



Conditional Random Fields for tubulin-microtubule segmentation in cryo-electron tomography

Charles Kervrann, Sophie Blestel, Denis Chrétien

► To cite this version:

Charles Kervrann, Sophie Blestel, Denis Chrétien. Conditional Random Fields for tubulin-microtubule segmentation in cryo-electron tomography. IEEE International Conference on Image Processing (top 10%” papers in ICIP 2014), Oct 2014, Paris, France. pp.2080 - 2084, 10.1109/ICIP.2014.7025417 . hal-01103330

HAL Id: hal-01103330

<https://hal.inria.fr/hal-01103330>

Submitted on 23 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONDITIONAL RANDOM FIELDS FOR TUBULIN-MICROTUBULE SEGMENTATION IN CRYO-ELECTRON TOMOGRAPHY

Charles Kervrann, Sophie Blestel

Denis Chrétien

Inria Rennes - Bretagne Atlantique, Serpico Team
Campus universitaire de Beaulieu
35 042 Rennes Cedex, France

UMR 6290 IGDR, CNRS - UR1, TIPs Team
Campus universitaire de Beaulieu
35 042 Rennes Cedex, France

ABSTRACT

Cryo-electron tomography allows 3D observation of biological specimens in their native and hydrated state at high spatial resolution (4-5 nanometers). Traditionally cryo-tomograms have very low signal-to-noise ratios and conventional image segmentation methods are limited yet. In this paper, we formulate the segmentation problem of both small tubulin aggregates and microtubules against the background as a two class labeling problem in the Conditional Random Field framework. In our approach, we exploit image patches to take into account spatial contexts and to improve robustness to noise. Because of the contrast anisotropy in the specimen thickness direction, each 2D section of the 3D tomogram is segmented separately with an optional update of reference patches. This method is evaluated on synthetic data and on cryo-electron tomograms of *in vitro* microtubules.

Index Terms— Cryo-electron tomography, segmentation, microtubules, Conditional Random Fields.

1. INTRODUCTION

Microtubules (MTs) are nanometer “tubes” made of the heterodimeric tubulin protein, also present in all eukaryotic cells. These components of the cytoskeleton as actin filaments are involved in many cellular processes (mitosis, vesicular transport, ...), and the understanding of their structure and assembly mechanisms is of major importance [8]. MTs can be polymerized *in vitro* from purified tubulin, and observed in their native state by cryo-electron tomography (cryo-ET) (see Fig.1). However 3D visualization of electron tomograms (ET) is hampered by noise [9, 12]. Low signal-to-noise ratios and low contrast of tomograms are due to several factors: i) low electron doses required to limit specimen damages; ii) close electron densities of the water and of the atoms that constitute the observed specimens; iii) reconstruction of tomograms from a limited number of 2D projection views with a limited angular range (“missing wedge” effect)[11].

Segmentation usually improves 3D visualization and allows interpretation of tomograms. Manual segmentation is

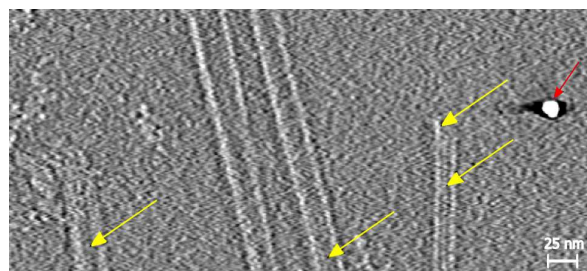


Fig. 1. 2D section of a 3D tomogram in cryo-ET showing low contrasted microtubules. Yellow arrows point to MT protofilaments (low atomic mass) and the red arrow to a highly contrasted gold particle (heavy atomic mass).

thus still employed and softwares have been developed to facilitate supervised segmentation [14, 20, 16]. Nevertheless manual segmentation is not satisfactory since it is largely subjective, very time consuming, and does not allow comparative analysis of many specimens. Several semi-automatic image segmentation methods have therefore been investigated specifically to segment electron microscope tomograms [10, 23, 19, 18] (see [22] for an overview). A commonly-used approach is the immersive watershed algorithm [21, 2]. To limit over-segmentation due to noise, boundary smoothness and shape priors have been proposed as well [23][19]. Other conventional approaches in image segmentation amount to detecting ridges from local structural information [17] or minimizing an energy functional established in the continuous [3] or discrete setting [10]. The energy is usually designed as a combination of several terms, each of them corresponding to a precise property which must be satisfied by the optimal solution [22].

In our study, segmentation of isolated tubulin aggregates and microtubules against the background is defined as a two class labeling problem. Segmentation is then achieved by computing the Maximum a Posteriori (MAP) estimation which amounts to minimizing a global energy functional. We propose an original data fidelity term based on a distance between patches [6] and we consider generalized Ising mod-

els for regularization [4]. Segmentation is then section by section, with an optional update of the reference patches used to represent two classes: “object” (tubulin aggregates and rod-like structures) and “background”.

The remainder of the paper is organized as follows. In Section 2 and Section 3, we present respectively the statistical modeling framework and our 3D segmentation method. The performance of the algorithm on real and synthetic data is demonstrated in Section 4.

2. SEGMENTATION AND CRF-MAP INFERENCE

This section describes the probabilistic model that forms the basis our work. Let $\mathbf{f} = \{f_s\}_{s \in \mathcal{S}}$ be the noisy observation data where $f_s \in \mathbb{R}$ denotes the observation at site s and $\mathcal{S} \subset \mathbb{Z}^2$ is the set of sites in the image, that is the regular grid of pixels. In our modeling approach, the background-object separation problem is expressed as a binary labeling problem, i.e. $u_s \in \{B, O\}$ where “B” and “O” denote “background” and “object” respectively. The random field is said to be a CRF with respect to a neighborhood if and only if it satisfies the Markovian property: $P(u_s | \mathbf{f}, \mathbf{u}_{\mathcal{S} \setminus s}) = P(u_s | \mathbf{f}, \mathbf{u}_{\partial s})$, where $\mathcal{S} \setminus s$ is the set of all pixels except the pixel s , ∂s denotes the set of neighbor pixels of s and $\mathbf{u}_{\partial s} = \{u_t : t \in \partial s\}$.

CRF, especially designed to directly model the *posterior* distribution [15], is known to be more flexible than usual Markov Random Fields (MRF) since it enables the capture of non local interactions between pixels and the mix of observable and hidden variables. Let $P(\mathbf{u} | \mathbf{f})$ be the conditional probability of the set of class assignments \mathbf{u} . According to the Markov-Gibbs equivalence, we have $P(\mathbf{u} | \mathbf{f}) = \frac{1}{Z} \exp -E(\mathbf{u}, \mathbf{f})$ where Z is the partition function and $E(\mathbf{u}, \mathbf{f})$ is the energy functional written in terms of unary and pairwise terms:

$$E(\mathbf{u}, \mathbf{f}) = \sum_{s \in \mathcal{S}} D(u_s, \mathbf{f}) + \sum_{s \in \mathcal{S}} \sum_{t \in \partial_s} I(u_s, u_t, \mathbf{f}). \quad (1)$$

The unary potential $D(u_s, \mathbf{f})$ is a data fidelity potential used for object detection and the potential $I(u_s, u_t, \mathbf{f})$ is an interaction term used to impose some regularity to the segmentation map. In CRF the unary potential at site s is a function of all the observation data \mathbf{f} as well as that of the label u_s . In addition, the pairwise potential for each pair s and t is a function of all data \mathbf{f} as well as of the labels u_s and $u_t, t \in \partial_s$. Note that, in MRF the pairwise potential represents prior and is independent of the observation [13]. The Maximum a Posteriori (MAP)-CRF estimation of \mathbf{u} is defined as: $\hat{\mathbf{u}} = \arg \max_{\mathbf{u}} P(\mathbf{u} | \mathbf{f}) = \arg \min_{\mathbf{u}} E(\mathbf{u}, \mathbf{f})$ and a global optimization algorithm is needed for computation.

3. CRF IN CRYO-ELECTRON TOMOGRAPHY

Segmentation of cryo-electron tomograms is a hard task due to the very low signal-to noise ratios, the average inten-

sity that varies from one section to another and the spatial anisotropy along the axial direction (specimen thickness direction). Since the spatial artifacts and contrast changes vary especially along the axial direction, it is more appropriate to process individually each 2D section of the 3D tomogram. We exploit information obtained from previous segmented sections to analyze the subsequent sections.

As cryo-ET images have very low signal-to-noise ratios, we consider a data fidelity potential based on a contextual dissimilarity measure between 2D image patches. Patches are pictorial elements (small sub-images) that enable to take into account semi-local spatial contexts in the 2D image [13]. In our model, the two classes of interest are characterized by several square patches of size $\sqrt{n} \times \sqrt{n}$ without loss of generality but isotropic circular patches could be considered as well. We define a potential $D(\cdot)$ that expresses the amount of similarity between the observed patch and the patches of the two classes of interest (“background” and “object”) as sketched in Fig. 2. The potential $D(\cdot)$ is defined as the Euclidean distance between vectorized image patches [6]:

$$D(u_s, \mathbf{f}) = \|\mathbf{z}_s(\mathbf{f}) - \mathbf{z}_s(u_s, \mathbf{f})\|_2^2 \quad (2)$$

where $\mathbf{z}_s(\mathbf{f})$ is the n -dimensional observed patch centered at pixel s , and $\mathbf{z}_s(u_s, \mathbf{f})$ is the patch that represents the class u_s at pixel s : $u_s = O$ for the “object” class and $u_s = B$ for the “background” class. In the following, the $\sqrt{n} \times \sqrt{n}$ observed patch centered at pixel s is denoted as: $\mathbf{z}_s(\mathbf{f})[t] \triangleq f_{s+t}, \forall t \in \left\{-\frac{\sqrt{n}-1}{2}, \dots, \frac{\sqrt{n}-1}{2}\right\}^2$. The reference patches are “learned” from the observations as explained below.

The regularization potential takes the form of a generalized Ising model and depending on the observation \mathbf{f} . This potential penalizes isolated pixels and imposes some regularization to the segmentation map [4]:

$$I(u_s, u_t, \mathbf{f}) = \beta (1 - \mathbb{1}[u_s = u_t]) \left(1 + \frac{e^{-(f_s - f_t)^2 / \alpha}}{\|s - t\|}\right) \quad (3)$$

where β is a positive constant and $\mathbb{1}[\cdot]$ is the indicator function: $\mathbb{1}[u_s = u_t] = 1$ if $u_s = u_t$ and 0 otherwise. In our modeling, we considered a 4-neighbor system: $t \in \partial_s$ if $\|s - t\| = 1$ and $\|\cdot\|$ denotes the Euclidean norm. The prior-contrast potential (3) mixes observed variables and hidden variables and represents a combination of an Ising prior for labeling spatial coherence together with a contrast term that acts to discount partially the coherence and to preserve edges. In this modeling, α enables to deal with noise and intra-class intensity variability. We defined α as $\alpha = \langle (f_s - f_t)^2 \rangle$ where $\langle \cdot \rangle$ denotes the mean over all the cliques of the image. In practice, the regularization parameter β is unchanged and set to 0.05. The MAP estimation is obtained by minimization of (1) wrt \mathbf{u} using a Min Cut/Max Flow algorithm [5].

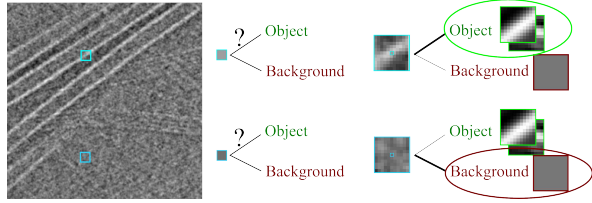


Fig. 2. At each pixel s , the current patch $z_s(\mathbf{f})$ (blue boxes) is compared to the sets of patches that represent the “object” and “background” classes.

3.1. “Background” and “object” reference patches.

As illustrated in Fig. 3, the “background” is a 2D slowly spatially varying intensity image assumed to be corrupted by an additive noise. The “background” reference patch $z_s(u_s = B, \mathbf{f})$ is thus location dependent. Therefore we define $z_s(u_s = B, \mathbf{f})$ as a nearly constant intensity patch (Fig. 3 (b)) whose intensity is equal to f_s^σ : $z_s(u_s = B, \mathbf{f})[t] \triangleq f_s^\sigma, \forall t \in \left\{-\frac{\sqrt{n}-1}{2}, \dots, \frac{\sqrt{n}-1}{2}\right\}^2$ where $f^\sigma = \{f_s^\sigma, s \in \mathcal{S}\}$ is the input 2D section blurred using a Gaussian filter having a large variance σ^2 with respect to the MT width. The Gaussian filter is a low-pass filter that only selects low frequencies. In our case, the microtubules are then oversmoothed and the slowly varying intensities of the background are unchanged.

In our approach, we propose to use a bag of patches picked up on the MT walls. The patches encode some geometric information about the tubulin aggregates and rod-like structures that form protofilaments in a 2D section. To reduce dimensionality while representing the “object” class, we form a set of N patches $\mathcal{Z}_O = \{\tilde{z}_1, \dots, \tilde{z}_N\}$ where \tilde{z}_l denotes a “mean” patch computed from a bag of patches selected along a curve $L_l, l = \{1, \dots, N\}$ passing through a MT wall. Formally, we have $\tilde{z}_l = \frac{1}{|L_l|} \sum_{s \in L_l} z_s(\mathbf{f})$, where $|L_l|$ is the number of pixels belonging to the curve L_l . Such curve can be semi-automatically extracted using the *Fast-Marching* algorithm [7], which enables to find the shortest path between two extremity points in the sense of a geodesic metric dependent on the input 2D image (Fig. 3(a)).

At a given pixel s the “object” patch $z_s(u_s = O, \mathbf{f})$ used in $D(\cdot)$ is the patch in \mathcal{Z}_O whose distance is the lowest, that is $z_s(u_s = O, \mathbf{f}) = \arg \min_{\tilde{z}_l \in \mathcal{Z}_O} \|z_s(\mathbf{f}) - \tilde{z}_l\|_2^2$. The pixels labeled as “object” ($u_s = O$) correspond both to isolated tubulin aggregates and MTs. Considering the same set of patches to represent both MTs and isolated tubulin aggregates is justified by the fact that they are made of tubulin and related intensity in cryo-ET images is similar.

3.2. Segmentation of 3D tomograms

The algorithm is initialized by manual selection of several couples of points on the MT walls. Small patches enable the detection of more MT walls but also isolated small tubulin

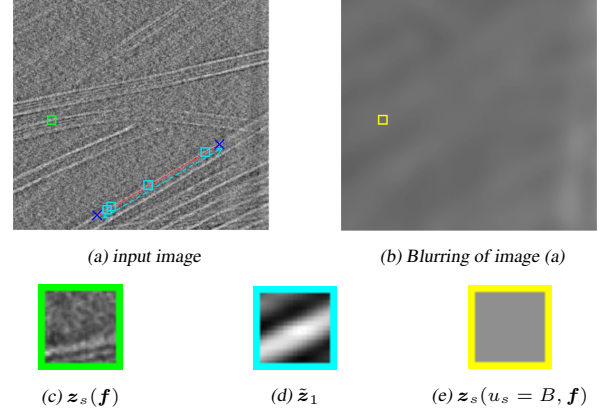


Fig. 3. Design of the “background” and “object” reference patches: the user selects two points (blue crosses) in a MT wall and a curve is extracted using the *Fast-Marching* algorithm [7]. An “object” reference patch (d) is computed from the set of patches selected along the computed curve. The “background” reference patch (e) at pixel s is a nearly constant intensity patch (blurring of the input image (a) using a Gaussian filter). A typical patch is shown in (c).

aggregates of interest. The minimization of $E(\mathbf{u}, \mathbf{f})$ provides a binary image. It turns out that large patches are more robust to noise but miss low contrasted parts of microtubules. By increasing β , isolated pixels are removed.

To extend the 2D method to the segmentation of a 3D tomogram, we segment separately each 2D section of the 3D tomogram because of contrast changes along the axial direction. We initialize the algorithm by selecting manually couples of points on one or a few arbitrarily chosen sections. To avoid a manual and tedious initialization that would consist in selecting landmarks on every MT walls and on each section, we detect automatically new landmarks on the subsequent sections as follows. A MT wall is approximately a long rod-like structure when they are observed longitudinally. A linear MT pattern on section $k + 1$ is detected by applying the Hough transform to the binary map estimated on the previous section k . We assume that the MT walls correspond to the longest lines in the segmented image, so they are detected in the Hough transform of the image by extracting the peaks of maximal value. Finally, the set \mathcal{Z}_O of “object” patches can be updated by collecting additional patches in the subsequent sections as performed for the first section.

4. EXPERIMENTAL RESULTS

The method is first demonstrated on synthetic tomograms obtained from a 3D density map of a microtubule. First, 2D projections of the 3D map are computed, with angles ranging from -60° to $+60^\circ$ with an increment of 2° [11]. To constitute ground truth, the projections are retro-projected and a simple thresholding is applied to the 3D reconstruction. Noisy tomograms are simulated by adding white Gaussian noise to the 2D projections with different standard devia-

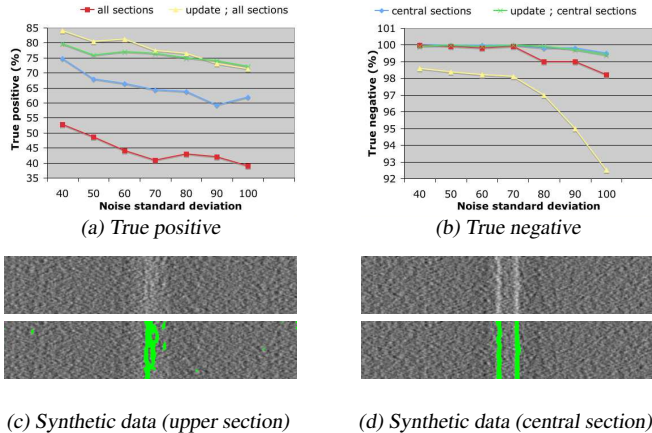


Fig. 4. Evaluation of the 3D segmentation method on a ground truth for different values of noise: (a)-(b): The two graphs display respectively the true positives and the true negatives of the 3D segmentation method, with and without update of the “object” patch set; (c)-(d): Two slices of the tomogram are represented (noise of standard deviation 70) (top) with the overprinted segmentations (bottom).

tions; since the contrast of the projections is 60, we add noise with standard deviations ranging from 40 to 100 intensity values. To simulate correlated noise observed in real cryo-ET images, the noisy projections are convolved with a Gaussian kernel of standard deviation 1 pixel. The tomograms are finally obtained by applying a weighted back-projection to the noisy projections. The two graphs in Fig. 4(a) and Fig. 4(b) display respectively the true positives (object pixels labeled as “object”) and the true negatives (background pixels labeled as “background”) of the automatic segmentation of the noisy tomograms. Updating the “object” patch set improves significantly the segmentation results and the method is quite robust to several signal-to-noise ratios (SNR). Nevertheless, the number of background pixels labeled as “background” decreases for low SNRs if we consider all the sections. This suggests that the method cannot correctly distinguish isolated tubulin aggregates from structures corresponding to correlated noise. More supervision is then required to reduce false detections in the foreground.

At present, about 80 views on several single tilt series were acquired over an angular range of -60° to $+60^\circ$ in very low electron doses conditions using a 200 kV LaB6 electron microscope equipped with a $2k \times 2k$ CCD camera. Specimens were vitrified in liquid ethane using a home made guillotine device, under controlled humidity and temperature conditions. Three-dimensional reconstructions were performed using the IMOD software and were visualized with UCSF Chimera. In this study, cryo-ET tomography of microtubules assembled *in vitro* from pure tubulin [8] reveal 3D features such as the outwardly curved shape of the growing microtubule ends, tubulin molecules trapped in the lumen during assembly (Fig. 5 (left)), or dislocations inside the lattice (Fig. 5 (right)). The algorithm is able to detect both tubular-shaped

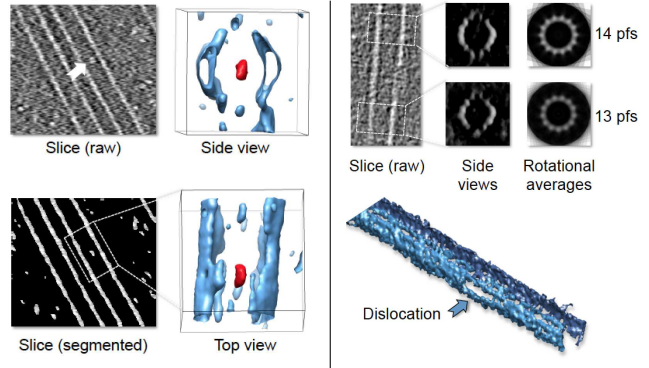


Fig. 5. Tubulin trapped (in red) inside MT during assembly (left); dislocation of protofilament number transitions (14 to 13 pfs) (right).

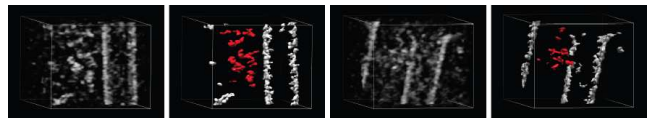


Fig. 6. Extraction of MAP globular domains (in red) and MTs inside the tomogram with our segmentation method (see text).

structures and isolated protein aggregates and the user can select manually the connected components of interest (in red in Fig. 5). For a set $\mathcal{Z}_O = \{\tilde{z}_1, \tilde{z}_2, \tilde{z}_3\}$ of “mean” 15×15 patches, processing a 1024×1024 section is lower than 3 mn.

5. CONCLUSION AND PERSPECTIVES

We have proposed a CRF-MAP estimation framework and a user-friendly algorithm for the segmentation of 2D sections of cryo-ET. The semi-automatic segmentation of 3D tomograms is achieved by performing the 2D method section by section and by updating the reference patches if the contrast changes abruptly from one section to another one. The decision is made by the user and depends on the acquisition conditions. The user needs only to specify the patches of interest for training. Comparisons with competitive methods have been performed but they need actually to be customized as we did to cope with contrast changes along the axial direction.

The proposed method is now used routinely by biologists to study more complex macromolecular assemblies. As illustrated in Fig. 6, cryo-ET allows typically visualization of a special class of Microtubule Associated Proteins (MAP) that remain attached to the growing ends of microtubules. It turns out that the specific structure that must be present at microtubule growing ends and which is recognized by the MAPs remains unknown yet. Comparison of samples prepared in the absence and presence of MAPs using the segmentation method allows an unambiguous localization of the protein densities inside the cryo-electron tomograms.

6. REFERENCES

- [1] B. Bajaj, Z. Yu, and M. Auer, "Volumetric feature extraction and visualization of tomographic molecular imaging," *J. Struct. Biol.*, vol. 144, pp. 132–143, 2003.
- [2] L.A. Baker and J.L. Rubinstein, "Edged watershed segmentation: a semi-interactive algorithm for segmentation of low-resolution from electron cryomicroscopy," *J. Struct. Biol.*, vol. 176, pp. 127–132, 2011.
- [3] A. Bartesaghi, G. Sapiro, and S. Subramaniam, "An energy-based three-dimensional segmentation approach for the quantitative interpretation of electron tomograms," *IEEE T. Image Proc.*, vol. 14, pp. 1314–1323, 2005.
- [4] Y. Boykov and M.P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D Images," in *Proc. Int. Conf. Computer Vision (ICCV'01)*, Vancouver, Canada, pp. 105–112, 2001.
- [5] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE T. Patt. Anal. Mach. Intell.*, vol. 26, pp. 1124–1137, 2004.
- [6] A. Buades, B. Coll, J.M. Morel, "A review of image denoising methods, with a new one," *SIAM J. Mult. Model. Sim.*, vol. 4, pp. 490–530, 2005.
- [7] L. Cohen and R. Kimmel, "Global minimum for active contour models: A minimal path approach," *Int. J. Comp. Vis.*, vol. 24, pp. 57–78, 1997.
- [8] F. Coquelle, S. Blestel, C. Heichette, I. Arnal, C. Kervrann, D. Chrétien, "Cryo-Electron Tomography of microtubules assembled in vitro from purified components," *Meth. Mol. Biol.*, vol. 777, pp. 193–208, 2011.
- [9] A. Frangakis and R. Hegerl, "Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion," *J. Struct. Biol.*, Vol. 135, pp. 239–250, 2001.
- [10] A.S. Frangakis and R. Hegerl, "Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis," *J. Struct. Biol.*, vol. 138, pp. 105–113, 2002.
- [11] A. Guesdon, S. Blestel, C. Kervrann and D. Chrétien, "Single versus dual-axis cryo-electron tomography of microtubules assembled in vitro: Limits and perspectives," *J. Struct. Biol.*, vol. 181, pp. 169–178, 2013.
- [12] P. van der Heide, X.-P. Xu, B.J. Marsh, D. Hanein and N. Volkman, "Efficient automatic noise reduction of electron tomography reconstructions based on iterative median filtering," *J. Struct. Biol.*, vol. 158, pp. 196–204, 2007.
- [13] C. Kervrann, F. Heitz, "A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics". *IEEE T. Image Process.*, vol. 4, pp. 856–862, 1995.
- [14] J. R. Kremer, D. N. Mastrorade, and J. R. McIntosh, "Computer visualization of three-dimensional image data using IMOD," *J. Struct. Biol.*, vol. 116, pp. 71–76, 1996.
- [15] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. Int. Conf. Machine Learning*, pp. 282–289, 2001.
- [16] M. Marko and A. Leith, "Stereo three-dimensional reconstructions from stereoscopic contouring," *J. Struct. Biol.*, vol. 116, pp. 93–98, 1996.
- [17] A. Martinez-Sanchez, I. Garcia and J.-J. Fernandez, "A ridge framework for segmentation of 3D in electron microscopy datasets," *J. Struct. Biol.*, vol. 181, pp. 61–70, 2013.
- [18] F. Moussavi, G. Heitz, F. Amat, L.R. Comolli, D. Koller, M. Horowitz, "3D segmentation of cells boundaries from whole cell cryogenic electron tomography volumes," *J. Struct. Biol.*, vol. 170, pp. 134–145, 2010.
- [19] H. Nguyen and Q. Ji, "Shape-driven three-dimensional watersnake segmentation of biological membranes in electron tomography," *IEEE T. Med. Imaging*, vol. 27, pp. 616–628, 2008.
- [20] G. Perkins, C. Renken, M. E. Martone, S. J. Young, M. Ellisman, and T. Frey, "Electron tomography of neuronal mitochondria: three-dimensional structure and organization of cristae and membrane contacts," *J. Struct. Biol.*, vol. 119, pp. 260–272, 1997.
- [21] G.D. Pintilie, J. Zhang, T. Goddard, W. Chiu and D.C. Gossard, "Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering and fitting of structures by alignment to regions," *J. Struct. Biol.*, vol. 170, pp. 427–438, 2010.
- [22] K. Sandberg, "Methods for image segmentation in cellular tomography," *Meth. Cell Biol.*, vol. 79, pp. 769–798, 2007.
- [23] N. Volkman, "A novel three-dimensional variant of the watershed transform for segmentation of electron density maps," *J. Struct. Biol.*, vol. 138, pp. 123–129, 2002.