Journal of Structural Biology 196 (2016) 525–533

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Fast and automatic identification of particle tilt pairs based on Delaunay triangulation

J.L. Vilas [a,*], J. Navas [a], J. Gómez-Blanco [a], J.M. de la Rosa-Trevín [a], R. Melero [a], I. Peschiera [b], I. Ferlenghi [b], J. Cuenca [a], R. Marabini [c], J.M. Carazo [a], J. Vargas [a,d], C.O.S. Sorzano [a,d,*]

[a] Biocomputing Unit, Centro Nacional de Biotectnologíay (CNB-CSIC), Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain
[b] GSK Vaccine, Via Fiorentina 1, Siena, Italy
[c] Escuela Politécnica Superior, Universidad Autónoma de Madrid, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain
[d] Bioengineering Lab, Escuela Politécnica Superior, Universidad San Pablo CEU, Campus Urb. Montepríncipe s/n, 28668, Boadilla del Monte, Madrid, Spain

ARTICLE INFO

Article history:
Received 17 May 2016
Received in revised form 14 October 2016
Accepted 16 October 2016
Available online 18 October 2016

Keywords:
Electron microscopy
Single Particle Analysis
Automatic picking
Random conical tilt
Orthogonal tilt

ABSTRACT

Random conical tilt (RCT) and orthogonal tilt reconstruction (OTR) are two remarkable methods for reconstructing the three-dimensional structure of macromolecules at low resolution. These techniques use two images at two different sample tilts. One of the most demanding steps in these methods at the image processing level is to identify corresponding particles on both micrographs, and manual or semiautomatic matching methods are usually used. Here we present an approach to solve this bottleneck with a fully automatic method for assigning particle tilt pairs. This new algorithm behaves correctly with a variety of samples, covering the range from small to large macromolecules and from sparse to densely populated fields of view. It is also more rapid than previous approaches. The roots of the method lie in a Delaunay triangulation of the set of independently picked coordinates on both the untilted and tilted micrographs. These triangulations are then used to search an affine transformation between the untilted and tilted triangles. The affine transformation that maximizes the number of correspondences between the two micrographs defines the coordinate matching.

1. Introduction

Single Particle Analysis has become extremely important for microscopy due to its ability to determine the structure of macromolecules at high resolution, in some cases as fine as 1.8 Å (Merk et al., 2016) or 2.2 Å (Bartesaghi et al., 2015). A key element in the three-dimensional (3D) reconstruction process is specification of an initial volume of the macromolecule, which is then refined through an iterative procedure. Random conical tilt (RCT) (Radermacher et al., 1986; Sorzano et al., 2015) and orthogonal tilt reconstruction (OTR) (Leschnizer and Nogales, 2006) are two methods that address the problem of producing low resolution initial maps based on experimental information. RCT makes use of tilt angles, usually from 0° to 70°, whereas OTR uses two sample positions with an angular difference of 90°, which alleviates the missing cone limitation of RCT (Leschnizer and Nogales, 2006). Knowledge of the tilt angle facilitates assignment of orientations, as it imposes a constraint on every pa ir of particles, allowing for reconstruction of an initial volume.

When tilt pairs of micrographs are acquired, the information about particle orientation provided by the tilt angle can only be used if the same particle is identified in the untilted and tilted micrographs, to form a "particle tilt pair". Several studies have addressed particle selection of particle tilt pairs. Excluding manual methods, which require visual identification of corresponding pairs and manual picking in the two micrographs, both semiautomatic and automatic methods have been proposed. For instance, Xmipp (de la Rosa-Trevín et al., 2013) uses a semiautomatic method to pick tilt pairs from micrographs, which requires manual picking of a small set of particle tilt pairs. Once several particle pairs have been picked manually in untilted and tilted micrographs, the algorithm finds an affine transformation compatible with these pairs. For each new coordinate, untilted or tilted, the program then suggests the corresponding coordinate in the other micrograph. Other methods are automatic, such as *TiltPicker* (Voss et al., 2009), which establishes particle pair matching by attempting to transform the untilted to the tilted image by rotations and translations. *MaverickTilt* (Hauer et al., 2013) determines tilt pair correspondence using an affine transformation based on the invariance of

* Corresponding authors at: Biocomputing Unit, Centro Nacional de Biotectnologíay (CNB-CSIC), Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain (C.O.S. Sorzano).
E-mail addresses: jlvilas@cnb.csic.es (J.L. Vilas), coss@cnb.csic.es (C.O.S. Sorzano).

http://dx.doi.org/10.1016/j.jsb.2016.10.007
1047-8477/

barycentric coordinates in this kind of transformation. A new method based on tilt invariant neighbors was recently proposed (Shatsky et al., 2014), which also explains the limitations of using affine transformations to identify particle tilt pairs. Despite all these procedures, however, the identification of particle tilt pairs remains a tedious, time-consuming step in RCT and OTR reconstruction workflow.

Here we present a new, rapid algorithm for automatic selection of particle tilt pairs, which has been integrated into the *Xmipp* (de la Rosa-Trevín et al., 2013) and *Scipion 1.0* (de la Rosa-Trevín et al., 2016) processing frameworks. The method comprises two steps, 1) particles are automatically detected in both micrographs independently, providing the coordinates of the untilted and tilted particles, after which 2) particle pairs are then automatically identified. In Step 1, any automatic picking tool can be used, such as *Xmipp3 - Autopicking* (Abrishami et al., 2013), *EMAN - boxer* (Tang et al., 2007) or (Hoang et al., 2013), among others. Picking can thus be performed and assignment of particle tilt pairs carried out using only the coordinates of the picked particles. Step 2 uses the coordinates obtained to create two Delaunay triangulations (untilted and tilted). Finally, the best affine transformation is sought according to the number of matched particle tilt-pairs between untilted an d tilted triangles. The core of this algorithm is the affine transformation; in the Appendix A we discuss the use of this type of transformation to assign particle tilt pairs. We show the performance of the algorithm with two experimental examples illustrating quite different scenarios of the use of the algorithm.

2. Methods

Let us consider two micrographs of the same sample recorded at two different orientations of the sample holder (untilted and tilted images), using an electron microscope. Without loss of generality, let us assume that the untilted position coincides with the horizontal plane. This condition will be discussed below (Section 2.2). The goal is to establish a one-to-one transformation between coordinates in the untilted and tilted micrographs. A natural choice for this correspondence is an affine transformation as shown in Sorzano et al. (2015). The search for this transformation is performed in several steps.

1. *Two previous, independent pickings*. Assume that both micrographs of the tilt pair are fully picked. Due to the large number of particles that could be present in the original micrographs and to the combinatorial nature of the optimization problem, it is impractical to determine the optimum transformation between untilted and tilted coordinates using an exhaustive search.
2. *Delaunay triangulations*. Two Delaunay triangulations are built with the coordinates identified in each micrograph (one for the untilted and other for the tilted micrograph). Delaunay triangulation presents several advantages for browsing tilt pairs, as it considerably helps in the search for the affine transformation; triangles simplify the combinatorial nature of the problem. In addition, given its relationship with the Voronoi diagram, efficient algorithms can be used for rapid computation of nearest neighbors and distances.
3. *Search for the affine transformation*. This problem could be approached by triangulating the set of particle coordinates of each micrograph and identifying corresponding triangles between both triangulation sets. Because the behavior of the affine transformation found is not perfect (the tilt coordinates predicted by the untilted coordinates usually lie on coordinates near the experimental position), this task requires the identification of those points close to the predicted coordinate. For this task we have also made use of the Delaunay triangulation (Mulchrone, 2003;Singh et al., 1996).

In summary, the algorithm begins with the two sets of coordinates (Fig. 1). Once the untilted and tilted particles have been fully and independently picked, a coarse search step establishes a rough first correspondence of particle tilt pairs, after which the refinement step will enhance this matching for a more accurate affine transformation.

2.1. Delaunay triangulation

Delaunay triangulation defines a type of triangulation such that any circumcircle to every triangle contains only the three vertexes of the triangle. An extensive mathematical treatment can be found in de Berg et al. (2008). For a given set of n points, the Voronoi diagram assigns a single polygon around every point of the set, where each point is interior to the polygon and verifies that this inner point is the closest point of the set to any other point in the interior of its polygon. Once the Voronoi areas are computed, the location of the closest point to a new input point is almost trivial, as it is only necessary to identify the Voronoi area that encloses this new point (see Fig. 2, which shows the Delaunay triangulation (continuous black) and the Voronoi diagram (dashed red) for a set of points.

The Voronoi areas of a triangulation can be computed easily once the Delaunay triangulation of the set of points has been built. In our implementation, incremental Delaunay triangulation was used to create the entire net (see Fig. 3) by adding a new point to the triangulation at each step of the algorithm and updating the triangulation so that all triangles fulfill the Delaunay conditions. When a new point is added, the triangle that contains this point is split into three new triangles, the edges of which must

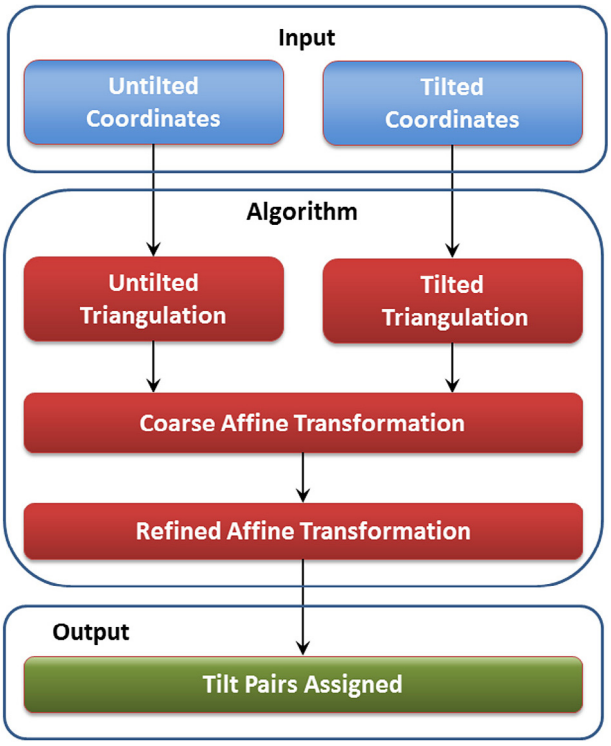

Fig. 1. Assignment Tilt Pairs algorithm workflow. Blue blocks represent the input coordinates, red blocks constitute the algorithm and green block is the result of the assigning process.

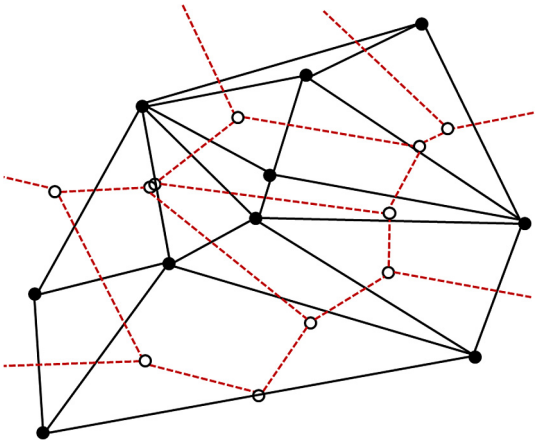

Fig. 2. Voronoi tesselation (dashed-red) created from a Delaunay triangulation (continuous-black).

be tested and the triangulation modified for all edges that do not fulfill the Delaunay conditions. The outcome of this part of the algorithm is a tree that contains the Delaunay triangulation. At the root of this tree is a large triangle that contains all points in the set. The leaves of this tree describe the final Delaunay triangulation. The nodes between the root and the leaves describe triangles of different sizes with a hierarchical (nested) relationship.

To add a new point, the algorithm searches from the root to the leaves of this tree, and for each node tests which of its three children encloses the new point. When a leaf node is reached, it means this node is the triangle that encloses the new point and it is split into three new triangles, its edges are checked and the triangulation modified, if necessary. On average, the path from root to node in the graph requires $O(\log n)$ operations and, as there are n points to be added, the incremental Delaunay triangulation requires $O(n \log n)$ operations (de Berg et al., 2008). The construction of the Voronoi areas require $O(n)$ operations; so that the overall complexity to build both structures (Delaunay and Voronoi) is $O(n \log n)$.

The final step of the algorithm involves searching for affine transformations between Delaunay triangulations in both the untilted and tilted micrographs, for which triangles from a micrograph must be found in the triangulation of the other micrograph. In Fig. 4, Delaunay triangulations for a set of particle tilt pairs are superposed to their corresponding micrographs.

2.2. Properties of an untilted-tilted affine transformation

A transformation that preserves the colinearity and ratios among elements (it is not necessary to maintain distances and angles, only the spatial relationship between their elements) is called an affine transformation. It can also be defined as a projective transformation between two planes, in which the projection vertex is located at infinity and in a certain direction. An affine transformation thus arises as the natural means to establish a

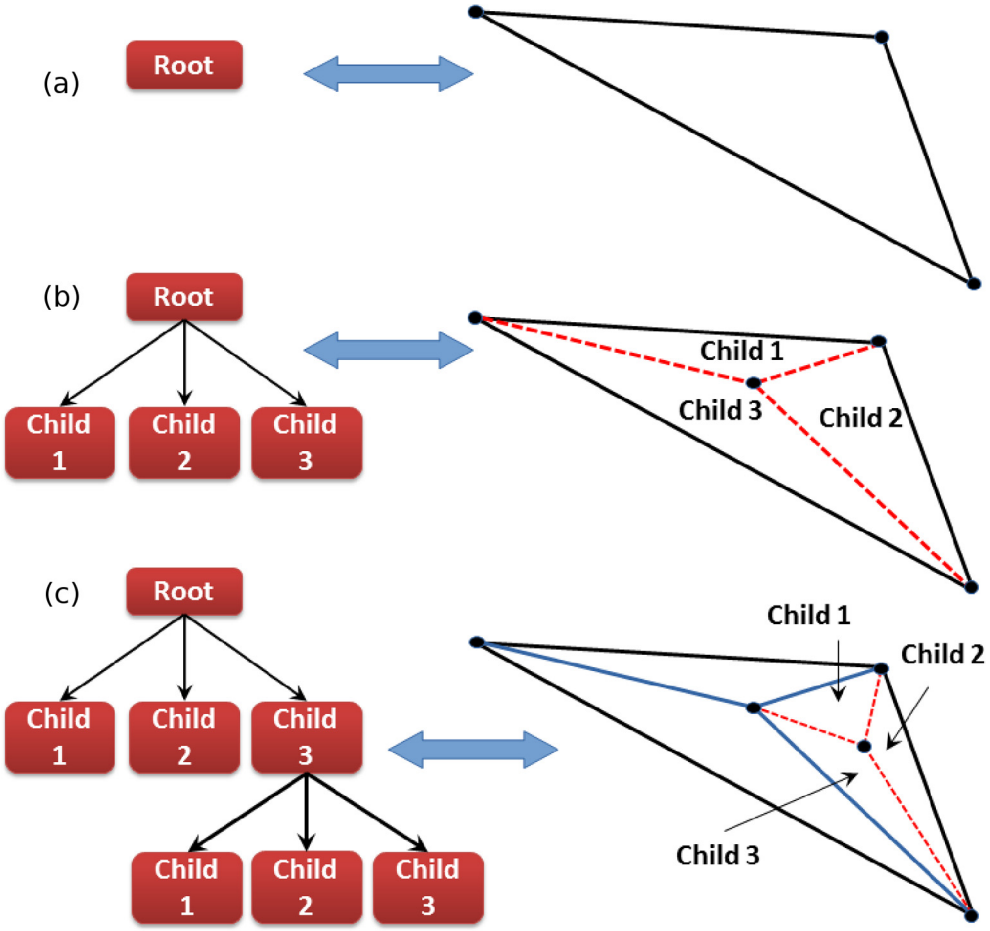

Fig. 3. Graph construction process of the Delaunay incremental algorithm. (a) Initial triangle, all points of the set are inside. (b) Triangulation state after first insertion. (c) Triangulation state after second insertion.

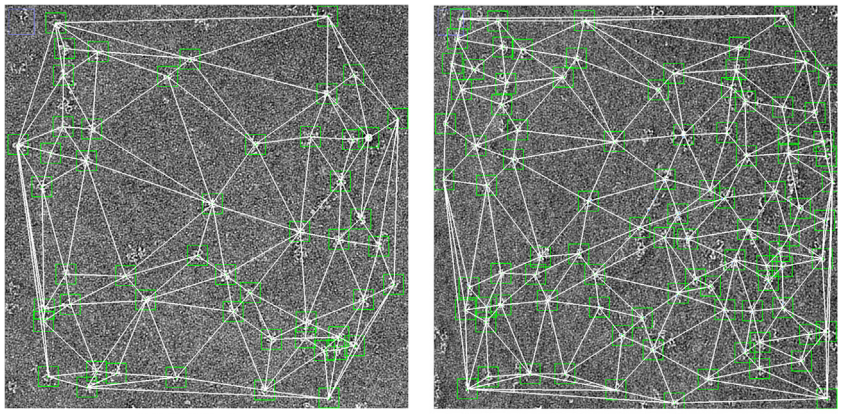

Fig. 4. Independently picked particles in the (a) untilted, and (b) tilted micrographs corresponging to Sample 2. Their respective Delaunay triangulation are shown as well.

one-to-one correspondence between two sets of coplanar coordinates (untilted and tilted). An affine transformation can be expressed mathematically as the combination of an arbitrary, non-degenerate linear transformation and a translation,

$$\mathbf{u} = A\mathbf{t} + \mathbf{s}, \tag{1}$$

where $\mathbf{u} = (u_x, u_y)^T$ and $\mathbf{t} = (t_x, t_y)^T$ are the coordinates of the untilted and tilted particles in their corresponding micrographs, T denotes the transpose operator, $\mathbf{s} = (s_x, s_y)^T$ is a translation vector, and A is a 2×2 matrix termed transformation or affine matrix. Hence, an affine transformation is completely defined by knowing the exact correspondence between three pairs of points (untilted-tilted). It involves the estimation of six parameters: two shifts s_x, s_y and four matrix elements, a_{ij}. As a consequence, triangulating the untilted and tilted sets of coordinates simplifies the search for affine transformations (we need only identify which triangle from the untilted triangulation corresponds to which triangle in the tilted triangulation). Given two triangles, there are 6 affine transformations that relate them, because of the 6 possible permutations of 3 vertices. In principle, the correct correspondence could be gi ven by any of these permutations and all must be explored.

The eigenvalues and eigenvectors of the affine matrix provide information about the tilt angle, θ, and its axis. This is the angle between the horizontal plane and the tilt plane. When the sample is tilted, the distances along the normal direction to the tilt axis shrink by a factor $\cos\theta$ (Guckenberger, 1982); the eigenvalues must therefore be 1 and $\cos\theta$. In addition, the affine transformation sought cannot have negative eigenvalues, since they would imply a mirror of one of the micrographs, which is impossible.

As a result, we can impose the following set of constraints on the eigenvalues and the determinant of the affine transformation found:

- The matrix A must be positive definite.
- The eigenvalues must be close to 1 and $\cos\theta$:

$$\left|\frac{1}{2}\left[\mathrm{Tr}A + \sqrt{\mathrm{Tr}A - 4\det A}\right] - 1\right| \leqslant \epsilon \tag{2}$$

$$\left|\frac{1}{2}\left[\mathrm{Tr}A - \sqrt{\mathrm{Tr}A - 4\det A}\right] - \cos\theta\right| \leqslant \epsilon \tag{3}$$

- The determinant must meet $|\det A - \cos\theta| \leqslant \epsilon$. This condition is a consequence of the previous one.

If the measurement conditions were ideal and without errors, ϵ would be zero, and the inequality becomes an equality. Nevertheless, perfect conditions are never achieved and, as a consequence, ϵ represents a small positive value. These conditions make the search for the affine transformation efficient, since they provide a quick test to check whether an affine transformation candidate is suitable for defining the untilted-tilted correspondence.

Finally, the reference system considers that the untilted micrograph coincides with the horizontal plane. The nomenclature "untilted" can be extended to other non–horizontal positions; nevertheless, calculation of the tilt angle and its axis requires a reference system. For this reason, in methods such as OTR, tilt pairs can also assigned, although the tilt angle cannot. In this case, the affine transformation will be an isommetry, that is, the identity and/or a translation. The tilt angle is therefore zero, but the tilt pairs would be correctly assigned. Estimation of the tilt angle in untilted positions outside the horizontal plane should thus be treated carefully.

2.3. Search of the affine transformation

The search to assign particle tilt pairs is divided into two steps. First, a coarse search tries to identify a suitable affine transformation by means of triangle pairs and second, the affine transformation calculated in the first step is refined to include more points and increase accuracy.

1. *Coarse search*: The goal of this step is to find a pair of corresponding triangles, that will allow us to determine the affine transformation. We know we have two corresponding triangles because the number of inliers (matched points) detected after applying the affine transformation induced by these two triangles is relatively large; that is, if we find two corresponding triangles, not only will their vertices match between the untilted and tilted micrographs, but many other points will also match, especially those around the seed triangles. Our goal, at this stage is thus to find at least one pair of corresponding triangles. To do so, we sort the untilted and tilted triangles by descending area, and 20% of the triangles with the smallest areas are discarded; this threshold attempts to speed up computation time and increase reliability. Each untilted triangle is compared to all tilted triangles whose areas are smaller than the area of the untilted triangle. F or each comparison, six possible affine transformations are calculated (corresponding to the six possible vertex matches) with $\theta < 70^\circ$. This restriction obeys to the fact that the tilt angle is measured experimentally in that range. For those transformations that meet the conditions of Section 2.2, we calculate the position of all untilted coordinates predicted by the transformation matrix and count the number of inliers (a predicted tilted coordinate is an inlier if its distance

to a picked tilted coordinate is less than a tolerance). The tolerance for identifying inliers is a given percentage of particle size, $20-50\%$ are typical values; justification of this criterion can be seen in the Appendix A. We thus look for the transformation with the maximal number of inliers. When two affine transformations have the same number of inliers, we retain the one whose predicted coordinates are closer to the picked tilted coordinates.
2. *Refinement*: In the previous phase, we determined a number of corresponding coordinate pairs, as many pairs as there are inliers. With these pairs, we solve Eq. (1) in a least squares sense using the Moore–Penrose pseudoinverse. The goal is to achieve a refined affine transformation, after which the updated transformation is used to identify more inliers, and the refinement step is repeated until no new inlier is found.

2.4. Automatic determination of the tilt angle

The relationship between the area of two corresponding triangles (one in the untilted micrograph and its corresponding triangle in the tilted one) must be $A_{tilt} = A_{untilt}\cos\theta$. Once a set of tilt pairs has been identified, an estimation of the tilt angle between the two micrographs is calculated by analyzing the areas of the matched triangles. The tilt axis direction is calculated by proposing an Euler transformation from the untilt to tilt coordinates with the tilt angle obtained. An angular optimization that minimizes the quadratic mean error is then performed. Hence, a better estimation of the tilt angle is achieved while a quality measure for this estimation is provided in the form of its standard deviation. A detailed information about how the tilt angle and tilt axes are calculated is given in Eq. (6) in Sorzano et al. (2015).

2.5. Robustness of the algorithm to false positives

The method uses two independent previous pickings to establish correspondence of tilt pairs. The algorithm is affected by the accuracy of each picking method. If one picking method detects a false positive in the untilted (or tilted) micrograph, the assignment of tilt pairs requires that:

1. The other picking finds the same false positive in the other micrograph, although the tilted micrograph shows defocus and less quality.
2. The affine transformation is compatible with the false positive tilt pair, in other words, a transformation compatible with tilt angle, the eigenvalues of the affine matrix, shift, and triangles area condition.

As a consequence, the algorithm rarely gives a false positive result. In contrast, if one picking detects a true particle and the other does not, the assignment cannot find the tilt pairs of the particles selected. This situation does not depend on the accuracy of the assignment algorithm, but rather on the input data, i.e. previous pickings and their accuracy. It is thus better to pick more points than an usual picking for a Single Particle Analysis project, even with the risk that the picking includes noise. This ensures that all particles are picked, as the algorithm will discard false positives.

3. Results

To test the algorithm, we used two experimental conditions, one in which the field of view contains many particles and one with very few (Fig. 5). We automatically identified particles in

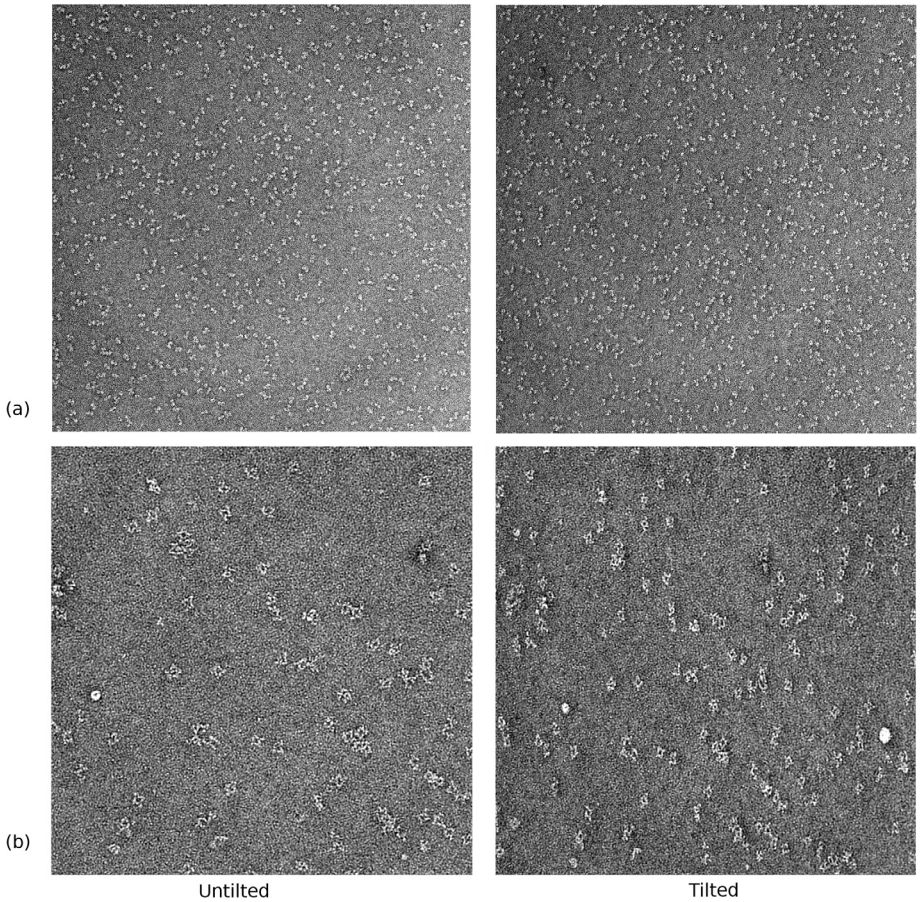

Fig. 5. Tilt pairs of micrographs for (a) the Eukaryotic primosome sample with a tilt angles of 40°, and (b) Immune complex with a tilt angle of 55°.

the untilted and tilted micrographs independently using *Xmipp3 - Autopicking* (de la Rosa-Trevín et al., 2013;Abrishami et al., 2013), an automatic particle picker able to learn from the user.

3.1. Sample 1: Eukaryotic primosome

The complex comprising the subunit CTD (C – terminal domain) of the polymerase (Pol α) and the primase in the eukaryotic primosome (Nuñez Ramirez et al., 2011) was selected as our first test case. It is a large macromolecule with a molecular weight of 170 kDa. Six tilt pairs of micrographs (4096 × 4096 pixels) were acquired automatically using a TVIPS F416 CMOS sensor, with a tilt angle of 40°, in a JEOL JEM-1230 TEM microscope with an acceleration voltage of 100 kV, a nominal magnification of 54,926, and 2.84 Å/pixel; remaining experimental details have been reported in Nuñez Ramirez et al. (2011). The mean number of particles per micrograph was 736 and 900 for the untilted and tilted micrographs with standard deviations of 33 and 42 particles, respectively. This data set shows a large shift between the untilted and tilted micrographs, as much as 40% of micrograph size, and the field of view contains many particles that could hinder manual identification of par ticle pairs. Our method found the correct correspondences for all six tilt pairs of micrographs. *Xmipp Autopicker* (Abrishami et al., 2013) identified 4416 and 5398 particles in the total data set of untilted and tilted micrographs, respectively, and our new algorithm found correspondences between 2237 particles, which approximately coincides with a manual analysis of these micrographs (Fig. 6). In addition, the algorithm automatically estimated the tilt angle for each tilt pair. The average tilt was calculated to be 38.97° and the standard deviation (SD) 0.54°, which coincides with the nominal tilt (numerical results are summarized in Table 1). A bar plot showing predicted tilt angles for all six micrographs indicated good agreement between experimental and measured tilt angle (Fig. 7a). Mean computation time was 50 s per tilt pair (using a laptop computer with a 4-core Intel *i*7 processor and 8 Gb RAM), using the parallelization available in *Scipion*.

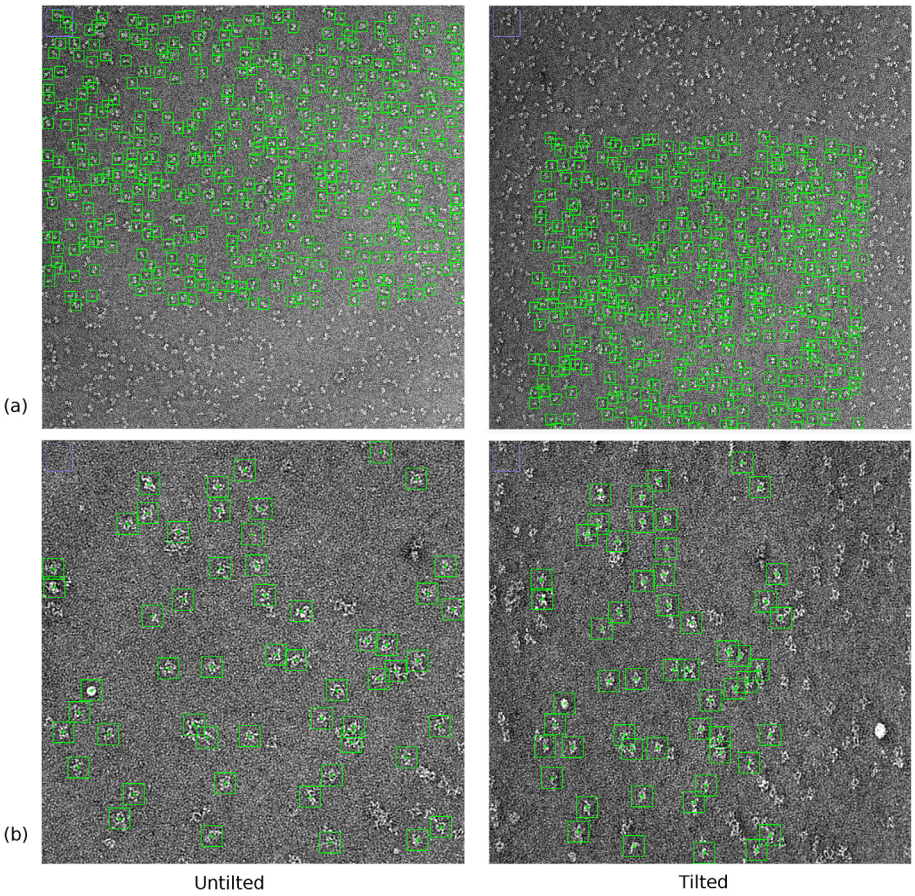

Fig. 6. Assignment of particle tilt pairs for both samples, at the left, untilted micrographs, at the right, tilted micrograph. (a) Eukaryotic primosome sample with a tilt angles of 40°, and (b) Immune complex with a tilt angle of 55°.

Table 1
Summary of information for samples 1 (eukaryotic primosome) and 2 (antibody-antigen complex).

Sample	Mean tilt(°)	SD (°)	Micrograph	Picked particles	Tilt pairs assigned
Sample 1	38.97	0.54	Untilted Tilted	4416 5,398	2229
Sample 2	53.47	0.44	Untilted Tilted	1927 2819	1051

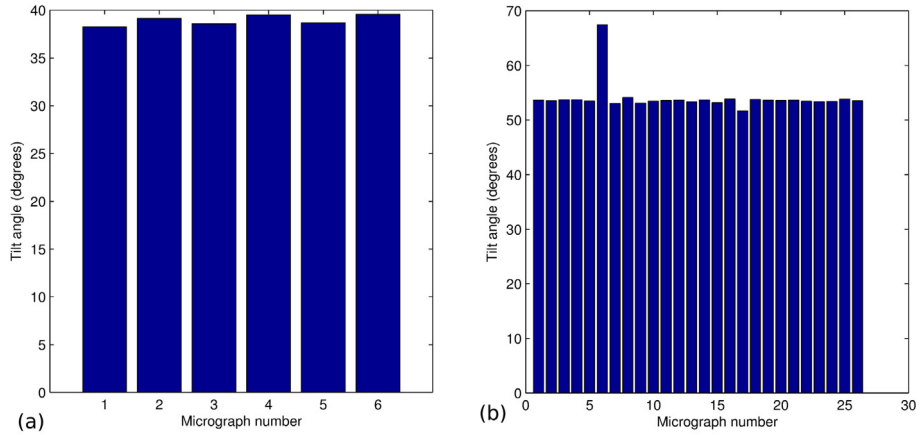

Fig. 7. Predicted tilt angle for each micrograph. (a) Sample 1: eukaryotic primosome, experimental tilt was 40°. (b) Sample 2: Immune complex, experimental tilt was 55°.

3.2. Sample 2: antibody-antigen complex

In our second example, we analyzed tilt pairs of micrographs of two human antibodies to the same antigen forming an immune complex, with a total molecular weight of 355 kDa. A data set of 26 tilt pairs of images with a tilt of 55° and size of 2048 × 2048 pixels were acquired on a Phillips CM200-FEG microscope operating at 200 kV equipped with a TVIPS TemCam-F224HD sensor. Nominal magnification was 50,000 and pixel size was 3.3 Å. The sample was chosen specifically to test the algorithm with a relatively small macromolecular complex and a small number of particles per micrograph; a representative tilt pair micrograph from this set is shown in Fig. 5. The mean number of particles per micrograph was 74.1 and 108.4 for untilted and tilted micrographs, with respective SD of 10.3 and 14.6. The total number of particles identified in the untilted and tilted micrographs was 1927 and 2819, respectively. Of the 26 tilt pairs, the alg orithm was able to pair 25 correctly (96.15% accuracy), with a total of 1051 particle pairs (Fig. 6). The tilt angle was also estimated automatically with a mean value of 53.47° and a standard deviation of 0.44°. These results are summarized in Table 1. The bar plot for the micrographs shows the predicted tilt angle and presents only one angle out of range (Fig. 7b). This algorithm, run on the same laptop above, took less than one minute to compute the 26 tilt pairs of micrographs.

4. Discussion and conclusions

In this study, we advanced in the automation and reliability of the processing of cryoelectron microscopy images by proposing an affine algorithm to pair particles automatically in untilted and tilted micrographs in a rapid, fully automated manner. In this way, particle pairs can be determined in a simple and quick manner, removing one of the bottlenecks of classical RCT and OTR workflows. This method helps to accelerate the first step in the workflow, to achieve 3D reconstruction of macromolecules using tilt pairs in the *Scipion 1.0* (de la Rosa-Trevín et al., 2016). In addition, assignment of tilt pairs completes the circle, with the recent publication of the mathematical fundamentals of RCT reconstruction (Sorzano et al., 2015), which provides new expressions for handling the geometric parameters without intermediate operations. *Scipion 1.0* capability for tilt pair processing is thus enhanced. The samples used for validation of the algorithm were chosen for macromolecular weight, tilt angle, number of particles per field of view, shift between untilted and tilted micrographs, geometry and distance between particles, to allow evaluation of more than one scenario and to determine the performance of the method. Sample 1, the eukaryotic primosome, has a large number of particles with small distances between them, whereas Sample 2, the antibody-antigen complex, had fewer, more separated particles. Our algorithm produced correct particle correspondence in both cases.

The algorithm input consists of two sets of coordinates, and the correspondence between these sets is determined automatically. If automatic picking is used to select the particles in the untilted and tilted micrographs, a particle might be selected in one micrograph but not the other. As long as particle selection is similar in both micrographs, our algorithm is relatively robust in this situation and allows identification of similar triangles in both micrographs.

The only two parameters of the algorithm are the maximum expected shift (which can be as much as 40% of micrograph size) and the tolerance for identifying a predicted tilted coordinate as an inlier. Our experience is the robustness of the algorithm to these two parameters. The first one mainly affects the algorithm speed, while the second might affect its accuracy if the tolerance is too small or too large; typically, a distance of 20–50% of particle size should be an appropriate choice.

The algorithm is fully automatic, although manual interaction with the results is possible at the end of the process, to allow individual particle pairs to be disabled if, for instance, the tilt angle is out of range and yields incorrect particle pairing.

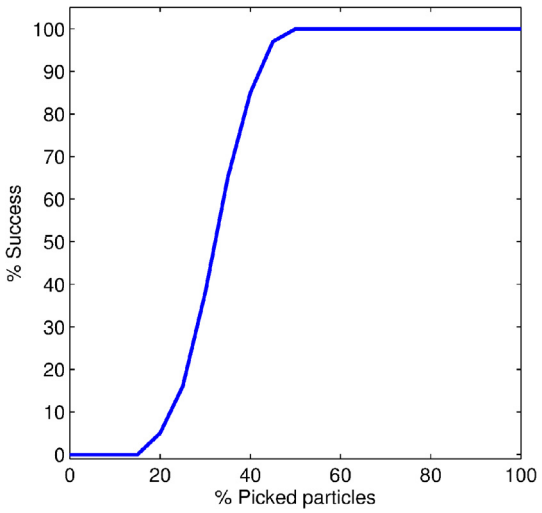

Fig. 8. Robustness test: succeed of the method under percentages of randomly removed particles in both micrographs, untilted and tilted.

The algorithm showed remarkable robustness in distinct picking conditions. To test this, a fixed percentage of particles was removed at random in one untilted and tilted micrograph pair for Sample 1. The algorithm was launched 100 times for each percentage of particles removed, from 0 to 100% of the total number of picked particles. This process is summarized in Fig. 8, in which the X-axis determines the percentage of particles removed and the y-axis indicates the percentage of success defined as the probability of matching. It can be seen that the algorithm shows was considerably robust with a success rate of 97% when more than half (55%) of the particles had been removed.

Beside the robustness of the algorithm, one of its main advantages is its speed. Two main features contribute to this end. The first is the early rejection of many affine transformations that do not meet the algebraic conditions that correspond to correct untilted-tilted transformation; this filter removes most of the possible affine transformations to be evaluated. For the remainder, most of the calculation time is spent in searching for the nearest neighbor of a predicted tilted coordinate. The Delaunay triangulation used here greatly reduces this search time, which is $O(\log n)$ in our implementation. This speed is a clear advantage with respect to other fully automatic algorithms, whose execution time can be in the order of hours, depending on particle numbers, in particular, the Particle Correspondence Tilt Pairs method (PTC, Hauer et al., 2013) applied to the Sample 1, took a mean time of 14.6 min oer micrograph while our algorithm took 50 s per micrograph.

The use of an affine transformation to identify tilt pairs was recently discussed (Shatsky et al., 2014), with the suggestion that sample curvature might introduce errors in the tilt axis calculation. This would lead to inexact correspondence of particle tilt pairs, and a shift from the predicted and picked positions. This possibility is mathematically modeled in the Appendix A, where we characterize the error due to the use of an affine transformation. Moreover, a criterion is proposed for determining when two positions can be considered to represent the same particle. Our algorithm considers the displacement due to the tolerance parameter for identifying, i. e, two points are equivalent if their distance is less than the tolerance. The drawback associated with use of the affine transformation is thus explicitly overcome.

The algorithm is publicly available from *Xmipp* (http://xmipp.cnb.csic.es) and has been integrated into *Scipion 1.0* (http://scipion.cnb.csic.es). A tutorial for this kind of analysis is found at https://github.com/biocompwebs/scipion/wiki/tutorials/scipion_tutorial_initialvolume.pdf

Acknowledgements

The authors would like to thank JJ de la Morena and C Mark for editorial assistance. JLV receives an FPI fellowship (BIO2013-44647-R) from the Spanish Ministry of Economy and Competitiveness (MINECO). COSS is the recipient of a Ramón y Cajal fellowship. This work was supported by grants from the Comunidad de Madrid (S2010/BMD-2305 to JMC), the MINECO – Spain (AIC-A-2011-0638 to JMC) and the Fundación General CSIC (Programa ComFuturo to JV).

Appendix A. Compatibility of the affine transformation with curved samples

An affine transformation preserves the planarity and the ratios between distances. If all untilted or tilted points lie on the same plane, then their corresponding pairs will also do. This scenario is nonetheless idealized, as the sample can present certain curvature or roughness. Even if we consider a perfectly flat experimental sample, it has thickness and the particles inside might not be arranged on the same plane. As a result, the reliability of an affine transformation for establishing correspondences between tilt pairs of particles must be considered.

To facilitate calculation, let us assume an affine transformation from the untilted to tilted positions. Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be a reference system of orthonormal vectors; then, taking sample curvature or displacement from flat positions into account, the untilted particle will be in position $\mathbf{u} = (x_u, y_u, z_u(x_u, y_u))$, where x_u, and y_u define the position of the particle in the horizontal plane, and $z_u = z_u(x_u, y_u)$ is particle height or displacement from the ideal flat sample. As a result of the tilting process, the tilt position of the same particle will be given by

$$\mathbf{t} = R\mathbf{u} + \mathbf{s}. \tag{4}$$

In this case, $\mathbf{s} = (s_1, s_2, s_3)$ is the shift vector and R is an Euler matrix in the ZYZ system, with angles α, θ, γ,

$$R(\alpha,\theta,\gamma) = \begin{pmatrix} \cos(\alpha)\cos(\theta)\cos(\gamma) - \sin(\alpha)\sin(\gamma) & -\cos(\gamma)\sin(\alpha) - \cos(\alpha)\cos(\theta)\sin(\gamma) & \cos(\alpha)\sin(\theta) \\ \cos(\alpha)\sin(\gamma) + \cos(\theta)\cos(\gamma)\sin(\alpha) & \cos(\alpha)\cos(\gamma) - \cos(\theta)\sin(\alpha)\sin(\gamma) & \sin(\alpha)\sin(\theta) \\ -\cos(psi)\sin(\theta) & \sin(\theta)\sin(\gamma) & \cos(\theta) \end{pmatrix}. \tag{5}$$

Finally, an image is acquired, which can be understood as the projection of the particles on the detector. Considering that the detector surface is parallel to the plane defined by the vectors $\mathbf{e}_1, \mathbf{e}_2$, the projector $P = \mathbf{e}_1\mathbf{e}_1^T + \mathbf{e}_2\mathbf{e}_2^T$ will define the coordinate. The projected untilted particles will then have coordinates $\mathbf{u}' = P\mathbf{u} = (x_u, y_u)$, and the tilted coordinates will be

$$\begin{aligned} \mathbf{t}' = P\mathbf{t} = {} & x_u(\mathbf{e}_1^T R\mathbf{e}_1)\mathbf{e}_1 + y_u(\mathbf{e}_1^T R\mathbf{e}_2)\mathbf{e}_1 + z_u(x_y, y_u)(\mathbf{e}_1^T R\mathbf{e}_3)\mathbf{e}_1 \\ & + y_u(\mathbf{e}_2^T R\mathbf{e}_2)\mathbf{e}_2 + x_u(\mathbf{e}_2^T R\mathbf{e}_1)\mathbf{e}_2 + z_u(x_y, y_u)(\mathbf{e}_2^T R\mathbf{e}_3)\mathbf{e}_2 \\ & + (\mathbf{e}_1^T\mathbf{s})\mathbf{e}_1 + (\mathbf{e}_2^T\mathbf{s})\mathbf{e}_2. \end{aligned} \tag{6}$$

Note that the products $\mathbf{e}_i^T R\mathbf{e}_j$ are the matrix elements r_{ij} therefore Eq. (6) can be rewritten in matrix form as

$$\mathbf{t}' = R_0\mathbf{u}' + \xi + \mathbf{s}', \tag{7}$$

where

$$R_0 = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \qquad \xi = z_u(x_y, y_u)\begin{pmatrix} r_{13} \\ r_{23} \end{pmatrix} \qquad \mathbf{s}' = (\mathbf{e}_1^T\mathbf{s})\mathbf{e}_1 + (\mathbf{e}_2^T\mathbf{s})\mathbf{e}_2. \tag{8}$$

The dependence $\xi = \xi(z_u)$ should be noted. Eq. (7) does not represent an affine transformation, because the shift $\xi\mathbf{v} + \mathbf{s}'$ depends on the particle height, z_u. So the choice of an affine transformation should thus be revised if it will be used to assign tilt pairs.

To establish a comparison between the affine transformation proposed for assigning particle tilt pairs and the exact solution derived in this appendix, the distance η between the predicted (pr – subindex) and the exact (ex - subindex) tilt positions can be estimated as follows

$$\eta^2 = ||\mathbf{t}_{ex} - \mathbf{t}_{pr}||^2 = ||R_0\mathbf{u} + \xi + \mathbf{s}_{ex} - A\mathbf{u} - \mathbf{s}_{pr}||^2, \tag{9}$$

$$\begin{aligned}\eta^2 = {} & \mathbf{u}^T R_0^T R_0\mathbf{u} - \mathbf{u}^T R_0^T A\mathbf{u} + \mathbf{u}^T R_0^T \xi + \mathbf{u}^T R_0^T \Delta\mathbf{s} + \mathbf{u}^T A^T A\mathbf{u} \\ & - \mathbf{u}^T A^T R_0\mathbf{u} - \mathbf{u}^T A^T \xi - \mathbf{u}^T A^T \Delta\mathbf{u} + \xi^T\xi + \xi^T R_0\mathbf{u} - \xi^T A\mathbf{u} \\ & + \xi^T\Delta\mathbf{s} + \Delta\mathbf{s}^T\Delta\mathbf{s} + \Delta\mathbf{s}^T R_0\mathbf{u} - \Delta\mathbf{s}^T A\mathbf{u} + \Delta\mathbf{s}^T\xi,\end{aligned} \tag{10}$$

where $\Delta\mathbf{s} = \mathbf{s}_{ex} - \mathbf{s}_{pr}$. Assuming $\Delta\mathbf{s} = 0$,

$$\begin{aligned}\eta^2 = {} & \mathbf{u}^T R_0^T R_0\mathbf{u} - \mathbf{u}^T R_0^T A\mathbf{u} + \mathbf{u}^T R_0^T \xi + +\mathbf{u}^T A^T A\mathbf{u} - \mathbf{u}^T A^T R_0\mathbf{u} \\ & - \mathbf{u}^T A^T \xi + \xi^T\xi + \xi^T R_0\mathbf{u} - \xi^T A\mathbf{u}\end{aligned} \tag{11}$$

The analysis of this expression can be easily performed if the tilt axis is almost aligned with the vertical axis ($\alpha \approx 0 \approx \gamma$), then $\sin\alpha \sim \alpha, \sin\gamma \sim \gamma, \cos\alpha = \cos\gamma \sim 1$, and second orders and higher can be removed. Substituting this approximation into Eq. (8) results in

$$R_0 = \begin{pmatrix} \cos\theta - \alpha\psi & -(\alpha + \psi\cos\theta) \\ \psi + \alpha\cos\theta & 1 - \alpha\psi\cos\theta \end{pmatrix} \qquad \xi = z_u(x_y, y_u)\sin\theta\begin{pmatrix}1\\ \alpha\end{pmatrix}. \tag{12}$$

The eigenvalues of the R_0 matrix in Eq. (12) are approximately 1 and $\cos\theta$, respectively, which agree with the proposed eigenvalues (if the tilt axis is not aligned with the vertical axis, the condition on its eigenvalues is still valid, although in this case the mathematical analysis is not so straightforward). As a result, in the coarse step, the matrices R_0 and A could be equal (assuming that the coordinate tilt pairs are correctly identified).

The error between the coordinates predicted by the affine transformation and the true coordinate due to the curvature of the sample is given by

$$\eta^2 = \xi^T\xi = z_u^2(x_u, y_u)\sin^2\theta, \tag{13}$$

At this point, we propose the following criterion:

Given an affine transformation from an untilted set S_u of points to another set called predicted points S_{pr}, and the set of exact tilted S_{ex} positions. Two points, $t_{pr} \in S_{pr}$ and $t_{ex} \in S_{ex}$, can be considered identical if the distance $\eta < 2r$, where r is the particle radius.

This criterion can be summarized as the condition

$$\frac{z_u(x, u, y_u)}{r} < \frac{1}{2}. \tag{14}$$

To illustrate the applicability of this criterion, let us consider a micrograph centered at X_0, with a parabolic shape given by $Z = \kappa(X - X_0)^2$, as was done in Shatsky et al. (2014) with $\kappa = 1/10,000$. The tilt axis is assumed to be coincident to the line $X = X_0$ axis. For this example, a detector with $4,096 \times 4,096$ pixels, a tilt angle of 40°, and particles size $r = 50$ pixel can be used. The maximum height of the sample is reached at its extremes, $z_u^{max} = 419$. This height implies an angle of 5.8° from the floor to the top. Although particles at the sample extremes do not meet the criterion, a broad region of the micrograph do meet it, specifically 61% of the area of the micrograph.

References

Abrishami, V., Zaldívar-Peraza, A., de la Rosa-Trevín, J.M., Vargas, J., Otón, J., Marabini, R., Shkolnisky, Y., Carazo, J.M., Sorzano, C.O.S., 2013. A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. Bioinformatics 29 (19), 2460–2468.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X.W., Milne, J.L.S., Subramaniam, S., 2015. 2.2 åresolution cryo-em structure of β-galactosidase in complex with a cell-permeant inhibitor. Science 348 (6239), 1147–1151.

de Berg, M., Cheong, O., van Kreveld, M., Overmars, M., 2008. Computational Geometry: Algorithms and Applications. Springer-Verlag.

de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., 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-Trevín, J.M., Quintana, A., del Cano, L., Zaldívar-Peraza, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castells, J., Cuenca, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J.N.A.P.C.A.M.K.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 (1), 93–99.

Guckenberger, R., 1982. Determination of a common origin in the micrographs of tilt series in three-dimensional electron microscopy. Ultramicroscopy 9 (1–2), 167–173.

Hauer, F., Gerle, C., Kirves, J.M., Stark, H., 2013. Automated correlation of single particle tilt pairs for random conical tilt and orthogonal tilt reconstructions. J. Struct. Biol. 181 (2), 149–154.

Hoang, T.V., Cavin, X., Schultz, P., Ritchie, D.W., Gempicker: a highly parallel gpu-accelerated particle picking tool for cryo-electron microscopy, BMC Struct. Biol. 13.

Leschnizer, A., Nogales, E., 2006. The orthogonal tilt reconstruction method: an approach to generating single-class columes with no missing cone for ab initio reconstruction of asymmetric particles. J. Struct. Biol. 153 (3), 284–299.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M.I., Pragani, R., Boxer, M.B., Earl, L.A., Milne, J.L.S., Subramaniam, S., 2016. Breaking cryo-em resolution barriers to facilitate drug discovery. Cell 165 (7), 1698–1707.

Mulchrone, K.F., 2003. Application of delaunay triangulation to the nearest neighbour method of strain analysis. J. Struct. Geol. 25 (5), 689–702.

Nuñez Ramirez, R., Klinge, S., Sauguet, L., Melero, R., Recuero-Checa, M.A., Kilkenny, M., Perera, R.L., Garcia-Alvarez, B., Hall, R.J., Nogales, E., Pellegrini, L., Llorca, O., 2011. Flexible tethering of primase and DNA Pol α in the eukaryotic primosome. Nucleic Acids Res. 39 (18), 8187–8199.

Radermacher, M., Wagenknecht, T., Verschoor, A., Frank, J., 1986. A new 3-d reconstruction scheme applied to the 50 s ribosomal subunit of E-coli. J. Microsc. 141 (1). RP1–RP2.

Shatsky, M., Arbelaez, P., Han, B.G., Typke, D., Brenner, S.E., 2014. Automated particle correspondence and accurate tilt-axis detection in tilted-image pairs. J. Struct. Biol. 187 (1), 66–75.

Singh, R.K., Tropsha, A., Vaisman, I.I., 1996. Delaunay tesselation of proteins: Four body nearest-neighbor propensities of amino acied residues. J. Comput. Biol. 3 (2), 213–221.

Sorzano, C.O.S., Alcorlo, M., de la Rosa-Trevín, J.M., Melero, R., Foche, I., Zaldívar-Peraza, A., del Cano, L., Vargas, J., Abrishami, V., Otón, J., Marabini, R., Carazo, J. M., 2015. Cryo-em and the elucidation of new macromolecular structures: random conical tilt revisited. Sci. Rep. 5, 14290.

Tang, G., Peng, L., Baldwin, P.R., Mann, D., Jiang, W., Rees, I., Ludtke, S.J., 2007. Eman2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157 (1), 38–468.

Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., Carragher, B., 2009. Dog picker and tiltpicker: software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. 166 (2), 205–213.