](#). However, with a correct partition of the fiducial markers and local drift correction, the two-stage algorithm can efficiently discover the underlying local drift.

In practice, for the tilt series alignment, we may track several neighbors of the micrographs and then compose the fiducial marker correspondences into the fiducial marker tracks. Here, for all the datasets, we matched the fiducial marker positions within 2-neighbors micrographs, i.e. the  $n$ th and  $(n+1)$ th, and  $n$ th and  $(n+2)$ th micrographs, by the GMM model, the naive sampling and the proposed algorithm. Within each matching operation, the ratio of the matched fiducial marker pairs to the total number of potential correspondence was calculated. We also tested RAPTOR on the datasets, based on the prealigned tilt series produced in IMOD. However, limited by the runtime and execution error, RAPTOR failed to run on the Nitrosop2 and Nitrosop3 datasets. All of the methods were run on a Fedora 25 system with 128 Gb memory and two E5-2667v4 (3.2 GHz) CPU.

Considering that the deformation usually happens in high tilt angles, we calculated the average detection ratio of micrograph pairs with tilt angle  $\geq 30^\circ$  or  $\leq -30^\circ$ . [Table 1](#) summaries the averaged detection ratio for each method and dataset. As shown in [Table 1](#), all the methods perform quite well on the Hemocyanin dataset, which has a relatively small field of view ( $2K \times 2K$ ) and contains about one hundred fiducial markers only. However, for the other datasets which contain larger numbers of fiducial markers, the two-stage algorithm performs much better than the other methods. For the datasets of Nitrosop2 and Nitrosop3, the accuracy was increased from  $\sim 94\%$  of the GMM model and the naive sampling method to  $\sim 99\%$  of the two-stage algorithm. Here, the gain of accuracy mainly comes from the model flexibility introduced by the

second stage of the algorithm. The experimental results demonstrate the effectiveness of the two-stage algorithm in fiducial marker correspondence for micrographs with wide field and numerous fiducial markers.

The efficiency is also a critical factor when a new algorithm is applied to the real-world datasets. For each matching operation, the runtime of different methods against the number of fiducial markers is recorded. Figure 8 illustrates the corresponding runtime for each dataset, where the x-axis represents the average number of fiducial marker positions for each matching operation, and the y-axis represents the runtime (ms) in the log scale (integrated within IMOD, RAPTOR's per image runtime is not available). Table 2 summarizes the total runtime for each dataset with different methods.

As shown in Figure 8A, all of the fiducial marker tracking methods have acceptable execution efficiency when there are not too many fiducial markers. However, with the increasing of the fiducial marker number, the runtime of RAPTOR and the naive sampling surge dramatically. As shown in Figure 8E and F, for a dataset with about five hundreds fiducial markers, the execution time of the naive sampling method has increased to  $\sim 8$  min per matching, reaching 21 and 12.6 h for the entire tilt series of Nitrosop2 and Nitrosop3, respectively. The GMM model has a relatively low cost compared to the naive sampling. Nevertheless, it still costs more than 15 and 10 min for the entire tilt series of Nitrosop2 and Nitrosop3,

respectively. On the contrary, the average runtime of the two-stage algorithm has been controlled within 100 ms per micrograph matching for almost all the datasets. Specifically, the two-stage algorithm only costs 21.1 s and 17.3 s for the Nitrosop2 and Nitrosop3 datasets, which is  $40\times$  faster than the GMM-based fast-tracking method and  $1000\times$  faster than the native sampling method.

### 3.2.3 Case study within the tilt series alignment workflow

Finally, we integrated the two-stage algorithm into the fully automatic alignment scheme (Han *et al.*, 2015) to further verify its application in an end-to-end workflow.

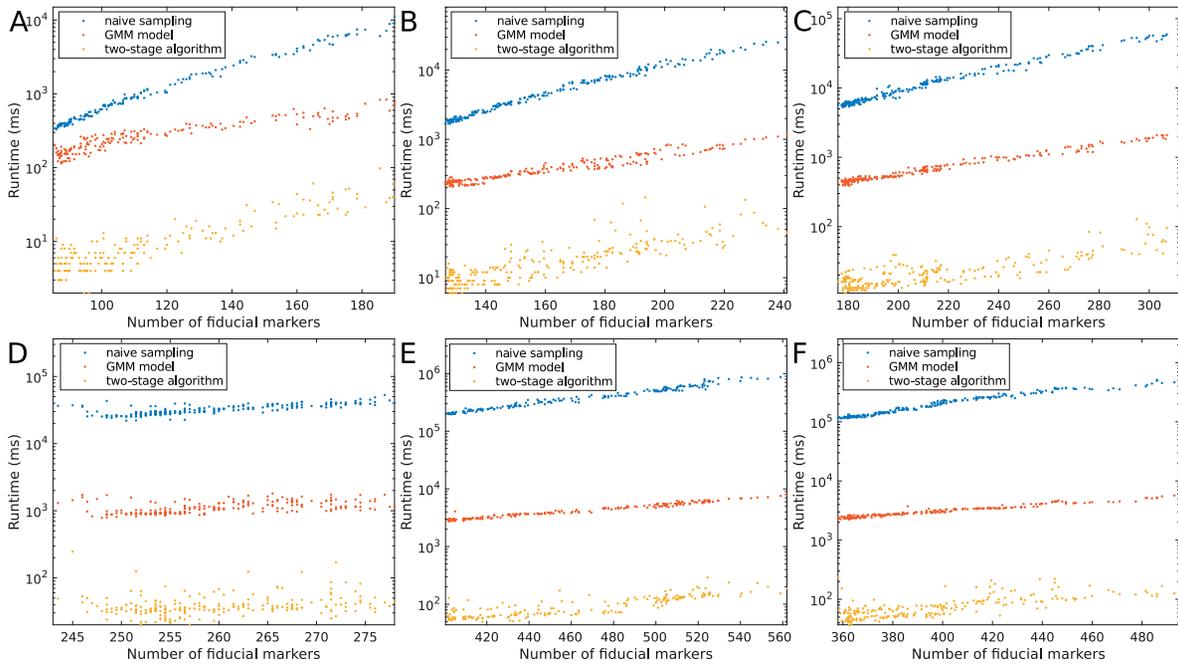
In the alignment scheme proposed by Han *et al.* (2015), the fiducial markers are firstly detected and refined. Then, with the refined fiducial marker positions, a fiducial marker correspondence algorithm is executed to guarantee the consistent matching of the fiducial marker positions from neighboring micrographs. When the required fiducial marker correspondences are obtained, a set of fiducial marker tracks will be generated and the projection parameters are able to be optimized with the configured tracks, to finally determine the geometric transformation of the micrographs in tilt series alignment. Here, the original marker correspondence algorithm will be replaced by our new two-stage algorithm, and the affine projection model is used in projection parameter optimization.

**Table 1.** The correspondence accuracy of the methods on high tilted micrographs

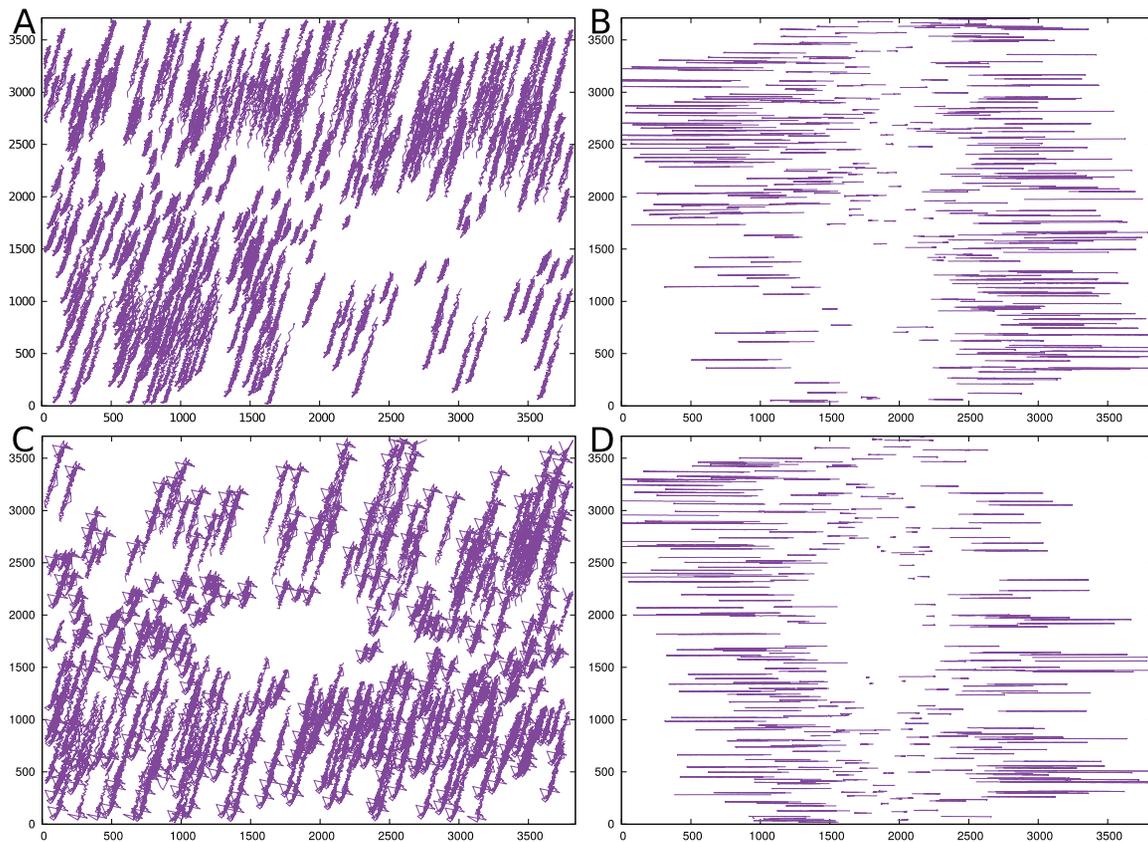
	RAPTOR	Naive sampling	GMM model	Two-stage algorithm
Hemocyanin	92.1%	97.5%	98.2%	99.2%
Vibrio1	84.3%	93.3%	95.4%	98.6%
Vibrio2	81.8%	94.3%	94.5%	99.1%
Nitrosop1	78.7%	93.1%	95.2%	98.7%
Nitrosop2	—	91.9%	94.6%	98.9%
Nitrosop3	—	92.6%	94.8%	99.2%

**Table 2.** The total runtime of fiducial marker tracking for each tilt series

	RAPTOR	naive sampling	GMM model	two-stage algorithm
Hemocyanin	14.5 min	6.4 min	58.1 s	2.6 s
Vibrio1	2.40 h	24.5 min	1.6 min	5.0 s
Vibrio2	2.51 h	1.01 h	3.0 min	5.8 s
Nitrosop1	3.36 h	2.16 h	4.5 min	10.0 s
Nitrosop2	—	21.01 h	15.2 min	21.1 s
Nitrosop3	—	12.60 h	11.4 min	17.3 s



**Fig. 8.** Runtime of the proposed two-stage algorithm (in yellow), the GMM-based fast tracking (in red) and the naive sampling (in blue) on (A) Hemocyanin, (B) Vibrio1, (C) Vibrio2, (D) Nitrosop1, (E) Nitrosop2 and (F) Nitrosop3. The x-axis represents the average number of fiducial markers and the y-axis represents the runtime (ms) in the log scale



**Fig. 9.** Illustration of tilt series alignment with the two-stage algorithm used for fiducial marker tracking. (A) Overlay of raw fiducial marker tracks ( $x$ - $y$  coordinates) extracted from the Nitrosop2 dataset. (B) Overlay of the aligned fiducial marker tracks of the Nitrosop2 dataset. (C) Overlay of raw fiducial marker tracks extracted from the Nitrosop3 dataset. (D) Overlay of the aligned fiducial marker tracks of the Nitrosop3 dataset

Here, we demonstrate the alignment results of the Nitrosop2 and Nitrosop3 datasets as the case study. In the tilt series alignment, only the tracks with enough length are used for projection parameter estimation. For the Nitrosop2 dataset, the two-stage algorithm produced 338 fiducial marker tracks that cover at least 70% of the entire tilt series, with an average track length of 105.9 (Fig. 9A). After the optimization of projection parameters, these fiducial marker tracks resulted in a mean alignment residual of 0.69 pixels (Fig. 9B). For the Nitrosop3 dataset, the two-stage algorithm produced 302 fiducial marker tracks that cover at least 70% of the entire tilt series, with an average track length of 105.3 (Fig. 9C). After the optimization of projection parameters, these fiducial marker tracks resulted in a mean alignment residual of 0.47 pixels (Fig. 9D). Here, the difference in residual error between the two datasets may be caused by the serious sample deformation in the Nitrosop2 dataset (Fig. 7), which can be solved by a higher-order projection model (Fernandez et al., 2018; Lawrence et al., 2006). However, judging from the horizontal fiducial marker tracks after the correction of in-plane rotation and translation (Fig. 9B and D), the orthogonal projection is still able to produce a successful alignment, based on the fiducial marker tracks produced by the two-stage algorithm.

#### 4 Conclusion and discussion

In this article, we proposed a novel two-stage algorithm for the fiducial marker correspondence in electron tomography. The aim of this work is to completely solve the fiducial marker tracking problem in a robust and ultrafast way. The algorithm combines both the robustness of the combinatorial search and the inherent flexibility of the probabilistic model, to further improve the accuracy of fiducial marker correspondence. Generally, the two-stage algorithm can

solve the fiducial marker correspondence with arbitrary initial positions of the micrographs within just a few seconds.

Currently, there are more than ten thousand of the tilt series within the ETDB (Ortega et al., 2019), and the database keeps exploding with the wide application of the ET technique, which raises the emergency demand for a well-designed, high-performance fully automatic software. The improved fiducial marker correspondence accuracy could be used to generate more complete fiducial marker tracks in a more efficient way. With the aid of a large number of well-tracked fiducial markers, a detailed study in projection model selection and validation is possible, especially on the datasets with a serious sample deformation. Cooperation with the world-famous groups (Fernandez et al., 2018; Lawrence et al., 2006) is expected to make follow-up research about robust large-scale bundle adjustment with non-linear projection model, which may further improve the accuracy of tilt series alignment. The related code of our method is also shared online, interested readers may utilize the code to build their own efficient and automatized software.

#### Funding

This work was supported by the National Key Research and Development Program of China [2020YFA0712400], the National Natural Science Foundation of China [62072280, 11931008, 61771009], the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Awards No. FCC/1/1976-17, FCC/1/1976-23, FCC/1/1976-26, URF/1/4098-01-01, URF/1/4352-01-01, URF/1/4379-01-01, REI/1/0018-01-01 and REI/1/4473-01-01.

*Conflict of Interest:* none declared.

## References

- Aiger, D. *et al.* (2008) 4-Points congruent sets for robust pairwise surface registration. *ACM Trans. Graph.*, **27**, 1–85:10.
- Amat, F. *et al.* (2008) Markov random field based automatic image alignment for electron tomography. *J. Struct. Biol.*, **161**, 260–275.
- Bentley, J.L. (1975) Multidimensional binary search trees used for associative searching. *Commun. ACM*, **18**, 509–517.
- Brandt, S. and Ziese, U. (2006) Automatic TEM image alignment by trifocal geometry. *J. Microsc.*, **222**, 1–14.
- Castaño-Díez, D. *et al.* (2007) Fiducial-less alignment of cryo-sections. *J. Struct. Biol.*, **159**, 413–423.
- Chen, Z. and Haykin, S. (2002) On different facets of regularization theory. *Neural Comput.*, **14**, 2791–2846.
- Fernandez, J.-J. *et al.* (2018) Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. *J. Struct. Biol.*, **202**, 200–209.
- Fernandez, J.-J. *et al.* (2019) Consideration of sample motion in cryo-tomography based on alignment residual interpolation. *J. Struct. Biol.*, **205**, 1–6.
- Fischler, M.A. and Bolles, R.C. (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, **24**, 381–395.
- Frank, J. (2006) *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*. Springer.
- Han, R. *et al.* (2015) A novel fully automatic scheme for fiducial marker-based alignment in electron tomography. *J. Struct. Biol.*, **192**, 403–417.
- Han, R. *et al.* (2018) A fast fiducial marker tracking model for fully automatic alignment in electron tomography. *Bioinformatics*, **34**, 853–863.
- Hauer, F. *et al.* (2013) Automated correlation of single particle tilt pairs for random conical tilt and orthogonal tilt reconstructions. *J. Struct. Biol.*, **181**, 149–154.
- Himes, B.A. and Zhang, P. (2018) emclarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat. Methods*, **15**, 955–961.
- Jensen, G. (2019) Scientific Challenges for ET. <https://etdb.caltech.edu/challenges>. [ETDB online-2019].
- Jensen, J.L.W.V. (1906) Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.*, **30**, 175–193.
- Jian, B. and Vemuri, B. (2011) Robust point set registration using Gaussian mixture models. *IEEE Trans. Pattern Anal.*, **33**, 1633–1645.
- Koenderink, J.J. and van Doorn, A.J. (1991) Affine structure from motion. *J. Opt. Soc. Am. A*, **8**, 377–385.
- Kremer, J.R. *et al.* (1996) Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.*, **116**, 71–76.
- Lawrence, A. *et al.* (2006) Transform-based backprojection for volume reconstruction of large format electron microscope tilt series. *J. Struct. Biol.*, **154**, 144–167.
- Lawrence, M. (1992) Least-squares method of alignment using markers. In: Frank, J. (ed.) *Electron Tomography*. Springer, US, pp. 197–204.
- Mastrorarde, D.N. and Held, S.R. (2017) Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.*, **197**, 102–113.
- Myronenko, A. and Song, X. (2010) Point set registration: coherent point drift. *IEEE Trans. Pattern Anal.*, **32**, 2262–2275.
- Ortega, D.R. *et al.* (2019) ETDB-Caltech: a blockchain-based distributed public database for electron tomography. *PLoS One*, **14**, e0215531.
- Qu, H.-B. *et al.* (2017) Probabilistic model for robust affine and non-rigid point set matching. *IEEE Trans. Pattern Anal.*, **39**, 371–384.
- Redner, R.A. and Walker, H.F. (1984) Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, **26**, 195–239.
- Rigort, A. *et al.* (2010) Micromachining tools and correlative approaches for cellular cryo-electron tomography. *J. Struct. Biol.*, **172**, 169–179.
- Vilas, J. *et al.* (2016) Fast and automatic identification of particle tilt pairs based on delaunay triangulation. *J. Struct. Biol.*, **196**, 525–533.
- Wan, W. and Briggs, J. (2016) Chapter thirteen - cryo-electron tomography and subtomogram averaging. In: Crowther, R. (ed.) *The Resolution Revolution: Recent Advances in cryoEM, Volume 579 of Method. Enzymol.*, pp. 329–367.
- Zheng, S.Q. *et al.* (2017) Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.