Chapter 7
Alignment of Tilt Series

A. Verguet, C. Messaoudi, C. O. S. Sorzano and S. Marco

Abstract Computing of three-dimensional reconstructions from images obtained by transmission electron tomography needs three main steps: data acquisition, projection alignment, and 3D reconstruction. In this chapter we will focus on the process of alignment moving from the justification of its need to the study of the different classical approaches (cross-correlation, use of added fiducial markers) that have been commonly used in this alignment process. We will also discuss the most recent algorithms (multiscale registration, invariant feature recognition) as they have been adapted to Electron Tomography and improved to increase the accuracy and resolution of the final tomograms.

7.1 Why Do We Need to Align Tilt Series?

We may consider that tomography consists in determining a numerical representation in three dimensions of an object from their projections which requires the mathematical combination of the acquired projections by a process called reconstruction. In the case of transmission electron tomography (TET), these projections

A. Verguet
INSERM U1196, Centre Universitaire, Bat. 112, 91405 Orsay Cedex, France
e-mail: amandine.verguet@ssji.net

C. Messaoudi
CNRS UMR9187, Centre Universitaire, Bat. 112, 91405 Orsay Cedex, France
e-mail: cedric.messaoudi@curie.fr

C. O. S. Sorzano
Biocomputing Unit, National Center of Biotechnology (CSIC). C/Darwin,
3, Campus Universidad Autonoma (Cantoblanco), 28049 Madrid, Spain
e-mail: coss@cnb.csic.es

S. Marco (✉)
Institut Curie, Centre de Recherche, Centre Universitaire,
Bat. 112, 91405 Orsay Cedex, France
e-mail: sergio.marco@inserm.fr

are generated by tilting the object to be reconstructed around an axis. Therefore, the reconstruction process requires an accurate determination of the orientation of this axis, which is common to all projections, to combine them in a correct way. In addition, mathematical combination of projections needs that each one of the values on the 3D numerical representation of the reconstructed object has been estimated from data representing the same object at different tilt angles used during the acquisition process. In case that the orientation of the tilt axis is not precisely defined for each projection or that the values combined to compute the reconstruction do not correspond to an equivalent position in the original object, the mixed information will result into inaccurate results and artifacts. Some classical artifacts are blurred borders with small shifts or object deformation in "banana" or "star" shapes (see Fig. 7.1). A detailed explanation of geometrical artifacts can be found in [1].

In practice, as during the acquisition of the projection images the specimen is placed in a holder, which is physically tilted by a goniometer inside the imaging system, the precision and stability of motors to keep sample at the same exact position during acquisition is unattainable at the nanometer scale expected for TET. To compensate the lack of perfect tilting, automated acquisition tracks position and focus of sample. However, despite the high performance achieved, this compensation is still not enough and gives rise to shifts and rotations. In addition, the sample can be damaged during the acquisition process by suffering shrinkage or simply the compensation process can slightly modify the position of the focal plane.
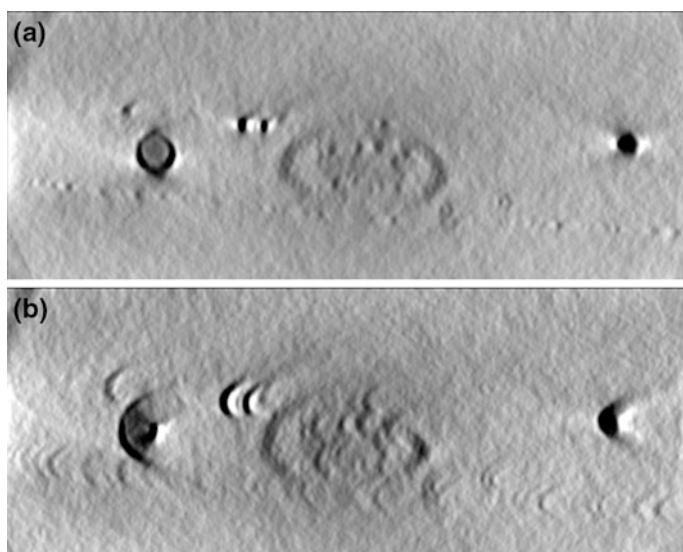

Fig. 7.1 Reconstructions of *Pyrodictium abyssi*. **a** XZ from a reconstruction with correct alignment of tilt series images. **b** XZ from a reconstruction with incorrect alignment of tilt series images. It can be seen that spherical shape are elongated to the right in a shape similar to banana

This can lead to magnification changes or deformations in the recorded projections. The result is that the geometric relationships between the object and the obtained projections are not precisely known initially. Thus, a step of image alignment of these projection images is mandatory prior to compute accurate reconstructions. In this chapter, we will discuss different approaches to perform image alignment used with the aim of correcting most of the deviations from an ideal projection geometry. This means to describe methods to determine the geometrical relationships existing between the different projections of a tomographic tilt series.

Different approaches have been used to perform image alignment. The most classical one, frequently used to correct shifts between projections, is based on the maximisation of the cross-correlation existing between images [2–4]. However, this approach does not consider the existence of magnification changes or deformations cross-correlation, so that it is not good enough to align projections when shrinkage or change of focus occurs during the acquisition process which is not unusual in biological samples which are sensible to damages induced by the electron beam. A second way to perform image alignment, frequently used in biological samples, is to add fiducial gold colloidal markers to the sample. As gold beads can be localized very accurately, due to their spherical shape and high contrast, the alignments base on these markers are very accurate. In addition, if a large number of beads is used, the errors in their localization are averaged [5]. The marker-based methods have another advantage as it will generate a 3D model of their positions which guarantees a consistent alignment among the images from the full range of tilt angles. This 3D model can also be adapted to correct deformations induced in the sample during imaging. However, it is not always possible to use fiducial gold colloidal markers as they can interfere with reconstruction process (streak artifact for example [6]). Moreover, even if markers are added during sample preparation before observation, sometimes they are not uniformly distributed, being absent (or numbering not enough) in the region of interest for the reconstruction. To deal with this problem, approaches based on feature recognition have been developed. Thus, instead of adding external markers to the sample characteristic features are automatically extracted and tracked along the projection images of tilt-series before building 3D models prior to determine alignment parameters ().

7.2 Standard Alignment Process

As aforementioned, the acquisition of tilt series under the transmission electron microscope suffers from the goniometer and sample instabilities which main effects leads to shifts and in-plane rotations. To evaluate the accuracy of different alignment processes and algorithms it is frequent to use numerical phantoms which values are perfectly determined and which precisely simulate the different shifts, rotations or deformations to be corrected. Thus, it is possible to compare the parameters determined by algorithms or process to the simulated values included in
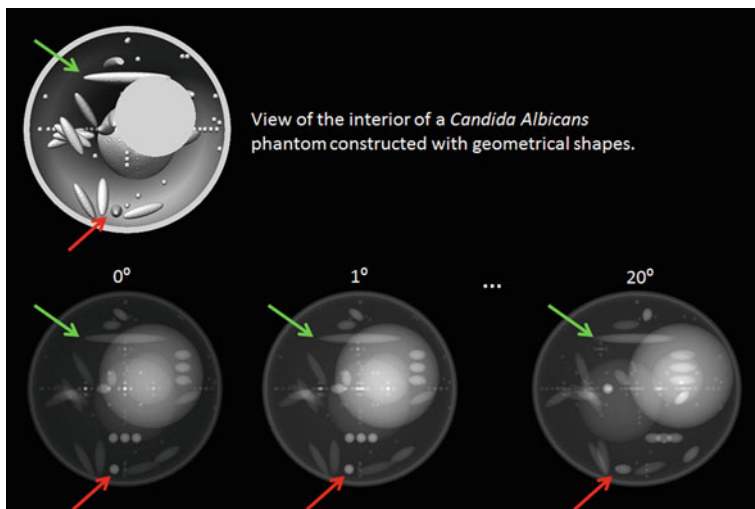
Fig. 7.2 Phantom of a *Candida albicans* cell. In the top row, the cell has been virtually cut to show its interior. In the bottom row, projections at 0°, 1° and 20° are shown. Typically, a tilt series go from −60° to 60°

the phantom. Therefore, we show in Fig. 7.2 the phantom of a *Candida albicans* cell and 3 of its projections at different angles (0°, 1°, and 20°) that we will use for to illustrate how alignment approaches can identify the shift and rotation between two projection images and correct them as if they would have been acquired using a perfect goniometer.

As previously discussed, we can distinguish to main families of methods to align projections in a tomographic tilt series. The first family, which historically was the first used, considers all the information present in the projections, while the second family focus on a few confident points (points that can be precisely identified along the tilt series, see the arrows in Fig. 7.2). Since the first family of methods are frequently less accurate than those from the second family, they are usually used to precenter projections. This is done prior to the refinement of the alignment and to the determination of the orientation of the tilt axis, which is commonly performed by using methods based on 3D models of the positions of characteristic features.

7.2.1   Precentering the Tilt Series

Maybe, one of the first attempts to correct for the shifts was performed in 1982 by [4]. The idea was to exploit the similarity between projections at different tilts. Actually, if the difference between tilt angles is not large, the difference between the images is rather small (see tilt projections at 0° and 1° in Fig. 7.2) [4] performed a very detailed analysis of the modifications of the correlation function needed to

account for the fact that the two images being compared are coming from the same 3D object and are related by a single tilt operation. However, let us give one step back and introduce the correlation as a measure of similarity between two images.

Let us assume that we have two identical images whose relationship between both of them is a simple shift (see Figs. 7.1, 7.2 and 7.3). The cross-correlation function (called auto-correlation when only one image is used) between Figs. 7.1 and 7.2 may be defined as:

$$R_{1,2}(\Delta x, \Delta y) = \sum_{x,y} I_1(x - \Delta x, y - \Delta y) I_2(x, y)$$

This function is maximum when the two images maximally overlap. In this way, we may identify the shift $\mathbf{\Delta}_{1,2} = (\Delta x, \Delta y)$ required to go from Fig. 7.1 to Fig. 7.2, or viceversa $\mathbf{\Delta}_{2,1} = -\mathbf{\Delta}_{1,2}$ (see Fig. 7.3). An interesting formula of this similarity estimator is that the location of the maximum is insensitive to linear transformations of the image intensities. Fortunately, this function can be calculated very quickly due to a property of the Fourier transform

$$R_{1,2}(\Delta x, \Delta y) = FT^{-1}\{FT\{I_1\}(FT\{I_2\})\}$$

Therefore, all we have to do to determine the position that maximize the overlap between two images is to transform them to Fourier space, multiply the Fourier transform of one image by the complex conjugate of the other, and come back to real space. The simplicity of this operation has made that most packages to perform three-dimensional reconstructions from transmission electron microscopy
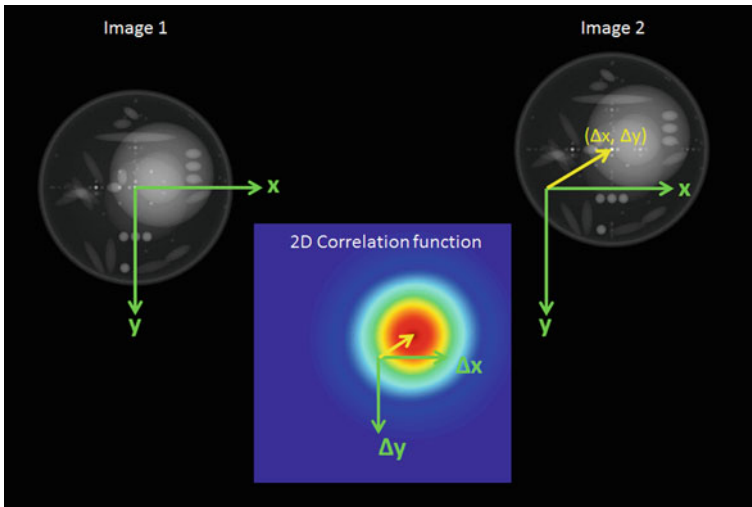

Fig. 7.3 Two identical images related by a shift, note that Image 2 is a shifted version [shifted by a vector displacement $(\Delta x, \Delta y)$] of Image 1. The correlation function is maximum at the shifts required to go from Image 1 to Image 2

projections, include a tilt series alignment by cross correlation, as just described in this chapter and not in its modified version proposed by [4]. When this approach is applied to an experimental TET tilt series, the cross-correlation between two consecutive images, I_i and I_{i+1} is computed and the corresponding displacements $(\mathbf{\Delta}_{i,i+1})$ identified. This estimated displacement is a good estimation of the real shifts since in a tilt series, when the increment on the tilt angle is low, two consecutive images are almost the same (except for a shift, as illustrated in Fig. 7.2). Once the consecutive shifts are identified, we may find the relative shift between any two images i and j. For that, all images are translated to be centered with respect to the 0-tilt image. Let us illustrate this by representing a tilt series from $-\theta_{max}$ to θ_{max} degrees with increments of $\Delta\theta$ degrees (e.g., from $-60°$ to $60°$ in steps of $1°$), with indexes from $-i_{max}$ to i_{max} $(i_{max} = \theta_{max}/\Delta\theta)$. The shift needed to align an image with negative tilt to the 0-tilt image is just the accumulation from $-i$ to 0 of all the consecutive shifts:

$$\mathbf{\Delta}_{-i,0} = \mathbf{\Delta}_{-i,-i+1} + \mathbf{\Delta}_{-i+1,-i+2} + \cdots + \mathbf{\Delta}_{-1,0} = \sum_{j=-1}^{-1} \mathbf{\Delta}_{-j,-j+1}$$

On the other side, for positive tilts, we need to accumulate the shifts in reverse order

$$\mathbf{\Delta}_{i,0} = -\mathbf{\Delta}_{i-1,i} - \mathbf{\Delta}_{i-2,i-1} - \cdots - \mathbf{\Delta}_{0,1} = -\sum_{j=0}^{i-1} \mathbf{\Delta}_{j,j+1}$$

With these translations we may produce a new set of centered images $\tilde{I}_i$. Unfortunately, reality is not that easy for several reasons:

- Experimental images are extremely noisy, resulting in a noisy correlation function whose maximum may be spuriously misplaced (see Fig. 7.4).
- If one of the images is in-plane rotated with respect to the other, the correlation function is distorted with respect to the unrotated correlation function. This distortion may produce a totally incorrect estimation of the displacement vector (see Fig. 7.5).
- Local differences in the illumination conditions or the presence of a persistent illumination pattern totally distorts the correlation pattern (see Fig. 7.6).

Acknowledging these difficulties, we may try to robustly estimate the shifts between any two images. First, we can bandpass filter the images to remove any persistent illumination pattern, smooth local illumination variations (low frequency) as well as noise and small image details (high frequency) unnecessary to globally align two images. Once the images are bandpass filtered, they can be safely down-sampled, to reduce their size and speed-up calculations. Then, we may construct a polar 2D correlation function (the correlation function when the images are expressed in polar form). The location of the maximum in this polar correlation map indicates the optimal rotation [7]. In this way, we can identify both rotations and translations by alternating between looking for the best shift, then for the best rotation, and iterating several times this sequence till convergence.
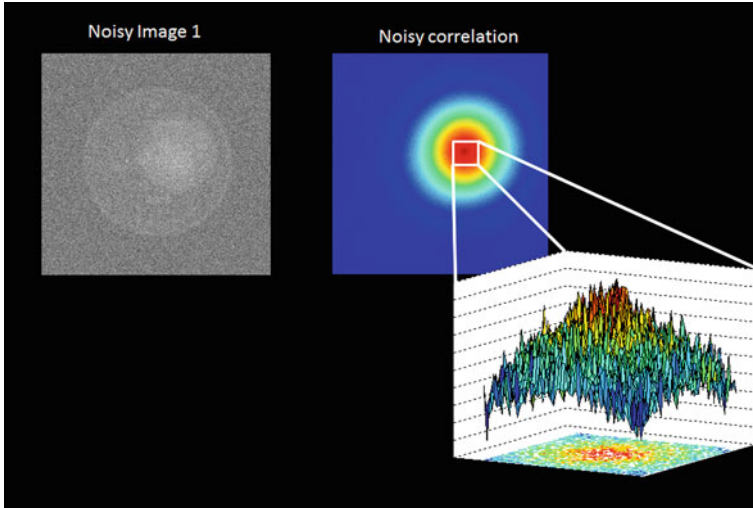
Fig. 7.4 Noisy Image 1, the corresponding correlation function with the Noisy Image 2, and a zoom of the summit of the correlation function
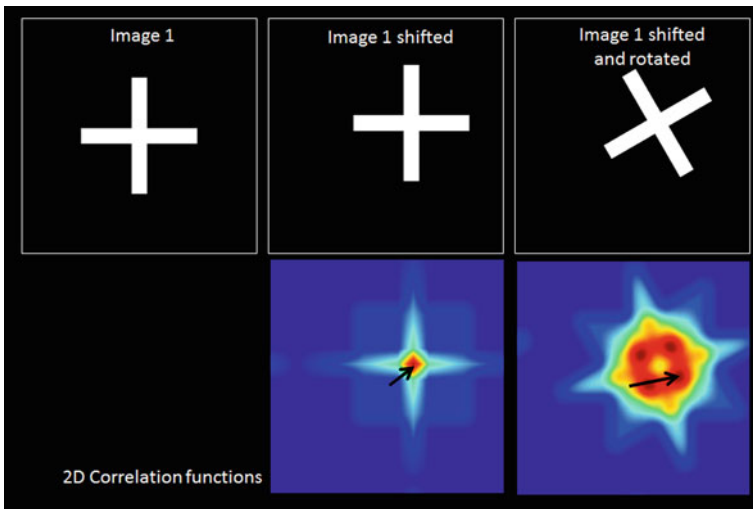

Fig. 7.5 Reference image (left), shifted image (middle) and shifted and rotated image (right), and their corresponding correlation functions. In the only shift case, the maximum of the correlation function correctly identifies the displacement vector, but this is not the case if there is also a rotation. Even small rotations of 2°–5° may totally break the structure of the correlation function depending on the specific object being imaged
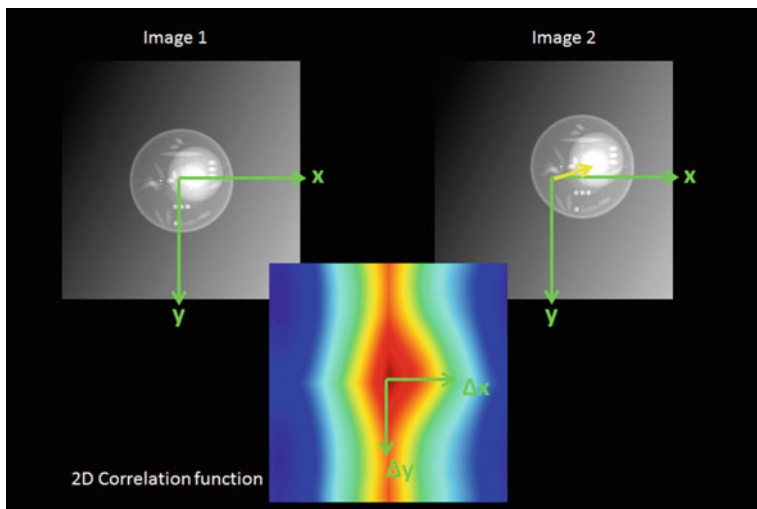
Fig. 7.6 The presence of a non-uniform illumination background, totally distorts the correlation structure

Correlations between images separated by small tilts (1°–3°) are still good approximations to the hypothesis that the images being compared are the "same" simply related by an in-plane operation.

Combining all these strategies, we may safely calculate vector displacements between any two images i and j, $\mathbf{\Delta}_{i,j}$ ($i < j$ and $|j - i|\Delta\theta < \varepsilon$, for a user selected ε). Then, we can solve the overdetermined system of equations

$$\mathbf{\Delta}_{i,j} = \sum_{k=i}^{j-1} \mathbf{\Delta}_{k,k+1}$$

as was done for Direct Detector frame alignment [8].

7.2.2 Determining the Orientation of the Tilt Axis

Once projection images on a tilt series are prealigned, they share a common orientation of the tilt axis. Thus the orientation of the tilt axis along the tilt series should not change and each projection in the tilt series should correspond to a different projection of the specimen at a given tilt angle. Therefore it is now possible to estimate the direction of this tilt axis used to generate the projections (see Fig. 7.7). Once identified, it is usually aligned with the vertical axis. Although this is not absolutely required to perform 3D reconstruction, it is convenient since it converts the 3D reconstruction problem into a 2D reconstruction problem conveying a speed-up of the reconstruction process by more than 1 order of magnitude.

Fig. 7.7 Tilt axis shown on a projection of *Pyrodictium abyssi* taken at the Electron Microscope. Ideally, the visualization of the whole tilt series should show a smooth transition from one image to the next in which the location and orientation of the tilt axis is fixed
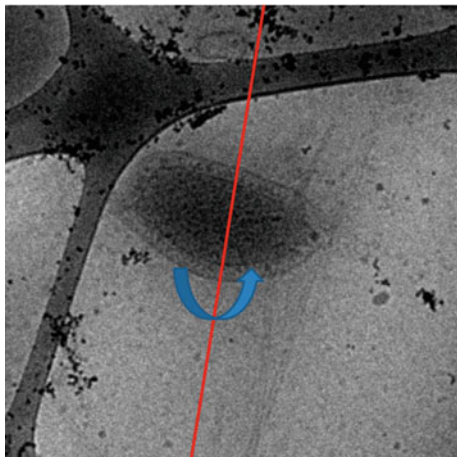
The reason is that each slice can be reconstructed independently using the pixel values observed in each one of the rows of the aligned images.

Nevertheless, the accurate estimation of the direction of the tilt axis is not always possible from prealigned projections for different reasons: (1) the goniometer instabilities can induce some in-plane rotations which make the orientation of the tilt axis to be slightly different in each image; (2) our precentering of the images (previous section) is normally far from a perfect algorithm and although major movements have been corrected, there is still a non-negligible amount of shift between successive images; (3) more subtle, small errors during the centering of images may accumulate along the tilt series, so that there is an important accumulated drift from the beginning of the tilt series to its end. This latter error is the responsible of the "banana" shapes observed at the 3D reconstruction of misaligned tilt series.

Therefore, the accurate determination of the tilt axis direction in each one of the images is the most delicate process in the alignment process. For doing so, we need to identify corresponding points along the tilt series (see Fig. 7.8). These corresponding points in the projections are called 2D landmarks and they can be identified by different approaches:

- Manual, where markers are selected in each image of the tilt-series. The interest is the manual validation of corresponding points. However, the location of points are imprecise (a few pixels precision) and the process is not reproducible and time-consuming.
- Semi-automatic, where landmarks are manually selected on one image and the corresponding location on the other images is performed automatically. The interest is the manual selection of interesting points to track with an improvement of reproducibility and time consumption.
- Fully automatic, where the 2D landmarks are identified as any outstanding image feature like local maxima or minima [9] (Chap. 6), Harris corners [10], or any other feature detection algorithm could be employed (Fig. 7.8 show local
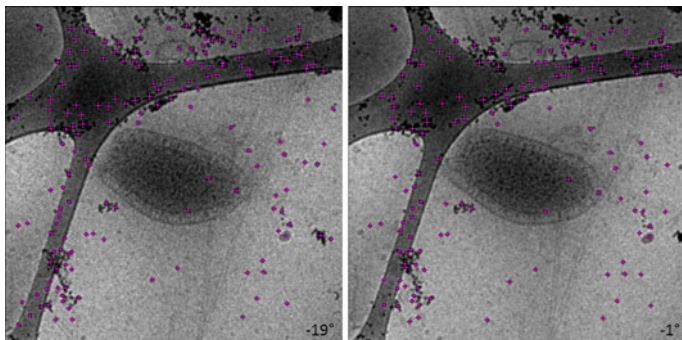
Fig. 7.8 Corresponding points between two projections at different tilt angles

minima features detected by TomoJ). Even small pieces of the image have been used once an initial alignment is manually provided by the user [11].

In many experiments, gold beads (seen as black dots in the images due to the high electron density of gold with respect to the biological material) are added to the sample. If case that they do not aggregate and that their distribution is good enough and uniform over the region of interest, they can be used as fiducial markers that can be easily tracked along the tilt series and so improve quality of alignment. However, adding gold beads is not strictly required for the alignment process because, even if gold beads are available, points that can be safely tracked along the tilt series can be frequently identified in the dataset. For instance, some of the points shown in Fig. 7.8 correspond to gold beads, but some others do not.

Once 2D landmarks (usually between 15 and 100) are identified on each image, the next step is to establish the correspondences between points in different images to obtain 2D landmark chains, i.e., tracking the same landmark along the tilt series (see Fig. 7.9). A landmark chain is allowed to have gaps (this means that the algorithm cannot find the landmark in a given projection, but it does in the projections before that image and after that image). The maximum gap length is normally selected by the user and frequently it is a number between 1 and 5 images; the larger the gap length, the higher the probability of tracking the wrong point. Each landmark chain is supposed to be related to a single (unknown) 3D landmark. This 3D landmark, when projected onto the different collected images, is located at different places. The coordinate of their projections depend as

$$\mathbf{p}_{ij} = A_{i,tiltAxis}\mathbf{r}_j + \mathbf{s}_i \qquad (7.1)$$

That is, the coordinate of the j-th 3D landmark in image i, $\mathbf{p}_{ij} \in \mathbb{R}^2$, can be calculated using a matrix that depends on the orientation of the tilt axis in the i-th image and its tilt angle, $A_{i,tiltAxis}$, the 3D location of the landmark, $\mathbf{r}_j \in \mathbb{R}^3$, and a shift of the projection, $\mathbf{s}_i \in \mathbb{R}^2$ (this shift should be ideally zero if the precentering was perfectly performed). This projection model is at the basis of a number of
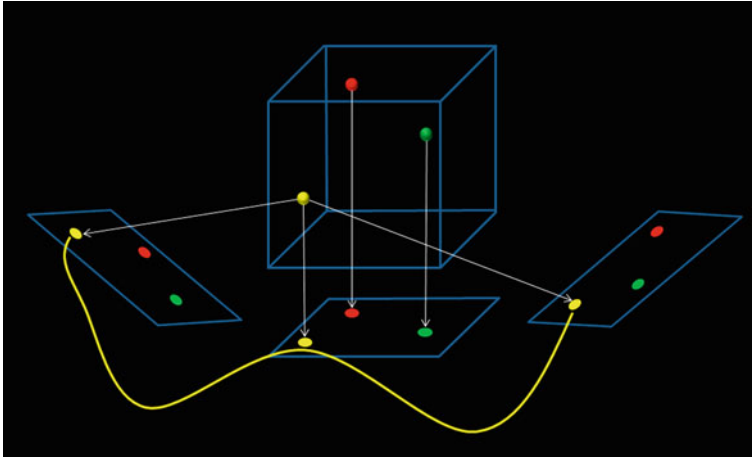
Fig. 7.9 The 2D landmarks observed in each of the projections correspond to the projection of a 3D landmark (whose exact location has to be estimated). A landmark chain (like the yellow one highlighted in the figure) is formed by the set of 2D locations of the same 3D landmark projected at different tilt angles

works [10, 12–17]. A much more detailed description of the projection mathematics can be found at [18].

After having determined the landmark chains, it is possible to precisely compute the orientation of the tilt axis by finding the matrices $A_{i,tiltAxis}$ and residual shifts $\mathbf{s}_i$. If there are $N_{projections}$ in the tilt series, we need to estimate $N_{projections}$ angles (the small rotations suffered by each projection), $2N_{projections}$ shift parameters (x, y shifts per image), and 2 parameters for the orientation of the tilt axis with respect to the electron beam. This makes a total of $3N_{projections} + 2$ parameters. However, the process requires the estimation of the auxiliary variables $\mathbf{r}_j$. Thus, if we have an average of N_{chains} landmark chains per image, each with a (x, y, z) location, there would be a total of $3N_{chains}N_{projections}$ extra parameters to estimate. Each point in the landmark chain brings 2 equations for an overdetermined equation system, and let us assume that on average the length of a chain is L_{chain}. Then, there are $2N_{chains}N_{projections}L_{chain}$ equations and $3N_{projections}(N_{chains} + 1) + 2$ unknowns. The ratio between the number of equations and the number of unknowns is approximately $2/3L_{chain}$. A typical value of L_{chain} is between 10 and 50, meaning that this is a highly overdetermined equation system. Therefore, this equation system can be easily solved by Least Squares [10, 12, 16], or by Least Squares combined with some statistically robust technique [14, 18].

An important problem of the construction of the landmark chains is the problem of landmark occlusion (illustrated in Fig. 7.10). When the projection of two landmarks overlaps, algorithms have difficulties in deciding which 2D landmarks go with which, and landmark chains are sometimes misconstrued mixing projections from several 3D landmarks. A robust resolution of the equation system at (7.1) tends to mitigate this effect.

Fig. 7.10 The projections of two 3D landmarks (green and yellow in the figure) may come close to each other as the tilt is increased. Eventually they overlap, and as the tilt keeps increasing they separate again. If instead of two 3D landmarks, the 3D landmarks tend to aggregate (as shown in the experimental image), then the problem is much more aggravated
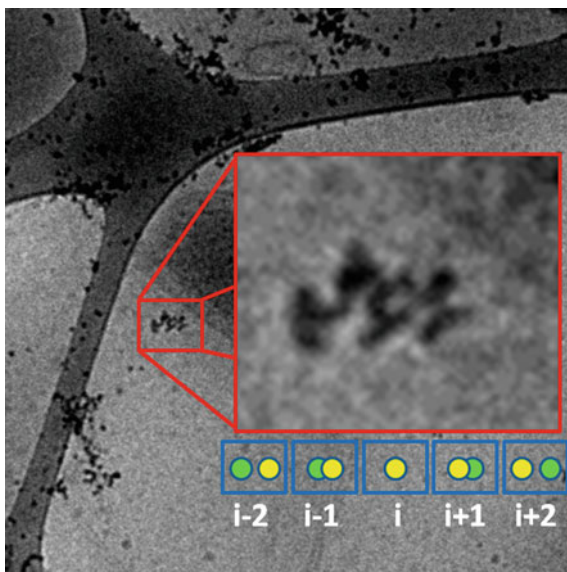
Another source of problems during the alignment is the uneven distribution of landmarks along the tilt series. Normally at high tilts it is more difficult to distinguish 2D landmarks due to the thicker sample and resulting low contrast. Additionally, as the tilt increases, there are more chances of overlapping projections of different 3D landmarks resulting in a lack of reliable landmark chains covering high tilt angles. However, in spite of these difficulties, projection alignment and determination of the tilt-axis orientation is frequently feasible, but in some cases some advanced alignment methods are required.

7.3 Advanced Alignment Methods

In this section, we will explain and evaluate the interest of some improvements for image registration between two images suitable for the alignment of tilt series in electron tomography. Some of these improvements use global image information by applying a multiscale scheme associated to change in similarity measure. Others are based on registration of unique descriptors for points of interest.

7.3.1 Multiscale Registration

Most recent advances in image registration are based on multiscale scheme associated to gradient descent algorithms to determine the transformation between two

images. The main principle of multiscale scheme is to separate information in images at different scale. Coarse scales would then consist to main shapes and general features of the images while fine scales consist to details and noise. The decomposition allows a coarse-to-fine registration in which an initial transformation is found for the coarsest scales of the images and then it is refined using finer scales images. The use of this technique has many advantages:

- large transformations between two images can be found on coarser scales of images
- the signal-to-noise-ratio increases with coarser scales thus making it easier the finding of alignment
- as images at coarse level have less possible local minima, the robustness is increased
- part of the computation is done on smaller data and it need less time than computing everything on full size data

The multiscale decomposition can be done in several different ways. The simpler one is Gaussian pyramid where images size are reduced after applying a Gaussian filter in a pyramidal manner (Fig. 7.11). But other possibilities have also been proposed such as Laplacian pyramids [19], wavelets or BV, L^2 decomposition [20].

The multiscale approach allows the use of different models for transformation: translation only, rigid transformation (translation and rotation), affine transformation (global deformations where straight lines stays straight) or non-rigid transformation (combination of global and local deformation). With translation only or rigid transformation the gain over classical cross-correlation is not so clear. The main interest is the increased robustness and by extension, a lower error propagation at the cost of higher computation time compared to Fourier calculations. However, translation, as aforementioned is not enough to correct most of deviation from the ideal projection geometry. The addition of deformation in the transformation could seem a good idea to correct deformations in projection geometry
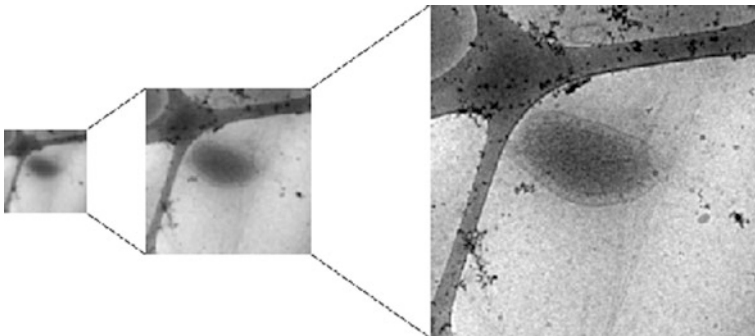

Fig. 7.11 The gaussian pyramid is formed from original image (on the right), filtered with Gaussian filter and reduced in size to form coarser scale of image. The coarser scale is used first for correction of large transforms and then finer scale is used to further improve transform

(change in magnification or sample shrinkage for example). The problem arising is that images are easily aligned 2 by 2 without taking into account global projection geometry but then some deformations in images are due to change of tilt angle. It is then necessary to add some prior information such as tilt axis position and tilt angle of images to prevent some wrong correction in projection geometry. When information about tilt axis position and tilt angle is not known a priori, the possibility of deformation is unusable.

Once the type of transformation is fixed, the multiscale approach also allows different optimization techniques to find the best transformation. It can be brute force testing, where an exhaustive but limited set of possibilities is taken at each scale. It helps to stay around the local minima found at coarsest scale. Better optimization scheme involves the use of gradient descent algorithms, often powell or convex. The interests of these approaches were well debated in the medical imaging field [21–24] to register data in 2D and 3D from computer axial tomography (CAT), positron emission tomography (PET), magnetic resonance imaging (MRI) or ultrasound.

To estimate the better result of alignment, there is need of a metric. Many algorithms use mean (or sum) of square difference or correlation (see part 7.1 for more concise description), but other metrics exists such as mutual information. Mutual information was designed to allow the registration of images with very different contrasts [25, 26]. Formally, the mutual information of two discrete random variables X and Y can be defined as:

$$I(X; Y) = \sum_{x,y} p(x, y) log \frac{p(x, y)}{p(x)p(y)}$$

Where $p(x, y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. In the case of 3D chemical mapping by EFTEM, mutual information was designed to align energy filtered images together between them or with zero-loss or plasmon images which represents the ultra-structural information. So, mutual information is needed to obtain better alignment between chemical and structural images (two modalities). Even if some studies demonstrated higher robustness and lesser sensitivity to noise [23], the higher computation cost over cross-correlation makes the use of mutual information approach questionable to align images coming from a single modality (ultra-structural tilt series from standard tomographic approach for example).

Multiscale registration allows to have sometimes a better alignment than cross-correlation methods but it does not compensate the fact that it is not sufficient as it will still align image two by two and so some propagation of errors will still occurs even if smaller. There is also need to take into account the fact that the information inside the image is a projection from a 3D sample and so the improvement is not competitive enough compared to feature-based registration. Therefore, multiscale registration is suitable for prealignment but it will frequently require further refinement based on landmarks.

7.3.2 Invariant Feature Recognition

A way to further improve alignment of tilt series is to optimize the detection of corresponding points in images of the tilt series. For this purpose, new approaches in object detection or panoramic stitching, based on the concept of Scale Invariant Feature Transform (SIFT), introduced by Lowe [27], have been proposed for alignment in electron tomography [28].

Feature-based alignment methods require several steps which are performed by separate algorithms. Here we will describe these steps which correspond to the common strategy used for image alignment using SIFT. Additionally to the emergence of this method many other algorithms appeared, providing alternatives for the different steps.

- Detection of points of interest

First step is the detection of points of interest. To this purpose the algorithm has to choose points in the image which will be easy to locate in other images of the same sample. These points of interest are determined where the algorithm is able to recognize a feature which is based on mathematical properties, such as local extrema, and may not match physical features of the sample. This feature detection step is performed independently on all images of a series and leads to the creation of unique descriptors for each point of interest.

- Creation of a unique descriptor for each point of interest

With a set of point of interest on each image, alignment requires that they correspond to the same physical location on the feature to be matched. One way to achieve that is to rely on distinctive features. Most algorithms attach a single descriptor to each interest point for that purpose.

- Comparison of interest points between two or more images

The next step is to identify the matches in the sets of points of interest by comparing their descriptors. Once the point of interest, associated to the same feature, are grouped together they can be used to deduce the space transformation which occurred during the tilt-series acquisition.

7.3.2.1 Methods to Detect the Points of Interest

Different mathematical methods can be used to perform the detection step which yields various shape of interest on the image: point, curve or area. The methods proposed in detection algorithms are essentially based on contours detection, i.e. the detection of variation of intensity levels on the area near the location of interest. In the case of the SIFT detection algorithm, each point of interest is identified by its location on the image (coordinates x, y), its gradient orientation, the scale factor and
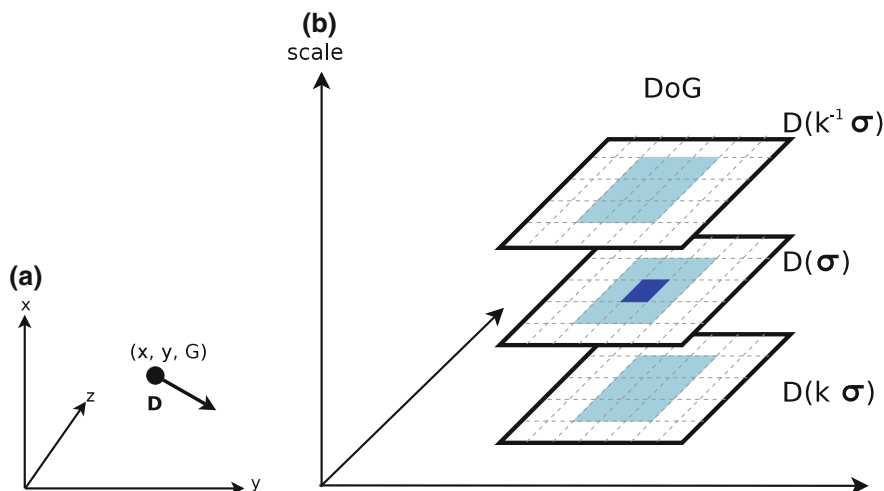
Fig. 7.12 a An interest point is composed by its location, the gradient information and an unique descriptor noted D. **b** Diagram showing an example of the extremum detection during the computing of the difference of Gaussian for one point. The point of interest (dark blue) is evaluated from its neighborhood (light blue) between 3 different scales

finally a descriptor noted as D. The Fig. 7.12a is an example of the composition of an interest point.

In order to determine the location of the points of interest, the initial step is their detection at various image scales. For fulfilling this task, first the original image is reduced several times to obtain different scale levels, called octaves. For each image scale, a Gaussian blur is applied several times.

Second, in order to detect the local extrema, the computation of the Gaussian gradients is required at the same scale (Fig. 7.12b). Differences of Gaussians (DoG) are used on adjacent Gaussian in an octave. The different blur and octave can be represented as a pyramid of images as shown on Fig. 7.13.

Then, all local extrema are searched on all DoG at all scales. This defines the location of the interest points. It is stored along with the scale and convolution level that led to its detection. Afterwards, to obtain a position with a sub-pixel resolution it is possible to interpolate using a Taylor model. Moreover, the points with a low contrast or on an edge without curvature are eliminated.

Additionally to the location and scale, an orientation needs to be computed for each point of interest. This is obtained by filtering the gradients of multiple areas near the interest point. Once multiple gradients have been computed for some points of the neighborhood, an histogram is created. This histogram categorizes the orientations in a fixed number of classes which are weighted by the amplitude of gradients. At the end, the retained orientation of the point of interest is defined by the major orientation in the histogram (Fig. 7.14b).
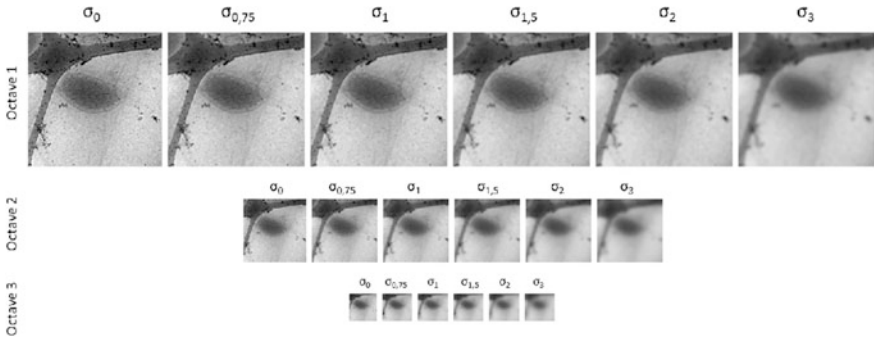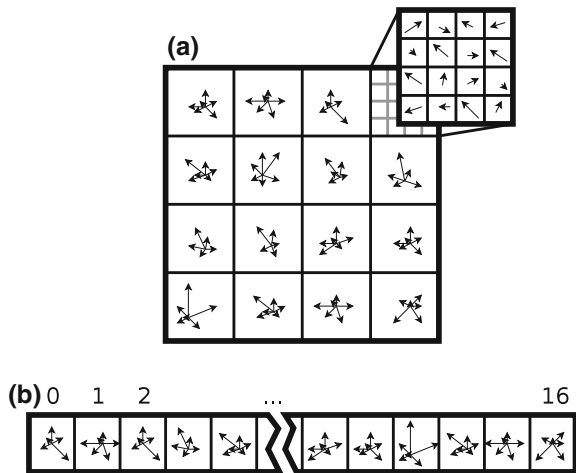
Fig. 7.13 This pyramid is obtained with a projection of *Pyrodictium abyssi* at several gaussian blurs and scales also called octave

Fig. 7.14 a This step is the creation of the future descriptor, with the computing of each histogram of orientation in the neighborhood. The histogram is shown as the sum of vectors. **b** Each histogram composes the final descriptor

There are several other alternate algorithms available for detection. Some are similar to the SIFT detection as they rely on the pyramid concept but using a differing convolution kernel. Here is a list of notable detection algorithms:

- Most blob detection methods are based on local differences in an image, as is the case for the search of local extrema. Several properties of the region of interest, like contrast and pixel intensity, are compared against surrounding regions. The difference of Gaussian presented previously is one of these methods. However, it is possible to use the Laplacian of Gaussian (LoG) or the difference of Hessian (DoH) instead.
- Some algorithms, like Harris corner and FAST (Features from Accelerated Segment Test [29]), are based on corners detection. They work by detecting corners which are the intersection of two edges on an image. This method is based on finding small patches in the image which are not similar to neighbor

patches. When a patch is located on a low contrast area, all close patches of a given size are similar. When the patch is located on a line or edge, close patches along the edge are similar too. This similarity is measured with the sum of squared differences of multiple patches being compared.

- Affine variants of the previous methods can be used to identify similar regions between images with scale change, rotation and shearing. Harris affine and Hessian affine are the most common.

7.3.2.2   Construction of the Descriptor

Once interest points are detected, they must be made distinctive. To do so, one approach is to build a vector which depends on a small image patch near the interest point. There are two common methods used to compute a descriptor: SIFT and SURF. The SIFT descriptor is based on orientation histograms. The SURF (Speed-Up Robust Features) [30] descriptor is based on the sum of the Haar wavelet response around the point of interest.

In the SIFT algorithm, the calculation of the descriptor is similar to the last step of the detection process, but it is more computationally intensive. The descriptor is computed from an image area around the interest point, which is first transformed according to the scale and orientation computed during the detection phase. This ensures that the content of the descriptor is not sensitive to the image scaling and rotation. Thus, the descriptor only depends on the orientation and the amplitude of the gradient in multiple neighborhood areas of the point of interest. In order to create a descriptor having the aforementioned properties a common way to proceed is to determine the orientations in different 4×4 pixels patches in the 16×16 neighborhood of the interest point. Nevertheless, other patch sizes may be used. The Fig. 7.14 shows an example of gradient in each patch around an interest point. An histogram of orientations is then created for each patch. The final descriptor is composed by a vector containing values from all histograms.

The final size of the descriptor depends on the choice of patch size and the number of orientations classes in the histogram. For example, for a patch size of 4×4 pixels and a number of orientation classes of 8, the histogram created will have a size of 36 like in the Fig. 7.14.

In this example, the values in each histogram of the 16 patches are organized according to 8 angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) for a final size vector of 128 values. Then the vector is normalized in sum unit to obtain contrast invariance. To be sure that the vector of the interest point will be invariant in front of the local affine transformations, each value is weighted in the histogram. The method allows having a descriptor more robust and unique for each interest point on the image.

As in the detection section, several other alternatives at the SIFT descriptor are available since some years. These alternatives are presented here as a non-exhaustive list:

- The ORB algorithm [31] is like SIFT divided in two major parts: the detection and the creation of a descriptor. ORB is based on the Features from Accelerated Segment Test (FAST) algorithm for the interest points detection and on the BRIEF algorithm [32] for the descriptor creation. The FAST algorithm relies on the corner detection method.

The BRIEF algorithm itself is based on the DAISY algorithm [33] which uses another type of descriptor which is a binary vector which created by following 3 steps:

- The research of a pattern around the interest point,
- The selection of several couple of points a
- The creation of the descriptor itself starting by a comparison of point intensity in each couple. If the value of the first point intensity is higher of the second, the value returned is 1, else it is 0. This way allows to obtain a binary chain whose the size is depending of the number of couples selected.

In the case of the binary descriptor, the pattern used can be different according to the algorithm.

- BRISK [34] is based on FAST algorithm for detection and inspired of BRIEF/DAISY for the descriptor creation, it used a concentric pattern to determined the neighborhood of the interest point.
- The Descriptor-Nets or D-net [35] is based on SIFT for detection, but guided random position are also proposed. The descriptor creation relies on paths connecting interest points in graph. Thus, it uses the information between nodes to create the descriptor instead of using neighborhood of interest points.

7.3.2.3 Comparison of Interest Points Between Two or More Image: Match and Alignment

Once the points of interest of two images are detected, the next step is to establish the match between the points of the first image with those of the second image.

This match between is determined by computing the Euclidian distance. The couple of interest points with the smallest distance is preserved. The difference between the two Euclidian distances of the interest points descriptor selected is then computed. The Fig. 7.15 is a diagram of the possible matches between three descriptors of three interest points.

The difference obtained is next compared with a user defined threshold and the match is considered strong when the difference computed is higher than the threshold. In this case, the point of interest of the first image and the best of the points selected in the second image are considered the same. This method of matching is considered as brute-force mechanism.
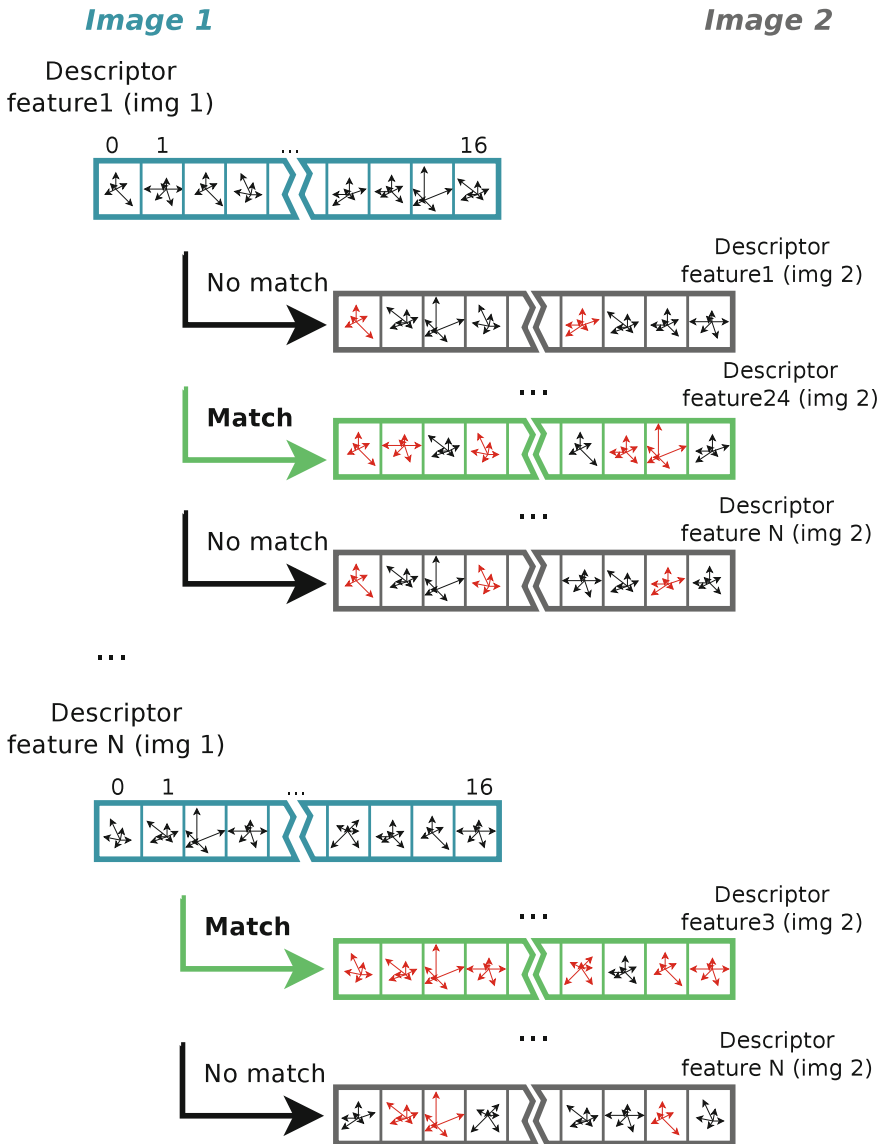
Fig. 7.15 This diagram shows the research of the best match between each descriptor of the first image with each descriptor in the second image. In this example The red orientations are the cases of match between the descriptor of interest in the first image and each descriptor of the second

In the case of SIFT, the algorithm used to determine the parameters of the alignment of two images is the algorithm RANdom SAmple Consensus (RANSAC), an iterative algorithm proposed by Fischler and Bolles.

Summarizing, the algorithm SIFT is used to find common features between two images. These features are then used to find the parameters to align these two images. Since, tomography involves the alignment of a series of images acquired with different tilts angles, the detection step and the descriptor creation step must be performed for each image in the tilts series. The alignment step occurs once the matching of the descriptors associated to interest points have been determined for each couple of images in the tilts series. To align the images in the tilts series, an additional step of chain creation is needed. This mechanism consists in the tracking of each point of interest based on the results of the matching. The obtained chain can begin and end on any image. The chains of each point of interest along the sequence of images are used during the final alignment. The alignment model, allowing to determine the orientation of the tilt axis, that can be used in tomography is detailed in Sect. 2.2.

7.4 Extension to Data Other than Tilt Series

7.4.1 Data Alignment in Multiple-Axis Tomography

The impossibility to acquire a full 180° tilt series under the electron microscope generate a lack of information. To compensate this limitation two approaches are frequently used: N-axis tomography and sub-tomogram averaging. Both of then consist in combine reconstructed volumes in a single final reconstruction but they differ in the origin of data. In N-axis tomography several tomograms of a single object are combined whereas in sub-tomogram averaging tomograms of several objects are fused.

7.4.1.1 N-Axis Tomography

N-axis tomography, allows to get back some of the missing information by acquiring multiple tilt series. The simplest approach is dual-axis tomography in which two tilt series are recorded with perpendicular tilt-axes [36, 37]. The use more than two axis has also been proposed [38, 39], but the higher dose and complexity of post-processing, compared to the gain of information is not valuable enough.

The reconstruction of data from N-axis tomography is performed by independent alignment and reconstruction of each tilt series followed by the determination of the geometric relationships between the different reconstructions. These relationships are computed using methods similar to those described in this chapter: correlation-based or landmarks-based approaches, but in three dimensions. Most of the time, software, such as imod [40], uses gold beads correspondences between the reconstructions to solve shifts, rotations and deformations such as shrinkage. Once aligned, tomograms

are combined in the Fourier space to take advantage of the precise positioning of the missing information in the Fourier transform. Thus, intensities for each frequency are generated by averaging non-null values, in the Fourier space, from tomograms to be combined.

7.4.1.2 Sub-Tomogram Averaging

Sub-tomogram averaging consists in performing many independent reconstructions of biochemically identical or structurally similar objects (often macromolecular complexes) and then to average them to get a final reconstruction having higher resolution, less noise and without missing information artifacts [41]. To this purpose the objects are extracted from a single or from several tomograms. Since each one of the extracted object does not have the same orientation with respect to tilt axis, their missing information is different. Therefore, when combined, the lack of information of one extracted object is compensated by other objects. Such in N-axis tomography the information is combined in the Fourier space.

The computational process for sub-tomogram averaging is mainly adapted from single particle analysis approach [42]. This requires data alignment, classification and reconstruction averaging. Because of the missing information occurring in different orientation in each sub-tomogram, the major difficulty is the 3D alignment which need to consider this lack of information to compute the correlation functions in the Fourier space. Presently there are two main software devoted to sub-tomogram averaging: Dynamo [43] and Relion [44].

7.4.2 Serial Images Alignment

Nowadays an important effort is realized to get 3D information by other tomographic methods which are not based on the acquisition of tilt-series. In structural biology these methods are based on sample serial sectioning followed by image acquisition of each section. The most widespread methods are the serial block face (SBF) and dual-beam (FIB–SEM) in scanning electron microscopes and serial sectioning in transmission electron microscopy (ssTEM) or array tomography in scanning electron microscopes or light microscopy. The main differences between the first two methods and the latest ones are based on how sections are obtained. For SBF and FIB–SEM, resin embedded biological samples are directly sectioned and imaged in the electron microscope whereas in the other cases sections are produced before being deposited onto a support which is transferred to the microscope. Once serial images acquired, from the image processing point of view, the reconstruction process is identical independently of the acquisition method: images are superposed to produce a 3D reconstruction.

The limitation of this technique is associated to the anisotropy in the voxel size because X, Y dimensions depends on the electron microscope whereas Z dimension

relays on the sectioning system. In addition, the sectioning process lead to a rupture of the matter continuity in the sample and to deformations induced by sectioning forces when ultramicrotomes are used. Deformations and anisotropy should be taken in consideration to align images. Thus, a pre-alignment based on rigid alignment approaches is needed and, when deformations occur, it should be followed by elastic transformations computation to correct the deformations induced by sectioning. In both cases the methods used for axial-tomography described in previous sections of this chapter are valuable for both, rigid and elastic alignment. However, specific methods, based on squared difference [45, 46] or landmarks defined either manually or automatically using specific definition [47] or using SIFT [48], have been proposed for serial image alignment.

7.5 Conclusion

Alignment of images in tomography is a crucial step to succeed in accurate reconstructions process. In spite that several methods and algorithms exist, while tomography trends towards new frontiers such as 3D chemical mapping, there is an active research on this field of image processing. Thus, from classical approaches, based on cross-correlation and on the use of gold-beads as fiducial markers, we move nowadays to more robust methods allowing the correction of image deformations or the automatic determination of inherent landmarks. Thus, it is expected that in near future, the advances in image alignment will contribute to the development of high-through flow tomography integrated in a multimodal and multiscale 3D imaging approach.

References

1. B. Turoňová, L. Marsalek, P. Slusallek, On geometric artifacts in cryo electron tomography. Ultramicroscopy **163**, 48–61 (2016)
2. J. Frank, B.F. McEwen, M. Radermacher, J.N. Turner, C.L. Rieder, Three-dimensional tomographic reconstruction in high voltage electron microscopy. J. Electron Microsc. Tech. **6**, 193–205 (1987)
3. J. Frank, B.F. McEwen, Alignment by cross-correlation. Electron Tomogr (Springer Science, 1992), pp. 205–213
4. R. Guckenberger, Determination of a common origin in the micrographs of titl series in three-dimensional electron microscopy. Ultramicroscopy **9**, 167–174 (1982)
5. S. Brandt, J. Heikkonen, P. Engehardt, Automatic alignment of transmission electron microscope tilt series without fiducial markers. J. Struct. Biol. **136**, 201–213 (2001)
6. M. Cao, H.-B. Zhang, Y. Lu, R. Nishi, A. Takaoka, Formation and reduction of streak artefacts in electron tomography. J. Microsc. **239**(1), 66–71 (2010)
7. Y. Cong, J.A. Kovacs, W. Wriggers, 2D fast rotational matching for image processing of biophysical data. J. Struct. Biol. **144**(1–2), 51–60 (2003)

8. X. Li, P. Mooney, S. Zheng, C.R. Booth, M.B. Braunfeld, S. Gubbens, D.A. Agard, Y. Cheng, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods **10**(6), 584–590 (2013)

9. M. Petrou, C. Petrou, *Image processing: The Fundamentals*. (John Wiley & Sons, 2010)

10. S. Brandt, J. Heikkonen, P. Engehardt, Multiphase method for automatic alignment of transmission electron microscope images using markers. J. Struct. Biol. **133**, 10–22 (2001)

11. H. Winkler, K.A. Taylor, Accurate marker-free alignment with simultaneous geometry determination and reconstruction of tilt series in electron tomography. Ultramicroscopy **106** (3), 240–254 (2006)

12. S. Brandt, J. Heikkonen, Optimal method for the affine F-matrix and its uncertainty estimation in the sense of both noise and outliers. In *Proc. of the 8th IEEE Intl. Conf. on Computer Vision (ICCV)* (Vancouver, Canada, 2001), pp. 166–173

13. F. Cantele, E. Paccagnini, G. Pigino, P. Lupetti, S. Lanzavecchia, Simultaneous alignment of dual-axis tilt series. J. Struct. Biol. **169**(2), 192–199 (2010)

14. D. Castaño-Díez, A. Al-Amoudi, A.M. Glynn, A. Seybert, A.S. Frangakis, Fiducial-less alignment of cryo-sections. J. Struct. Biol. **159**(3), 413–423 (2007)

15. D. Castaño-Díez, A. Seybert, A.S. Frangakis, Tilt-series and electron microscope alignment for the correction of the non-perpendicularity of beam and tilt-axis. J. Struct. Biol. **154**, 195–205 (2006)

16. M.C. Lawrence, Least-squares method of alignment using markers. Electron tomogr. (Springer, 1992), pp. 197–204

17. P. Penczek, M. Marko, K. Buttle, J. Frank, Double-tilt electron tomography. Ultramicroscopy **60**(3), 393–410 (1995)

18. C.O.S. Sorzano, C. Messaoudi, M. Eibauer, J.R. Bilbao-Castro, R. Hegerl, S. Nickell, S. Marco, J.M. Carazo, Marker-free image registration of electron tomography tilt-series. Bmc Bioinf. **10**(1), 1 (2009)

19. P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code. Commun. IEEE Trans. **31**(4), 532–540 (1983)

20. E. Tadmor, S. Nezzar, L. Vese, A multiscale image representation using hierarchical (BV, L 2) decompositions. Multiscale Model. Simul. **2**(4), 554–579 (2004)

21. S. Klein, M. Staring, J.P.W. Pluim, Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. Image Process. IEEE Trans. **16** (12), 2879–2890 (2007)

22. V.R.S. Mani, S. Arivazhagan, J.J. Braino, Multimodal image fusion using multiresolution techniques. Elixir Adv. Engg. Info. A. **55**, 13160–13163 (2013)

23. D.C. Paquin, D. Levy, E. Schreibmann, L. Xing, Multiscale image registration. Mathematics **5** (2006)

24. P.J. Slomka, R.P. Baum, Multimodality image registration with software: state-of-the-art. Eur. J. Nucl. Med. Mol. Imaging **36**(1), 44–55 (2009)

25. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, G. Marchal, Automated multi-modality image registration based on information theory. Inf. Process. Med. Imaging 263–274 (1995)

26. W.M. Wells, P. Viola, H. Atsumi, S. Nakajima, R. Kikinis, Multi-modal volume registration by maximization of mutual information. Med. Image Anal. **1**(1), 35–51 (1996)

27. D.G. Lowe, Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60** (2), 91–110 (2004)

28. R. Han, F. Zhang, X. Wan, J.-J. Fernández, F. Sun, Z. Liu, A marker-free automatic alignment method based on scale-invariant features. J. Struct. Biol. **186**(1), 167–180 (2014)

29. E. Rosten, T. Drummond, Machine learning for high-speed corner detection. Comput. Vision–ECCV 2006 (Springer, 2006), pp. 430–443

30. H. Bay, T. Tuytelaars, L. Van Gool, Surf: speeded up robust features. *Proc. Computer Vision-ECCV* (2006)

31. E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: an efficient alternative to SIFT or SURF. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 2564–2571 (IEEE) (2011)

32. M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, Pascal Fua, BRIEF: computing a local binary descriptor very fast. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1281–1298 (2012)

33. Engin Tola, Vincent Lepetit, Pascal Fua, Daisy: an efficient dense descriptor applied to wide-baseline stereo. Pattern Anal. Mach. Intell. IEEE Trans. **32**(5), 815–830 (2010)

34. S. Leutenegger, M. Chli, R.Y. Siegwart, BRISK: binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on,* pp. 2548–2555 (IEEE) (2011)

35. F. Von Hundelshausen, R. Sukthankar, D-nets: beyond patch-based image descriptors. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2941–2948 (IEEE) (2012)

36. S. Lanzavecchia, F. Cantele, P.L. Bellon, L. Zampighi, M. Kreman, E. Wright, G.A. Zampighi, Conical tomography of freeze-fracture replicas: a method for the study of integral membrane proteins inserted in phospholipid bilayers. J. Struct. Biol. **149**(1), 87–98 (2005)

37. D.N. Mastronarde, Dual-axis tomography: an approach with alignment methods that preserve resolution. J Struct. Biol. **120**(3), 343–352 (1997)

38. S. Hata, H. Miyazaki, S. Miyazaki, M. Mitsuhara, M. Tanaka, K. Kaneko, K. Higashida, K. Ikeda, H. Nakashima, S. Matsumura, and others, High-angle triple-axis specimen holder for three-dimensional diffraction contrast imaging in transmission electron microscopy. Ultramicroscopy. **111**(8), 1168–1175 (2011)

39. C. Messaoudi, N.Garreau Loubresse, T. Boudier, P. Dupuis-Williams, S. Marco, Multiple-axis tomography: applications to basal bodies from Paramecium tetraurelia. Biol. Cell **98**(7), 415–425 (2006)

40. J.R. Kremer, D.N. Mastronarde, J.R. McIntosh, Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. 71–76 (1996)

41. J.A.G. Briggs, Structural biology in situ–the potential of subtomogram averaging. Curr. Opin. Struct. Biol. **23**(2), 261–267 (2013)

42. J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State* (Oxford Univ. Press, New York, USA, 2006)

43. D. Castaño-Dez, M. Kudryashev, M. Arheit, H. Stahlberg, Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. J. Struct. Biol. **178**(2), 139–151 (2012)

44. A.M.B. Tanmay, J.R. Christophe, L. Jan, A.P. Lori, H.W.S. Sjors, Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. Structure **23**(9), 1743–1753 (2015)

45. I. Arganda-Carreras, C.O.S. Sorzano, J. Kybic, C. Ortiz de Solórzano, bUnwarpJ: consistent and Elastic Registration in ImageJ. Methods and Applications. *Proc. Conference Image J Users-Developers*, Luxembourg (2008)

46. C.O.S. Sorzano, P. Thévenaz, M. Unser, Elastic registration of biological images using vector-spline regularization. IEEE Trans. Biomed. Eng. **52**(4), 652–663 (2005)

47. C.-W. Wang, E.B. Gosno, Y.-S. Li, Fully automatic and robust 3D registration of serial-section microscopic images. Sci. R. **5** (2015)

48. S. Saalfeld, R. Fetter, A. Cardona, P. Tomancak, Elastic volume reconstruction from series of ultra-thin microscopy sections. Nat. Methods **9**(7), 717–720 (2012)