A Novel Method to Detect Microtubules in High-Noisy Cryo-EM Micrographs

Huixin Tong¹, Linhua Jiang*¹, Songwen Pei¹, Wenjie Chen², Xiao Lin¹, Guanhua Zhang², Haibin Cai*²

¹Shanghai Key Lab of Modern Optical Systems, University of Shanghai for Science and Technology, Shanghai, 200093, China
²Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, 200062, China

*Corresponding author: L.Jiang(lhjiang@ust.edu.cn) ; H.Cai(hbcai@sei.ecnu.edu.cn)

Abstract—Microtubule (MT), a major component of the cytoskeleton, becomes a hot research topic in the field of structural biology in recent years. Cryo-Electron Microscopy (Cryo-EM) is an effective tool to observe the nano-structure of microtubule, while segmentation and recognition of the micrographs of MTs are mostly done manually because of the extreme noises and massive shade of other cell components in Cryo-EM micrographs. Here we present a novel automated method to segment and detect microtubules based on morphological image operation and Meanshift clustering algorithm combining with a tubular enhancement filter. The purpose is to determine the location and the shape of the specific microtubular structure. Our tests verified that the new algorithm can effectively identify the microtubules in the high-noisy Cryo-EM micrographs.

Keywords—morphological operation; Meanshift; high-noisy; Cryo-EM micrograph; microtubules;

I. INTRODUCTION

As a component of the cell cytoskeleton, microtubules are rope-like polymers of tubulin. It plays an important role in maintaining of cell structure, providing platform for intracellular transport, forming the spindle together with other proteins, etc [1] [2]. The study of the function and structure of microtubule is of great scientific significance. For example, it can be judged whether the patient has cancer or not, according to the function and morphology of the microtubule.

Cryo-Electron Microscopy (Cryo-EM) allows scientists observe the nature of cell straightforwardly in the scale of 6-7 nanometers [4]. However, Cryo-EM micrographs are usually highly affected by serious noise while the signal to noise ratio is rather low (less than 0.1). Generally speaking, Cryo-EM micrographs have the following characteristics: (1) low contrast and there are massive noises in the image. (2) Interior components in Cryo-EM samples are complex, not only the target objects, but also the presence of ice, broken particles, etc; (3) Contours of microtubules are not obvious, and some target objects do not have clear and complete contours. In addition, there is no regular texture in the region of the microtubules. Together with sharply varied features and obscure contours of the target objects, all these difficulties greatly hinder the digital processing of Cryo-EM micrograph. Therefore, manual recognition and segmentation of electron micrographs are still prevalent in recent years. Nevertheless, along with continuously research, the amount of collected data of micrographs are increasing rapidly. Manual work is obviously time-consuming and inaccurate and can’t fulfill the requirement.

Scientists recently use computer and digital processing methods to analyze Cryo-EM micrographs. A template-matching method to identify target objects is presented by Bohm J [5], particles are identified by cross-correlation by using a set of three-dimensional templates. Achilleas and Reiner Hegerl [6] firstly introduced Nonlinear Anisotropic Diffusion Filter. Different to the traditional filtering techniques it tends to weaken boarders and contours. Their results are exciting because nonlinear solution enables obvious boarders of information in images to be preserved, while noises within the contours are eliminated. Furthermore, Frangakis [7] used wavelet filters to enhance the high noisy tomography photos. M. Kass, A. Witkin, and D. Terzopoulos [8] put forth a revised Snake method to identify target objects.

Although the works mentioned above provide efficient methods to pick out targets from Cryo-EM micrographs, almost all of them are to segment single particles instead of long microtubules. In my micrographs, microtubules are the major information, while contaminants from other cellular organs and noises take up relatively larger areas than the microtubules. It is worth mentioning that Wen yan put forward a new method to detect the microtubules in Cryo-EM micrographs in his paper [4].

Fig. 1. A: High-noisy Cryo-EM micrograph. B: Gray scale distribution in the horizontal dotted line in A.

As we have noted that in most micrographs the microtubules and their surrounding contaminants can be distinguished more intuitively by their morphological differences. Microtubules are often long and similar to rectangles or gray tubes, while most contaminants take the shape of round disks. We make full use of this morphological differences in segmentation task and we think that
introduction of prior shape information of microtubules is an effective help to improve the segmentation. A tubular shape enhancement filter based on Hessian matrix is an effective enhancement method for tubular areas and it can significantly enhance the tubular region in the image while suppressing the non-tubular object area [9][10].

In the proposed algorithm, a series of operation were applied. Firstly we used the morphological method to smooth the image, and a tubular shape enhancement filter based on Hessian matrix was constructed; then three types of information, such as shape filtering value, gray value, and spatial location were introduced to feature space; finally the boundary information was extracted using morphological operator and drawn in the original images to mark microtubules in the Cryo-EM micrographs.

II. METHODOLOGY

A. Specimen Preparation and Electron Microscopy

Biological preparation and pre-research are handled in [2]. Mouse 3T3 fibroblasts were cultured and before their exposure to electron microscope, 3T3 cells were cryo-fixed by plunge freezing the coverslips in liquid ethane. Cryo-immobilized samples were analyzed using a CM-10 Transmission Electron Microscope. For quantitative analysis, various microtubules alongside with their plus ends were scored and categorized. Projection of a microtubules showed that two dark lines caused by the superposition of proto filaments in the direction of projection (i.e., the "sides" of the microtubules), have a higher signal than the "bottom" and "top" of the microtubules [4].

In total, micrographs containing 70 microtubules were selected from the results in [2] to be tested by our method.

B. Preprocessing

Preprocessing is the initial step of detecting the microtubules. It aims to attenuate noises without blurring the images. Cryo-EM micrographs contain serious noise: the noise is a little bigger than the average particle noise, so it can be considered as a small thorn. The traditional Gaussian filter and Anisotropic Diffusion filter[6] can’t get good results with our high-noisy images, while morphological operations can easily remove the noises. Therefore, as the first step of the detection, we used a morphology based method to suppress the noises.

C. Construction of tubular enhancement filter based on Hessian matrix

Hessian matrix is a symmetric matrix composed of two 4 order partial derivatives, it can be used to identify the local shape and the structure of the image[9][10]. In image processing, the core idea of the tubular enhancement filter is to analyze the Hessian matrix of every pixel. It can be used to determine whether the pixel is a pixel in the target area. In this paper, the tubular enhancement filter is designed according to the eigenvalues of the matrix.

1) Hessian matrix and its eigenvalues

In this paper, the Hessian matrix is listed as following [9]:

\[
H = \begin{bmatrix}
    f_{xx} & f_{xy} \\
    f_{xy} & f_{yy}
\end{bmatrix}
\]

The two eigenvalues of Matrix H can be obtained by the following equations [9]:

\[
\lambda_1 = K + \sqrt{K^2 - Q^2} \tag{2}
\]

\[
\lambda_2 = K - \sqrt{K^2 - Q^2} \tag{3}
\]

\[
K = (f_{xx} + f_{yy})/2 \tag{4}
\]

\[
Q = \sqrt{f_{xx}^2 + f_{yy}^2 - f_{xy}^2} \tag{5}
\]

It is assumed that \( \lambda_1 < \lambda_2 \), and if \( \lambda_1 > \lambda_2 \), it should exchange the position of the two eigenvalues.

2) tubular enhancement filter

The tubular enhancement filter based on Hessian matrix is a function of the two eigenvalues: \( \lambda_1 \) and \( \lambda_2 \), and the expression is:

\[
Y(\lambda_1, \lambda_2) = \begin{cases}
    |\lambda_1| - |\lambda_2| & \text{if } \lambda_1 < 0 \\
    0 & \text{otherwise}
\end{cases} \tag{6}
\]

Eq. (6) is the function expression of tubular enhancement filter which is very sensitive to the shape, the output value in the tubular region in the image is large, while the output value is close to 0 in the other region.

D. Meanshift segmentation algorithm combined with tubular enhancement filter

Meanshift algorithm, a non parametric technique, can be used to solve many image processing and vision problems, which can overcome many limitations of traditional segmentation algorithms [12].

Gray value of some microtubules is close to other non-microtubule regions, and in the vicinity of microtubules, especially near the microtubules there are often a number of similar circular impurities, therefore if we use traditional Meanshift clustering which only uses gray-level information, we cannot get ideal segmentation results, the results often contain a variety of non-tubules. In particular, some contaminants which are closely next to microtubules are more difficult to be removed. To separate pollutants from microtubules in the process of Meanshift clustering, we integrate the output value of tubular enhancement filter into the feature space. The tubular filtering output value in the tubular region is relatively large while the output value is small in other regions, thus in the feature space, the distance between the two different shapes of regions with the same grey value is widened.

The construction process of the N+2 dimensional feature space of Meanshift is as following: Characteristic information of a pixel, such as coordinate of pixels, gray value information and the shape filtering output value constitute the 3 dimensional feature space of Meanshift clustering, and they are independent of each other, therefore a pixel in the feature space can be expressed as \( x = (x, y, f_3) \). The pixel density expression in the feature space is:
where \(K_{h_1, h_2}(x) = \frac{C}{h_1 h_2} K_{h_1}^{h_1} x - x \bigg| K_{h_2}^{h_2} x - x \bigg|^{h_1} \bigg| K_{h_2}^{h_2} x - x \bigg|^{h_2} \)

In Eq. (7), \(x^i\) is the spatial information, \(x^g\) is the gray value information and \(x^s\) is the shape information. \(C\) is the corresponding normalization constant.

**E. Segmentation process**

Input: The original high-noisy image \(I = (x, y)\);

Step1: Pre-treatment to remove small particles;

Step2: Build tubular enhancement filter based on Hessian matrix and applied to filter the original image;

Step3: The spatial information, the gray value information and the output value of the filter constitute the feature space of Meanshift clustering;

Step4: Using Meanshift clustering algorithm to handle the feature space until the algorithm converges;

Step5: Fusion over segmentation region, in which the fusion strategy is used in the “or” way;

- **region similarity fusion strategy:** if the distance between the two regions is less than \(h/2\), then fusion.
- **small area fusion strategy:** remove those small area.

Output: The final segmentation image \(s = (x, y)\).

Finally we extracted the edges of microtubules using morphological operation in the output and showed them in the original image.

**III. EXPERIMENT AND ANALYSIS**

In order to verify the proposed method, we have tested 70 microtubules in 50 separated micrographs (Size:1916*2191). 66 of the microtubules were successfully detected, with the success rate of 94.2%. But our method also marked 4 regions that were not microtubules, with fake-detection rate of 6.8%. Additionally we also compared with other segmentation algorithms. Some final results are shown in fig. 4.

**A. Compare with the conventional method**

Fig.2 compares the segmented microtubule results using conventional Meanshift clustering algorithm with the method proposed in this paper. Fig.2A and 2D show the high-noisy micrograph which contains serious noise and many different kinds of non-microtubule materials of various sizes. In particular, some non-microtubules with similar round shape, which are closely next to the microtubules, were troublesome to the segmentation of microtubules. Fig.2B and 2E show the segmentation results of the conventional Meanshift clustering algorithm. In the segmented images (fig.2B and 2E), microtubules have been well segmented, however at the same time, the output image contains other forms of non-microtubule materials, particularly the circular contaminants closely next to the circular were segmented together with the microtubules. While in the other segmented results (fig.2C and 2F) using the new proposed algorithm, microtubules were successfully segmented from the high-noisy micrograph. Thanks to the shape information of the microtubules added to the Meanshift clustering algorithm, the distance between microtubules and other forms of non-microtubule materials are expanded, it got the better results and helped segment the microtubules from the complex background even if the contaminants are close to the microtubules.

**B. Compare with the other method**

Fig.3 compares the segmentation result of the new algorithm with the algorithm used by Yan Wen [4]. As shown in fig.3b, many non-microtubules distributed in the background are not segmented out. By comparison, it is obvious that the results using our new algorithm are better than that using the proposed algorithm by Yan Wen [4].

**C. Some final recognition results**

Fig.4 shows the final recognition results of microtubules from some high-noisy images.

**D. Algorithm evaluation**

To further quantitatively evaluate the accuracy and effectiveness of the new algorithm, we did an objective, systematic and quantitative evaluation. The accuracy coefficient is given in [14]:

\[S = \frac{S_{seg} \cap S_{gold}}{S_{seg} + S_{gold}}\]

Where \(S_{seg}\) is the segmented result of our method, and \(S_{gold}\) is the result segmented by biologists manually.

From the fig.5 we can see the mean accuracy coefficient index for the micrographs is approximately 0.8763.
In this paper, we proposed a novel method which can effectively detect and identify microtubules in high-noisy Cryo-EM micrographs with complex background. It has three steps: (1) Pre-processing to remove the noise; (2) Image segmentation using Meanshift algorithm combined with tubular enhancement filter; (3) Edges of the microtubules detected using morphology operation which are used to mark the microtubule in the original images. This new algorithm can help the structural biology researchers in the initial step of study the structure of microtubules.

Based on the test results, it is proved that the new algorithm has better performance than the traditional Meanshift algorithm and other algorithms mentioned in the reference literatures to segment microtubules in high-noisy Cryo-EM micrographs with complex background.

ACKNOWLEDGMENT

The research was partly supported by the program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning and the Shanghai Municipal Natural Science Foundation under Grant No.15ZR1428600. We are grateful to the support from the 863 Program of China (2015AA015304), the Shanghai Natural Science Foundation (15ZR1410000), the Shanghai Municipal Natural Science Foundation (15ZR1428600).

REFERENCES


TABLE I. COMPARSED WITH METHOD PROPOSED WITH WENYAN [4]

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean accuracy coefficient index</th>
<th>Running time of the processing of 50 micrographs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref.[4]</td>
<td>0.7913</td>
<td>About 5 minutes</td>
</tr>
<tr>
<td>Ours</td>
<td>0.8763</td>
<td>About 5 minutes</td>
</tr>
</tbody>
</table>
