Journal of Structural Biology 217 (2025) 108153

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Research article

Automatic detection of alignment errors in cryo-electron tomography

F.P. de Isidro-Gómez [a,b], J.L. Vilas [a], J.M. Carazo [a], C.O.S. Sorzano [a,*]

[a] Biocomputing Unit, Centro Nacional de Biotecnologia (CNB-CSIC), Darwin, 3, Campus Universidad Autonoma, 28049 Cantoblanco, Madrid, Spain
[b] University Autonoma de Madrid, 28049 Cantoblanco, Madrid, Spain

ARTICLE INFO

Edited by Steven J. Ludtke

Keywords:
Cryo-electron tomography
Tilt-series alignment
Electron tomography
Electron microscopy
Image processing
Structural biology

ABSTRACT

Cryo-electron tomography is an imaging technique that allows the study of the three-dimensional structure of a wide range of biological samples, from entire cellular environments to purified specimens. This technique collects a series of images from different views of the specimen by tilting the sample stage in the microscope. Subsequently, this information is combined into a three-dimensional reconstruction. To obtain reliable representations of the specimen of study, it is mandatory to define the acquisition geometry accurately. This is achieved by aligning all tilt images to a standard reference scheme. Errors in this step introduce artifacts into the final reconstructed tomograms, leading to loss of resolution and making them unsuitable for detailed sample analysis. This publication presents algorithms for automatically assessing the alignment quality of the tilt series and their classification based on the residual errors provided by the alignment algorithms. If no alignment information is available, a set of algorithms for calculating the residual vectors focused on fiducial markers is also presented. This software is accessible as part of the Xmipp software package and the Scipion framework.

1. Introduction

The three-dimensional (3D) structural analysis of biological specimens is a significant milestone in modern biology. The integration of structural and functional information provides scientists with the necessary tools to understand the underlying biological organization of the sample. Cryogenic electron tomography (cryo-ET) is a sophisticated imaging technique extensively utilized in exploring biological complexes. It allows the study of the structure of macromolecules without losing the information of their biological context. This technique enables the study of various biological samples, from cellular environments to purified complexes.

As in classical cryo-electron microscopy (cryoEM), in its variant of single-particle analysis (SPA), the biological specimen in its native hydrated state is rapidly frozen and then imaged via a Transmission Electron Microscope (TEM) under cryogenic conditions, with the main difference that the sample is tilted in the image acquisition process. Thus, cryo-ET has become an essential tool for studying the structure and function of macromolecular complexes *in situ*, such as cell machinery and viruses. Tomography has allowed a higher level of detail of biological processes at the molecular level, with applications in multiple disciplines, such as structural biology, virology, cell biology, and drug discovery (Robertson et al., 2020; Van Drie and Tong, 2022).

In detail, cryo-ET allows the determination of the three-dimensional structure of the sample by combining projective information from different views. Thus, by rotating the sample holder inside the TEM, it is possible to acquire a set of two-dimensional projective images. Then, the projective information in this set of images is subsequently combined into the three-dimensional space, elucidating the internal structure of the sample under study. However, correcting the sample's relative movements and estimating the acquisition's geometry are mandatory before this reconstruction. This is the aim of the tilt series alignment algorithms. Any error introduced in this step will compromise the final reconstruction, resulting in the loss of detailed structural information. This work presents a set of algorithms to evaluate the alignment quality.

In some sample preparation protocols, gold bead particles are included as high-contrast markers acting as reference points in the tilt series acquisition. Some alignment algorithms exploit these points to make alignment calculations easier and more robust by tracking their position along the series and solving the geometry of the acquisition. These are known as fiducial-based algorithms (Sorzano et al., 2009; Castano-Diez et al., 2010; Mastronarde and Held, 2017; Fernandez et al., 2018; Fernandez and Li, 2021; Seifer and Elbaum, 2022; Sorzano et al., 2020; Coray et al., 2024; Xu et al., 2024). A different set of methods is employed in samples where these particles are absent, known as fiducial-less algorithms (Zheng et al., 2022; Galaz-Montoya et al., 2015). This family of algorithms is based on an iterative process

* Corresponding author.
E-mail address: coss.eps@ceu.es (C.O.S. Sorzano).

https://doi.org/10.1016/j.jsb.2024.108153
Received 25 July 2024; Received in revised form 25 November 2024; Accepted 2 December 2024
Available online 16 December 2024

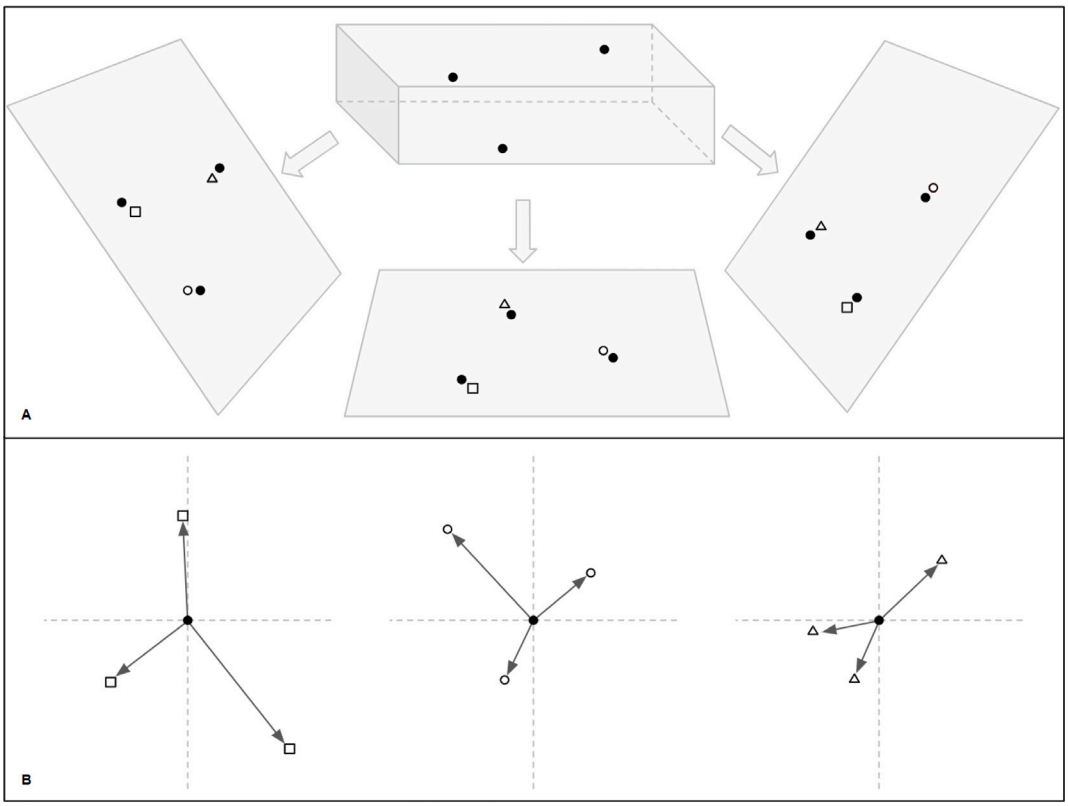

Fig. 1. Schematic definition of residual vectors. (A) represents the projection of the three-dimensional coordinate onto every tilt image (black dots) and the detected fiducials in each tilt image (circles, squares, and triangles). Thus, the residual is defined as the vector whose origin is the projection of the fiducial coordinate and the endpoint in the detected landmark. (B) represents the calculated set of residuals for each fiducial in the sample. A hypothetical case with only three fiducials and three tilt images is presented for simplicity.

of alignment and reconstruction in which the reconstructed volume is reprojected and matched to the acquired tilt series.

The study of the performance of the alignment algorithms followed in this work is based on the calculation of residual error vectors. A residual vector is a two-dimensional vector representing the detected marker's relative position in the tilt image compared to its calculated position after solving the series alignment. Thus, the set of residual vectors measures the quality of the computed alignment in a tilt series. A schematic of this definition is shown in Fig. 1.

The goal of residual-based algorithms is to minimize the magnitude of the residual vectors, thus minimizing the reprojection error. Using residual error vectors as a quality metric to report alignment errors is common, even for fiducial-less approaches. In cases where the alignment algorithm does not provide this measurement, this work introduces a set of algorithms for calculating these residual vectors. However, in that case, it will only be possible if the sample includes fiducial markers.

The alignment quality assessment algorithm is tested using the residual vectors from the alignment algorithm and those calculated by the new algorithm introduced in this work. This approach evaluates the algorithm's ability to classify tilt series accurately and the performance of the residual calculator to generate residuals that reflect alignment quality as effectively as the alignment algorithm.

A common practice in the field is to use heuristic techniques to detect the presence of misalignment. These techniques could include visual inspection of aligned tilt series, artifacts in the reconstructed tomogram, or visualization of the trajectories of high-contrast points in the protected tilt series. This work presents an automatic alternative to the misalignment detection problem that does not require user intervention and offers an autonomous quality alignment classification of input tilt series.

Previous work took advantage of the artifacts observed in the tomographic reconstruction of tilt series that present alignment errors (de Isidro-Gómez et al., 2024). This approach, although practical, presents a double pitfall: first, the calculation of the tomographic reconstruction is required, implying a computational payload; and second, the reconstruction process might shadow some of the most subtle alignment errors, becoming imperceptible in the tomogram. The algorithms presented in this work focus directly on the tilt series to assess the alignment quality, so these two pitfalls are avoided.

In summary, we address the need to automatically assess the performance of any tilt series alignment by working directly on the tilt series and avoiding tomographic reconstruction. Hence, the need for manual inspection of the alignment results is avoided, freeing the users from the manual, laborious, and error-prone task of analyzing the alignment quality of the series. The algorithm presented in this work has been implemented in Xmipp (de la Rosa-Trevín et al., 2013; Strelak et al., 2021). It is also accessible through the Scipion workflow engine (de la Rosa-Trevin et al., 2016), within its tomography environment ScipionTomo (Jimenez de la Morena et al., 2022) under the protocol name `xmipptomo - detect misaligned ts`.

2. Methods

The goal of residual-based algorithms is to minimize the reprojection error by reducing the magnitude of the calculated residuals (Sorzano et al., 2020)

$$E = \sum_i \sum_{j \in V_i} \|\mathbf{p}_{ij} - (A_i \mathbf{r}_j + \mathbf{d}_i)\|^2, \quad (1)$$

being $\mathbf{r}_j$ the jth three-dimensional coordinate and $\mathbf{p}_{ij}$ its projection onto tilt image i. The matrix A_i is the projection matrix accounting for the tilt around the tilt axis and a subsequent in-plane rotation,

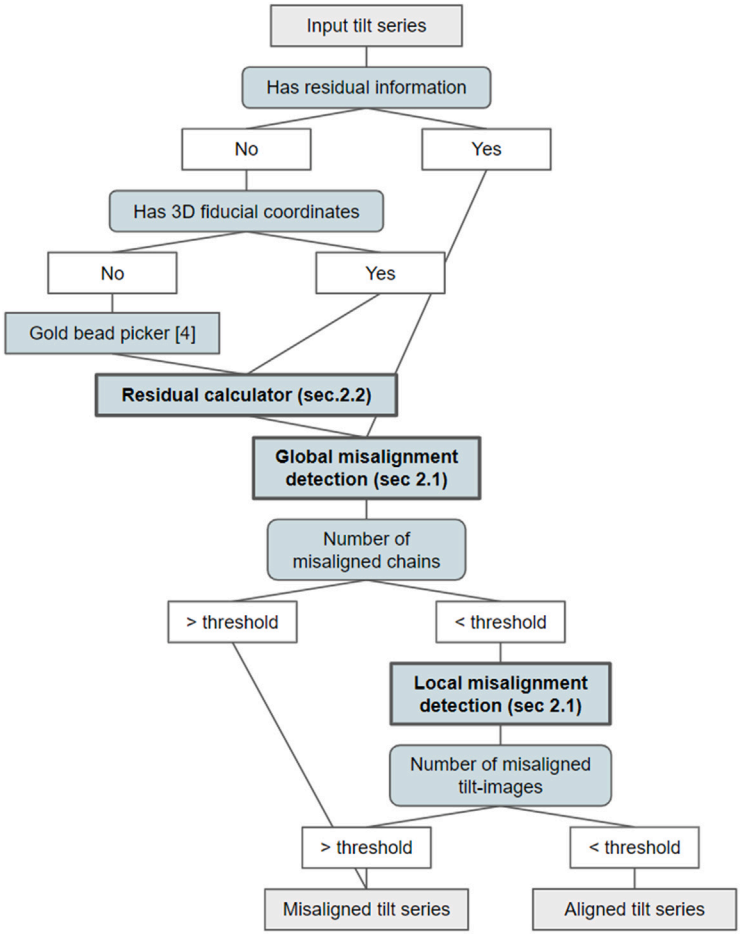

Fig. 2. Scheme of the workflow for tilt series alignment classification. Boxes in bold highlight the steps solved by the algorithms presented in this work.

while $\mathbf{d}_i$ is a 2D vector accounting for an in-plane shift of the ith image. Errors in minimizing this function with respect to A_i may lead to misaligned tilt series. Based on the authors' observations, errors in the pre-alignment of the tilt series (where only shifts are typically corrected via cross-correlation) and suboptimal detection of the landmarks in the tilt images are common sources of alignment errors, in addition to the presence of local minima typical of any optimization problem.

This section introduces the algorithms developed to detect alignment errors based on residual vectors and their characterization. First, we present an algorithm for analyzing these residuals in search of possible misalignments in the series. If the alignment algorithm does not provide the residual vectors, a set of algorithms for their calculation based on detecting fiducial markers in the sample is introduced. A schematic of the proposed workflow is introduced in Fig. 2.

2.1. Residual-based alignment errors detection

This algorithm aims to detect alignment errors through the study of residual vectors, as these provide a measure of the relative position of the detected fiducial in the tilt image with respect to its projected location after solving the alignment, thus providing a measure of the alignment quality. Software packages widely use residual errors in quality measure reports. Software packages widely use residual errors in quality measure reports. It presents the advantage that they are available using different alignment strategies, from fiducial-based algorithms (using gold beads as landmarks) to patch-tracking algorithms (used in fiducial-less samples).

In the development of this work, several measurements and statistics have been studied to characterize the alignment errors present in the tilt series from the reported residual vectors. All calculated statistics are saved and reported to make this information available to the user and any posterior data curing or statistic-based filtering of the analyzed tilt series. All these parameters might be calculated globally or per tilt. The statistics reported are as follows:

1. Convex hull: if the set of residuals represents a set of points C, the convex hull is defined as the unique and minimal convex subset containing C. The residual vectors are characterized by reporting the area and perimeter convex polygon defined by this subset of points (Gonzalez, 2009).
2. Binomial test: used to detect deviations in the sign of a set of residuals (Zar, 2010). The hypothesis is that residuals are randomly distributed, meaning there is no bias in the direction of the residuals, and their sign is equally likely ($p = 0.5$). This hypothesis is tested by defining as a statistic the probability of observing k positive-signed (or negative) residuals in a total set of n elements, as described in Eq. (2)

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{2}$$

   Since residuals are two-dimensional vectors, two tests can be performed, one in X and another in Y direction. Thus, this hypothesis is tested positive if the reported p-value of the binomial test is lower than a significance level (typically 0.05).
3. F-test: used to detect differences in the directional variances of a population of residuals. First, the covariance matrix of the set of residuals is calculated, defined in Eq. (3)

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{x,y} \\ \sigma_{y,x} & \sigma_y^2 \end{pmatrix}, \tag{3}$$

the eigenvectors of the covariance matrix determine the directions of maximum variance in the residual population, while the eigenvalues, λ_1 and λ_2, determine the dispersion of the residuals along these axes (Lowry, 2014). The hypothesis to be tested is that the difference in the dispersion over both directions is not significant. For this, the statistic of the F test is defined as the quotient of the two eigenvalues, defined in Eq. (4)

$$F = \frac{\lambda_1}{\lambda_2}, \tag{4}$$

this hypothesis is tested positive if the reported p-value of the F-test is less than a significance level (typically 0.05).

4. Augmented Dickey–Fuller test: to test if a population of residuals is stationary (it does not have a unit root). If the set of residuals is stationary, no trend or seasonal effect is present in the data. This means the residuals are randomly distributed, and the mean and variance should be constant along the tilt series. In other words, this test hypothesizes that the residuals present a random walk behavior (Dickey and Fuller, 1979). The statistic of the augmented Dickey–Fuller test is defined in Eq. (5)

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}, \tag{5}$$

where $\hat{\gamma}$ is the autoregressive term, a quantification of the influence of the previous value of the series on the current one, and $SE(\hat{\gamma} = \frac{\sigma}{\sqrt{n}})$ is the standard error of $\hat{\gamma}$ where σ is the standard deviation and n the number of observations of the sample. Thus, if the statistic value is more negative than the critical value, the hypothesis is rejected, which means that the series is stationary. Intuitively, this test aims to measure whether the residuals have a random walk distribution or exhibit some directional bias, implying a potential drift in the alignment.

5. Mahalanobis distance: a measure of the distance of each residual to a specific distribution defined by

$$D = \sqrt{(\mathbf{e}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{e}-\boldsymbol{\mu})}, \tag{6}$$

where $\mathbf{e}$ represents the residual vector, $\boldsymbol{\mu}$ represents the mean of the residual population, and $\boldsymbol{\Sigma}$ represents the covariance matrix (Mahalanobis, 2018). We assume a set of residual vectors follows a two-dimensional Gaussian distribution if the tilt image does not present a large misalignment. Note that $\boldsymbol{\mu} = 0$ for our specific case, since the Gaussian residual distribution has zero mean.

Thus, each residual is weighted based on its Mahalanobis distance to a two-dimensional isotropic distribution with a covariance matrix $\boldsymbol{\Sigma}$, whose variance is characterized by the fiducial size. This is defined in Eq. (7)

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}, \tag{7}$$

where $\sigma_x^2 = \sigma_y^2 = \sigma^2$, assuming that the residuals do not have any skew in their distribution. By default, a σ^2 value equal to one-third of the fiducial size is proposed, ensuring that 99.7% of the residuals fall within the fiducial radius in a Gaussian distribution. This threshold is meant to be modified if a more relaxed or strict scenario is pursued. If the residual vectors come from fiducial-less samples, this parameter may be adjusted to allow for the smallest relative movements that ensure a sufficiently accurate reconstruction.

However, for the automatic detection of alignment errors, it is essential to select a statistic (or a set of them) that allows for a robust assessment of the tilt series alignment. From the experiments performed to detect the statistics that hold the most information to characterize the residuals, it has been concluded that the Mahalanobis distance is the most resilient quality metric. This is further explained in the Results section.

Both global and per-tilt misalignment detection are driven once every residual vector is weighted by its Mahalanobis distance. First, a global analysis is performed, and every chain of landmark residuals is analyzed. Each chain comprises a set of residuals referenced to the same landmark at every tilt image, with as many chains as fiducials used in the alignment. If the proportion of chains that exhibit an average Mahalanobis distance greater than 1 exceeds a specified threshold, the tilt series is flagged as misaligned, and the analysis is concluded. This threshold is set at 0.8, meaning that at least 80% of the chains must be classified as aligned.

If global misalignment is not detected, an equivalent analysis is performed for each tilt image with two possible criteria to report misalignment. First, misalignment is flagged if the average Mahalanobis distance of all residuals in the image exceeds one. Alternatively, a voting criterion is used to determine whether the percentage of residuals with a Mahalanobis distance greater than one exceeds a certain threshold (by default 80%). Additionally, a parameter allows the user to specify the maximum number of tilt images that can exhibit misalignment. The entire series is flagged for global misalignment if this threshold is exceeded. If not, only the misaligned tilt images are excluded from further processing.

Finally, we would like to suggest modifying the previous approach, which is more resilient to noisy residual vectors. For this, we use the Z-score, defined in Eq. (8)

$$Z = \frac{|e| - \mu}{\sigma}. \tag{8}$$

Before computing the global and per-tilt average residual distances, residuals whose Z-score exceeds a given threshold (the default parameter is three standard deviations) are removed. This behavior is particularly interesting if the alignment algorithm does not provide the residual vectors and another algorithm must calculate them, as introduced in the next section. This additional filter makes the assessment of alignment quality more robust if the calculation of the residuals is noisy.

Apart from those mentioned above, any tilt series for which at least two three-dimensional coordinates are not provided is excluded from this analysis and classified as misaligned. This is because it is impossible to solve the alignment of a tilt image with only a single projection point if the tilt axis orientation has to be calculated. Calculating the shifts and rotations for every image to reference them to a common geometry is necessary.

2.2. Residual vectors calculation

Scientists do not always have access to alignment quality metrics, particularly the alignment residual vectors. Sometimes, they rely solely on alignment transformations or aligned tilt series. However, to our knowledge, only a few alignment algorithms do not provide any of this information (Zheng et al., 2022). Thus, we introduce algorithms to calculate the residual vectors needed to feed the algorithm proposed in the previous section. To achieve this, the three-dimensional coordinates of a subset of fiducials in the sample must be provided. In an earlier work (de Isidro-Gómez et al., 2024), we introduced an algorithm for detecting these gold beads in case this information is unknown to the user. However, this implies the drawback of calculating the tomographic reconstruction of the tilt series.

The robustness of this step is critical to detecting alignment errors in tilt series. The number of landmark coordinates provided and the quality of the calculated residuals will drive the performance of the misalignment detection algorithm. Nonetheless, several quality control steps have been introduced to ensure the algorithm's robustness to outliers.

This section presents two algorithms. The first algorithm aims to detect fiducial markers, a problem that remains an open challenge in the field (Hou et al., 2024). To achieve this, the steps described below will be executed on each tilt image of the series. The results obtained at each step of the algorithm are exemplified in Fig. 3.

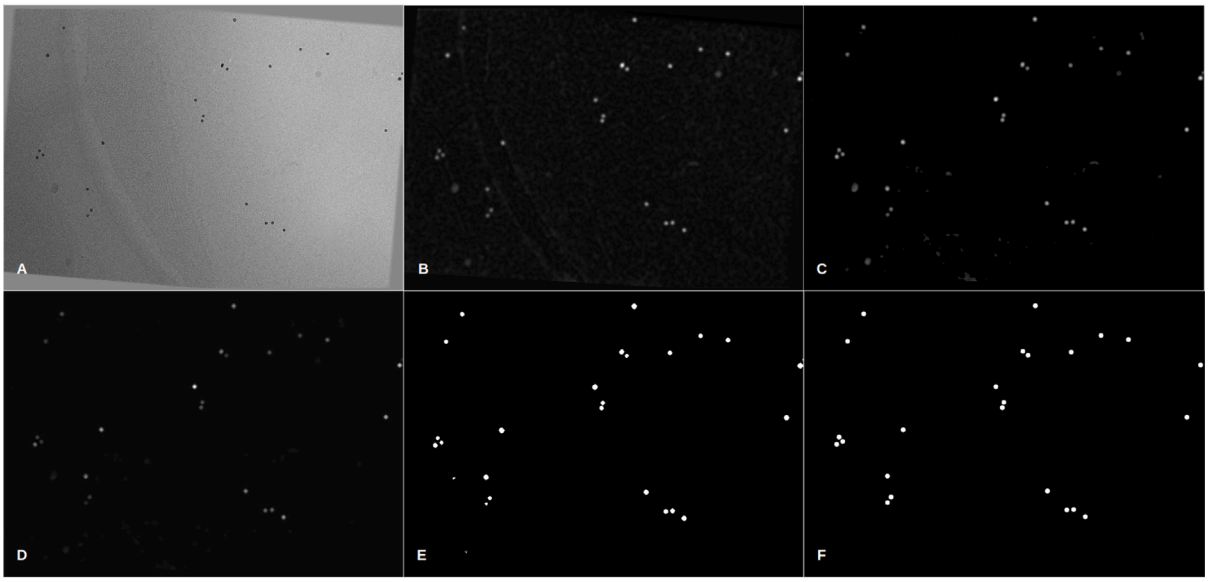

Fig. 3. Intermediate results at different stages of the fiducial detection algorithm: (A) original tilt image, (B) result after landmark enhancement and background subtraction, (C) result after Z-score thresholding, (D) result after maximum pooling and directional filter, (E) labeled regions of interest after preprocessing, and (F) inpainting of the resulting detected landmarks after filtering the regions of interest. This tilt image belongs to tilt series E48g4_30 from EMPIAR 11457 dataset at 0°. An equivalent figure at 60° its provided in the Supplementary Material.

1. Interpolation edges detection: Due to image interpolation during the alignment process, sharp edges are introduced in the images. Also, in some cases, artifacts are observed on the border of the tilt images (see Supplementary Material, Fig. 4 and Movie 1). Thus, when the images are aligned, these artifacts are reallocated in sensitive regions, even spoiling the postprocessing of the images. To solve this, we detect the background of the images, calculating the gradient magnitude image with the use of a Sobel filter defined in Eq. (9)

$$M(x,y) = \sqrt{(G_x \star I(x,y))^2 + (G_y \star I(x,y))^2}, \tag{9}$$

being G_x and G_y, the gradient kernels in the vertical and horizontal directions, and $\star$ the convolution operator. Then, the contour of the images is removed and set to the background value, removing any possible interpolation artifact.
2. Downsampling: The images are downsampled, targeting a specific size of the fiducials. This operation increases the signal-to-noise ratio and computational efficiency. Also, it allows the definition of convolutional kernels to be used in posterior steps.
3. Landmark enhancement: The downsampled images are convolved with a fiducial kernel for enhancement. Fiducials are modeled as a two-dimensional Gaussian distribution as defined in Eq. (10)

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right), \tag{10}$$

as in Eq. (7), $\sigma_x^2 = \sigma_y^2 = \sigma^2$, imposing no skew in the definition of the kernel. The sigma value is adjusted to one-third of the landmark target size from the previous downsampling step. Finally, a rolling-ball background subtraction is applied to the image, with the ball radius set by default to two times the landmark target size (Sternberg, 1983). Customization of this parameter is allowed, ensuring the preservation of fiducial candidates while removing background information. Intermediate results of this step are shown in Fig. 3 B.
4. Detect outlier elements. All the pixels inside the interpolation limits calculated in the first step are analyzed. Those presenting z-scores lower than a given threshold are masked from further analysis, keeping only the outlier values. Intermediate results of this step are shown in Fig. 3 C.

   Then, the images are morphologically dilated (comparable to a maximum pooling operation) to keep the landmark regions more homogeneous. After the thresholding, non-fiducial high-contrast elements might not have been masked out. A band-pass directional filter is applied to ensure the detection's robustness. The image is directionally filtered in as many directions as the user inputs (8 directions by default) with an angle amplitude of 10° for the complete cone. Each direction is also bandpass filtered, centered in the landmark target size. Finally, each direction is combined in a weighted mark applied to the image. Thus, round objects (landmarks) are preserved while removing those with a high signal-to-noise ratio but presenting different shapes (such as carbon edges or membranes). Intermediate results of this step are shown in Fig. 3 D.
5. Filter regions of interest. Finally, the pixels from the previous filtered and masked image with a z-score lower than the given threshold are removed from any further processing. Intermediate results of this step are shown in Fig. 3 D. The image is subsequently labeled, and two different criteria morphologically analyze each region:

   - Relative area: those regions significantly bigger or smaller than the expected area of the target landmark are removed.
   - Circularity: those regions whose shape differs significantly from a circle, measured as the ratio between the area of the regions and the area of its circumscribed circle are removed.

   Thus, after applying these filters, only those regions from the previous image that fulfill the morphological constraints are included as potential fiducials. Intermediate results of this step are shown in Fig. 3 F.
6. Coordinates centering. Optionally, coordinated can be centered by the maximum shift obtained from the Fourier correlation of each landmark with its mirror.

The second algorithm is fed with the landmarks' location at each tilt image. A set of residual vectors is calculated for each tilt image, computing the distance between the projection of each three-dimensional

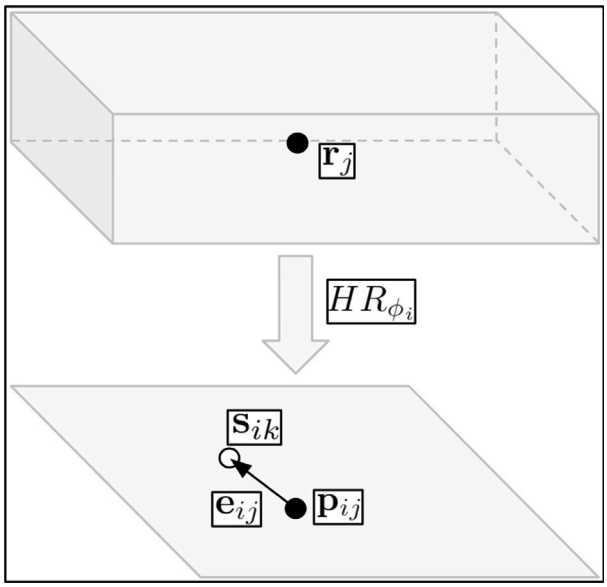

Fig. 4. Schematic of the residual calculation algorithm, showing the projection of the fiducial coordinate onto the tilt image (black dots) and defining the residual as the vector from the projection to the nearest detected landmark (empty dot).

coordinate to its closest detected fiducial. As with the previous algorithm, the set of operations applied to each tilt image is outlined in the following paragraphs.

1. Residual vector calculation. First, the set of three-dimensional coordinates is projected onto each tilt image. Let $\mathbf{r}_j$ be the jth three-dimensional coordinate and $\mathbf{p}_{ij}$ its projection onto tilt image i

$$\mathbf{p}_{ij} = HR_{\phi_i}\mathbf{r}_j, \tag{11}$$

where H is the matrix that projects the three-dimensional coordinate into its X and Y components, and R_{ϕ_i} is a rotation matrix of ϕ_i degrees corresponding to the tilt angle of each tilt image. Notice that no parameter for in-plane shift or rotation is introduced in the projection, unlike in Eq. (1). Although these parameters are typically included in alignment models, errors in estimating these parameters lead to the misalignment we aim to detect. As illustrated in Fig. 4, this step involves calculating the projection of the three-dimensional coordinates onto each tilt image (black dots).
Being $\mathbf{s}_{ik}$ the closest detected landmark in tilt image i to the projection $\mathbf{p}_{ij}$, the residual vector is defined in Eq. (12)

$$\mathbf{e}_{ij} = \mathbf{s}_{ik} - \mathbf{p}_{ij}, \tag{12}$$

which is illustrated in Fig. 4 as the vector whose origin is in the detected projection (black dot) and ends in the closest detected landmark (empty dot).
2. Residual vector pruning. A robust detection of misalignment requires a good estimation of the residual vectors. However, suppose the fiducial corresponding to the projection of one of the three-dimensional coordinates is not detected. In that case, the resulting vector will not adequately characterize the quality of the tilt series alignment. This is especially pernicious in an aligned tilt series, as the vector modulus transitions from quasi-null to one of significant magnitude.
However, if it is an aligned series, the resulting vector will originate from a landmark (the one that has not been correctly detected) and end at another (the next closest one that has been detected). Note that this does not occur if the tilted image is not correctly aligned. In this way, the correlation of the regions at the origin and the end of the vector can be calculated. If the correlation is significant, it validates that the residual originates and ends in a fiducial, removing this residual from future processing.
3. Metadata generation. The vectors that survived the previous pruning are collected into metadata relating the three-dimensional coordinate, the landmark in each tilt image closest to each projection, and the residual vectors associating their relative positions. With this information, it is possible to feed the misalignment detection algorithm introduced in the previous section.

3. Results

This section presents the results of the introduced methods. First, an analysis of the quality metrics introduced in this work is provided, followed by the performance of these methods on three different public datasets available in the EMPIAR public archive (Iudin et al., 2022).

All datasets are processed uniformly before analysis. The tilt series are aligned using the IMOD software package (Mastronarde and Held, 2017) (correcting for both shift and angle) and then visually inspected for classification, segregating the series with the correct alignment from those without.

To consider a tilt series aligned, the transition between the tilt images that compose it must be smooth, and the common landmarks between the different images must describe rectilinear trajectories perpendicular to the tilt axis (which by convention is positioned vertically and centered in each image). Focusing on high-contrast elements common to all images and observing the trajectories they describe facilitate this identification. (See also Supplementary Material, Movies 1 and 2 and Fig. 3).

The maximum number of allowed misaligned tilt images is also set to zero. Thus, if a single image is labeled misaligned, the entire series is labeled misaligned. This criterion maximizes the algorithm's level of stringency. The rationale behind this approach is to ensure that in our experiments, no error in the classification algorithm is shadowed for any tilt image. Thus, all errors are evident in the confusion matrices in the following sections.

For each dataset analyzed, a confusion matrix is presented, divided into two sections based on whether the residual vector source that feeds the classification algorithm is the alignment algorithm (IMOD) or the residual vector calculation algorithm introduced in this work (Automatic).

This work does not aim to analyze the performance of the IMOD alignment algorithms. A better parameter tuning and tilt series preprocessing, such as dose filtering or CTF correction (which localizes the signal in the real space (Glaeser et al., 2021)), could lead to improved performance. Also, some of these tuned choices are sample-dependent. The aim is to provide algorithms that can detect alignment errors in an automatic processing pipeline, where the same parameters are used for all acquired tilt series.

Authors are aware of alternative methods in tomography for tilt series alignment. However, IMOD is considered a standard in the field and provides a comprehensive report of the residual alignment vectors, allowing performance comparison with the methods presented in this work. Thus, we consider this work of special relevance for pipelines that include IMOD as their tilt series alignment algorithm. All of this processing pipeline has been executed in the ScipionTomo framework, ensuring its reproducibility and validation.

Regarding the computational load of the presented algorithms, both exhibit very short execution times. Misalignment detection algorithm execution is almost instantaneous, while the calculation of residuals, although slightly more costly, takes only a few seconds. Additionally, there are no significant differences in execution time relative to image size, as a downsampling process is applied to the images.

3.1. Quality metrics analysis

In the Methods section, a set of metrics is introduced to evaluate the alignment quality of a tilt series that can be grouped into three different conceptual groups: statistical test-based metrics (binomial test, F test, and augmented Dickey–Fuller test), geometric metrics (perimeter and area of the convex hull), and distance metrics (Mahalanobis distance). These metrics are calculated for both chains of residuals associated with the same fiducial along the tilt series, as well as for each set of residuals belonging to the same tilt image.

A typical behavior observed in alignment algorithms is that only a subset of tilt images, particularly those at high tilt, are not well-referenced to the common geometry. This results in the metrics computed over chains of residuals overshadowing the subset of misaligned images if their number is not significant enough. This imposes the limitation that any metric selected for automatic quality assessment must accurately characterize both global misalignment of the tilt series and local misalignment, computed individually for each tilt image. This limitation is more significant than it might initially appear, as some alignment algorithms require only a few fiducials (sometimes just 2 or 3) to resolve the alignment, leading to a reduced set of residuals on which to perform the analysis.

Given these limitations, it has been observed that the Mahalanobis distance provides results that allow for the assessment of alignment quality, as demonstrated in the following results sections. In the ideal case where the alignment is perfectly solved, all residuals should have zero modulus. However, when the perfect solution is abandoned, the residual vectors consistently have a non-zero modulus. If the calculated solution begins to deviate from the ideal, the residual vectors will start to form a two-dimensional pattern. If the calculated alignment is sufficiently close to the ideal solution, it is expected to resemble a two-dimensional Gaussian distribution with a reduced standard deviation. As introduced in the methods section, the standard deviation of the typical distribution of residual vectors for an aligned series is set to one-third of the fiducial radius. This defines a distribution where the majority (99.7%) of the residual vectors are expected to have a modulus smaller than the fiducial radius. If the researcher wishes to be more stringent in the assessment of the tilt series alignment, this value can be reduced. Thus, the Mahalanobis distance effectively detects outlier elements that unexpectedly deviate from this expected distribution.

The effectiveness of the Mahalanobis distance in characterizing the alignment quality is observed in Fig. 5 (left), where the distribution of values is significantly different when comparing an aligned tilt series with a misaligned one. Establishing a threshold for images with an average distance of their residuals greater than one is also straightforward.

Among the metrics introduced in the Methods section, those based on statistical tests do not adapt well to the local alignment quality assessment. As mentioned above, in those cases where the number of fiducials used to assess misalignment is limited, it is not possible to obtain significant results from a statistical test since the sample is not sufficiently large. Furthermore, these tests search for an uneven distribution of the signs of the residual or a bias in its direction. Although these metrics are informative, they are prone to false positives (a misaligned tilt series with high-modulus residuals may present an isotropic distribution) and false negatives (an aligned tilt series with low-modulus residuals may present bias in its direction). The limitation of these metrics to characterize the quality of alignment is exemplified in Fig. 5 (left), where no significant differences in their behavior are evident when comparing the results obtained from an aligned series with those of a misaligned one.

Finally, metrics based on the geometry of the fiducials are absolute metrics. This introduces the limitation of complex thresholding when using these metrics to determine if a set of residuals belongs to a misaligned tilt series. This behavior can be observed in Fig. 5 (right), where, although there is a significant difference in the results obtained for these metrics (especially for the area of the convex hull), the problem of selecting a threshold is also evident. In addition, these metrics are sensitive to possible directional bias in the residuals. If the residual vectors point in a similar direction, the area and perimeter of the convex hull significantly decrease when compared to an isotropic distribution of the fiducials.

Although the Mahalanobis distance is chosen as the metric for automatic alignment quality assessment, all these metrics are reported for every tilt image and residual chain. This approach provides the user with the ability to perform additional filtering of the classifications obtained according to these metrics.

Table 1
Confusion matrix for EMPIAR-10453 dataset.

EMPIAR 10453		Predicted			
Source of residuals		IMOD		Automatic	
		Aligned	Misaligned	Aligned	Misaligned
Real	Aligned	175	0	175	0
	Misaligned	3	59	2	60

3.2. EMPIAR-10453

The first dataset analyzed corresponds to entry 10453 of the EMPIAR database (Turoňová et al., 2020). The sample presents a SARS-Cov-2 spike at a pixel size of 1.33 Åand gold beads of approximately 10 nm. This dataset comprises 237 tilt series and, after a one-by-one visual inspection, 61 present misalignment. Misalignment analysis using the presented algorithms was carried out, and the confusion matrices of the results obtained are presented in Table 1. In this table, the predicted classification and ground truth are compared. The table is divided based on the source of the residual vectors: the alignment algorithm (left) or the algorithm introduced in this work (right).

The F1 score is used to measure the performance of the classification algorithm. This metric is defined as

$$F_1 = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}, \tag{13}$$

where TP, FP, and FN are the number of true positives, false positives, and false negatives elements from the confusion matrices, respectively.

For completeness, the Jaccard index is included as an alternative performance measure. Unlike the F1 score, which tends to flatten, the Jaccard index exhibits rapid decay, making it more sensitive to classification errors. It is defined as:

$$J = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{14}$$

and, as in the previous case, TP, FP, and FN are the number of true positives, false positives, and false negatives elements from the confusion matrices, respectively.

For the IMOD residuals, three misaligned tilt series are misclassified, leading to an F1 score of 0.992 and a Jaccard index of 0.9831. In addition, when the residual vectors are provided by the algorithm presented in the work, two misaligned tilt series are misclassified; the F1 score is 0.995, and the Jaccard index is 0.988.

It should be noted that only one false positive is shared between both sources of residuals. This tilt series exhibits an unusual distribution of the fiducials, all clustered in one corner of the images (see Supplementary Material, Fig. 5). This distribution results in low-magnitude residuals, but some instability in the series alignment can be observed, becoming more pronounced in those regions away from the fiducial cluster. This is a clear example of the well-known importance of a homogeneous distribution of the fiducials.

Fig. 5. Metrics calculated for each tilt image from two series in the EMPIAR dataset 10453. The left column displays metrics from an aligned tilt series (TS_005), while the right column shows metrics from a misaligned tilt series (TS_233). For clarity and scale convenience, metrics are divided into three groups: the top row presents statistical tests, the middle row depicts geometrical measurements, and the bottom row illustrates Mahalanobis distances. Note that the geometrical metrics (convex hull area and perimeter) have different scales for clarity.

3.3. EMPIAR-11457

The second dataset analyzed corresponds to entry 11457 of the EMPIAR database (Ni et al., 2023). The sample presents ChAdOx spikes (AZD2816) with a pixel size of 2.18 Åand gold beads of approximately 12 nm. This dataset comprises two acquisitions at two different tilt axis angles (10° and 85° degrees) with 134 tilt series. For simplicity, the results are presented together. After a one-by-one visual inspection, 25 tilt series present misalignment. Both acquisitions are processed in parallel and follow the same steps. The confusion matrices for the classification based on the residual vectors provided by the alignment algorithm and those calculated by the algorithm introduced in this work are summarized in Table 2.

It can be observed that for the alignment algorithm residuals, two misaligned tilt series are misclassified, leading to an F1 score of 0.990 and a Jaccard index of 0.9802. Also, when the residual vectors are provided by the algorithm presented in the work, three misaligned tilt series are misclassified, and another four aligned ones are also misclassified, leading to an F1 score of 0.964 and a Jaccard index of 0.9314.

Table 2
Confusion matrix for EMPIAR-11457 dataset.

EMPIAR 11457		Predicted			
Source of residuals		IMOD		Automatic	
		Aligned	Misaligned	Aligned	Misaligned
Real	Aligned	99	0	95	4
	Misaligned	2	23	3	22

This dataset was particularly challenging, presenting a very low contrast in their high-tilt images. This significantly complicates fiducial detection, a key step in calculating residual vectors. Nonetheless, the false positive ratio obtained in both approaches does not differ significantly. Although the relative difference is high, the absolute rate is low for the false negative ratio. Thus, no tilt series presenting misalignment is included in any further processing.

It is interesting to note that two of the false negatives present in both classifications exhibit interesting behaviors. One presents a subtle movement along the entire tilt series, but the displacement never

Table 3
Confusion matrix for EMPIAR-10364 dataset.

EMPIAR 10364		Predicted			
Source of residuals		IMOD		Automatic	
		Aligned	Misaligned	Aligned	Misaligned
Real	Aligned	18	0	15	3
	Misaligned	0	0	0	0

exceeds the threshold. As for the other, the residuals also do not exceed the threshold since two very close three-dimensional coordinates were used to align the series (see Supplementary Material, Figure 6). This demonstrates that, at least for an extreme case like this, the residuals are not indicative of alignment quality and are compatible with those from an appropriately aligned series.

3.4. EMPIAR-10364

The third dataset analyzed corresponds to entry 10364 of the EMPIAR database (Burt et al., 2020). The sample presents a *Escherichia coli* minicells at a pixel size of 2.24 Å, presenting gold beads of approximately 9 nm. This dataset comprises 18 tilt series, and after their one-by-one visual inspection, there is no present misalignment. The confusion matrices for the classification based on the residual vectors provided by the alignment algorithm and those calculated by the algorithm introduced in this work are summarized in Table 3.

No tilt series is misclassified for the alignment algorithm residuals, leading to an F1 score and a Jaccard index of 1. When residual vectors are provided by the residual vector calculation algorithm presented in the work, three aligned tilt series are misclassified, leading to an F1 score of 0.909 and a Jaccard index of 0.833.

The alignment quality assessment shows better performance for this dataset when the algorithm is fed with residual vectors from the alignment algorithm. As observed in the previous dataset, there is a significant loss of contrast, particularly at high tilt angles. This loss complicates the detection of gold beads throughout the series, especially those classified as misaligned (see Supplementary Material, Figure 7). Evidence of this is that for the three misclassified tilt series, only the image at −60° is flagged as misaligned.

4. Discussion and conclusion

The increase in throughput in cryo-electron tomography acquisition has significantly increased the amount of data that scientists must analyze to elucidate the structures of interest in samples. The immediate consequence in the field is the automation of image processing methods to handle this enlarged data volume. This is especially critical in the initial steps of the pipeline, particularly up to the tomographic reconstruction step, which is essential for any tomography processing application (segmentation, subtomogram averaging, per particle per tilt, etc.).

Among all the steps performed up to this point, tilt series alignment is one of the most crucial and unstable. It is also the most time-consuming task due to data curation and manual quality checks. Therefore, in this work, we present an automatic tool for data curation at this step of the pipeline.

This new software directly analyzes the tilt series before tomographic reconstruction. This contrasts with our previous work (de Isidro-Gómez et al., 2024), which required tomograms, providing an advantage in computational load and data handling. Additionally, this approach allows for detecting alignment errors that might be obscured in the tomographic reconstruction.

To achieve this purpose, the software has been designed to require minimal configuration, which requires only data known to the user. Essentially, only a few options are of interest for modification: first, applying the robust choice for pruning the residual vectors, and second, choosing between voting or the mean as criteria to determine if the tilt series presents misalignment.

The datasets analyzed in this publication aim to present good variability in sampling rate and gold bead size, offering a comprehensive scope of the characteristics of available public cryo-electron tomography datasets. In our testing, we compared the performance of the classification algorithm when fed with either the residuals provided by the alignment algorithm or the residuals calculated by the algorithm introduced in this work. The results show that the performance of the classification algorithm is subtly enhanced when the source of the residual vectors is the alignment algorithm in cases where fiducial detection is compromised (for example, the case of dataset EMPIAR 10364), using the alignment algorithm as a source of residuals offers the advantage that the reported information only considers those residual vectors involved in the alignment. If a landmark is undetected in a tilted image and excluded from the alignment, its residual vector is not reported. This is possible because some alignment algorithms use multiple partial fiducial chains without requiring a single chain covering the entire tilt series.

Additionally, applying robust pruning of residuals or using the voting criteria is recommended if the distribution of vectors is noisy, as this complicates the alignment assessment. For scientists using this software in the ScipionTomo framework, this behavior is configured by default when the residual calculation algorithm sources vectors, although users are always free to change it. Both ways of accessing the software are explained in a guide included in the Supplementary Material.

CRediT authorship contribution statement

F.P. de Isidro-Gómez: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **J.L. Vilas:** Writing – review & editing, Supervision, Software, Formal analysis, Conceptualization. **J.M. Carazo:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **C.O.S. Sorzano:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the financial support from the Ministry of Science, Innovation, and Universities, Spain (BDNS n. 716450) to the Instruct Image Processing Center (I2PC) as part of the Spanish participation in Instruct-ERIC, the European Strategic Infrastructure Project (ESFRI) in the area of Structural Biology. This work was also supported by Grant PID2022-136594NB-I00 funded by MICIU/AEI/10.13039/501100011033 and "ERDF A way of making Europe", by the European Union and Comunidad Autónoma de Madrid, Spain through Grant S2022/BMD-7232. Additional support came from the European Union (EU) and Horizon 2020 through the HighResCells grant (ERC - 2018 - SyG, Proposal: 810057) and from the European Union (EU) and Horizon Europe through the Fragment Screen grant (Proposal: 101094131).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jsb.2024.108153.

Data availability

Data and code are open to the public.

References

Burt, A., Cassidy, C.K., Ames, P., Bacia-Verloop, M., Baulard, M., Huard, K., Luthey-Schulten, Z., Desfosses, A., Stansfeld, P.J., Margolin, W., et al., 2020. Complete structure of the chemosensory array core signalling unit in an E. coli minicell strain. Nature Commun. 11 (1), 743.

Castano-Diez, D., Scheffer, M., Al-Amoudi, A., Frangakis, A.S., 2010. Alignator: A GPU powered software package for robust fiducial-less alignment of cryo tilt-series. J. Struct. Biol. 170 (1), 117–126.

Coray, R., Navarro, P., Scaramuzza, S., Stahlberg, H., Castaño-Díez, D., 2024. Automated fiducial-based alignment of cryo-electron tomography tilt series in Dynamo. Structure.

de Isidro-Gómez, F.P., Vilas, J.L., Losana, P., Carazo, J.M., Sorzano, C.O.S., 2024. A deep learning approach to the automatic detection of alignment errors in cryo-electron tomographic reconstructions. J. Struct. Biol. 216 (1), 108056.

de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M., Sorzano, C.O.S., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. 184 (2), 321–328.

de la Rosa-Trevin, J.M., Quintana, A., del Cano, L., Zaldivar-Peraza, A., Foche, I., Gutierrez, J., Gomez-Blanco, J., Burguet-Castells, J., Cuenca, J., Abrishami, V., Vargas, J., Oton, J., Sharov, G., Navas, J., Conesa, P., Vilas, J.L., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2016. Scipion: a software framework toward integration, reproducibility, and validation in 3D electron microscopy. J. Struct. Biol. 195, 93–99.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Amer. Statist. Assoc. 74 (366a), 427–431.

Fernandez, J.J., Li, S., 2021. TomoAlign: A novel approach to correcting sample motion and 3D CTF in CryoET. J. Struct. Biol. 213 (4), 107778.

Fernandez, J.J., Li, S., Bharat, T.A.M., Agard, D.A., 2018. Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. J. Struct. Biol. 202 (3), 200–209.

Galaz-Montoya, J.G., Flanagan, J., Schmid, M.F., Ludtke, S.J., 2015. Single particle tomography in EMAN2. J. Struct. Biol. 190 (3), 279–290.

Glaeser, R.M., Nogales, E., Chiu, W., 2021. Single-Particle Cryo-EM of Biological Macromolecules. IOP Publishing.

Gonzalez, R.C., 2009. Digital Image Processing. Pearson Education India.

Hou, G., Yang, Z., Zang, D., Fernández, J.-J., Zhang, F., Han, R., 2024. MarkerDetector: A method for robust fiducial marker detection in electron micrographs using wavelet-based template. J. Struct. Biol. 216 (1), 108044.

Iudin, A., Korir, P.K., Somasundharam, S., Weyand, S., Cattavitello, C., Fonseca, N., Salih, O., Kleywegt, G.J., Patwardhan, A., 2022. EMPIAR: the electron microscopy public image archive. Nucleic Acids Res. 51 (D1), D1503–D1511.

Jimenez de la Morena, J., Conesa, P., Fonseca, Y.C., de Isidro-Gomez, F.P., Herreros, D., Fernandez-Gimenez, E., Strelak, D., Moebel, E., Buchholz, T.O., Jug, F., Martinez-Sanchez, A., Harastani, M., Jonic, S., Conesa, J.J., Cuervo, A., Losana, P., Sanchez, I., Iceta, M., del Cano, L., Gragera, M., Melero, R., Sharov, G., Castano-Diez, D., Koster, A., Piccirillo, J.G., Vilas, J.L., Oton, J., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2022. ScipionTomo: Towards cryo-electron tomography software integration, reproducibility, and validation. J. Struct. Biol. 214 (3), 107872.

Lowry, R., 2014. Concepts and applications of inferential statistics. Mater. Sci. Appl. 6 (6).

Mahalanobis, P.C., 2018. On the generalized distance in statistics. Sankhyā Ind. J. Stat. A (2008-) 80, S1–S7.

Mastronarde, D.N., Held, S.R., 2017. Automated tilt series alignment and tomographic reconstruction in IMOD. J. Struct. Biol. 197 (2), 102–113.

Ni, T., Mendonça, L., Zhu, Y., Howe, A., Radecke, J., Shah, P.M., Sheng, Y., Krebs, A.-S., Duyvesteyn, H.M.E., Allen, E., et al., 2023. ChAdOx1 COVID vaccines express RBD open prefusion SARS-CoV-2 spikes on the cell surface. Iscience 26 (10).

Robertson, M.J., Meyerowitz, J.G., Skiniotis, G., 2020. Cryo-EM as a powerful tool for drug discovery. Bioorgan. Med. Chem. Lett. 30 (22), 127524.

Seifer, S., Elbaum, M., 2022. ClusterAlign: A fiducial tracking and tilt series alignment tool for thick sample tomography. Biol. Imag. 2, e7.

Sorzano, C.O.S., de Isidro-Gómez, F., Fernández-Giménez, E., Herreros, D., Marco, S., Carazo, J.M., Messaoudi, C., 2020. Improvements on marker-free images alignment for electron tomography. J. Struct. Biol. X 4, 100037.

Sorzano, C.O.S., Messaoudi, C., Eibauer, M., Bilbao-Castro, J.R., Hegerl, R., Nickell, S., Marco, S., Carazo, J.M., 2009. Marker-free image registration of electron tomography tilt-series. BMC Bioinformatics 10, 124.

Sternberg, S.R., 1983. Biomedical image processing. Computer 16 (01), 22–34.

Strelak, D., Jiménez-Moreno, A., Vilas, J.L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., de Isidro-Gómez, F.P., Horacek, J., Myska, D., Horacek, M., Conesa, P., Fonseca-Reyna, Y.C., Jiménez, J., Martínez, M., Harastani, M., Jonić, S., Filipovic, J., Marabini, R., Carazo, J.M., 2021. Advances in xmipp for cryo–electron microscopy: From xmipp to scipion. Molecules 26 (20), 6224.

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J.H., Welsch, S., Blanc, F.E.C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al., 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science 370 (6513), 203–208.

Van Drie, J.H., Tong, L., 2022. Drug discovery in the era of cryo-electron microscopy. Trends Biochem. Sci. 47 (2), 124–135.

Xu, Z., Li, H., Wan, X., Fernández, J.-J., Sun, F., Zhang, F., Han, R., 2024. Markerauto2: A fast and robust fully automatic fiducial marker-based tilt series alignment software for electron tomography. Structure.

Zar, J.H., 2010. Biostatistical Analysis. Prentice Hall.

Zheng, S., Wolff, G., Greenan, G., Chen, Z., Faas, F.G.A., Bárcena, M., Koster, A.J., Cheng, Y., Agard, D.A., 2022. AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. J. Struct. Biol. X 6, 100068.

Automatic detection of alignment errors in cryo-electron tomography

Supplementary material

F.P. de Isidro-Gómez[1,2], J. L. Vilas[1], J.M. Carazo[1*], C.O.S. Sorzano[1*]

November 23, 2024

[1] Biocomputing Unit, Centro Nacional de Biotecnologia (CNB-CSIC), Darwin, 3, Campus Universidad Autonoma, 28049 Cantoblanco, Madrid, Spain
[2] Univ. Autonoma de Madrid, 28049 Cantoblanco, Madrid, Spain
[*] Corresponding author

1 User guide

A simplified user guide to use the software presented in this work is introduced in this section. This guide includes both available sources available to the user: Xmipp standalone command-line mode and its integration in the Scipion framework.

1.1 Xmipp standalone

To work with the tools presented in the standalone version, only the Xmipp software package must be installed. Through the command line, the user has access to the full functionality of this software. The instructions for installing Xmipp can be found in github repository of Xmipp.

The xmipp program `xmipp_tomo_calculate_landmark_residuals` is responsible of detecting misaligned tilt series based on its residual vectors. The parameters of this program are summarized in Table 1.1. This is an example command for the execution of this program:

```
xmipp_tomo_detect_misalignment_residuals -i inputTs.mrcs --inputResInfo vResMod.xmd -o
alignmentReport.xmd --samplingRate 3.00 --fiducialSize 100.00
```

Parameter	Description
–inputResInfo	Input file containing residual models
-o	Output location for alignment report
–samplingRate	Pixel size of the input tilt series
–fiducialSize	Gold bead size in Angstroms
–numberTiltImages	Number of tilt-images in the series
–removeOutliers	Remove outlier residuals (robust mode)
–voteCriteria	Use a votting criteria (instead of average)

Table 1: Xmipp detect misalignment from residuals algorithm parameters.

If residual vectors are not provided, they can be calculated using the xmipp program `xmipp_tomo_calculate_landmark_` The parameters of this program are summarized in Table 1.1. This is an example command for the execution of this program:

```
xmipp_tomo_calculate_landmark_residuals -i inputTs.mrcs --tlt inputTlt.tlt --inputCoord
inputFiducialCoordinates.xmd -o vResMod.xmd --samplingRate 3 --fiducialSize 100.00 --thrSDHCC
3.00 --targetLMsize 8.00
```

1.2 Scipion framework

To work with the tools presented inside the scipion framework, it is necessary to have installed both the Xmipp and the Scipion software packages. This procedure is simplified since the installation of Xmipp

Parameter	Description
-i	Input tilt series
–tlt	Input tilt angle file
–inputCoord	Input 3D coordinates file
-o	Output location for residual models file
–samplingRate	Pixel size of the input tilt series
–fiducialSize	Gold bead size in Angstroms
–thrSDHCC	Threshold Z-score to consider a pixel value an outlier
–numberFTdirOfDirections	Number of directions to apply the bandpass filter
–targetLMsize	Target fiducial size for downsampling

Table 2: Xmipp calculate residuals from landmarks algorithm parameters.

is triggered when installing Scipion. The instructions for installing Scipion with Xmipp can be found in official webpage of Scipion. Since this is a simplified tutorial, the input information needed to feed the presented protocols is assumed to be available inside the Scipion project. However, more extensive documentation and tutorials on tomography data processing can be found in Scipion documentation landing page.

In particular, the "Tomogram Reconstruction" tutorial explains to the user how to obtain a set of aligned tilt series from a raw set of movies.

All algorithms presented in this work are maintained under the same Scipion protocol called `xmipptomo - detect misaligned TS`. Depending on the information that the user inputs to the protocol, it will trigger the different algorithms needed to assess the quality of the alignment.

Some software provide the residual information needed to directly study the alignment quality of the tilt series. This is, for example, the case of IMOD that is also the one used in the aforementioned tutorial. If the software provides this information, it is available inside the Scipion framework and thus can be input directly to the protocol. If this is the use case, an example of the configuration of the configuration of the protocol is available in Figure 1.

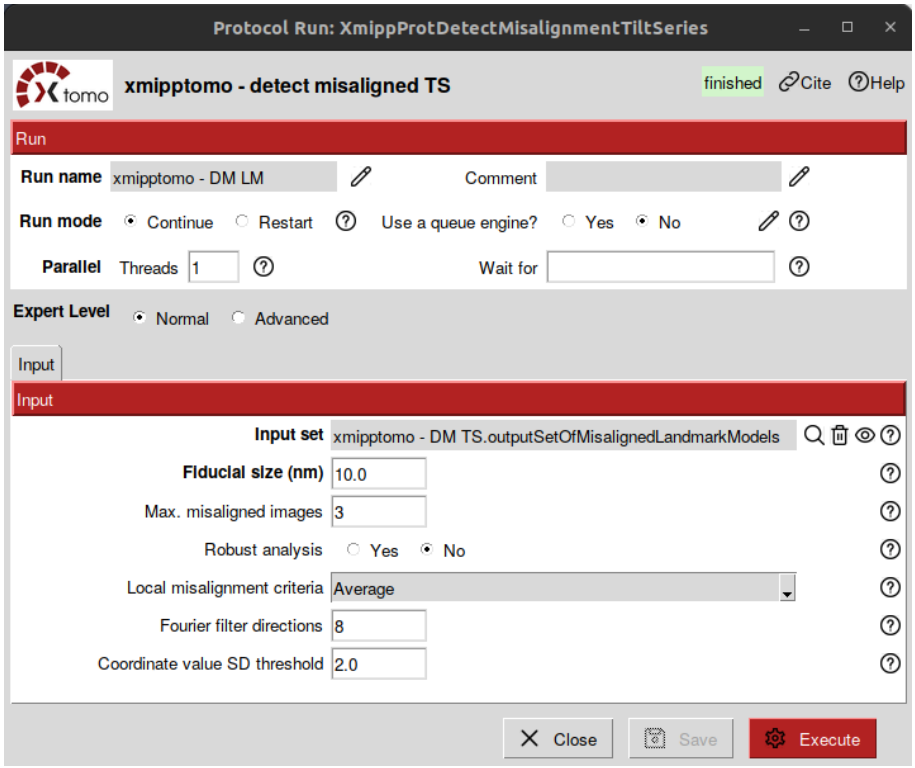

Figure 1: Detect misalignment protocol using residual vector models as input.

In case the residual information is not available is still possible to assess the quality of the alignment using the tilt series and the three dimensional coordinates of the fiducials. If this is the use case, an

example of the configuration of the configuration of the protocol is available in Figure 2.

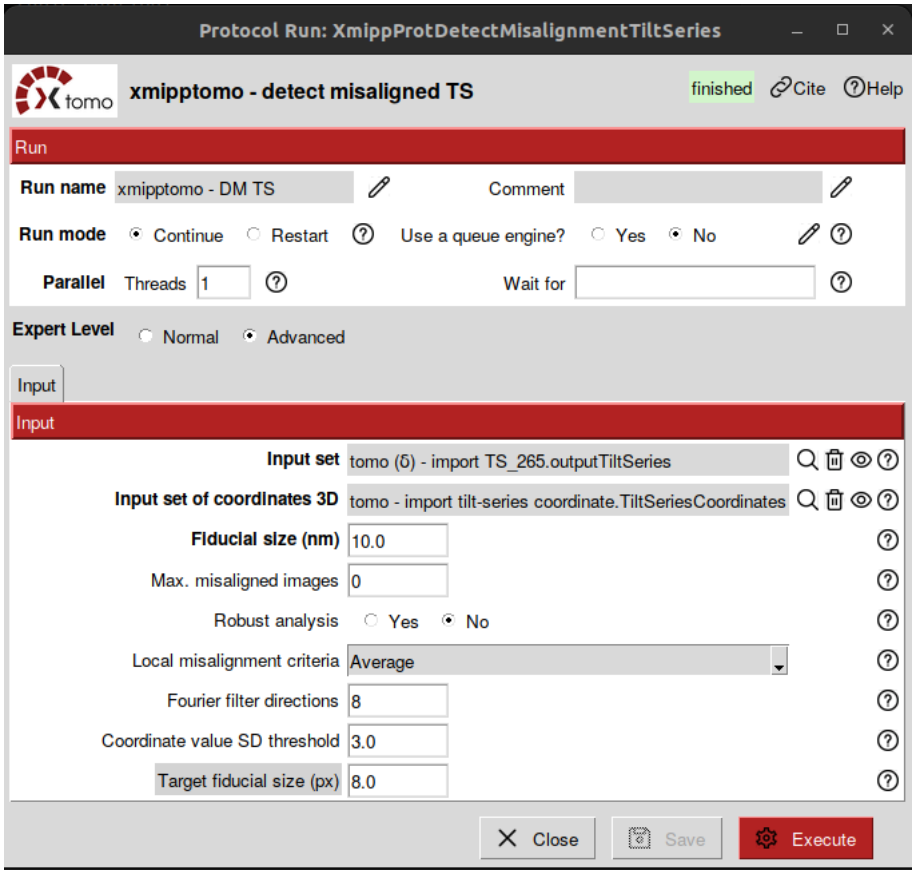

Figure 2: Detect misalignment protocol using the tilt series and the three dimensional coordinates of the fiducials as input.

The only option available in the protocol that is not in the Xmipp standalone version is the Maximum number of misaligned images. This option is inherent to the Scipion framework and sets the tilt series as misaligned if the number of detected misaligned tilt images exceeds this threshold.

2 Supplementary Movies

Two movies have been appended to the Supplementary Material to provide a clearer comparison between a tilt series with properly estimated geometry and one without:

1. Movie 1 presents and example of an aligned tilt series. This is tilt series TS_019 from EMPIAR 10453 dataset.

2. Movie 3 presents and example of a misaligned tilt series. This is tilt series TS_229 from EMPIAR 10453 dataset.

In addition, the calculated average Mahalanobis distance for each of these tilt series is shown in Figure 3.

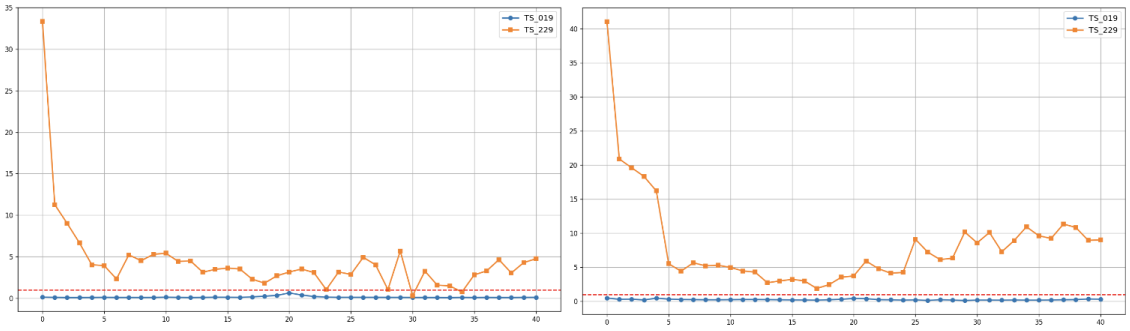

Figure 3: Associated average Mahalanobis distance for each tilt image belonging to Movies 1 and 2. The left plot shows the distribution of distances for each tilt image when the Mahalanobis distances are calculated over the residuals provided by IMOD. The right plot shows the same information when the residual calculation algorithm introduced in this work is the source of the residuals.

3 Supplementary Figures

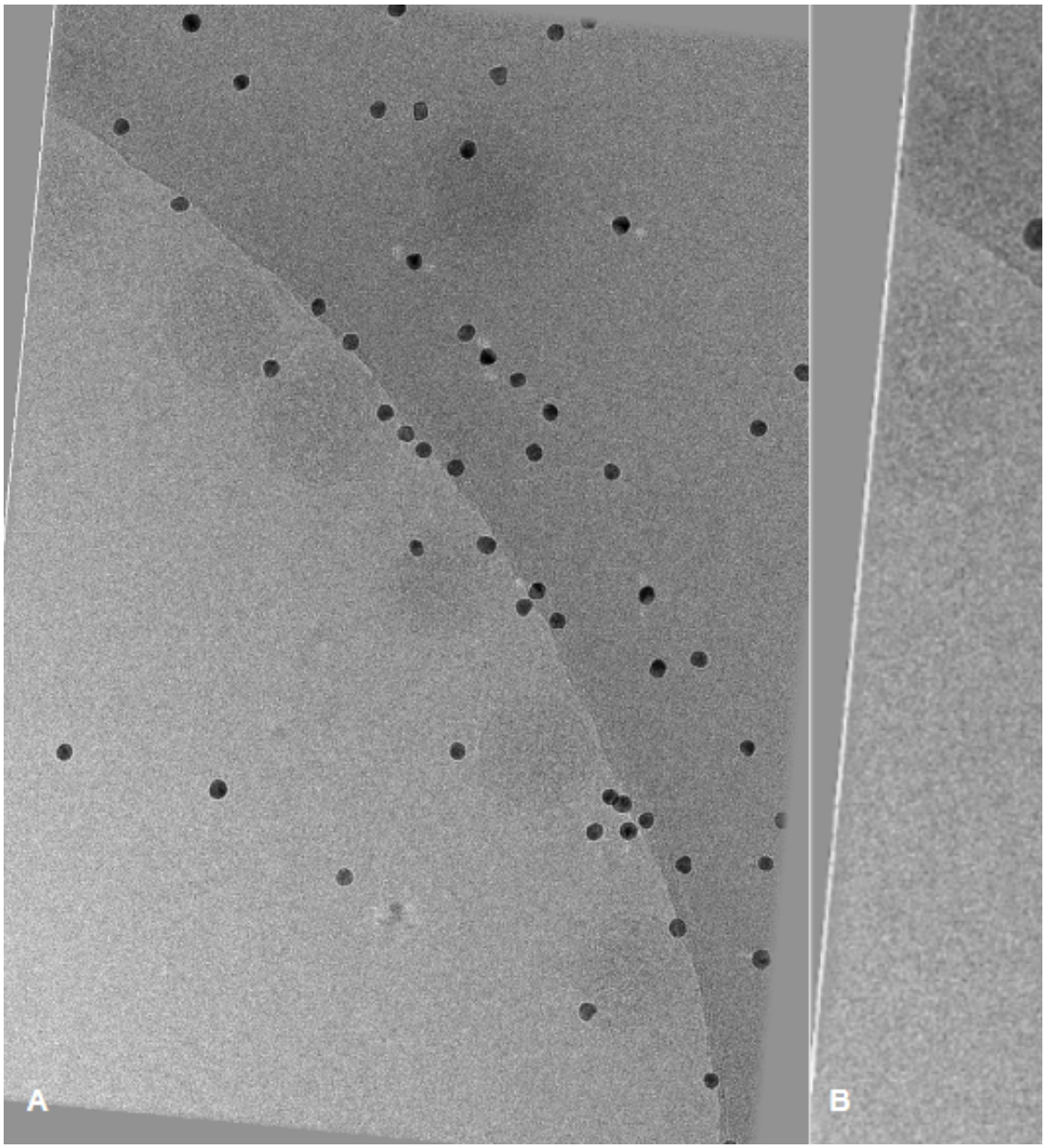

Figure 4: Example of border artifacts. (A) shows a tilt image presenting these artifacts in tilt series TS_005 from EMPIAR 10453, and (B) shows the detail.

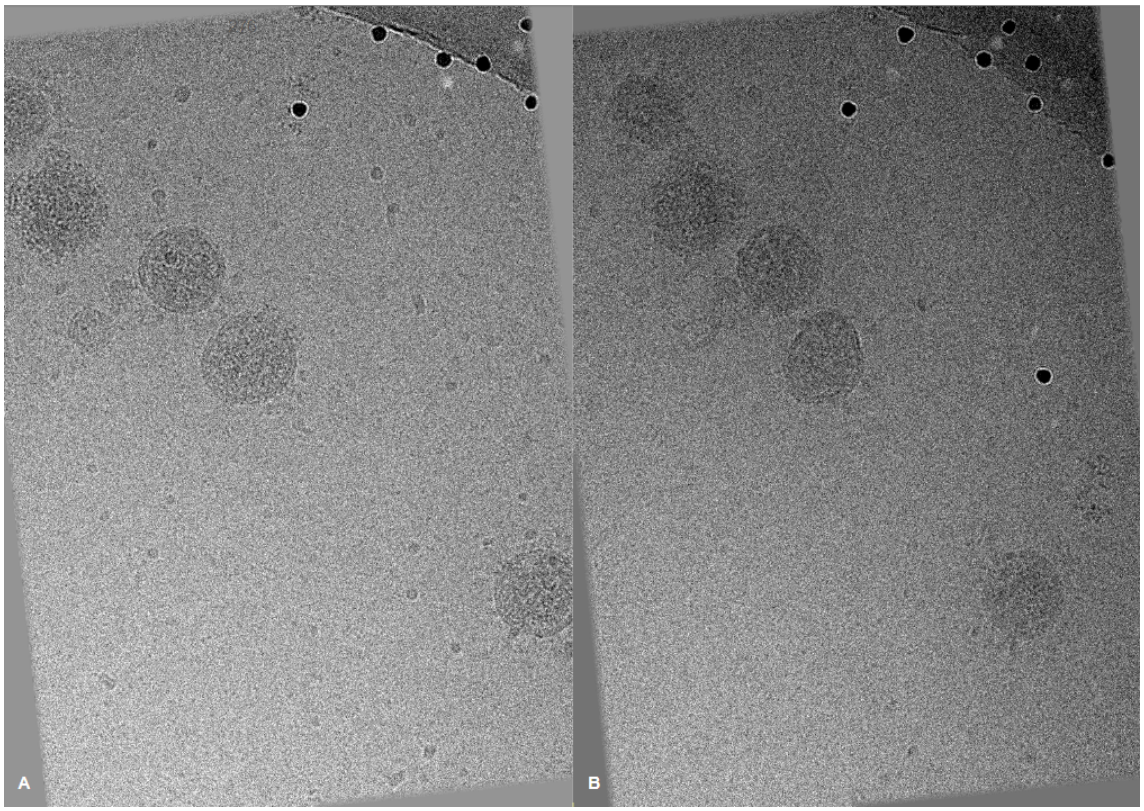

Figure 5: Tilt series TS_276 from EMPIAR 10453 dataset, false positive example from alignment quality assessment algorithm. This tilt series shows an unexpected distribution of residuals, being all of them clustered in one corner of the sample. (A) and (B) shows tilt images at 0° and 42° tilt angle respectively.

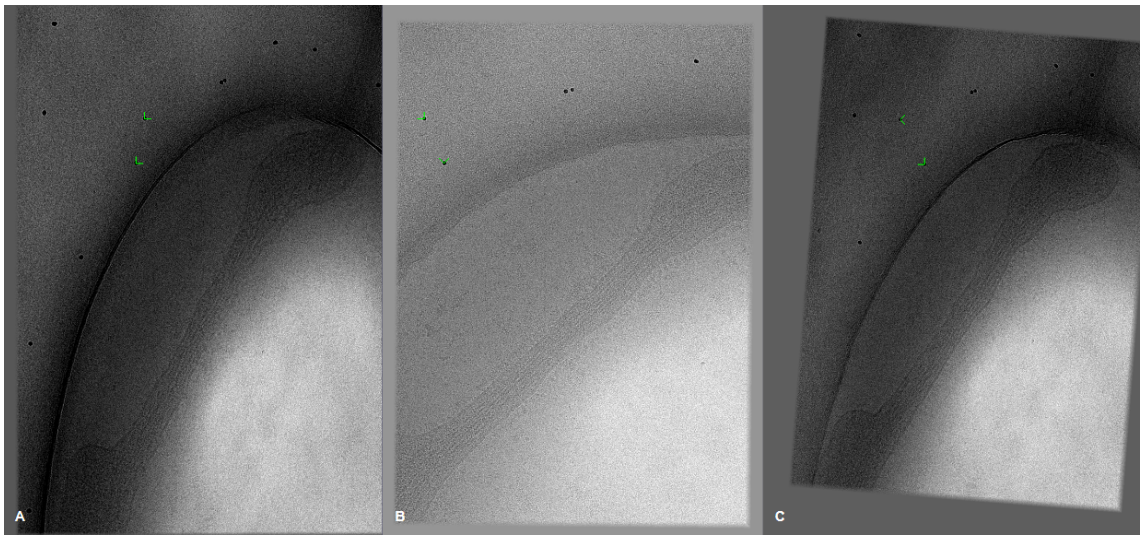

Figure 6: Tilt series E48g4_18 from EMPIAR 11457 dataset, false positive example from alignment quality assessment algorithm. This tilt series was aligned only using two very close fiducials, leading to low modulus residual vectors along the series. Residuals are plotted as yellow arrows but since their modulus is low, only the arrowhead is visible.

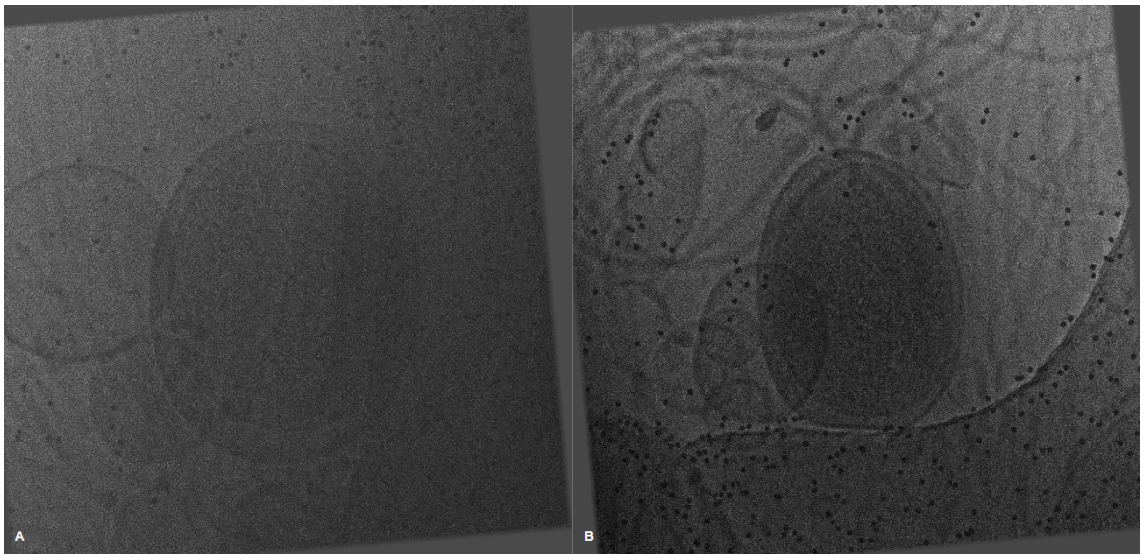

Figure 7: Tilt series TS002 (A) and TS017 (B) from EMPIAR 10364 dataset, both tilt images at -60°. Tilt series TS002 is a false negative example from the alignment quality assessment algorithm. The low contrast at high tilt for this tilt series difficult the fiducial detection and the posterior assessment of the alignment. The reduction in contrast is compared to TS017 tilt series, which has been correctly classified.

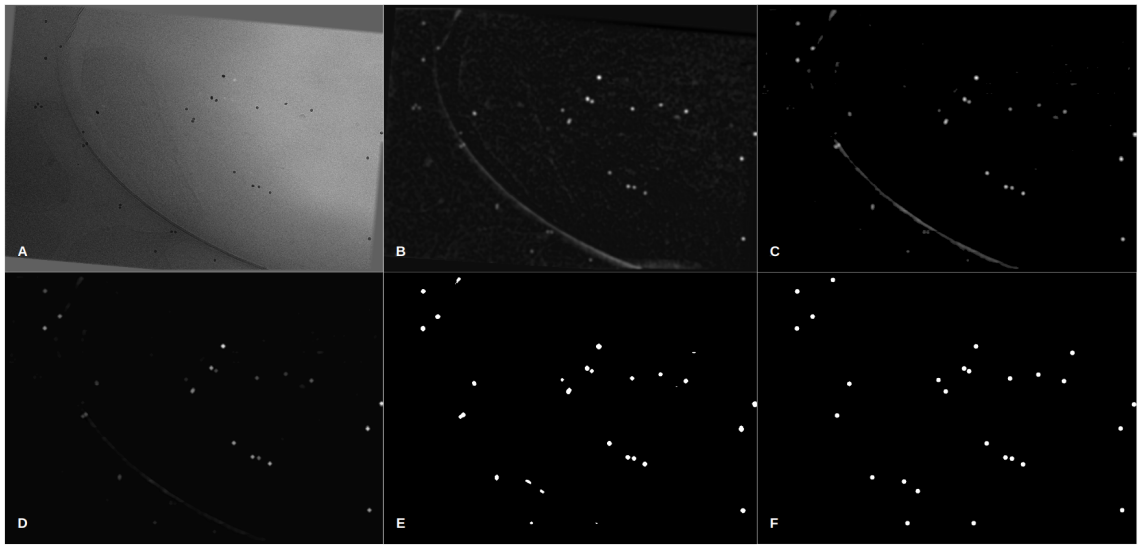

Figure 8: Intermediate results at different stages of the fiducial detection algorithm: (A) original tilt image, (B) result after landmark enhancement and background subtraction, (C) result after Z-score thresholding, (D) result after maximum pooling and directional filter, (E) labeled regions of interest after preprocessing, and (F) inpainting of the resulting detected landmarks after filtering the regions of interest. This tilt image belongs to tilt series E48g4_30 from EMPIAR 11457 dataset at 60°.