Journal of Structural Biology 189 (2015) 163–176

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Alignment of direct detection device micrographs using a robust Optical Flow approach

Vahid Abrishami [a,1], Javier Vargas [a,1], Xueming Li [b], Yifan Cheng [c], Roberto Marabini [d], Carlos Óscar Sánchez Sorzano [a], José María Carazo [a,*]

[a] Biocomputing Unit, Centro Nacional de Biotecnología-CSIC, C/ Darwin 3, 28049 Madrid, Spain
[b] School of Life Sciences, Tsinghua University, Beijing 100084, China
[c] The Keck Advanced Microscopy Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158-2517, USA
[d] Escuela Politécnica Superior, Universidad Autónoma de Madrid, C/ Francisco Tomás y Valiente, 28049 Cantoblanco, Madrid, Spain

ARTICLE INFO

Article history:
Received 7 October 2014
Received in revised form 29 January 2015
Accepted 3 February 2015
Available online 12 February 2015

Keywords:
Direct detection devices
Beam induced motion
Single particle analysis
Electron microscopy

ABSTRACT

The introduction of direct detection devices in cryo-EM has shown that specimens present beam-induced motion (BIM). Consequently, in this work, we develop a BIM correction method at the image level, resulting in an integrated image in which the in-plane BIM blurring is compensated prior to particle picking. The methodology is based on a robust Optical Flow (OF) approach that can efficiently correct for local movements in a rapid manner. The OF works particularly well if the BIM pattern presents a substantial degree of local movements, which occurs in our data sets for Falcon II data. However, for those cases in which the BIM pattern corresponds to global movements, we have found it advantageous to first run a global motion correction approach and to subsequently apply OF. Additionally, spatial analysis of the Optical Flow allows for quantitative analysis of the BIM pattern. The software that incorporates the new approach is available in XMIPP (http://xmipp.cnb.csic.es).

1. Introduction

The single-particle analysis (SPA) technique is able to yield three-dimensional (3D) structural information for biological complexes at near atomic resolution by combining many thousands of projection images obtained using transmission electron microscopy (TEM) (Frank, 1996). To achieve high-resolution results in SPA, the characteristics of the image-recording medium are of great importance. Traditionally, electron microscopy images were either recorded on photographic film or with scintillator-based charge-coupled device (CCD) cameras. Each of these two types of detectors offered certain advantages and disadvantages. Film was the preferred recording medium for high-resolution information but required manual scanning of the micrographs, thus limiting automation of processing. Scintillator-coupled CCDs permitted high-throughput image acquisition, allowing full integration between the electron microscope and image-processing software packages. However, scintillator-coupled CCDs record photons, not electrons, and the conversion of electrons to photons comes at the expense of resolution loss at high spatial frequencies (Frank, 2006). This "status quo" limitation has been overcome recently by the new generation of "direct detection devices" (DDDs), first introduced as academic prototypes in 2005 (Milazzo et al., 2005) and offered commercially in 2010 (Jin and Bilhorn, 2010). These sensors detect electrons directly and provide sharper images and higher signal-to-noise ratios (SNRs) (Bammes et al., 2012). Additionally, the fast image acquisition rate of these DDD detectors, ranging from 16 to 400 images per second (Bai et al., 2013; Bammes et al., 2012; Li et al., 2013b), makes it possible to study the behavior of frozen hydrated specimens as a function of electron dose and rate. Therefore, it has become clear that biological specimens in a solid matrix of amorphous ice move during imaging, resulting in "beam- induced motion" (BIM) (Brilot et al., 2012), which is a critical experimental "resolution barrier" in cryo-electron microscopy (Glaeser and Hall, 2011). The subsequent introduction of DDDs has cleared a path to obtaining reconstructions at close-to-atomic resolution for a broad range of specimens. However, the number of reported works that use DDDs is currently not large, and therefore, certain basic questions on BIM characterization remain unanswered. In general, BIM is expected to induce patterns of local movement, although the degree of locality and the

* Corresponding author. Fax: +34 585 4506.
E-mail address: carazo@cnb.csic.es (J.M. Carazo).
[1] Co-first authors.

http://dx.doi.org/10.1016/j.jsb.2015.02.001
1047-8477/

extent of the movement itself are varying widely in different reports (Bai et al., 2013; Booth et al., 2004; Campbell et al., 2012; Li et al., 2013a,b).

In Bai et al. (2013), results for the Falcon II (FEI) are presented using two test samples: the prokaryotic and eukaryotic ribosome. The authors report a movie processing workflow in which an initial 3D map (which disregards BIM) is first obtained and subsequently used to estimate an initial alignment from different frames by applying a statistical refinement method (Scheres, 2012). In their work, the authors report a high degree of local sample movements (see Fig. 1b in Bai et al. (2013)). The algorithm proposed in Bai et al. (2013) produces high-resolution information, but it requires the specimen images to be detected and picked up from the initial video frames, which is a challenge for small particles. However, in Li et al. (2013b), the authors showed results for the K2 Summit (Gatan) direct electron-detection camera and achieved ~3.6-Å resolution in their 3D map of an archaeal 20S proteasome (~700 kDa and dihedral D7 point group symmetry) using 10,000 particles. The alignment method used in Li et al. (2013b) consists of a pure in-plane drift correction in which a step of the sub-frame translational alignment is introduced by dividing each frame into a number of sub-frames (normally 3×3 sub-frames, each of 2000×2000 pixels). This approach is fast if running on GPUs (typically, it takes approximately 10–20 s to process 16 frames of 3876×3876 pixels), and at the end, an "average" micrograph is generated for each movie via the summation of all corrected frames. The output is easy to connect with standard processing workflows in use in the field because the DDD "video" is transformed into a "micrograph". The method is certainly appropriate for global sample movements but is not the best option if the sample motion is local. However, by achieving close to atomic resolution, Li et al. (2013b) convincingly showed that for their data, the majority of the motion to be corrected was global, especially if the first few frames were discarded (see Fig. 4b and e in Li et al. (2013b)). At the same time, we note the difference from the results in Bai et al. (2013) and Campbell et al. (2012) in which the authors report a high degree of local sample movements, as previously mentioned. Recently, in Scheres (2014), another method for BIM compensation was introduced that operates over sets of previously picked particles by fitting them to line trajectories along the frames of the stack. Moreover, Wang et al. (2014) introduced an approach that corrects the motion of boxed particles using the running averages of the frames and calculating the cross-correlations between each frame and the sum of the previously aligned frames, starting at the end of the exposure and working backwards towards the beginning. In this work, we do not enter into a discussion of the nature of BIM itself, but instead, we concentrate on a new image- processing approach that aims to achieve the following objectives: (1) obtain an in-plane "BIM corrected" image that integrates all frames and is computed directly from the stack without performing a particle picking step, and (2) provide fast, objective and quantitative characterization of BIM that accounts for both global and local BIM patterns. Naturally, this integrated image can be used in any of the standard image-processing workflows in cryo-EM as if it were a traditional micrograph.

The proposed method is based on an advanced Optical Flow approach (abbreviated as "OF" in this work) using a pyramidal implementation of the Lucas–Kanade (LK) algorithm (Lucas and Kanade, 1981) with iterative refinement (Bouguet, 2001), which makes the approach quite robust to high levels of noise (Vargas et al., 2014). In essence, OF works best at a local level and is therefore particularly suited for those cases in which the BIM pattern presents a high degree of local movements, as in the Falcon II data sets used in this work. If the BIM pattern is characterized primarily by global movements, OF will have only a minor effect on the final average. Still, even for those latter cases, we have found it advantageous to use the Li et al. (2013b) method combined with OF, by running the Li method followed by the second method to obtain an additional level of refinement and a highly intuitive graphical representation of the total BIM pattern.

2. Methods

Our proposed method is based on a regularized Optical Flow approach. The input is a video composed of a set of unaligned low dose frames, and the output is a single image obtained by averaging the resulting motion-corrected frames.

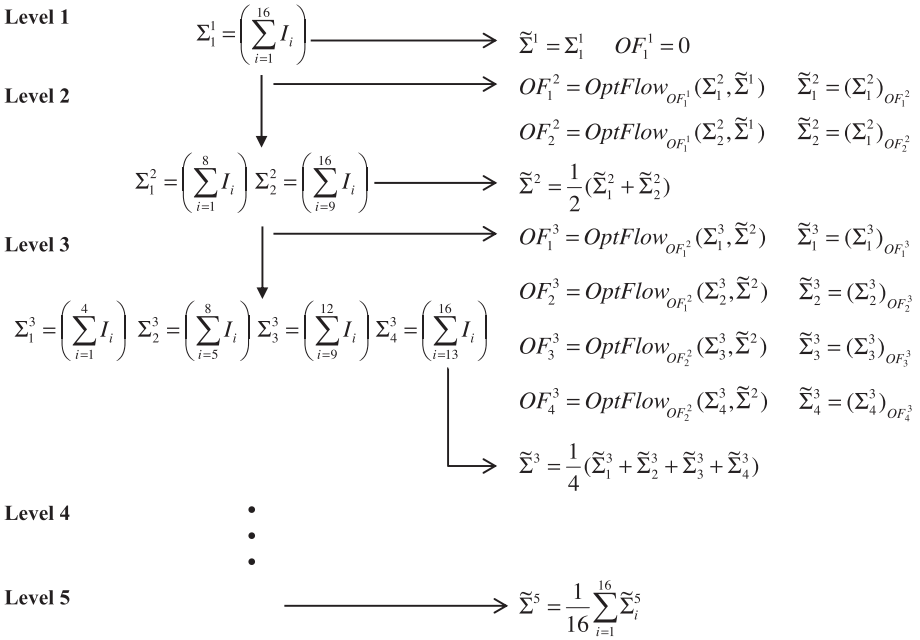

Fig.1. Diagram of the proposed Optical Flow method: The required steps for alignment of a video with 16 frames using the proposed Optical Flow method.

2.1. Optical Flow approach

Optical Flow is a method used in computer vision to obtain the local motion field between two image frames taken at times t and $t+\Delta t$ (Horn and Schunck, 1981). Let us assume that an arbitrary pixel at time t, with coordinates given by (x, y) and intensity $I(x, y, t)$ moves between two consecutive frames to a new position $(x+\Delta x,\ y+\Delta y)$ with changed intensity $I(x+\Delta x,\ y+\Delta y,\ t+\Delta t)$. Assuming that the movement is small, we can expand the intensity map at $t+\Delta t$ as given by:

$$I(x+\Delta x, y+\Delta y, t+\Delta t) \cong I(x,y,t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \tag{1}$$

We assume that brightness constancy equation holds, which is given by:

$$I(x+\Delta x, y+\Delta y, t+\Delta t) = I(x,y,t) \tag{2}$$

This restriction assumes that the brightness of an object does not significantly change between two consecutive frames, an assumption that is not problematic in our case because the exposure of each frame within the movie is the same. Expression (1) can be rewritten using the brightness constancy equation as:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{3}$$

where $u = \frac{\Delta x}{\Delta t}$ and $v = \frac{\Delta y}{\Delta t}$ correspond to the velocity components or local shift components per unit time (assuming that t is only a frame index in our case, but taking into account the temporal information, we can actually retrieve the velocity components). The LK method assumes that the displacements between images are small and approximately constant within a neighborhood of the point $(x_0,\ y_0)$ under consideration. Thus, the Optical Flow equation can be assumed to apply for all pixels $(\tilde{x},\ \tilde{y})$ within a window or neighborhood centered at $(x_0,\ y_0)$, and in that case, the motion (velocity) vector $(u,\ v)$ at the point $(x_0,\ y_0)$ must satisfy:

$$I_x(\tilde{x},\ \tilde{y})u + I_y(\tilde{x},\ \tilde{y})v = -I_t(\tilde{x},\ \tilde{y}) \tag{4}$$

with I_x, I_y and I_t equal to $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$ and $\frac{\partial I}{\partial t}$, respectively. Note that for a window or neighborhood composed of $N_w = N_x \times N_y$ pixels with N_w larger than two, Eq. (3) gives an over-determined set of N_w equations. Therefore, we obtain the displacements between images t and Δt at $(x_0,\ y_0)$ by:

$$\begin{pmatrix} I_x(\tilde{x}_1,\ \tilde{y}_1) & I_y(\tilde{x}_1,\ \tilde{y}_1) \\ \cdots & \cdots \\ I_x(\tilde{x}_N,\ \tilde{y}_N) & I_y(\tilde{x}_N,\ \tilde{y}_N) \end{pmatrix} \begin{pmatrix} u(x_0,\ y_0) \\ v(x_0,\ y_0) \end{pmatrix} = \begin{pmatrix} -I_t(\tilde{x}_1,\ \tilde{y}_1) \\ \cdots \\ -I_t(\tilde{x}_N,\ \tilde{y}_N) \end{pmatrix} \tag{5}$$

which is of the form $A\bar{x} = \bar{b}$. We can obtain $\bar{x}$ through the Moore–Penrose pseudo-inverse of A, described concisely as:

$$\bar{x} = (A^TA)^{-1}(A^T\bar{b}) \tag{6}$$

We observe that Eq. (6) is valid only if A^TA is invertible, which is true if I_x and I_y are different from zero in the neighborhood of the point $(x_0,\ y_0)$, a condition that is satisfied in our case because of the high noise in our frames. From a practical point of view, to avoid instabilities in the inversion of A^TA, it is recommended that a small number (approximately 0.1) is added to the diagonal elements (Bouguet, 2000).

2.2. Optical Flow with pyramidal decomposition and iterative refinement

The Optical Flow approach presented is the standard LK approach (Lucas and Kanade, 1981). This approach works well if the pixel displacements between the frames are sufficiently small to allow approximation using the first-order Taylor expansion shown in Eq. (3). However, this restriction is severe and can almost never be verified. A possible method for overcoming this limitation is to use the LK approach multiple times, leading to a new Lucas–Kanade Optical Flow approach with iterative refinement. The conceptual workflow of this iterative process is presented in the following for the case of two consecutive images I_1 and I_2:

(1) Estimate the local shifts Δx and Δy at each pixel using the LK Optical Flow approach presented in Section 2.1.
(2) Warp one image toward the other using the previously estimated shifts.
(3) Repeat steps (1) and (2) N_i times, where N_i is the number of iterations selected.

Suppose that the LK approach is applied between two consecutive images I_1 and I_2 and that shifts Δx and Δy are computed between them. Next, we focus on a certain pixel (i, j), and the corresponding shifts at that pixel are $\Delta x(i, j) = 2$ and $\Delta y(i, j) = 3$ pixels, for example. In this case, applying warping to $I_2(i, j)$ corresponds to changing this intensity value to $I_2(i+\Delta x(i, j), j+\Delta y(i, j)) = I_2(i+2, j+3) = I_2^1(i, j)$, where the superindex I denotes one Optical Flow iteration. Obviously, this warping process must be performed for all pixels using the associated shifts Δx and Δy, which is practically performed using cubic interpolation. If the so-obtained values of Δx and Δy are accurate, after applying the second step (warping), the two images I_1 and I_2^1 will be approximately equal. Therefore, if the LK algorithm is applied again on these two images (step 3), we obtain $\Delta x \cong 0$ and $\Delta y \cong 0$ at every pixel. Note that following our notation, we must rewrite these shift maps as $\Delta x^1 \cong 0$ and $\Delta y^1 \cong 0$ because we have applied one extra LK iteration. Moreover, we observe that in experimental cases, it is usually not possible to obtain $\Delta x^1 \cong 0$ and $\Delta y^1 \cong 0$, and additional iterations will be necessary. Therefore, after the kth iteration, the brightness constancy equation is expressed as follows:

$$I((x+\Delta x^k)+\Delta x^{k+1}, (y+\Delta y^k)+\Delta y^{k+1}, t+\Delta t) = I(x,y,t) \tag{7}$$

Without loss of generality, we assume that $\Delta t = 1$ and that $\Delta x^k = u^k\Delta t = u^k$ and $\Delta y^k = v^k\Delta t = v^k$. The final displacement vectors computed from this iterative scheme correspond to:

$$\Delta x = \sum_{k=1}^{N_i}\Delta x^k, \quad \Delta y = \sum_{k=1}^{N_i}\Delta y^k \tag{8}$$

with N_i as the number of iterations used. The LK with iterative refinement allows the algorithm to provide good results in cases in which the movements are not sufficiently small to directly use the Taylor expression shown in Eq. (1).

An important consideration for the LK Optical Flow algorithm with iterative refinement presented is that the window or neighborhood size must be specified. As intuitively expected, the choice of this window size introduces a trade-off between accuracy and robustness of the approach. In this work, accuracy relates to the local sub-pixel systematic error that is achieved, and robustness refers to the insensitivity of the approach with respect to disturbances, i.e., noise and outliers. In a nutshell, large window sizes provide more robustness but less accuracy. Additionally, to handle large motions between the images, it is necessary to use large integration windows. To solve this trade-off, Bouguet (2000) presented a pyramidal implementation of the iterative LK algorithm introduced above that substantially reduces the dependency on the window size. This approach is based on performing an iterative LK Optical Flow approach recursively over different resolution representations of the input images, known as pyramids, which are

obtained by successively down-sampling the images. We rank these image pyramid representations from coarse to fine resolutions (note that the images with highest resolution are the input images) and subsequently perform the iterative LK approach with a fixed window size, beginning with the images with lowest resolution and moving towards those with highest resolution. This approach first provides a coarse and rather robust estimation of the motion (velocity) vectors $(u_0,\ v_0)$ from the low-resolution pyramids. These vector fields are used as an initial guess for the next Optical Flow estimation using the next higher-resolution pyramidal representation images. This process is followed recursively until the highest-resolution images (input images) are used. Observe that this pyramidal implementation of the LK Optical Flow algorithm with iterative refinement provides a clear advantage because the integration window size can remain fixed and small, thus ensuring high accuracy and robustness. In Section 3.3, we present the details of the strategy we use in this work to select the window size for each case.

2.3. Proposed Optical Flow approach for DDD frame alignment

In cryo-EM, the types of images in which we are interested are characterized by notably low SNRs. As a consequence, we must add additional robustness against noise to the standard LK Optical Flow with the pyramidal decomposition and iterative refinement (LKPDI) presented in Section 2.2 (Bouguet, 2000). To this end, we have designed a sequential processing workflow, which attempts to maximize the signal and minimize the noise at each step of Optical Flow estimation.

Because the movie frame images are characterized by notably low SNRs, it is not possible to directly obtain reliable shift maps between consecutive frames. To improve the Optical Flow results, we follow a coarse-to-fine alignment process. To this end, we first obtain the average of all frames within the movie, which is initially established as the first estimation of the corrected movie average image. Subsequently, the frames are divided into two sets by grouping frames that are consecutive in time, and the average image of each set is computed. Observe that the first estimation of the corrected movie average as well as the obtained average images of each set have high SNRs at low resolution. As a consequence, if we align these average images with respect to the available corrected movie average, we obtain robust but coarse shift estimations. Using these shifts, a new corrected average can be estimated, and each of the frame sets can be divided in two again. Therefore, we once again perform the same process between the obtained average images of each group and the available corrected movie average using the previously obtained shifts as initial guesses. As a result, we obtain improved shift estimation in each refinement process. Therefore, the proposed alignment approach consists of a coarse-to-fine estimation of the shifts, which improves the robustness of the method with respect to noise in the images.

In the following, we introduce a more detailed description of the proposed method from a mathematical point of view. We first introduce the notation and subsequently describe the algorithm. Let I_i be the ith frame in the stack, Σ_j^k be the unaligned average of the jth subgroup of the frames in the kth iteration (how these subsets are formed will be described later), $\widetilde{\Sigma}^k$ be the aligned average in the kth iteration obtained using all images and, finally, let $\widetilde{\Sigma}_j^k$ be the aligned average of the images belonging to the subset j in iteration k. Next, the Optical Flow algorithm is described as follows:

1. For $k = 1$, the algorithm computes the average of all N unaligned frames $\Sigma_1^1 = \left(\sum_{i=1}^{N} I_i\right)$ and initializes the refined average estimation as $\widetilde{\Sigma}_1^1 = \Sigma_1^1$. Next, k is updated as $k = k + 1$, and for $k > 1$, the following iterative process is executed.
2. Images are split into 2^{k-1} subgroups, and the unaligned average of each group is computed. Therefore, there are 2^{k-1} averages with $N/2^{k-1}$ images in each average such that $\Sigma_j^k = \left(\sum_{i=(j-1)s^k+1}^{js^k} I_i\right)$, where $s^k = N/2^{k-1}$. Note that for $k = 2$, the set is divided into two groups, and for each group, we obtain the unaligned averages Σ_1^2 and Σ_2^2.
3. Next, we compute the LKPDI optical flows between the new set of unaligned averages Σ_j^k and the corresponding set of aligned images computed in the previous iteration $\widetilde{\Sigma}^{k-1}$. The result, which is a map of shift vectors centered at each pixel that aligns Σ_j^k with $\widetilde{\Sigma}^{k-1}$, is known as OF_j^k. The estimation of OF_j^k requires an initial guess; our algorithm uses $OF_{j/2+j\bmod 2}^{k-1}$ for $k > 2$ and a zero-valued matrix of displacement or shift vectors for $k = 2$.
4. Using the obtained shift vector matrix (OF_j^k), the averages Σ_j^k are aligned; we refer to these aligned averages as $\widetilde{\Sigma}_j^k$. A new corrected average is computed as $\widetilde{\Sigma}^k = \sum_{j=1}^{2^k}(\widetilde{\Sigma}_j^k)/2^k$. If the groups contain more than one frame, Steps 2 to 4 are repeated.

Fig. 1 shows a diagram of the proposed approach for a movie composed of 16 frames. This procedure is applicable to movies with a number of images that is not a power of 2 by simply readjusting the appropriate indices. Note that this alignment strategy improves the robustness of the approach against noise, and at the same time, it reduces other potential issues, such as the validity of the brightness constancy equation due to radiation damage or changes in brightness resulting from mass loss during exposure.

3. Results

This section presents a collection of tests using both simulated and experimental data. To obtain the optical flow, we used the OpenCV (Bradski, 2000) (Open Source Computer Vision Library) implementation of the Lucas–Kanade dense Optical Flow (in which the Optical Flow is computed for all pixels), which provides both CPU and GPU execution modes. For the LK algorithm, we set the number of iterations to 10 and the number of scales to 6 as constant values for all datasets and also separately optimized the window size for each dataset (the details of how to obtain the window size are described in Section 3.3). The algorithm was executed on a single GeForce GTX 690 GPU card with two Kepler GPUs as well as in CPU mode. Typically, the GPU-based execution time is only approximately three times faster than in CPU execution mode, clearly showing that I/O has a large detrimental impact on the GPU performance.

3.1. Simulations

We used computer simulations to test a subset of the basic features of the newly proposed algorithm. In particular, in this simplified system, we aimed to study two different situations: (1) global versus local movements, i.e., the ability of the method to compensate for both global and local movements, and (2) random versus drift-like movements, i.e., the ability to recover information when the movement in between frames is random and when it displays a certain directionality (a drift-like effect).

In all cases, we compared the results obtained by OF with the results provided by the Li et al. (2013b) alignment approach for two different cases. Note that because the output of our method is an integrated image, we cannot easily compare the proposed method with that of Bai et al. (2013), for which the output is a

3D reconstruction. In the first case, we applied the Li et al. (2013b) alignment method using the entire collection of image frames (Li Global), and in the second, we divided the images (i.e., the frames) into nine overlapping sub-images and aligned each of them independently, as suggested in Li et al. (2013b), to render the method more robust with respect to local sample movements (Li Local). For the simulation, we used one frame of a movie provided in Bai et al. (2013) with a size of 4096 × 4096 pixels that were cropped to 3876 × 3876 pixels (because of the presence of artifacts at the borders). Depending on the type of the simulation, we applied global or local shifts (random or drift-like) to this frame, and added a random Gaussian noise to each simulated frame, thus creating a series of movies. In all cases, the shifts were randomly drawn from a uniform distribution with zero mean and a variance of 2 pixels. The SNR of each frame in the different simulated movies is 0.5.

3.1.1. Global versus local movements

3.1.1.1. Global movements. In this case, we added the same movement to each pixel of the simulated frames. Thus, we composed a simulated movie formed by sixteen unaligned frames. The different applied shifts are shown in Table S1 with reference to the shifts of the tenth frame, i.e., $\Delta\widetilde{x}_i = (\Delta x_i - \Delta x_{10})$ and $\Delta\widetilde{y}_i = (\Delta y_i - \Delta y_{10})$. Note that for a stack of size 16, Li's implementation considers the tenth frame as the reference by default and provides all other shifts with respect to this frame. For this movie, we applied OF using a window size of 25 pixels as an input parameter. In turn, the alignment approach presented in Li et al. (2013b) was executed using the entire image and also dividing the frames into nine overlapping sub-images of size 2048 × 2048 pixels, as suggested in Li et al. (2013b). The resulting shifts (mean shifts) are quite similar to the theoretical shifts, as shown in Table S1. An analysis of the root-mean-square error (*rms*) between the computed and the theoretical displacements provides average *rms* values of 0.0048 and 0.011 for Li Global, 0.0027 and 0.011 for Li Local, and 0.032 and 0.042 for OF, respectively, in pixel units. As expected, in this case, Li's method provides slightly better results than OF because it is essentially a global alignment approach, whereas OF is a local approach. However, we observe that even in this unfavorable case, the proposed OF approach provides notably accurate and precise results (on the order of 10^{-2} pixels), indicating that the two methods behave the same in practical terms.

3.1.1.2. Local movements. For this case, we used the same starting image as in the previous case but introduced local shifts between frames. For simplicity, the local shifts were introduced only in x as given by the expression:

$$\Delta x_i = \begin{cases} 0, & i = 1, \\ 4\sin\left(2\pi\left(\frac{2x}{N_C} + \frac{i}{N_R}\right)\right), & i \neq 1 \end{cases} \quad (9)$$

where Δx_i are the local shifts, (N_C, N_R) are the number of rows and columns, i is the number of the frame within the movie, and N is the number of frames that compose the movie. The reason for this simulation with only local movements in the x-axis was to provide a precision estimation for the alignment approaches (how well the shifts are recovered when the theoretical shifts were exactly zero), thus avoiding such synthetic effects as quantization and interpolation while maintaining the added noise.

Fig. 2(A) displays a graphical plot of the theoretical shifts introduced by Eq. (9), for $i = 1$, with reference to the shift of the tenth frame given by $\Delta\widetilde{x}_1 = (\Delta x_1 - \Delta x_{10})$. In the same manner, Fig. 2B presents the shifts $\Delta\widetilde{x}_i$ recovered by OF, and Fig. 2C displays the results from Li Local. In this simple graphical representation, it is clear that 2B is rather similar to 2A, but 2C is markedly different. In more quantitative terms, we computed the mean of the *rms* errors between the theoretical and obtained shifts, obtaining 5.26 pixels for Global Li, 4.04 pixels for Local Li; 0.1 pixels for OF along x; and 0.19 pixels for Global Li, 0.20 pixels for Local Li, and 0.02 pixels for OF along y. As observed from this analysis and Fig. 2, in this local-shift case, the results obtained from OF are considerably better than those provided by the Li approach, as expected by the relationship between the two methods. Additionally, we observe that the OF results along the y-axis are significantly better than the results along the x-axis because of quantization and interpolation effects and because the OF approach assumes the shifts are the same within a window, which is obviously true for the shifts along the y-axis but not true along the x-axis.

3.1.2. Random versus drift-like movements

The goal of this test was to evaluate the performance of the Optical Flow approach in those cases in which the movement in between frames was random compared with the cases in which the movement had a certain level of directionality along the stack. We were particularly interested in this question because both the OF and Bai et al. (2013) method use a first step of averaging the sets of frames without correction, and the quality of this uncorrected average would be quite different if the frames moved randomly than if the frame movement had a certain directionality. In the former case, the uncorrected average would "only" appear as an isotropically blurred version of the ideal image, a rather intuitive type of degradation that is easy to address computationally, whereas in the latter case, the result would be a rather artifactual image that might preclude further steps of processing. For simplicity, we considered global movements (i.e., in which the entire frame moves in the same direction) such that the only variable remaining was the directionality of the movement along the stack. We studied three cases: random movements, movements with restricted directionality, and drift-like movements.

The first case addressing random movements was discussed under Section 3.1.1.1, and clearly, our newly proposed method was shown to perform quite well. The second case (movements with restricted directionality) was simulated by applying different shifts between frames, where the directions of the shifts were

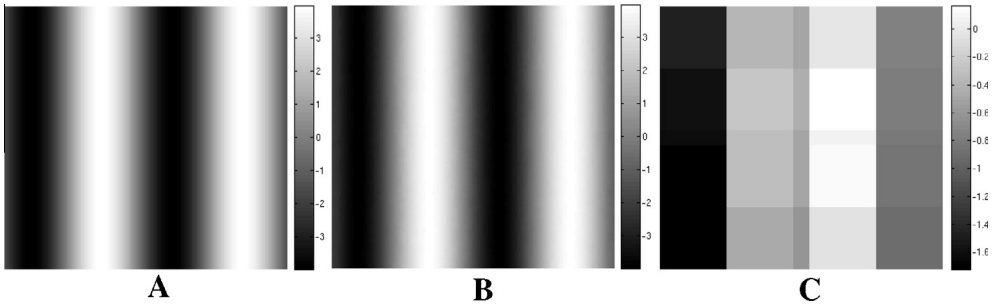

Fig.2. Recovered simulated local shifts using optical flow: graphical representation of theoretical and recovered local shifts along x referred to the shift of the tenth frame $(\Delta x_1 - \Delta x_{10})$. Theoretical shifts in x (A). Shifts obtained by OF (B) and Li Local (C), respectively.

drawn from a Gaussian distribution with a zero mean and a standard deviation of 0.4 radians or 23 degrees to generate a stack of 16 frames. Note that the shifts along the stack are cumulative; each shift is the sum of the previous shifts and the current shift. The Optical Flow method with different window sizes was applied to this stack. For a window size of 50 pixels, we obtained the best *rms* on the *x*- and *y*-axes of 0.2518 and 0.0146 pixels, respectively. Note that although we were still able to recover the shifts with good accuracy, the results were worse than in the previous case. Finally, in the third case, we considered the situation in which all shifts accumulated in exactly the same direction, and in this case, we clearly were not able to recover the movements, even if large window sizes were applied (note that the Li method was not affected by this drift-like effect and performed well in this case).

3.2. Experimental results

In this section, we present the results obtained by applying our proposed algorithm to different datasets, each taken from different specimen samples imaged on different microscopes and cameras, to assess the robustness of the proposed method under experimental conditions.

A key issue in comparing the performance of different algorithms is the use of a good comparison metric. This metric should provide objective results while ideally offering simple and rapid calculation such that multiple tests can be performed. Thus far, most of the published works on DDD have concentrated on exploring the best resolution achievable from a certain dataset when processed in a number of different ways. Indeed, this ideal describes the "ultimate" comparison metric, but it is clearly neither simple nor fast to calculate. Additionally, it is difficult to distinguish whether an improvement originates from the movie processing procedure or the 3D reconstruction iterative process. As a method of mitigating this problem, we analyzed the effect of misalignment (initial or "residual" misalignment) on the power spectral density (PSD) of the final "corrected" average. We present this analysis in the Appendix. An interesting result is that the shapes of the CTF envelope and of the spectral signal-to-noise ratio (SSNR) represent notably good performance metrics. Indeed, residual errors in the alignment correction translate into a dampening of these metrics such that we can judge the quality of the correction (how small the residual misalignment was) by comparing them and looking for the case in which this dampening is most reduced. Assuming axial astigmatism to be minimal in these high-quality datasets, we calculated frequency radial averages, resulting in 1D curves that can be easily analyzed. In all cases, we compared the results of OF with those of both the Local and Global Li approaches in terms of alignment accuracy and execution speed.

3.2.1. Ribosome dataset

In this case, we used two different movie stacks previously reported in Bai et al. (2013) and made publicly available at (http://www.ebi.ac.uk/~ardan/aspera/em-aspera-demo.html). The names of the movie stacks as deposited are "15_movie_gc_window.mrcs" and "205_movie_gc_window.mrcs", and both correspond to images of the *Saccharomyces cerevisiae* 80S ribosome. The images were recorded on a FEI Falcon II direct electron detector with nominal magnification of 59,000× at a working voltage of 300 kV, resulting in a pixel size of 1.77 Å (the size of the ribosome at this sampling rate is approximately 150 pixels).

We processed these movies with both the Li Global and Li Local approaches as well as with the proposed OF method using a window size of 150 pixels. The results are presented in Fig. 3A and B, which show the 1D profiles of the CTF envelope and SSNR functions obtained for "205_movie_gc_window.mrcs" and "15_movie_gc_window.mrcs", respectively. As expected, both the Li and OF approaches improve the results compared with the direct average of the unaligned frames. However, OF clearly provides better results (higher envelope and SSNR) than the Li approach for low and high frequencies. In Fig. 3C, we present a comparison between the periodogram (i.e., the modulus-squared of the discrete Fourier transform) of the initial uncorrected average (left) and that obtained by the application of the Optical Flow approach (right).

As a new piece of information, in Fig. 4, we present the vector fields obtained by OF corresponding to the movements that each pixel undergoes between frame 1 and frame 2 for the two test cases (see Movie S1 for a representation of the vector fields for all consecutive frames). In this figure, the information of the shifts is coded using the Hue, Saturation and Value (HSV) of the color. As the hue value varies from 0 to 1.0, the corresponding colors vary from red, through yellow, green, cyan, blue, and magenta and back to red such that red values actually occur at both 0 and 1.0. As the saturation varies from 0 to 1.0, the corresponding colors (hues) vary from unsaturated (shades of gray) to fully saturated (no white component). As value or brightness varies from 0 to 1.0, the corresponding colors become increasingly brighter. In our case, the hue gives information on the direction of the movement, and the saturation gives information for the magnitude. As observed from these movies, the orientation map is locally smooth, and the displacement pattern is clearly non-global, as originally reported (Bai et al., 2013). A simple and intuitive analysis of the movements among all of the movie frames is presented in Fig. 5 (as applied to the results shown in Movie S1). The information on the pixel movement is coded both in Cartesian and polar representations in which a number of circles are shown. The position of the center of each circle corresponds to the mean displacement between two consecutive frames calculated by averaging the pixel's displacements in *x* and *y* over all of the pixels within the frame as provided by the corresponding vector fields. In turn, the area of the circle represents the compound standard deviation in *x* and *y*, i.e., $\sqrt{\sigma_x^2+\sigma_y^2}$, where σ_x^2 and σ_y^2 are the variances in the *x*- and *y*-axes, respectively. Note that for the sake of this representation, we consider that the shifts along the *x*- and *y*-axes behave as random variables that are independent such that we can obtain the standard deviation of the sum as the square root of sum of the variances. Consequently, the position of the center of the circle represents the "global" (mean) displacement, and its area represents the pixel-to-pixel differences with respect to this mean displacement. In other words, the area is related to the amount of local versus global movements between two frames. It is clear from Fig. 5A and B that the global movements between frames are within a fraction of a pixel and that there is a certain directionality in the between-frame movement. We comment on this observation and subsequent plots later in this manuscript together with the results for additional cases presented.

The execution time for a stack of this dataset is 123 s using OF, 50 s using the Li Local approach, and 20 s using the Li Global method.

3.2.2. Influenza virus RNP dataset

In this experiment, we used a stack of influenza virus ribonucleoprotein (RNP) (unpublished data) consisting of 70 frames collected with a FEI Falcon II direct electron detector at the MRC with a pixel size of 2.26 Å. The total exposure time was 4 s.

As in the previous experiment, we obtained the integrated image using the simple average, Li Local, Li Global, and OF. A comparison of these methods using the CTF envelope as well as the SSNR is shown in Fig. 6A. As shown in this figure, OF presents a higher accuracy compared with that of the other methods. In Fig. 6B, the uncorrected (left) and corrected (right) periodograms are shown together, and in Fig. 6C, an analysis of inter-frame

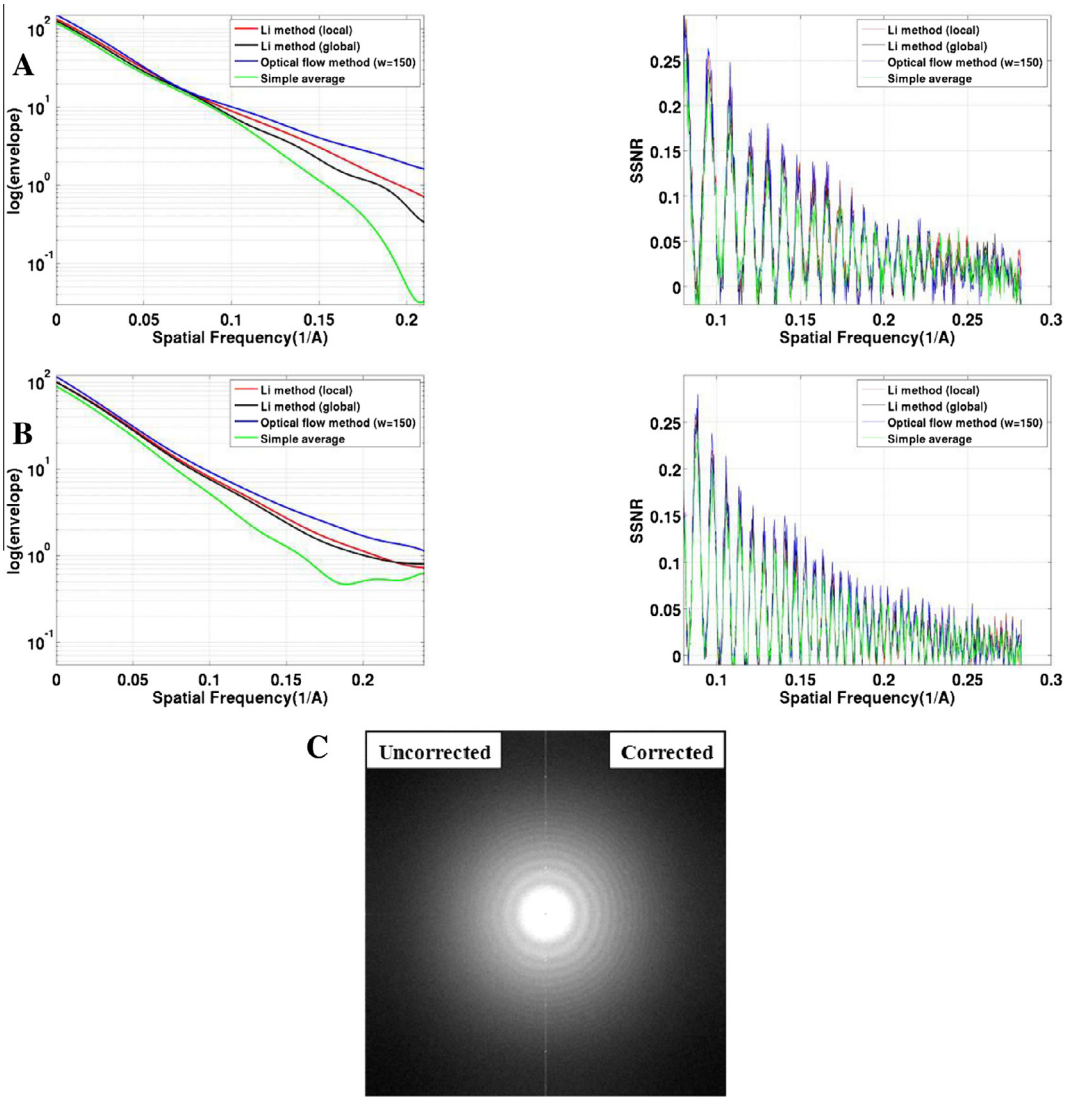

Fig.3. Comparison of the alignment results of OF and Li method for two videos of Bai et al. (2013): envelopes of the contrast transfer function (in logarithmic scale) as well as SSNR functions obtained from the unaligned average and the aligned averages using Li Global, Li Local and Optical Flow for the movies "205_movie_gc_window.mrcs" (A) and "15_movie_gc_window.mrcs" (B). Comparison of the periodograms (i.e., the modulus-squared of the discrete Fourier transform) of the uncorrected (left) and Optical Flow aligned average image (right) corresponding to "15_movie_gc_window.mrcs" (C).

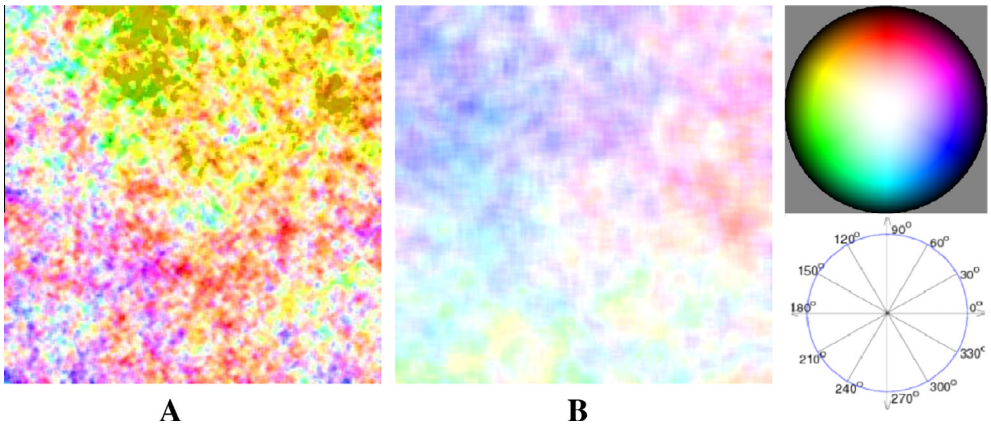

Fig.4. Vector fields for two videos of Bai et al. (2013): vector fields obtained by the Optical Flow approach corresponding to the movements for each pixel between frame 1 and frame 2 for the stacks "205_movie_gc_window.mrcs" (A) and "15_movie_gc_window.mrcs" (B) (see also Movie S1). The information of the shifts is coded using the Hue, Saturation and Value (HSV) of the color. The hue gives information on the direction of the movement, and the saturation gives information for the magnitude.

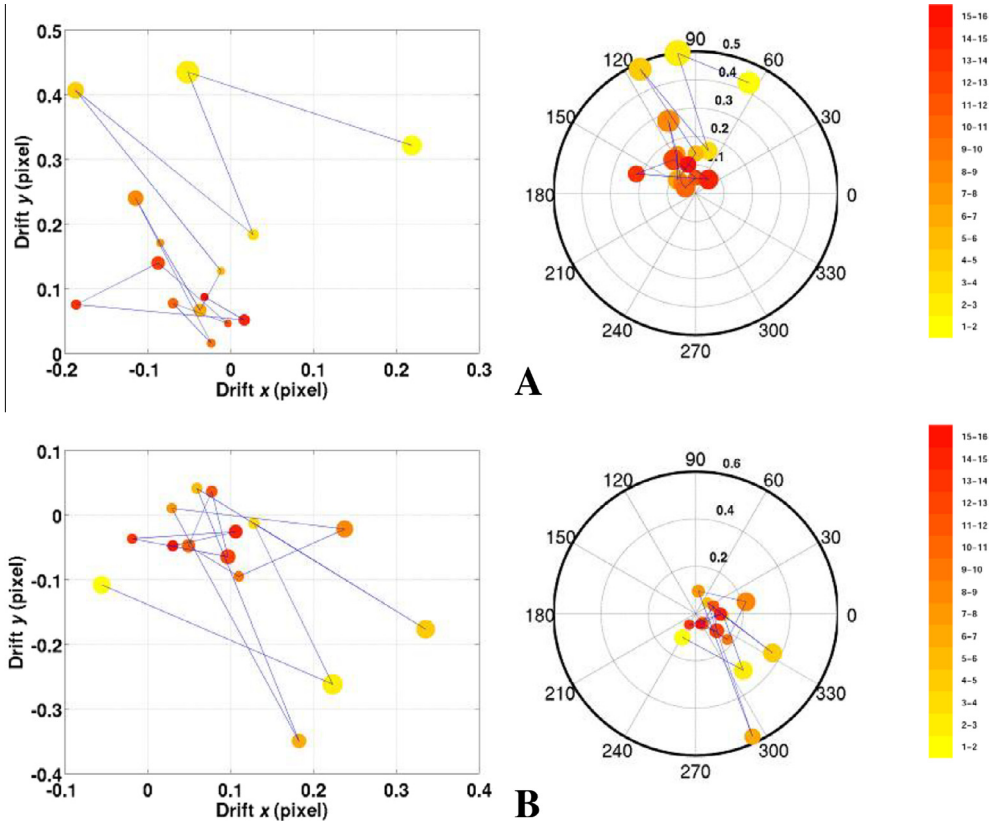

Fig.5. Movement analysis for two videos of Bai et al. (2013): analysis of the movements between consecutive frames of "205_movie_gc_window.mrcs" (A) and "15_movie_gc_window.mrcs" (B). In these Cartesian and polar representations, the displacements correspond to frame shifts as measured in pixels such that the center of each circle corresponds to the mean displacement between two consecutive frames, and the area of each point represents the compound standard deviation in the *x*- and *y*-axes.

movements is presented. Fig. 6C clearly shows that in this stack, the movements between frames are virtually random.

The total processing time for this stack is 5 min and 17 s for OF, 1 min and 25 s for Li Global, and 4 min and 17 s for Local Li.

3.2.3. Archeal 20S proteasome

In this work, we also tested the OF procedure on image stacks recorded using a K2 Summit camera operated in super-resolution counting mode. Images were recorded from frozen-hydrated archaeal 20S proteasome, as described previously (Li et al., 2013a). Briefly, each image stack consists of 24 sub-frames with a pixel size of 1.22 Å. The total exposure time is 5 s. Because it is an electron counting camera, the K2 Summit requires that images are recorded with a low dose rate in the camera, resulting in a long exposure time (Li et al., 2013a). Different from the previously presented images recorded with Falcon II, images recorded with K2 Summit in this work often contain drift-like motion in addition to charge-induced motion.

As in all previous cases, the stacks were directly processed with the newly proposed Optical Flow approach. However, in this case, we obtained results quite similar to those presented in Section 3.1.2 for the case of drift-like movements. In essence, the shape of the envelope function did not improve despite the use of rather large windows. At this stage, we explored the use of the Li et al. (2013b) method as a pre-processing step prior to OF (note that Li method is not perturbed by the possible accumulation of shifts in similar directions, as shown in Section 1.3.2). Furthermore, we incorporated into OF the information on inter-frame global movements provided by the Li et al. (2013b) method, resulting in a combined new method that provides a clear and intuitive tool for representing the complete BIM pattern from the vector fields.

We extended this analysis to the processing of a large set of K2 frame stacks and consistently found that the corresponding BIM pattern had a substantially higher level of global versus local movements. Naturally, for those cases in which the degree of local BIM pattern was higher, the improvements due to the application of OF were more significant. Still, in all cases, the application of OF after the Li method provided a quantitative understanding of the BIM pattern of the stack via analysis of the corresponding vector fields, as presented in previous plots in polar and Cartesian coordinates.

Following the previous reasoning, we first applied the Li Global method to remove global movements, followed by OF to correct for possible remaining local movements. The resulting shapes of the CTF envelope and SSNR functions are shown in Fig. 7A for a window size of 150 pixels, and clearly, the application of OF after Li Global improved the results only slightly. A comparison between the periodograms corresponding to the uncorrected (left) and fully corrected (right) stack average is shown in Fig. 7B. Furthermore, this dataset illustrates the power of the vector fields as a new and convenient diagnostic tool for analysis of BIM patterns. Additionally, in Fig. 7C, we show Cartesian and polar plots in which it is immediately noticeable that there is a strong directionality of the frame-to-frame global movement (the points are concentrated in a small slice in polar coordinates) and that the amount of this movement is larger than in any of the previous cases. Additionally, the area of the circles is notably small, even considering the change of scale in the radial variable required to properly represent larger shifts, further indicating that the ratio of global versus local movements was higher in this case than in previous ones.

Application of Li Global followed by Optical Flow on a typical stack of this dataset required 5 and a half minutes of processing.

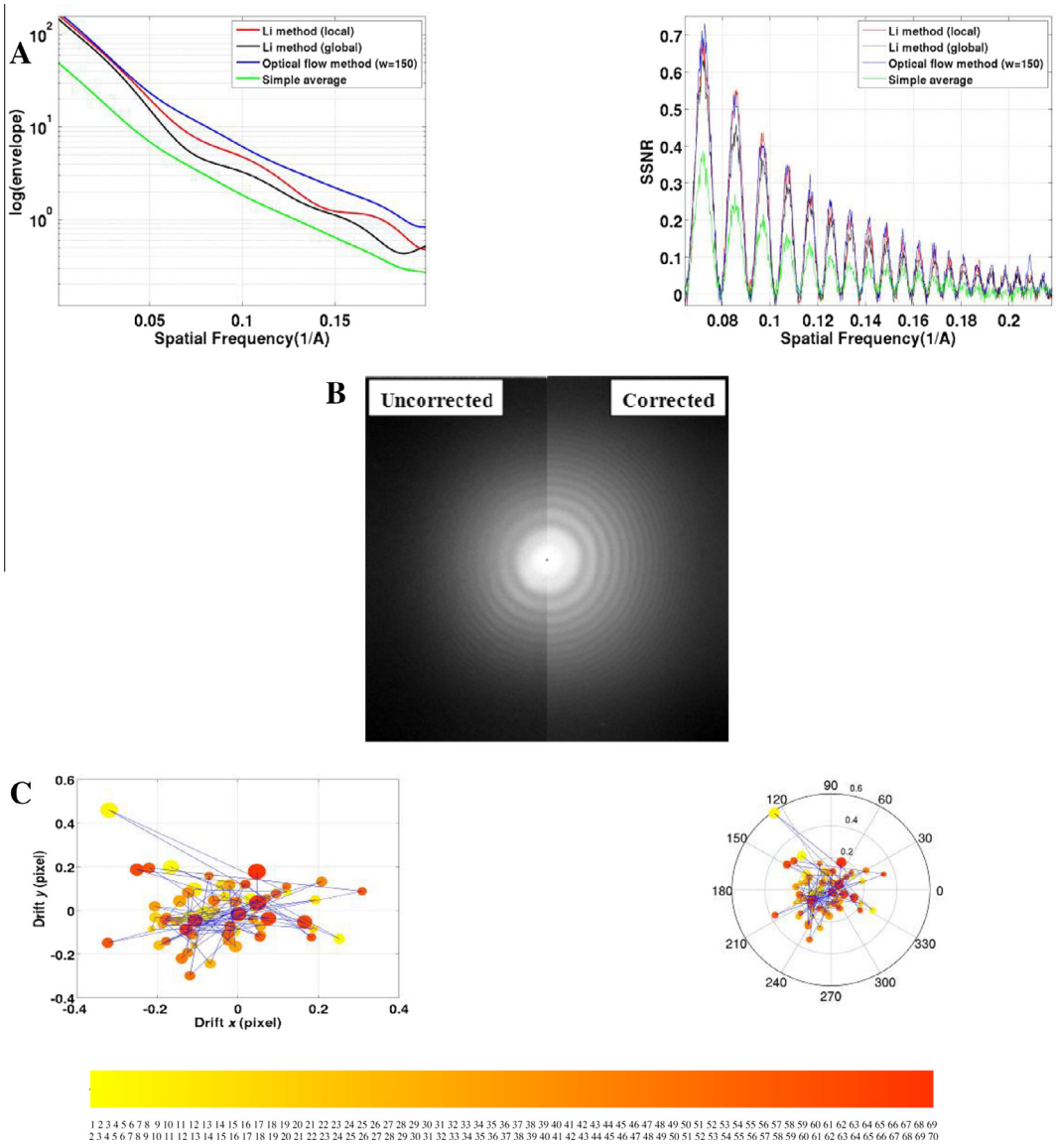

Fig.6. Comparison and analysis of the alignment result for the influenza virus RNP dataset: comparison of the envelopes of the CTF (in logarithmic scale) as well as the SSNR functions corresponding to the average image obtained from the summation of unaligned frames and aligned frames using Li Local, Li Global and Optical Flow with a window size of 150 pixels (A). The periodogram for the uncorrected (left) and corrected averages (right) (B), and the analysis of the movements between consecutive frames (C).

3.2.4. Specimen images presenting fixed-noise patterns

A fixed-noise pattern is present in a non-negligible proportion of the stacks (Li et al., 2013b; Shigematsu and Sigworth, 2013), making pertinent the analysis of the behavior of the newly proposed Optical Flow approach under these circumstances. According to first principles, we expect that a pattern present in the same manner in all frames would pose problems for the Optical Flow approach because it could drive the local movement detection towards the detection of no movement at all. We envisioned two possible solutions: (1) perform a type of filtering to attenuate the fixed-noise pattern or (2) first use Li Global to perform a coarse frame alignment that would "unalign" the fixed-noise pattern between the frames and subsequently apply Optical Flow. We concentrated on the second approach and obtained good results. Fig. 8A presents a comparison of the CTF envelopes and SSNR using Li Global, Li Local and Li Global + OF (window size = 150 pixels), and Fig. 8B shows the periodogram of an uncorrected stack average that presents a clear fixed-noise pattern that is removed by the combined procedure Li Global + OF. Finally, Fig. 8C displays an analysis of the movements between frames of this stack. Note that in cases with strong fixed pattern noise but without a significant global motion in each stack, our approach will not provide good results.

3.3. Selection of the window size parameter

All methods contain parameters, and their selection generally has important effects on the method's behavior. For our Optical Flow approach, and as initially introduced in the mathematical presentation, the most important parameter is the window size. Indeed, the window size represents the (square) neighborhood used to calculate the Optical Flow at that point, and movements are expected to be constant within that window.

The practical approach to calculating the window size for each set of videos has been to first run the algorithm on several videos using different window sizes ranging from 25 to 500 pixels, and second, to choose the window size that provides the highest integral over the 1D profile of the CTF envelope function. As an example, we consid-

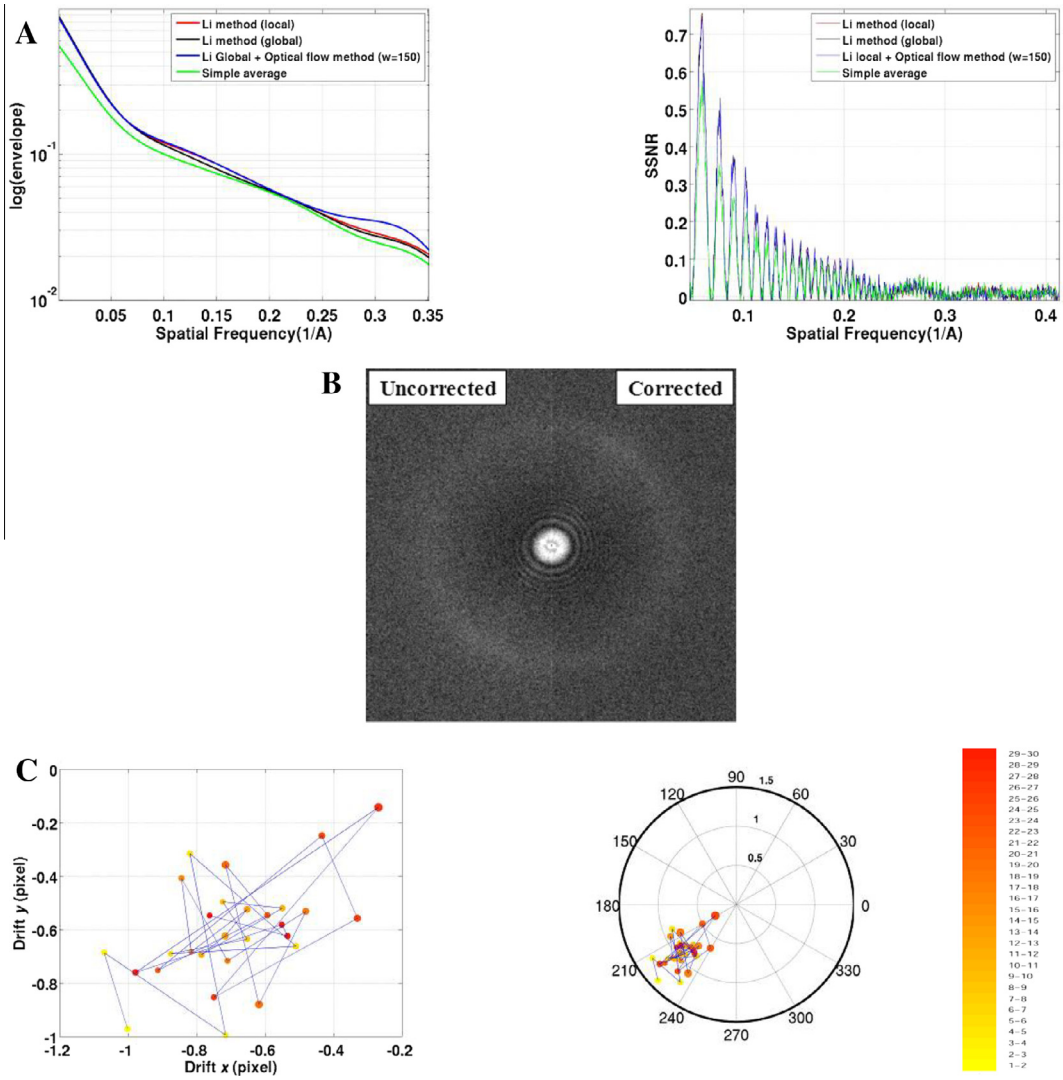

Fig.7. Analysis of 20S proteasome images obtained on a K2: 1D profiles corresponding to the envelope of the CTF function (in logarithmic scale) as well as the SSNR functions for the average image of a representative stack obtained using simple average (green), Li Local (red), Li Global (black) and, finally, Li Global + Optical Flow (blue) with a window size of 150 (A). Comparison of the periodograms of the uncorrected (left) and Li Global + Optical Flow (right) aligned averages (B), and analysis of the movements in between consecutive frames of the stack (C).

ered one micrograph (“205_movie_gc_window.mrcs”) of the dataset from Bai et al. (2013). In Fig. 9, we show the values of this integral for different window sizes. As observed in this figure, the maximum integral value can be obtained for a window size of 150, which is set as the “default” window size of the method because it works well in most practical cases.

4. Discussion and conclusion

Direct detection devices represent a significant step forward for 3DEM, and at the same time, they allow high-resolution and automation to be achieved. These sensors provide sharper images with higher SNRs. Additionally, the rapid image acquisition rates of these DDDs detectors allow us to “see” biological specimens moving in their solid matrix of amorphous ice, a phenomenon referred to as beam-induced motion (BIM). Several approaches have been proposed to correct for these sample movements (Bai et al., 2013; Campbell et al., 2012; Li et al., 2013b; Wang et al., 2014). The method shown in Li et al. (2013b) consists of a frame drift-like correction with a degree of robustness to local movements provided by the use of sub-frame alignment; the output of each processed movie is a corrected averaged micrograph obtained from the aligned frames. In (Bai et al., 2013) the alignment method is quite different because it works with the individual specimens detected and extracted from the frames, as does the method of Scheres (2014). This approach allows in-plane and out-of-plane local movement correction at the expense of requiring individual specimen detection.

In this work, we proposed a frame alignment approach that can correct for local sample movements in a rapid and efficient manner. The approach is based on a robust Optical Flow workflow, which has been tested with simulated and experimental data and compared with the method of Li et al. (2013b) and obtained satisfactory results. The input of the proposed algorithm is a set of unaligned frames (movie), and the output is a corrected averaged micrograph, as in the Li et al. (2013b) approach. The approach assumes that the particle movement is local, invertible and smooth, which in mathematics is known as a diffeomorphism. Note that the proposed model assumes that the displacements between images are equal within a window that typically has a value of approximately 150 pixels.

Additionally, to characterize the BIM pattern in the sample, a movie with the vector field of the movements between frames is also provided, which can be further analyzed using intuitive 2D

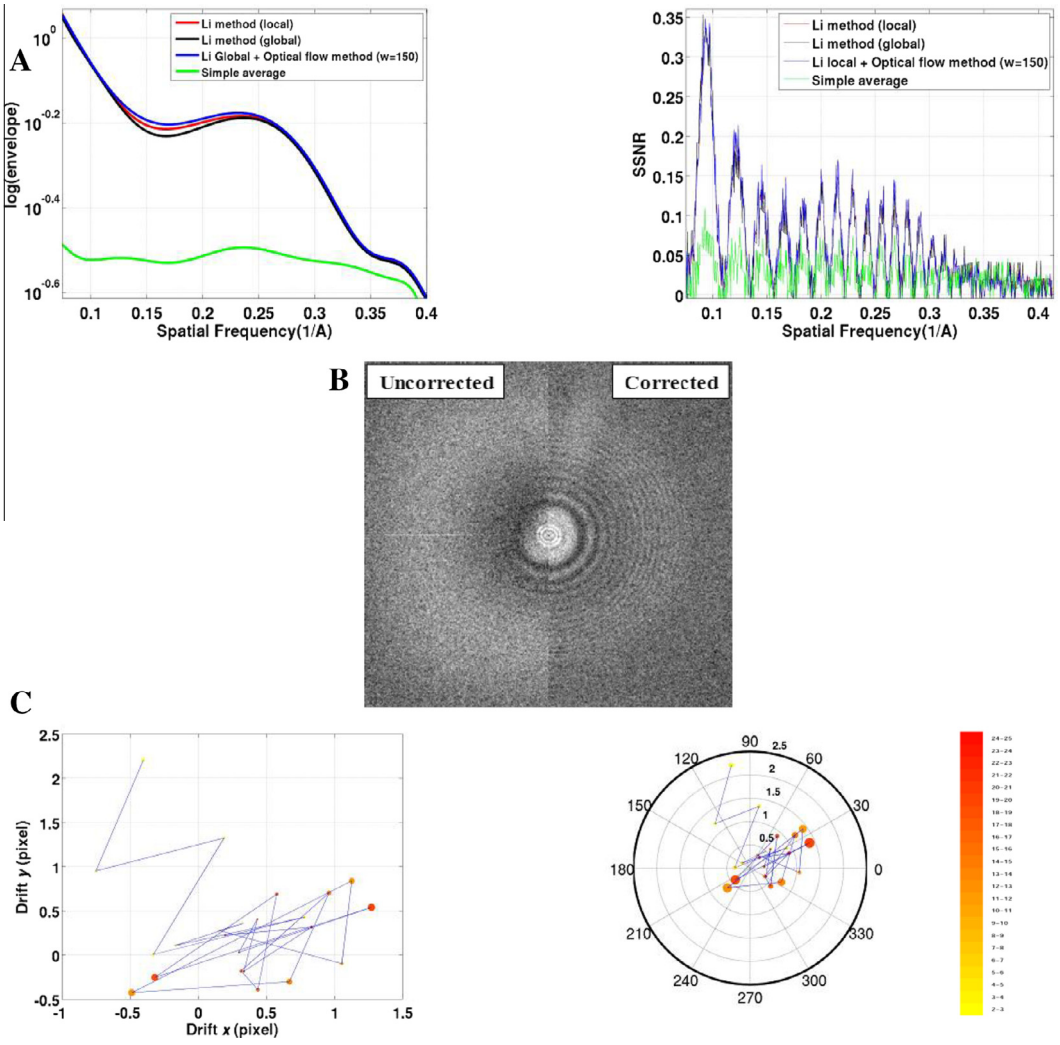

Fig.8. Analysis of fixed-noise pattern case from K2 data: Envelope of the CTF (in logarithmic scale) as well as the SSNR function corresponding to the corrected average after applying simple average (green), Li Local (red), Li Global (black) and Li Global + Optical Flow (blue) with a window size of 150 pixels (A). Comparison of the periodograms of the uncorrected (left) and Li Global + Optical Flow (right) aligned averages (B), and the analysis of the movements between consecutive frames (C).

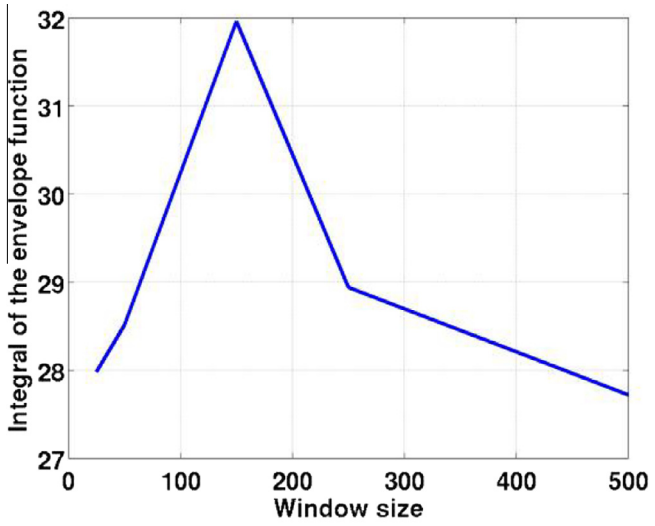

Fig.9. Obtaining the window size for optical flow: integrated values of the 1D profile of the CTF envelope function for different window sizes for the case of "15_movie_gc_window.mrcs". The maximum integral value is obtained for a window size of 150 pixels.

representations that show frame-to-frame movements, where the location of the center of each circle is given by the mean of the movements in the *x*- and *y*-axes and the area is described by the square root of the sum of variances along these axes.

A shortcoming of the proposed method is that it is especially sensitive to the consistency of the frame-to-frame direction of movement. Indeed, if the movements between frames are similar to a unidirectional drift, then the average of the uncorrected frames creates an effect that might translate into averaging of a particle in a frame with background in another frame in the extreme case of small particles and large movements. Still, this effect can be recovered by combining the methods of Li and Optical Flow such that by applying the Li method, we first calculate the global movement between frames, and using the Optical Flow approach, we (1) compensate for local movements and (2) calculate a BIM pattern characterization with the use of the vector fields.

Note that our approach (as well as Li's) corrects for local in-plane movements at the image level, and it cannot incorporate out-of-plane movements (i.e., known in practical terms as rocking). However, the expected extent of out-of-plane rotations and their impact on resolution may not be large, especially for small specimens. Indeed, for a specimen whose outer diameter was 150 Å, rocking of ±2.5 degrees would translate into a 3-Å shift at the periphery. Furthermore, the experience presented in Li et al.

(2013b) indicates that for small specimens, it may be impossible to apply particle-based correction methods (i.e., Bai et al. (2013)) due to the inability to detect the particles in the individual frames, making the approach presented in this work the method of choice to: (1) correct for in-plane local displacements prior to specimen detection, especially for local movements, and (2) characterize the BIM pattern with the aid of the vector fields. In much the same way, for larger particles for which rocking might be a problem, the initial use of OF might provide partially corrected frames that could significantly increase the speed of convergence of any individual specimen image-based correction approach.

In addition to its ability to efficiently correct for local movements, OF also provides a rather direct, simple and intuitive characterization of the BIM pattern with the aid of the vector fields and their associated graphical polar representations. Indeed, this analysis can indicate the amount of global versus local movements as well as its magnitude and directionality. For instance, a simple comparison of the plots shown in Figs. 5–7 shows that the ratio of global versus local movements is much larger for the 20S proteasome recorded on a K2 than for both the ribosomes or the influenza RNPs recorded on a Falcon II. Whether this observation represents a general trend associated with certain imaging conditions cannot be determined from this reduced set, but we are certainly provided with the appropriate tools to perform this type of wide-range analysis in a systematic and clear manner. Still, the accumulated experience gathered in the different laboratories contributing to this work tends to indicate that indeed, the Falcon II data present a larger degree of local distortions than the K2 data, making the Optical Flow approach particularly powerful for Falcon II data sets, whereas for K2 data, the additional improvement is reduced. We observe that this contrasting behavior may not be so unexpected because the detection mechanisms of both cameras are different. The K2 camera operates with low dose rates and large acquisition times, whereas Falcon II works with high doses obtained at short camera acquisition times.

Moreover, new figures of merit have been proposed to compare the different alignment results (uncorrected average, Li-Global and Local, and OF). Indeed, and as presented in the Appendix, the effect of summing the unaligned frames is an additional modulating term in the PSD. Therefore, the contrast transfer function (CTF) envelope and the SSNR functions are sensitive to these between frame misalignments typical of DDDs.

Finally, in future work, we will explore the possibility of using the proposed alignment approach in tomography, thus improving the quality of each tilt-pair image by alignment of the frames composing the movie. Note that this case is especially challenging because of the reduced SNR of these images.

All of the methods presented are publically available as a component of XMIPP (http://xmipp.cnb.csic.es) (de la Rosa-Trevín et al., 2013).

Author contribution

V.A. and J.V. developed the OF method and obtained the results, Y.C. and X.L. provided K2 images and contributed to results analysis, C.S. and R.M. contributed to algorithmic developments, and J.M.C. focused on analysis of results and writing.

Acknowledgments

The authors acknowledge financial 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 CAM (S2010/BMD-2305), and a postdoctoral "Juan de la Cierva" Grant with reference JCI-2011-10185 to Javier Vargas. Vahid Abrishami holds a La Caixa scholarship, and C.O.S. Sorzano is the recipient of a Ramón y Cajal fellowship. This work was partially funded by Instruct, a component of the European Strategy Forum on Research Infrastructures (ESFRI), and supported by national member subscriptions. Critical reading by Joachim Frank is greatly appreciated.

Appendix A

In this Appendix, we analyze the effect of frame misalignment on the PSD of the image resulting from summation of all frames in a stack. Because different noise models could be considered, we differentiate between "deterministic" and "random" noise. In the first case, we show that the envelope of the CTF is a notably good indicator of the extent of misalignment, whereas in the second case, both the envelope of the CTF and the SSNR are good indicators of frame misalignment.

Deterministic noise

In this section, we analyze the effect of frame misalignment for the case in which noise is modeled as a deterministic signal that affects all frames, e.g., the situation for certain of the ice contributions to the frame image. We consider an ideal "micrograph" given by $I(x, y)$. However, $I(x, y)$ is never experimentally measured in a DDD movie; instead, its recording is fractionated into N frames with local and smooth shifts between frame i and $i+1$ given by $g_{x,i}(x, y)$ and $g_{y,i}(x, y)$. Therefore, we can recover $I(x, y)$ using a corrected average of the different frames as:

$$I(x, y) = (1/N)\sum_{i=1}^{N} I(x - g_{x,i}^{-1}(x, y), y - g_{y,i}(x, y)). \tag{A.1}$$

Note that if the shifts are smooth, we consider that they are constant in small image patches. Therefore, if we divide the images into M small pieces, each image patch given by index $m \in [1, M]$ can be calculated as

$$I^m(x, y) = (1/N)\sum_{i=1}^{N} I_i^m(x - x_{0i}^m, y - y_{0i}^m). \tag{A.2}$$

Next, we calculate the Power Spectrum Density (PSD) of each patch

$$\begin{aligned} PSD^m(X, Y) &= \left| (1/N)\sum_{i=1}^{N} FT\left[I_i^m(x - x_{0i}, y - y_{0i})\right] \right|^2 \\ &= \left| (1/N)\xi(X, Y)\sum_{i=1}^{N} e^{-j\left(\frac{2\pi}{N_C}Xx_{0i} + \frac{2\pi}{N_R}Yy_{0i}\right)} \right|^2 \end{aligned} \tag{A.3}$$

where we assumed that $FT\left[I_i^m(x, y)\right] = \xi(X, Y)$, N_C and N_R are the number of columns and rows of the image, respectively, and (X, Y) are the frequency coordinates. If we further develop Eq. (A.3), we obtain

$$\begin{aligned} PSD^m(X, Y) &= \left((1/N)\xi(X, Y)\sum_{i=1}^{N} e^{-j\left(\frac{2\pi}{N_C}Xx_{0i} + \frac{2\pi}{N_R}Yy_{0i}\right)} \right) \\ &\left((1/N)\xi(X, Y)\sum_{i=1}^{N} e^{-j\left(\frac{2\pi}{N_C}Xx_{0i} + \frac{2\pi}{N_R}Yy_{0i}\right)} \right)^* = (1/N^2)|\xi(X, Y)|^2 \\ &\left(N + \sum_{i=1}^{N}\sum_{i'=1, i'\neq i}^{N} e^{j\left(\frac{2\pi}{N_C}(x_{0i'}^m - x_{0i}^m)X + \frac{2\pi}{N_R}(y_{0i'}^m - y_{0i}^m)Y\right)} \right) = (1/N^2)|\xi(X, Y)|^2 \\ &\left(N + 2\sum_{i=1}^{N}\sum_{i'>i}^{N} \cos\left(\frac{2\pi}{N_C}(x_{0i'}^m - x_{0i}^m)X + \frac{2\pi}{N_R}(y_{0i'}^m - y_{0i}^m)Y\right) \right) \end{aligned} \tag{A.4}$$

In Periodogram averaging, we assume that the PSD of the entire micrograph is the average of the PSD of all patches:

$$PSD(X,Y)=(1/M)\sum_{m=1}^{M}PSD^m(X,Y)=1/M\sum_{m=1}^{M}\left((1/N^2)|\xi(X,Y)|^2\left(N+2\sum_{i=1}^{N}\sum_{i'>i}^{N}\cos\left(\frac{2\pi}{N_C}(x_{0i'}^m-x_{0i}^m)X+\frac{2\pi}{N_R}(y_{0i'}^m-y_{0i}^m)Y\right)\right)\right)$$
$$=(1/MN^2)|\xi(X,Y)|^2\sum_{m=1}^{M}\left(\left(N+2\sum_{i=1}^{N}\sum_{i'>i}^{N}\cos\left(\frac{2\pi}{N_C}(x_{0i'}^m-x_{0i}^m)X+\frac{2\pi}{N_R}(y_{0i'}^m-y_{0i}^m)Y\right)\right)\right) \quad (A.5)$$

If there is no shift between frames, the ideal PSD is given by $\widetilde{PSD}(X,Y)=|\xi(X,Y)|^2$.

Thus, we can rewrite the actual PSD (Eq. (A.5)) as:

$$PSD(X,Y)=\Psi(X,Y)\widetilde{PSD}(X,Y) \quad (A.6)$$

where $\Psi(X,Y)$ collects all of the terms that are different from the ideal PSD. Note that if there are no shifts between the different frames, all cosines in Eq. (A.5) are equal to 1, and $\Psi(X,Y)=1$. Otherwise, if there are shifts, then necessarily $0\leqslant\Psi(X,Y)\leqslant 1$. Taking into account that the ideal PSD is given by

$$\widetilde{PSD}=\widetilde{PSD}_b(X,Y)|CTF(X,Y)|^2+\widetilde{PSD}_a(X,Y) \quad (A.7)$$

where $CTF(X,Y)$ corresponds to the CTF, and $\widetilde{PSD}_b(X,Y)$ and $\widetilde{PSD}_a(X,Y)$ refer to the PSD before and after CTF, respectively (Sorzano et al., 2007; Vargas et al., 2013), by substituting Eq. (A.7) into Eq. (A.6), we obtain:

$$PSD(X,Y)=\Psi(X,Y)\left(\widetilde{PSD}_b(X,Y)|CTF(X,Y)|^2+\widetilde{PSD}_a(X,Y)\right)$$
$$=\widetilde{PSD}_b\Psi(X,Y)|CTF(X,Y)|^2+\Psi(X,Y)\widetilde{PSD}_a(X,Y) \quad (A.8)$$

From Eq. (A.8), we can extract several conclusions. The first and most obvious is that the misalignments have a direct effect on the CTF consisting of an additional modulation term. The effective CTF is given by:

$$CTF(X,Y)=\sqrt{\Psi(X,Y)}E(X,Y)\sin\chi(X,Y) \quad (A.9)$$

Moreover, observe that $\Psi(X,Y)$ is not isotropic in general. Additionally, we note from Eq. (A.8) that the misalignment modulation function $\Psi(X,Y)$ affects the CTF envelope and the background noise $\widetilde{PSD}_a$ at the same time, although there is no effect on the CTF phase term $\chi(X,Y)$.. We observe that the SSNR (Booth et al., 2004), which is estimated from the radially averaged CTF envelope divided by the noise curve (the baseline that passes through the zeroes in the CTF) is not adequate for characterizing good or bad frame alignments because it is insensitive to $\Psi(X,Y)$. Consequently, a good approach to analyzing the frame alignment quality is to study the decay in an "effective" CTF envelope function that corresponds to

$$\hat{E}(x,y)=\sqrt{\Psi(X,Y)}E(X,Y) \quad (A.10)$$

In the remainder of the paper, we refer to this "effective" CTF envelope as the CTF envelope

Random noise

We consider this case if noise is modeled as a random event added to each frame. Up to (A.2), there is no difference between the deterministic and the random case. However, when we calculate the PSD of the *m*th patch, we should calculate it in the correct manner for random signals. Let us assume that each patch has a deterministic component $I_{i,d}^m(x,y)$ and a random component $\varepsilon_i^m(x,y)$,

$$I^m(x,y)=(1/N)\sum_{i=1}^{N}I_{i,d}^m(x-x_{0i}^m,\,y-y_{0i}^m)+(1/N)\sum_{i=1}^{N}\varepsilon_i^m(x-x_{0i}^m,\,y-y_{0i}^m). \quad (A.3')$$

Thus

$$PSD^m(X,Y)=\left|(1/N)\sum_{i=1}^{N}FT[I_{i,d}^m(x-x_{0i},\,y-y_{0i})]\right|^2+(1/N)PSD_\varepsilon(X,Y)=(1/N^2)|\xi(X,Y)|^2\left(N+2\sum_{i=1}^{N}\sum_{i'>i}^{N}\cos\left(\frac{2\pi}{N_C}(x_{0i'}^m-x_{0i}^m)X+\frac{2\pi}{N_R}(y_{0i'}^m-y_{0i}^m)Y\right)\right)+(1/N)PSD_\varepsilon(X,Y) \quad (A.4')$$

Next, we calculate the PSD of the entire micrograph

$$PSD(X,Y)=(1/M)\sum_{m=1}^{M}PSD^m(X,Y)=(1/MN^2)|\xi(X,Y)|^2\sum_{m=1}^{M}\left(\left(N+2\sum_{i=1}^{N}\sum_{i'>i}^{N}\cos\left(\frac{2\pi}{N_C}(x_{0i'}^m-x_{0i}^m)X+\frac{2\pi}{N_R}(y_{0i'}^m-y_{0i}^m)Y\right)\right)\right)+(1/N)PSD_\varepsilon(X,Y) \quad (A.5')$$

Compared with the deterministic case, we write

$$PSD(X,Y)=\Psi(X,Y)\widetilde{PSD}(X,Y)+(1/N)PSD_\varepsilon(X,Y)=\widetilde{PSD}_b\Psi(X,Y)|CTF(X,Y)|^2+\Psi(X,Y)\widetilde{PSD}_a(X,Y)+(1/N)PSD_\varepsilon(X,Y) \quad (A.6')$$

The SSNR becomes

$$SSNR(X,Y)=\frac{\widetilde{PSD}_b\Psi(X,Y)|CTF(X,Y)|^2}{\Psi(X,Y)\widetilde{PSD}_a(X,Y)+(1/N)PSD_\varepsilon(X,Y)}=\frac{\widetilde{PSD}_b(X,Y)|CTF(X,Y)|^2}{\widetilde{PSD}_a(X,Y)+\frac{1}{N\Psi(X,Y)}PSD_\varepsilon(X,Y)} \quad (A.7')$$

The degree of misalignment is encoded in the $\Psi(X,Y)$ term. As the misalignment increases, $\Psi(X,Y)$ approaches 0, as does the $SSNR(X,Y)$. In this case, both the envelope and the SSNR serve as good measures of misalignment. As suggested by Wang et al. (2014), note that the noise appears to be a mix of deterministic and random noise for DDD movies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2015.02.001.

References

Bai, X.-C., Fernandez, I.S., McMullan, G., Scheres, S.H., 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife 2 (e0046).

Bammes, B.E., Rochat, R.H., Jakana, J., Chen, D.-H., Chiu, W., 2012. Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. J. Struct. Biol. 177, 589–601.

Booth, C.R., Jiang, W., Baker, M.L., Hong Zhou, Z., Ludtke, S.J., Chiu, W., 2004. A 9 Å single particle reconstruction from CCD captured images on a 200 kV electron cryomicroscope. J. Struct. Biol. 147, 116–127.

Bouguet, J.-Y., 2000. Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the Algorithm.

Bouguet, J.-Y., 2001. Pyramidal Implementation of the Affine Lucas Kanade Feature Tracker Description of the Algorithm, Vol. 2. Intel Corporation (3).

Bradski, G., 2000. {The OpenCV Library}. Dr. Dobb's Journal of Software Tools.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., Grigorieff, N., 2012. Beam-induced motion of vitrified specimen on holey carbon film. J. Struct. Biol. 177, 630–637.

Campbell, M.G., Cheng, A., Brilot, A.F., Moeller, A., Lyumkis, D., Veesler, D., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., 2012. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. Structure 20, 1823–1828.

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

Frank, J., 1996. Three-Dimensional Electron Microscopy of Macromolecular Assemblies. Academic Press.

Frank, J., 2006. Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State. Oxford University Press.

Glaeser, R.M., Hall, R.J., 2011. Reaching the information limit in cryo-EM of biological macromolecules: experimental aspects. Biophys. J. 100, 2331–2337.

Horn, B.K., Schunck, B.G., 1981. Determining optical flow. In: Technical Symposium East. International Society for Optics and Photonics, pp. 319–331.

Jin, L., Bilhorn, R., 2010. Performance of the DDD as a direct electron detector for low dose electron microscopy. Microsc. Microanal. 16, 854–855.

Li, X., Zheng, S.Q., Egami, K., Agard, D.A., Cheng, Y., 2013a. Influence of electron dose rate on electron counting images recorded with the K2 camera. J. Struct. Biol. 184, 251–260.

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

Lucas, B.D., Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. In: IJCAI, Vol. 81, pp. 674–679.

Milazzo, A.-C., Leblanc, P., Duttweiler, F., Jin, L., Bouwer, J.C., Peltier, S., Ellisman, M., Bieser, F., Matis, H.S., Wieman, H., 2005. Active pixel sensor array as a detector for electron microscopy. Ultramicroscopy 104, 152–159.

Scheres, S.H., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530.

Scheres, S.H., 2014. Beam-induced motion correction for sub-megadalton cryo-EM particles. eLife 3 (e03665).

Shigematsu, H., Sigworth, F., 2013. Noise models and cryo-EM drift correction with a direct-electron camera. Ultramicroscopy 131, 61–69.

Sorzano, C., Jonic, S., Núñez-Ramírez, R., Boisset, N., Carazo, J., 2007. Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. J. Struct. Biol. 160, 249–262.

Vargas, J., Otón, J., Marabini, R., Jonic, S., De La Rosa-Trevín, J., Carazo, J., Sorzano, C., 2013. FASTDEF: Fast defocus and astigmatism estimation for high-throughput transmission electron microscopy. J. Struct. Biol. 181, 136–148.

Vargas, J., Restrepo, R., Belenguer, T., 2014. Shack-Hartmann spot dislocation map determination using an optical flow method. Opt. Express 22, 1319–1329.

Wang, Z., Hryc, C.F., Bammes, B., Afonine, P.V., Jakana, J., Chen, D.-H., Liu, X., Baker, M.L., Kao, C., Ludtke, S.J., 2014. An atomic model of brome mosaic virus using direct electron detection and real-space optimization. Nat. Commun. 5.