Journal of Structural Biology 196 (2016) 515–524

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Foil-hole and data image quality assessment in 3DEM: Towards high-throughput image acquisition in the electron microscope

J. Vargas [a,*], E. Franken [b], C.O.S. Sorzano [a], J. Gomez-Blanco [a], R. Schoenmakers [b], A.J. Koster [c], J.M. Carazo [a]

[a] Biocomputing Unit, Centro Nacional de Biotecnología-CSIC, C/ Darwin 3, 28049 Cantoblanco (Madrid), Spain
[b] FEI Company, Achtseweg Noord 5, 5651 GG Eindhoven, The Netherlands
[c] Koster Lab, Department of Molecular Cell Biology, Section Electron Microscopy, Leiden University Medical Center, Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form 6 October 2016
Accepted 6 October 2016
Available online 8 October 2016

Keywords:
Cryoelectron microscopy
Data collection
Image quality

ABSTRACT

Automatic or semiautomatic data collection approaches on a transmission electron microscope (TEM) for Single Particle Analysis, capable of acquiring large datasets composed of only high quality images, are of great importance to obtain 3D density maps with the highest resolution possible. Typically, this task is performed by an experienced microscopist, who manually decides to keep or discard images according to subjective criteria. Therefore, this methodology is slow, intensive in human work and subjective. In this work, we propose a method to automatically or semiautomatically perform this image selection task. The approach is based on some simple, fast and effective image quality descriptors, which can be computed during acquisition, to characterize foil-hole and data images. The proposed approach has been used to evaluate the quality of different datasets consisting of foil-hole and data images obtained with a FEI Titan Krios electron microscope. The results show that the proposed method is very effective evaluating the quality of foil-hole and data images, as well as predicting the quality of the data images from the foil-hole images.

1. Introduction

Transmission Electron Microscopy applied to the three-dimensional (3D) characterization of vitrified biological specimens, a field referred to as cryo-EM, is experiencing a profound revolution. First, individual electrons can be counted directly by new Direct Electron Detectors (DEDs) that allow fast acquisition of movies instead of images, and better sensitivity than older CCD detectors. Additionally, great efforts have been made in electron microscope automation and throughput due to software which, after a short setup time, automates image acquisition for hours or even days. Combined, these new capabilities open the venue to a much larger data processing capabilities than currently possible making feasible high throughput data collection and processing (HTDCP). However, for HTDCP to be really possible, much intelligence has still to be incorporated into the microscope image acquisition software. Currently, there are different software high throughput data collection solutions which allow various degrees of automation and robustness in the electron microscope. These software solutions are developed to emulate some of the decisions and actions of a highly trained microscopist in collecting data from vitreous ice specimen. In the following, we review the basic characteristics of some of these solutions.

One of the earliest high throughput data collection solutions was Leginon (Carragher et al., 2000; Potter et al., 2001; Suloway et al., 2005). Leginon is a multi-scale imaging system to automatically acquire large numbers of images from transmission electron microscopes. EPU is the automated Single Particle Analysis (SPA) acquisition application from FEI Company, which similarly enables automated collection of large numbers of vitrified particles from pre-selected areas on a grid. The main steps of these automated SPA acquisition software packages are:

(a) First, an array of very low magnification images is acquired to obtain a so-called *atlas* of the entire grid. This atlas is used to select appropriate grid square targets, manually or by a semiautomatic process and according to an intensity-based criterion.
(b) After this, an image of each pre-selected grid square is obtained at higher magnification and a set of appropriate foil-holes are selected manually or semiautomatically.
(c) The imaging system is then magnified (adjusting automatically imaging parameters as defocus and astigmatism) and centered for each selected foil-hole, obtaining a new image of it. According to the ice quality, the imaged foil-holes

* Corresponding author.
E-mail address: jvargas@cnb.csic.es (J. Vargas).

http://dx.doi.org/10.1016/j.jsb.2016.10.006
1047-8477/

within the grid square can be then selected manually or semiautomatically using some image feature metrics, such as intensity mean and standard deviation. In the semiautomatic mode the definition of proper values for these feature metrics have to be determined previously by the user from high quality foil-hole images.

(d) High magnification images (data images) are automatically acquired from the previously selected foil-holes according to an acquisition template. In Leginon, the quality of these data images is determined from their power spectrum in order to remove micrographs affected by drift, among other problems. It is important to mention that the acquisition of the final data images takes a considerable amount of time because of processes such as autofocus and microscope stage stabilization, among others.

A scheme of this acquisition process is summarized in Fig. 1, where the different magnified images (a–d) refer directly to steps (a-d) previously outlined. EPU has shown to be useful in different works (Bai et al., 2013; Meyerson et al., 2014). Other solutions aimed at automated image collection for cryo-EM are JADAS, the JEOL Automated Data Acquisition System (Zhang et al., 2009), TOM (Nickell et al., 2005), which is an open MATLAB toolbox for the automated data acquisition in electron tomography, Auto3-DEM (Yan et al., 2007) that is more specialized for acquiring images of viruses, SerialEM (Mastronade, 2005) a widely used data collection software for cryo-electron tomography applications, UCSF Tomography (Zheng et al., 2006) or SAM (Shi et al., 2008).

All these acquisition systems concentrate on automatic instrument control allowing obtaining large numbers of images from transmission electron microscopes with minimal human interaction. However, obtaining only (or mostly) high quality images is still an open problem. As a consequence, the quantity of bad quality images collected when the microscope is commanded in automatic mode is potentially large, limiting the throughput and requiring a posterior image screening task. It is in this context in which this work is formulated, aiming at developing new image processing algorithms that could allow a more intelligent selection of sets of good images during automatic acquisition, with minimal user intervention and preferably only during setup phase. To this end, we aim at significantly increasing our current capacity to detect high quality regions from low magnification images (foil-hole images), so that data collection at high magnification (data images) will be focused only on the highest quality areas, maximizing the throughput of the instrument and the final quality of the so-obtained biological information, at the same time assuring a faster data collection. In Fig. 2 we show a data flow diagram of the new acquisition methodology with the novel steps indicated in red. Acquisition time reduction is achieved by the "Skip" branch. Data quality improvement is achieved thanks to both assess steps.

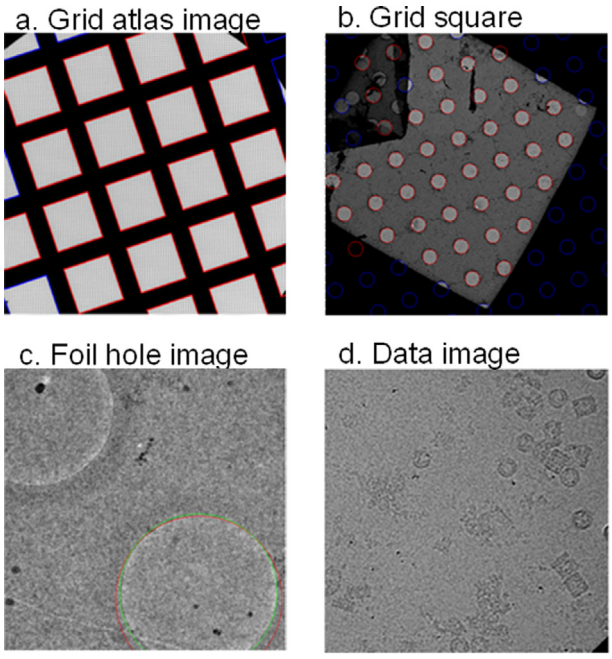

Fig. 1. Scheme of the acquisition process from the very low magnification Atlas image (a) to the high magnification data image (d).

2. Proposed method

The proposed method is based on defining a set of image quality descriptors to characterize foil-hole as well as data images. Our objective is the detection of bad quality images affected by well-known image quality deteriorating effects such as problems inside the foil-hole, like the presence of inhomogeneous ice, contaminations, empty images without macromolecular particles, completely

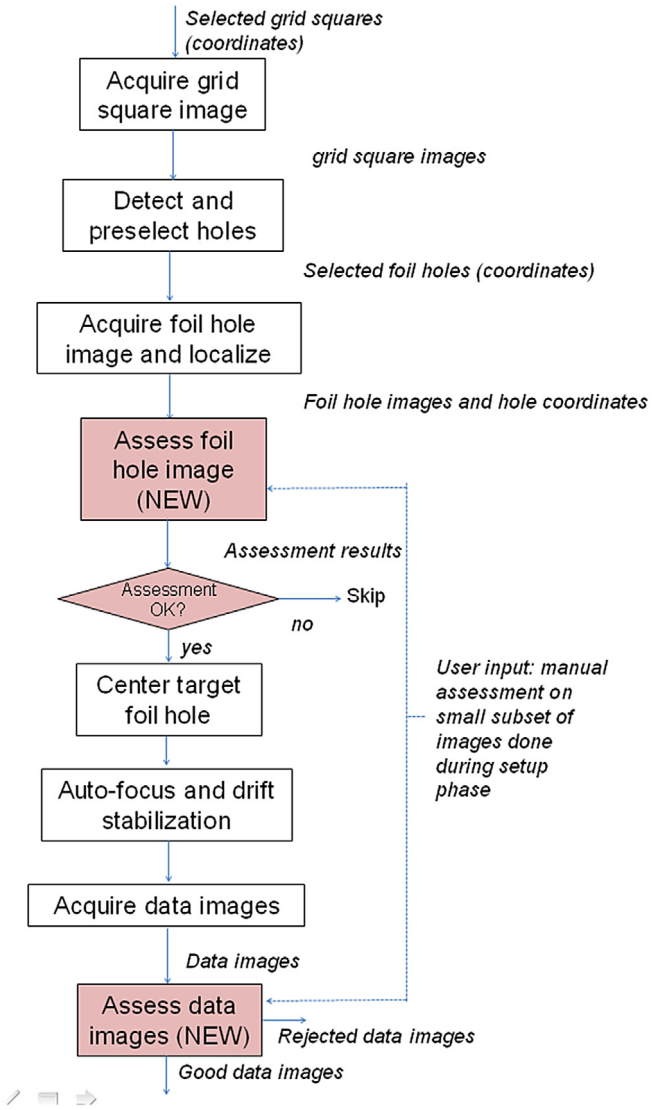

Fig. 2. Scheme of the data flow diagram of the new acquisition methodology with the novel steps indicated in red.

empty holes without ice, or the presence of a too thin or thick ice layer. After defining a set of fast and adequate image descriptors, our approach can work in two different ways. The first way is a semiautomatic process, where a short number (typically around 10 or 20) of manually supervised calibration images allow a classification approach to learn what high or low image quality means. After this learning process, the system is able to predict the goodness or badness of all acquired images. In the second way, the approach runs in a totally automated and unsupervised fashion. In Fig. 3, we show examples of typical problems in foil-hole as well as data images. Note that in this figure the high resolution images (data images) are obtained at the center of the corresponding foil-hole images. In the first column of Fig. 3 we show a good quality foil-hole, as well as data images. On the other hand, in the second, third and fourth columns of Fig. 3, we show bad quality images affected by the absence of particles, too thin ice layer, inhomogeneous ice and presence of contaminants, respectively. In the next section we present our proposed image quality descriptors to characterize foil-holes images and data images.

2.1. Foil-hole image quality descriptors

To characterize foil-hole images, the first step consists of the localization of the foil-hole in order to compute the quality descriptors only within this region. The image analysis algorithms for automatic localization of foil holes are already built in automated data collection systems like EPU, Leginon, SerialEM, JANDAS or UCSF Tomography as this process is crucial for accurate targeting of the final data image acquisitions. On the pixels that are inside the foil-hole, the following image quality descriptors are computed:

1) Histogram based indicators: for ice thickness characterization and contaminant detection.
2) Local homogeneity ice descriptors: for ice homogeneity characterization.
3) Autocorrelation descriptors: for accurate ice contaminants detection.

2.1.1. Histogram based descriptors

The histogram of gray values provides effective information to classify foil-hole images as high or low quality. First, note that the histogram of the ice background noise follows a Gaussian distribution (Sorzano et al., 2004). Therefore, the intensity histogram should have a significantly different shape when the imaged foil-holes have good ice compare to the situation when the ice is too thin or affected by contaminants. In order to quantify these differences, we have fitted the obtained histograms to Gaussian distributions, obtaining as parameters the mean, the standard deviation, and the fitting error. In Fig. 4 we show a diagram of the process. First, the foil-hole is detected and a square image is cropped from the foil-hole, which is indicated in Fig. 4 with a red square. On this subimage the intensity histogram is obtained and fitted to a Gaussian distribution, which corresponds to the red curve in Fig. 4. In Fig. 5(a), (b), (c) and (d) we show the resultant histogram-based descriptors for the foil-hole images shown in Fig. 3. As can be seen from Fig. 5, the proposed descriptors are sensitive to typical image deteriorating effects.

2.1.2. Local homogeneity ice descriptors

The presence of non homogenous ice within the foil–holes is an important and difficult to detect case. Usually, the ice layer is thinner in the foil-hole center than at the borders, and an example of this problem is shown in the third column of Fig. 3. In order to detect non homogenous ice within the foil–holes, we propose custom descriptors based on obtaining the intensity median $\mu(r)$ and intensity dynamic range $\xi(r)$ between the 10th and 90th percentiles along rings with different radii r, given by

$$\begin{aligned} \mu(r) &= p_{50}[I(r)] \\ \xi(r) &= p_{90}[I(r)] - p_{10}[I(r)] \end{aligned} \tag{1}$$

Therefore, for each foil-hole image we obtain two curves with information of the intensity median and dynamic range at different radius. From these curves we compute as descriptors the mean, dynamic range and standard deviation. In Fig. 6 we show a diagram of how these curves are computed. As can be seed from Fig. 6, for each foil-hole image a set of binary circular masks with different radii are defined and within these masks we compute the intensity median and dynamic range.

2.1.3. Autocorrelation descriptors

We also propose to use autocorrelation-based descriptors to accurately detect the presence of ice contaminants within the foil-hole as shown in the fourth column of Fig. 3. In these cases, the autocorrelation peak should be higher and wider than when there are no contaminants, because of the presence of a stronger spatial coherence in these images. Therefore, to quantify these spatial coherence differences we obtain as descriptors the mean and

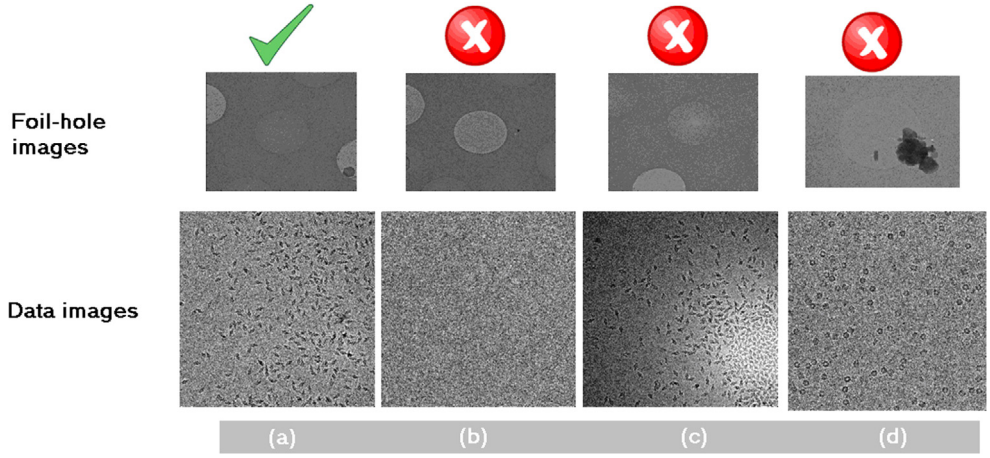

Fig. 3. Typical examples of foil-hole (first row of images) and data images (second row of images). In the first column, good quality foil-hole images as well as data images are shown. Additionally, in the second, third and fourth columns we show cases of bad quality images affected by absence of particles, too thin ice layer, inhomogeneous ice and existence of contaminants, respectively.

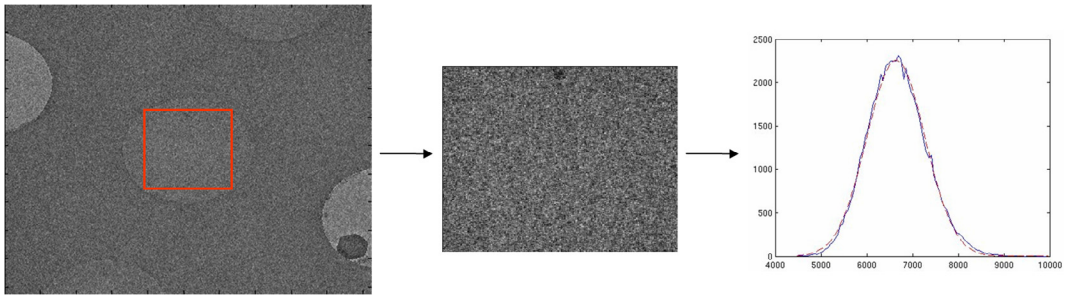

Fig. 4. Diagram of the process to compute the gray value histogram-based quality parameters from the foil-hole images. First, on the left hand side, the foil-hole is detected and a square image is cropped from the foil-hole, indicated with a red square. On this subimage, central column, the intensity histogram is obtained and fitted to a Gaussian distribution, which corresponds to the curve on the right hand side.

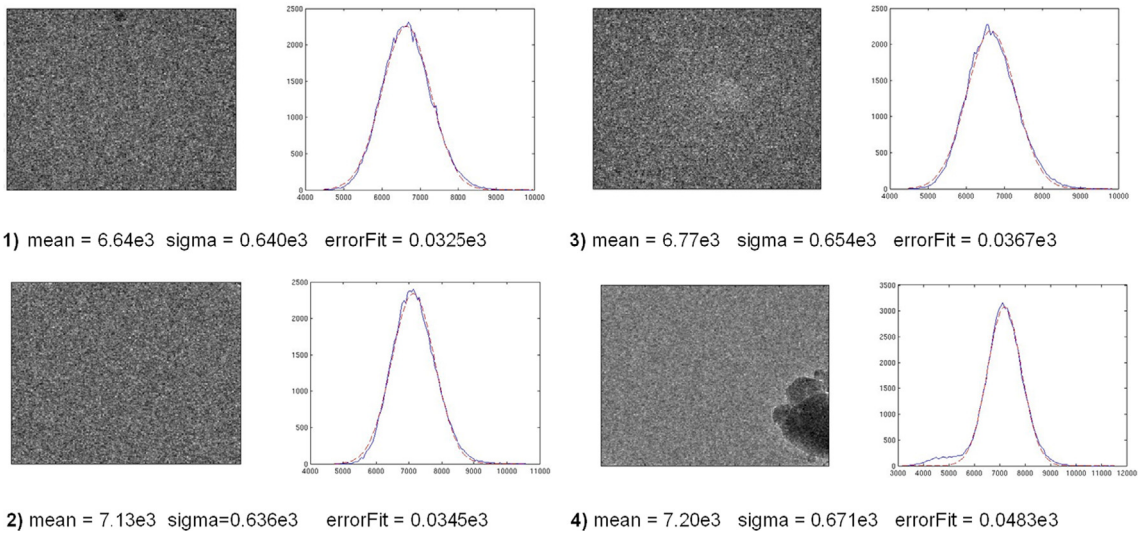

Fig. 5. Resulting histograms and histogram-based descriptors for the foil-hole images shown in Fig. 1 and labelled as (a), (b), (c) and (d).

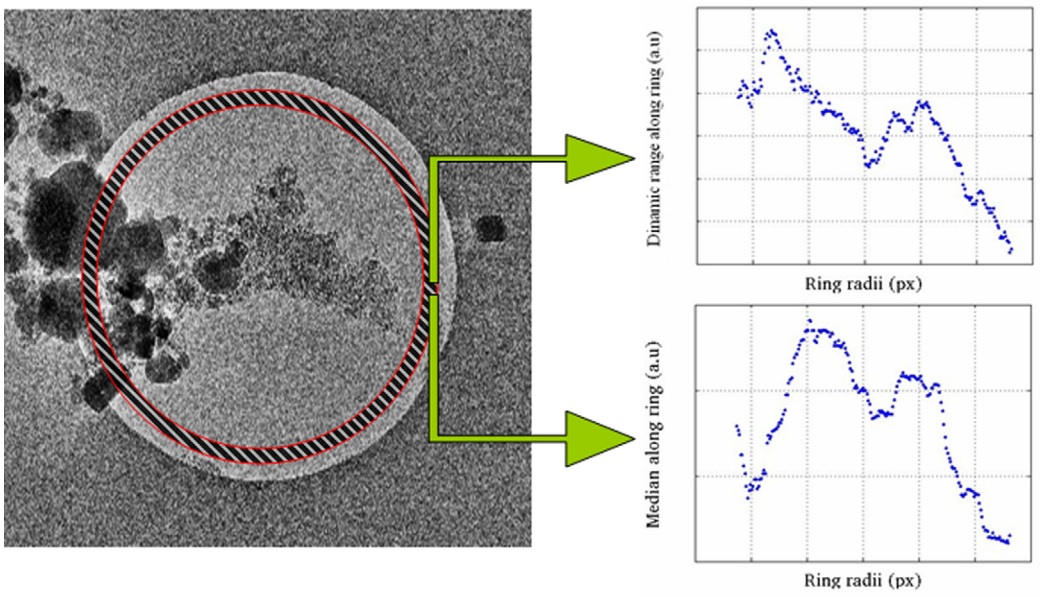

Fig. 6. Diagram of the process of obtaining the local homogeneity ice descriptors. A set of binary circular masks are defined in the foil-hole images with different radii and within these masks, the median and the dynamic range are determined. The mask is indicated by a red disk in the Figure.

the standard deviation of the autocorrelation map calculated over the cropped square image.

2.2. Data image quality descriptors

In order to characterize the high resolution data images, we use all the same descriptors outlined above plus an additional new descriptor which proved not to be feasible on the foil-hole images, aiming at detecting data images with non-homogeneous particle distributions as shown in Fig. 3(c).

2.2.1. Particle density-based descriptors

In order to detect non homogeneous particle distributions, first the data images are band-pass filtered using an approximate estimation of the particle size in order to enhance the low resolution particle signal. After this, the enhanced image is divided into 9 regions of similar size and the histogram of each region as well as of the whole image is computed. All these histograms are normalized by their total number of counts, and then, if the particle distribution within a data image is homogeneous, these curves must be approximately the same. In order to quantify this "particle homogeneity", we subtract the different normalized region histograms from the normalized whole image histogram, obtaining the standard deviation of these differences. Obviously, a small standard deviation is an indicator of a homogenous particle distribution.

2.3. Classification

Once the different image quality descriptors have been obtained, the next step consists in classifying the data according to its quality. We propose two ways of performing this task: supervised and automatic. On one hand, in the supervised mode, a small set of data previously labelled by an experienced user either as high or low quality is used to teach the classifier what high or low quality means. We have used a decision tree classifier (Breiman et al., 1984), which defines simple rules from the supervised data to predict responses to data. On the other hand, in the automatic mode, the classification is performed in a blind manner without any supervision. Here a z-score index is assigned to each image based on the PCA method to detect outliers based on similarity analysis. In this case, it is assumed that the number of outliers are smaller compared with the number of inliers in the dataset (Filzmoser, 2005; Vargas et al., 2013).

3. Experimental results

In order to check the proposed approach, we have developed multiple tests using different samples with different number of images, which are shown in Table 1. All the images were acquired without beam induced movement correction, as the image quality descriptors are computed on uncorrected average images due to the computational cost associated to frame-alignment during acquisition, and because these descriptors do not require high-resolution images, as the features we are interested in are clearly visible at low resolution.

3.1. First experiment: supervised training with all images

First, we supervised manually all the foil-hole and data images (denoted in Table 1 as FHI and DI, respectively) obtained by the EPU software, obtaining the resultant percentage of bad quality images for both cases. In all the cases the supervision process was performed by the same experienced user, who defined high and low quality images. As can be seen from Table 1, these percentages are in some cases very high, mainly because of too thin ice layers and the existence of contaminants. Second, we trained the classifier with all the available data, subsequently determining for each image if it was of high or low quality. Note that using the complete dataset for training the classifier and then using it again to classify the images does not represent a normal workflow of analysis. However, this test is intended to check that the quality image descriptors can retrieve enough information to classify diverse data in a consistent manner, and furthermore to detect consistency between FHI and DI images. In this favourable case, the differences between the predicted and manually determined quality were very low in all cases. These differences are shown in Table 1 as ErrorFHI and ErrorDI, representing the percentage of data and foil-hole images for which the classifier obtains a different quality value than the experienced user. Additionally, another interesting check that we did was to determine the data image quality from the foil-hole quality descriptors using the classifier. These results are shown in the last column of Table 1, which shows (in percentage) very small differences between the determined quality values with respect to the quality values obtained manually for the data images. This means that the quality of the data image can be obtained from the FHI descriptors.

In Fig. 7 we show the obtained decision trees for the foil-hole (a) and data images (c) of GroEL sample when the complete dataset was used for training. Observe that the vector $\mathbf{x}$ is composed by the different image quality descriptors presented in Section 2. The decision trees are formed by simple rules that classify the data from their quality image indicators. The decision trees are constructed from the training set and the corresponding feature vectors. From our experiments, we have checked that these rules are changed when we modify the training set without a significant modification in the classification error. This means that probably our image quality descriptors provide redundant information for our classification process. However, observe that this set of image

Table 1
Results on different datasets used in the experimental results section. Each dataset corresponds to a different sample (first column) and is composed by a different number of images (second column). In the third column the percentage of low quality images estimated manually (DI means data image and FHI means foil-hole image) is shown, while in the forth and fifth columns we present the accuracy results obtained when the foil-hole and data images descriptors are used to determine the quality of the foil-hole and data images, respectively. The sixth and last column presents the classification accuracy when the quality of foil-hole images is used to predict the quality of their corresponding data images. Note that the complete dataset was used to train the classifier.

Dataset	Num	% Bad	ErrorDI	ErrorFHI	FHI → DI
β-Gal	43	(DI)81% (FHI)81%	0%	2%	0%
	41	(DI)32% (FHI)34%	0%	3%	3%
	156	(DI)5% (FHI)6%	0%	1%	1%
GroEL	128	(DI)21% (FHI)27%	0%	1%	0%
	105	(DI)32% (FHI)27%	1%	5%	0%
KLH	140	(DI)1% (FHI)3%	0%	2%	0%
Hemoglobin	204	(DI)12% (FHI)11%	1%	3%	2%

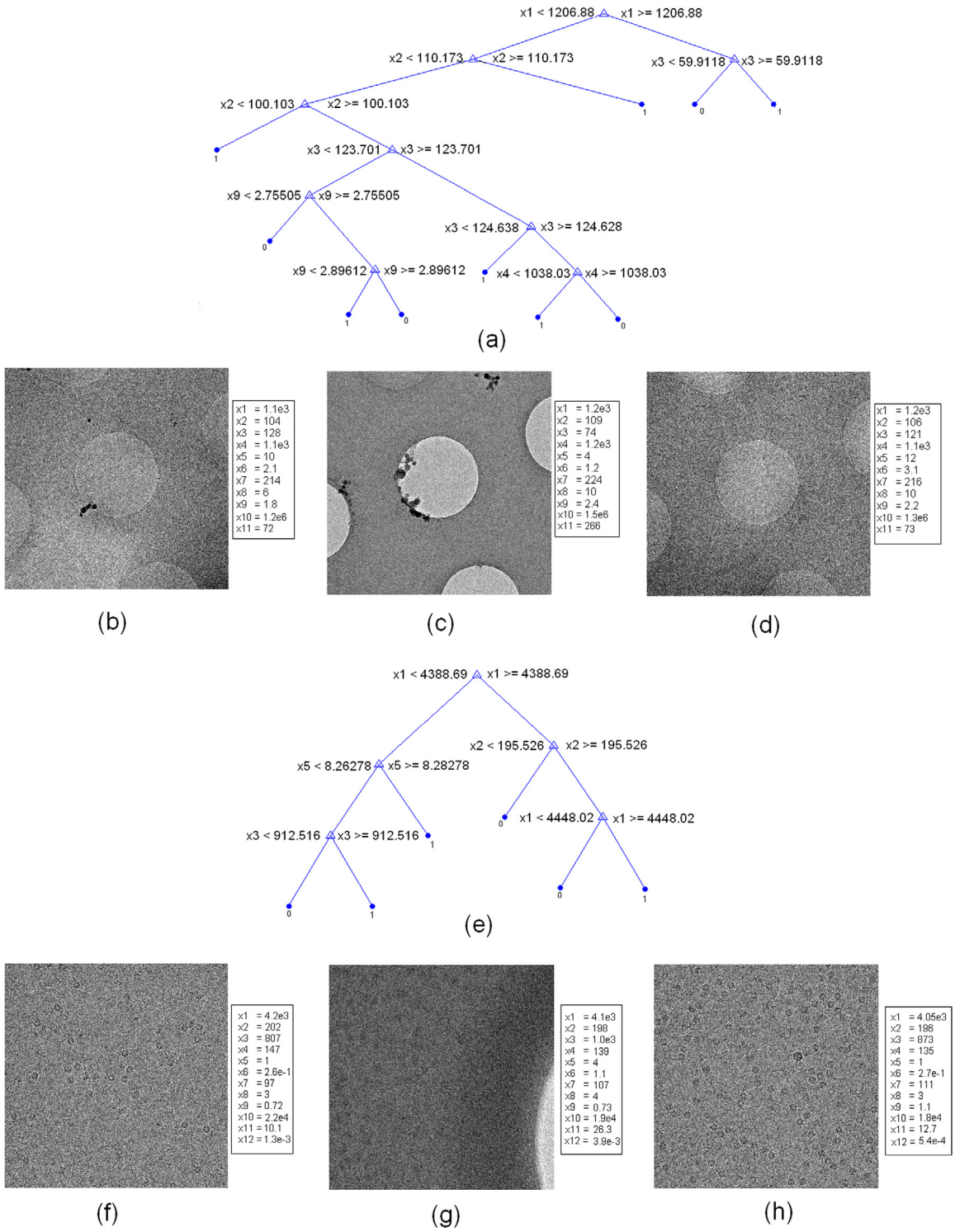

Fig. 7. Obtained decision trees for the foil-hole (a) and data images (c) of GroEL sample when the complete dataset was used for training. In these decision trees 0 and 1 corresponds to high and low quality images respectively. The feature vector **x** is composed by the following descriptors: *x*1, *x*2 and *x*3 corresponds to the mean, standard deviation and fitting error of the fitted Gaussian distribution (introduced in Section 2.1.1); *x*4, *x*5 and *x*6 refers to the median, dynamic range and standard derivation of μ(r) signal and *x*7, *x*8 and *x*9 are the median, dynamic range and standard derivation of ξ(r) (introduced in Section 2.1.2); *x*10, *x*11 corresponds to the mean and standard derivation of the autocorrelation map (introduced in Section 2.1.3), and finally for the data images *x*12 is the standard deviation of the differences between the region histograms and the normalized whole image histogram (introduced in Section 2.2.1). In addition (b), (d) and (e) correspond to failed and successful low and high quality FHI detections. Moreover, (f), (g) and (h) correspond to failed and successful low and high quality DI detections.

quality descriptors is inexpensive in terms of computing time and the classification precision is directly related with the richness of the information provided. Therefore, there is not any reason to try to reduce this set.

3.2. Second experiment: supervised training with 10 images

In the second experiment, the previous analysis was repeated, but this time only ten images from each dataset were used for

training, which were selected manually as being representative of each of the datasets. The obtained results are shown in Table 2. Obviously, in this case the prediction accuracy of the classifier is worse than in the previous case, but still the success rate is quite high, especially when the quality of the data image is predicted from the foil-hole quality descriptors. Additionally, we tested the behaviour of the proposed approach when the ten images used for training the classifier were selected randomly instead of being manually selected. For this purpose, we picked randomly ten foil-hole and data images from each dataset to train the classifiers, and then, the quality of all the foil-hole and data images were predicted, respectively. The obtained prediction accuracy results for the β-Gal (a–c), GroEL (d–f) and KLH (g–i) are shown in Fig. 8, where we performed these experiments 500 times, i.e. we obtained the prediction accuracy 500 times for each case, selecting 10 training images randomly. In the case of the β-Gal and GroEL sets the prediction accuracy was using all the images (the different datasets shown in Tables 1–3 were merged). As can be seen from Fig. 8, in that case in which the 10 training images are representative, providing good ensemble information to the classifier, the obtained accuracy results are good, and in most of the cases below 10%. On the other hand, if the 10 randomly selected training images are not representative of the dataset –for instance, when they are all very similar–, the classifier can not extract significant information from them and, therefore, the obtained accuracy is not good.

3.3. Third experiment: automatic training

We also tested on the same data set our proposed approach in a blind automatic mode, obtaining the results shown in Table 3. As can be seen from Table 3, we have not reported results for the first case of the β-Gal. In this case, the number of bad quality images is so high that it has not sense to use a method based on outlier's detection as in this case the outliers are indeed the good quality images. In this table we also report the percentage of false negative detection by the automatic classification approach (images that are classified as low quality when they are good quality). These percentages are presented in Table 3 by the numbers within the parenthesis in columns four, five and six. As can be seen from Table 3, the accuracy results provided by the classifiers are worse than in the cases shown before. However, there is still a significant improvement with respect to the accuracy reported without any quality classification process, which is shown in the second column of Table 3. Additionally, observe that most of the quality classification errors are due to the presence of false negatives, which on one hand reduces the number of good quality images, but on the other, it reinforces that those images being classified are indeed of good quality.

3.4. Fourth experiment: 3D reconstruction evaluation

Finally, we studied the effect of using our proposed approach on a 3D reconstruction when the data is pruned and not pruned according to their estimated quality value. The case we focused on was the one in which those data images classified as bad quality, using as input their corresponding foil-hole images, were not taken into account in the reconstruction process. To this purpose, we used a dataset of 389 micrographs of GroEL sample, partially composed by the images used on Experiments 3.1, 3.2 and 3.3, and we obtained their CTF using the Ctffind4 method (Rohou and Grigorieff, 2015). We proceeded by picking 27,347 particles using the Xmipp semi-automatic particle picking method (Abrishami et al., 2014), that were classified using Relion (Scheres, 2012) into 200 classes with the goal of removing bad images. From these selected classes we used RANSAC (Vargas et al., 2014) for the initial volume determination. Additionally, we run our proposed approach to classify the low resolution foil-holes images, with the classifier initially trained using 20 images. We used 20 images instead of 10 images as before to add some extra robustness to the classifier before going to 3D reconstruction. In this way only 150 high magnification data images were labelled as good quality, which is a 38% of the initial 389 micrograph set. From this reduced data set we select the particles included in the data images labelled as good quality obtaining 13,245 picked particles, which were all contained in the largest dataset of 27,347 particles. Our objective here is to show that the smaller set of 13,245 particles is one of the subsets with highest quality from the larger set of 27,347 particles. To show this point we performed 3D reconstructions using RELION and gold-standard procedure using the smaller set of 13,245 particles classified by the FHI images and ten different subsets of size 13,245 particles randomly picked from the larger set. The different FSC curves obtained from each 3D reconstruction by gold-standard are shown in Fig. 9. As can see from Fig. 9, the results obtained from the smaller set of 13,245 particles classified as high quality by the FHI images, has the best FSC values and the highest resolution. Therefore, it can be concluded that using our proposed approach the amount of high quality data obtained from the electron microscope can be improved (relaxing subsequent micrograph screening processes) without compromising the acquisition time and enhancing the throughput. Finally, in Fig. 10 we show FSC curves obtained from the smaller "good quality" set of 13,245 particles (red curve) and from the complete set of 27,347 particles (green curve). These curves show that essentially at high-resolution both reconstructions have the same amount of information, while the high quality set has 52% fewer particles than the other. The GroEL data used in the paper was acquired with a DDD detector (Falcon II) but without beam induced movement correction. In addition, our sample preparation procedure suffered from some small conformational continuous heterogeneity in the sample. These points explain that the resolution achieved is around 7 Å in both cases. Note that this comparison is not fair in the sense that using our proposed method during data collection we may have the same number of particles that without using this approach after similar microscope time. Therefore, the fair comparison would be between reconstructions obtained with similar number of particles as was done in Fig. 9. In Fig. 2 the main workflow of automated acquisition was shown, with the new steps indicated in red. The improved throughput of the data acquisition

Table 2
As Table 1, but using only ten images to train the classifiers.

Dataset	Num	% Bad	ErrorDI	ErrorFHI	DI → FHI
β-Gal	43	(DI)81% (FHI)81%	14%	9%	9%
	41	(DI)32% (FHI)34%	5%	5%	7%
	156	(DI)5% (FHI)6%	2%	2%	2%
GroEL	128	(DI)21% (FHI)27%	0%	1%	0%
	105	(DI)32% (FHI)27%	4%	9%	2%
KLH	140	(DI)1% (FHI)3%	0%	5%	2%
Hemoglobin	204	(DI)12% (FHI)11%	8%	9%	5%

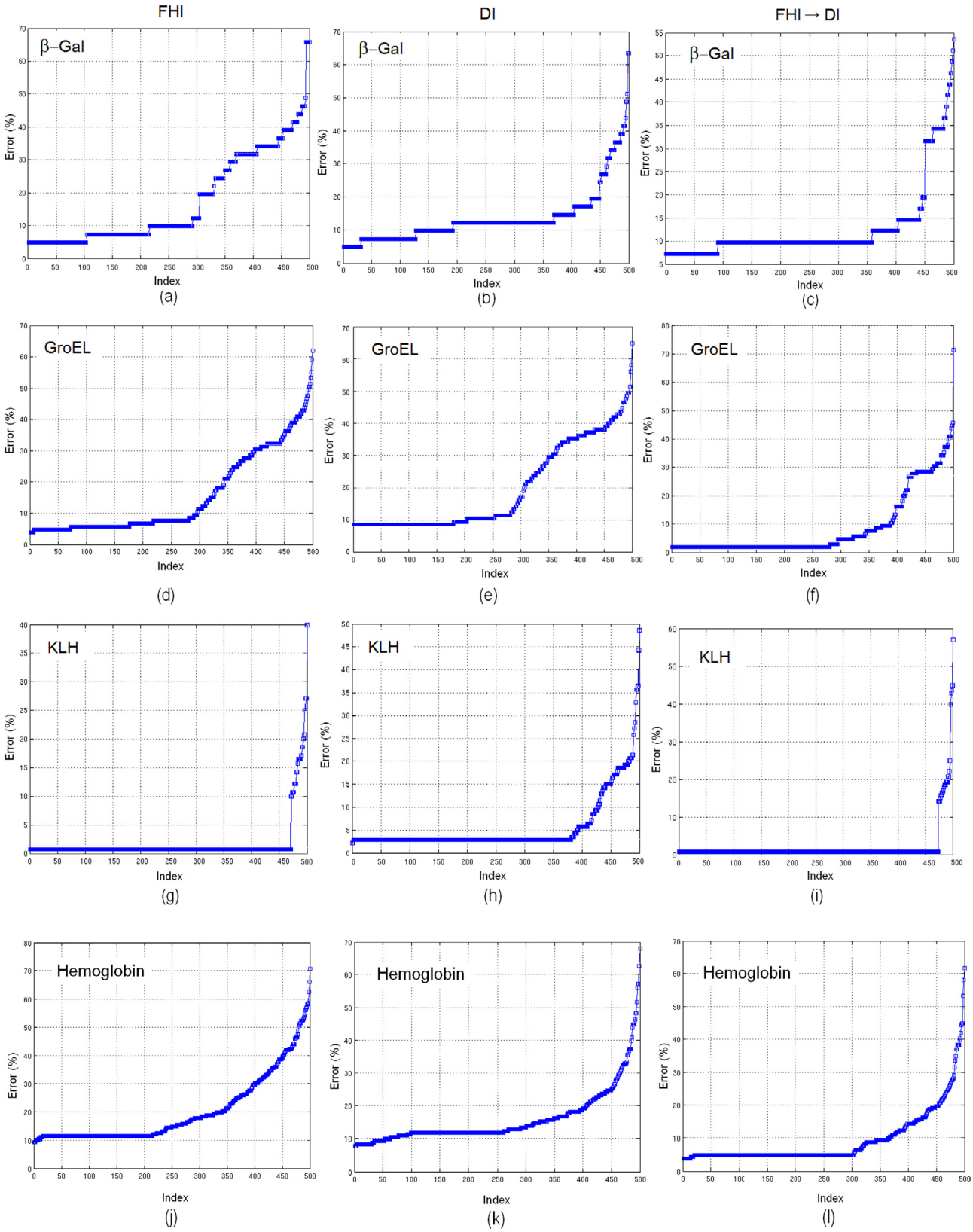

Fig. 8. Prediction accuracy results for the β-Gal (a–c), GroEL (d–f), KLH (g–i) and Hemoglobin (j–l), where the prediction accuracy was obtained selecting 10 training images randomly 500 times for each case (cases are ranked left to right in increasing error). For the case of the β-Gal and GroEL the datasets shown in Tables 1–3 were merged.

process is obtained because after rejecting a foil-hole image, the next steps are skipped. In the example above, we went from 389 to 150 data images rejecting 127 foil-hole images (33%) from a total of 278. The acquisition of the entire GroEL dataset with EPU took 19 h. The steps of moving the stage, centering the hole, auto-focus, waiting for drift stabilization, and final acquisition on

Table 3
As Tables 1 and 2, but using the automatic classification approach. The numbers within the parenthesis in columns four, five and six represent the percentage of false negatives (images that are classified as low quality while they are actually good quality) detected by the automatic classification approach.

Dataset	Num	% Bad	ErrorDI	ErrorFHI	FHI → DI
β-Gal	43	(DI)81% (FHI)81%	–	–	–
	41	(DI)32% (FHI)34%	3% (100%)	10% (25%)	7% (34%)
	156	(DI)5% (FHI)6%	17% (92%)	33% (87%)	34% (91%)
GroEL	128	(DI)21% (FHI)27%	12% (98%)	10% (100%)	10% (100%)
	105	(DI)32% (FHI)27%	11% (75%)	6% (50%)	9% (34%)
KLH	140	(DI)1% (FHI)3%	56% (100%)	59% (100%)	59% (100%)
Hemoglobin	204	(DI)12% (FHI)11%	40% (90%)	41% (91%)	41% (90%)

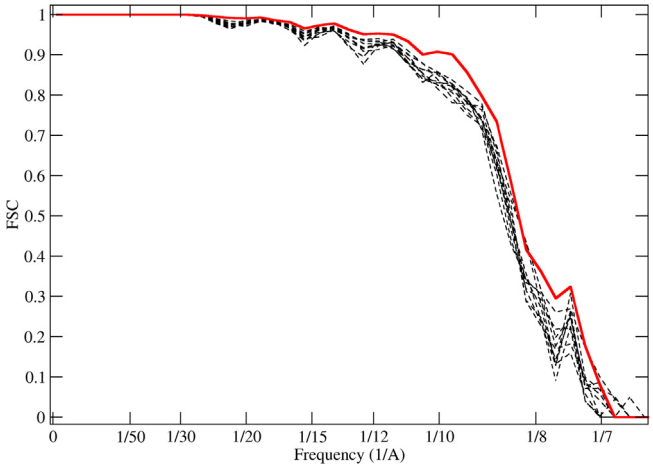

Fig. 9. The dashed black lines represent FSC curves obtained by gold-standard procedure computed using 13,242 particles randomly picked from the set of 27,347. The red curve is the respective FSC curve obtained by the 13,242 particles picked from the selected high quality set of micrographs.

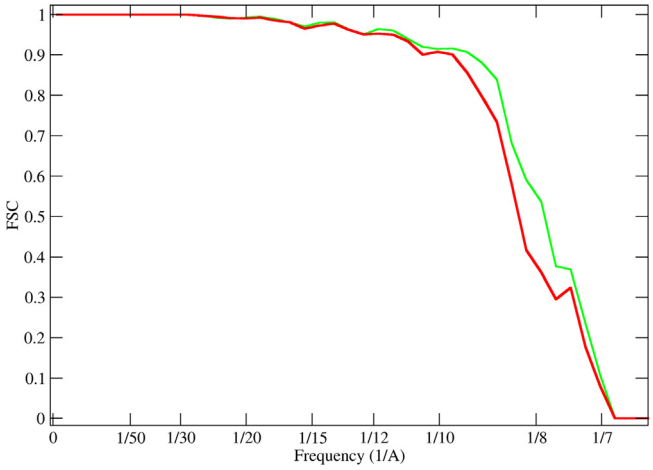

Fig. 10. FSC curves obtained by gold-standard procedure computed using the smaller "good quality" set of 13,245 particles (red curve) and the complete set of 27,347 particles (green curve).

average took 175 s per foil hole. If a foil hole would be directly rejected after the first foil hole image acquisition, only the stage move, a single acquisition, and execution of the image assessment algorithm would be required, in total taking approximately 10 s. This means that when the used EPU version would be patched to skip rejected foil holes, the saved acquisition time is #rejected foil holes $\times$ time saving = 127 * (175 – 10) = 5 Hours and 50 min. For this dataset the total acquisition time would therefore go down from 19 to 13 h, a speed up of about 30%. Note that besides a speed up of acquisition time, the reconstruction time will also go approximately 33% faster, since 33% of the images in the original dataset don't need to be processed anymore. Finally in Table 4 we provide details about the sample preparation and image acquisition of the different datasets used above in the first, second and third experiments.

4. Conclusion

Nowadays, automated single particle acquisition software packages allow a high degree of automation and robustness in the operation of electron microscopes. These software packages let managing the microscope in an automated manner for hours and even days without any human interaction. However, little effort has been put into the development of automatic or semiautomatic approaches to acquire only high quality data. Obtaining high quality sample preparation is a difficult process which should be optimized probably for each sample (Vinothkumar and Henderson, 2016). As a consequence not all the obtained foil holes are indicated to be imaged at high magnification. Typical problems that compromise the quality of the data images are the absence of particles in the amorphous ice, presence of inhomogeneous ice layers or presence of contaminants. All of these problems limit the quality of the collected data, requiring subsequent data curing efforts and limiting the amount of high-quality data to be collected from the electron microscope in a limited period of time. Nonetheless, obtaining high-quality data is of great importance in 3DEM. Firstly, the resolution of the final reconstruction will be affected by the quality of the input data. Secondly, microscope acquisition time is expensive, and therefore, users must obtain as many good quality data as possible in their restricted period of acquisition time.

Table 4
Details about the sample preparation and image acquisition of the different datasets used on the Experimental Results section.

Sample	β-Gal	GroEL	KLH	Hemoglobin
Grid Type	Quantifoil	Quantifoil	Quantifoil	Quantifoil
Hole size	2.5 μm	2.5 μm	2.5 μm	2.5 μm
Spacing	4.25 μm	4.25 μm	4.25 μm	4.25 μm
Vitrification by	Vitrobot	Vitrobot	Vitrobot	Vitrobot
TEM	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios
Detector	Falcon II	Falcon II	CETA 1	Falcon I
Data image size	4096 $\times$ 4096	4096 $\times$ 4096	4096 $\times$ 4096	4096 $\times$ 4096
Data image Field-of-view	465 nm	465 nm	929 nm	570 nm
Foil hole image field-of view	7.23 um	7.23 um	7.50 um	25 um

In this work, we have proposed a data quality classification approach based on a set of easy-to-compute and fast image quality descriptors, which can be calculated on the fly while the microscope is acquiring data. Two procedures have been proposed: supervised and automatic. The preferred procedure to use depends on the quality of the sample preparation. First, the automatic processing procedure gives more flexibility as no input is required from the microscopist. However, this approach should be performed only in cases where the number of low quality images is significantly smaller than the high quality ones. Therefore, this procedure should be used only with optimally prepared samples. Observe from Table 3 that the presence of false positives is lower and almost zero in the datasets with low percentages of low quality images. Second, the supervised method yields higher success rate but requires a training procedure that should be performed once for each collecting session. This process requires around 10 or 20 foil-hole and data images and takes a few minutes. This case is more useful in cases more affected by low quality images. Our experimental results show that a significant improvement can be made using our proposed method, illustrated by a GroEL reconstruction using approximately 40% less data we obtained at the end practically the same resolution as a reconstruction with no data removed.

Acknowledgements

The authors would like to acknowledge economical support from the Spanish Ministry of Economy and Competitiveness through Grants AIC-A-2011-0638 and BIO2013-44647-R, the Comunidad de Madrid through Grant (S2010/BMD-2305), as well as, a ComFuturo grant funded by the Fundación General CSIC and a EMBO short-stay grant with reference with ASTF 264-2014 to Javier Vargas, C.O.S. Sorzano is recipient of a Ramón y Cajal fellowship. We also want to acknowledge the NIH Living lab for providing the β-Gal and GroEL data, Felix de Haas (FEI) for providing the KLH dataset, and Sacha de Carlo (NECEN, FEI) for providing the Hemoglobin dataset.

References

Abrishami, V., Zaldivar-Peraza, A., de la Rosa-Trevín, J.M., Vargas, J., Otón, J., Marabini, R., Shkolnisky, Y., Carazo, J.M., Sorzano, C.O.S., 2014. A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. Bioinformatics 30 (20), 2891–2898.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. CRC Press.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R.A., Potter, C.S., Pulokas, J., Reilein, A., 2000. Leginon: an automated system for acquisition of images from vitreous ice specimens. J. Struct. Biol. 132 (1), 33–45.

Filzmoser, P., 2005. Identification of multivariate outliers: a performance study. Austrian J. Stat. 34, 127–138.

Shi, J., Williams, D.R., Stewart, P.L., 2008. A Script Assisted Microscopy (SAM) package to improve data acquisition rates on FEI Tecnai Electron Microscopes equipped with Gatan CCD Cameras. J. Struct. Biol. 164 (1), 166–169.

Mastronade, D.N., 2005. Automated electron microscope tomography using robust prediction of specific movements. J. Struct. Biol. 152, 36–51.

Potter, C.S., Chu, H., Frey, B., Green, C., Kisseberth, N., Maddedn, T.J., et al., 2001. Leginon: a system for fully automated acquision of 1000 electron micrographs a day. Ultramicroscopy 77, 153–161.

Rohou, A., Grigorieff, N., 2015. CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B., 2005. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 151 (1), 41–60.

Bai, X.C., Fernandez, I.S., McMullan, G., Scheres, S.H.W., 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife 2, e00461.

Meyerson, J.R., Kumar, J., Chittori, S., Rao, P., Pierson, J., Bartesaghi, A., Mayer, M.L., Subramaniam, S., 2014. Structural mechanism of glutamate receptor activation and desensitization. Nature 514, 328–334.

Nickell, S., Förster, F., Linaroudis, A., Net, W.D., Beck, F., Hegerl, R., Baumeister, W., Plitzko, J.M., 2005. TOM software toolbox: acquisition and analysis for electron tomography. J. Struct. Biol. 149 (3), 227–234.

Scheres, S.H.W., 2012. A Bayesian view on Cryo-EM structure determination. J. Struct. Biol. 415 (2), 406–418.

Sorzano, C.O.S., de la Fraga, L.G., Clackdoyle, R., Carazo, J.M., 2004. Normalizing projection images: a study of image normalizing procedures for single particle three-dimensional electron microscopy. Ultramicroscopy 101, 129–138.

Vargas, J., Abrishami, V., Marabini, R., de la Rosa-Trevín, J.M., Zaldivar, A., Carazo, J. M., Sorzano, C.O.S., 2013. Particle quality assessment and sorting for automatic and semiautomatic particle-picking techniques. J. Struct. Biol. 183, 342–353.

Vargas, J., Alvarez-Cabrera, A.L., Marabini, R., Carazo, J.M., Sorzano, C.O.S., 2014. Efficient initial volume determination from electron microscopy images of single particles. Bioinformatics 30 (20), 2891–2898.

Vinothkumar, K.R., Henderson, R., 2016. Single particle electron cryomicroscopy: trends, issues and future perspective. Q. Rev. Biophys. 49, e13.

Yan, X., Sinkovits, R.S., Baker, T.S., 2007. AUTO3DEM-an automated and high throughput program for image reconstruction of icosahedral particles. J. Struct. Biol. 157 (1), 73–82.

Zhang, J., Nakamura, N., Shimizu, Y., Liang, N., Liu, X., Jakana, J., Marsh, M.P., Booth, C.R., Shinkawa, T., Nakata, M., Chiu, W., 2009. JADAS: a customizable automated data acquisition system and its application to ice-embedded single particles. J. Struct. Biol. 165 (1), 1–9.

Zheng, S.Q., Keszthelyi, B., Branlund, E., Lyle, J.M., Braunfeld, M.B., Sedat, J.W., Agard, D.A., 2006. UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. J. Struct. Biol. 165 (1), 1–9.