Published in final edited form as:

J Microsc. 2015 April ; 258(1): 13–23. doi:10.1111/jmi.12209.

Localizing and Extracting Filament Distributions from Microscopy Images

Saurav Basu^{*}, Chi Liu[†], and Gustavo Kunde Rohde[‡]

^{*}S. Basu is with the Center for Bioimage Informatics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213 USA

[†]C. Liu is with the Center for Bioimage Informatics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213 USA

[‡]G. K. Rohde is with the Center for Bioimage Informatics, as well as the Department of Biomedical Engineering and the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA. Ph: (412) 268-3684. Fax: (412) 268-9580

Abstract

Detailed quantitative measurements of biological filament networks represent a crucial step in understanding architecture and structure of cells and tissues, which in turn explain important biological events such as wound healing and cancer metastases. Microscopic images of biological specimens marked for different structural proteins constitute an important source for observing and measuring meaningful parameters of biological networks. Unfortunately, current efforts at quantitative estimation of architecture and orientation of biological filament networks from microscopy images are predominantly limited to visual estimation and indirect experimental inference. Here we describe a new method for localizing and extracting filament distributions from 2D microscopy images of different modalities. The method combines a filter-based detection of pixels likely to contain a filament with a constrained reverse diffusion-based approach for localizing the filaments centerlines. We show with qualitative and quantitative experiments, using both simulated and real data, that the new method can provide more accurate centerline estimates of filament in comparison to other approaches currently available. In addition, we show the algorithm is more robust with respect to variations in the initial filter-based filament detection step often used. We demonstrate the application of the method in extracting quantitative parameters from confocal microscopy images of actin filaments and atomic force microscopy images of DNA fragments.

Keywords

Biological filament networks; curvilinear structures; local network topology; centerline curvature

1 Introduction

Curvilinear structures with complex network topology can be observed in a wide variety of biological systems including the extracellular matrix and cytoskeletal filaments. The diameter of these filaments is typically in nanometer range, while the filaments typically have persistence lengths (μm to mm) that are much greater than the resolution of

fluorescent imaging (200 nm) and form complex networks. For example, the actin cytoskeleton is required for mechanical integrity, force generation, cytokinesis and motility. The cytoskeleton also contributes to a wide range of cellular mechanisms including intracellular signaling and differentiation (Disanza et al., 2005). Widefield, confocal, and total internal reflectance fluorescence (TIRF) microscopy have been extensively used in recent years to image actin with the end goal of studying the effect of external perturbations on the structural alterations of such networks. The measurement of several architectural parameters of biological networks such as filament lengths (Lichtenstein et al., 2003; Shariff et al., 2010), persistence length (Ott et al., 1993), local orientation (Petroll et al., 1993; Thomason et al., 1996; Karlon et al., 1999; Weichsel et al., 2010), curvature distribution and local filament organization (Fleischer et al., 2007) are critical for quantifying the role of network geometry in cellular and tissue functions.

Recent advances in imaging capabilities for visualizing biological structures, including microscopy hardware and computational post-processing, have allowed increased resolution at nanometer length scales (Huang et al., 2010; Selvin et al., 2010). However, quantitative analysis of cellular filaments *in situ* by microscopic imaging is complicated by optical blurring, noise, clutter, as well as the geometric complexity of such dense networks inside cells. As such, for most experimental scientists, the process of identifying filament distributions from microscopy images is largely qualitative due to the lack of accurate quantitative evaluation of architecture, orientation and topology of specific networks.

The complete enumeration and characterization of biological filament networks from microscopy images remains a challenging problem. Partial solutions such as local orientation (Petroll et al., 1993; Thomason et al., 1996; Karlon et al., 1999; Weichsel et al., 2010) and total filament length (Lichtenstein et al., 2003) have been proposed in the recent past. However, a successful methodology that localizes centerlines of individual filaments (or coherently oriented filament bundles) despite the confounding factors associated with diffraction-based blurring and complicated filament architecture remains a largely unsolved problem. Local thresholding methodologies (Gonzales and Woods, 1992) or filament enhancement schemes followed by some sort of binary thinning (Loss et al., 2011), for example, are often inadequate solutions to the above-mentioned question since the thresholding parameter is difficult to guess and binary thinning ignores the intensity profile of the enhanced images. Luck et al. (Lück et al., 2010) describe a complete system for localizing and recovering the network topology of intermediate filaments from scanning electron microscopy images. The localization procedure in this case is also based on estimating linear structures from a binary (threshold)-based estimate of the filament locations.

Several researchers have pursued approaches for measuring partial topological parameters of filament networks from different calculable image properties. Petroll et al. (1993) use Fourier methods to estimate actin stress fiber orientation and Thomason et al. (1996) employ fractals to analyze cytoskeletal structure. The accuracy of these methods is difficult to test due to the complete decoupling of the biological components of a filament network from the actual pixel-wise image properties. Karlon et al. (1999) propose an improved orientation measurement compared to (Petroll et al., 1993; Thomason et al., 1996) by accumulating

image gradients into histograms defined over local image windows. Weichsel et al. (2010) propose a similar method to Karlon et al. (1999) where they calculate local coherency of the structure tensor in order to estimate the principal orientation of filaments. Although these estimated orientations have higher order information, calculations are independent of any actual segmentation of the actin fibers and are derived from image properties that relate to network topology only indirectly.

Lichtenstein et al. (2003) develop a somewhat generative model for detection of filament pixels in fluorescence microscope images. This process is statistically amenable, but it does not explicitly address network geometry information. Shariff et al. (2010) also investigate a generative approach combined with indirect (inverse) estimation of the generative model to estimate basic parameters (number, mean length) from live and fixed cells. Fleischer et al. (2007) propose an interesting methodology for measuring actin network morphology by fitting geometric tessellation models to actin network images. Finally, Xu et al. (2011) use multiple open active contours to segment *in-vitro* actin filament populations. This method can provide individual filament (bundle) information. However, contour merging and splitting rules are difficult to prescribe.

Finally, we mention that there exist a few commercially available software packages that perform some form of filament tracing, including Bitplane (2012); Wearne et al. (2005); Longair (2010). We note, however, that the functionality of these softwares is limited to tracing structures similar to neurons where the resolution of the image is higher relative to the filaments (neurons) being traced. As a result the neuronal branches are clearly separated and the geometry of the branch cross sections are relatively uniform. In contrast, our images are populated with dense filament networks where the filament thickness is much below the optical resolution of fluorescent imaging. The result is optical blurring, low signal-to-noise ratio, and ambiguity in delineating junctions and intersections in a dense network. Most of the commercially available softwares use techniques such as finding unambiguous and clearly separated seed *points* along well separated filament branches, which are infeasible at the resolution of confocal images we work with. Additionally, manual placement of seed points is often a preprocessing step in these software (Longair, 2010), an involved and time consuming process that we strive to avoid.

Our Algorithm and Contributions

Here we describe an image analysis method that can take as input microscopy images of biological filament networks and, despite confounding factors such as background fluorescence, clutter, and dense placement of filaments or filament bundles, can accurately localize centerline pixels of filaments. This layout of centerline filaments can be subsequently used for various important computations such as filament curvature distributions, local connectivity and topology, orientation distributions, placement within the cell, as well as total estimated lengths. We note that our method has an important limitation in that it is only able to recover the location of filament or coherently oriented filament bundles in an image. In many cases (e.g. confocal microscopy images of actin filaments) such crests can correspond to several filaments made inseparable given image resolution limitations.

More specifically, we present a new constrained diffusion-based filament centerline localization that accurately estimates filament centerlines in fluorescence confocal microscope images even in the presence of high filament density, background clutter, filament intersections and bifurcations. The algorithm works by utilizing a filament orientation map, which in our case is estimated with a matched filter approach, followed by a ‘constrained’ local mean shift approach, to estimate the location of filaments. We have applied our algorithm to estimate filaments from simulated data, real confocal microscopy images of actin filaments, as well as DNA filament images obtained using atomic force microscopy. We compare our results with other existing methods (Chang et al., 2001; Donoho et al., 2001). We show, qualitatively and quantitatively, increased performance in comparison to existing methods.

2 Method

We illustrate the sequential blocks in our algorithm and give examples along the way to demonstrate the intuition behind each step. The three main steps in our procedure are, in order, (1) matched filter-based filament likelihood estimation, (2) constrained diffusion-based thinning to localize filament center locations and (3) the determination of the local connectivity and topology of the filament distribution via a minimum spanning tree method. From the output of step three we are able to compute local connectivity maps, as well as several important filament network-related quantities such as: curvature, orientation, as well as local filament topology. Figure 1 contains an example of each step applied to an image of actin filaments (Wang, 2007).

2.1 Filament Likelihood Estimation

Any filament recovery process has to first distinguish the fluorescence signal from acquisition noise, and also determine whether a fluorescence signal belongs to a filamentous protein or is simply background clutter from very short filaments. There are several possible approaches for testing whether a particular pixel contains an underlying filament. Methods range from pixelwise differential operators such as (Frangi et al., 1998), morphology based detectors (Chang et al., 2001) to more global matched filter-based approaches (Donoho et al., 2001). We note that several of these approaches also apply to edge delineation in image data, though here we focus our discussion to filament distributions. In our implementation, we adopt a matched filter-based estimation procedure because we are sufficiently certain of the local filament width (given by the point spread function of the imaging instrument being used), and therefore can use locally accurate templates that can accumulate evidence from surrounding pixels.

The idea is to consider local evidence for the presence or absence of a filament. We do this in a given image by finding regions that at the same time 1) have the ‘appearance’ of a filament type structure and 2) have relatively high intensity values (a fluorescent filament has higher intensity than a background pixel). To that end, we create a set of filament filters modeled by a quadratic function. Let $\phi(x) = ax^2$, with a the curvature parameter. In addition, given a sub-window of size s , we sample $\phi(x)$ at the set of grid coordinates $(x_0, \dots, x_{s-1})$ provided by the sub-window. Then the set of coordinates $\vec{x}_i = [x_i, \phi(x_i)]$ is rotated by a

prescribed fixed angle θ . Using these operations we are able to create a sub window that has intensity equal to one at sampled coordinates $\vec{x} = [x_i, \phi(x_i)]$ (which can be rotated by some angle θ) and zero elsewhere. This binary sub window is finally convolved with an estimate point spread function (PSF) for the image(s) whose filaments are to be localized. In our work we utilize a Gaussian function approximation of the PSF. More accurate estimates of the PSF can be estimated using known imaging parameters such as fluorescence wavelength, and numerical aperture, for example (Goodman, 1996). Figure 2 contains a few example filters we created using this procedure. We denote a particular sub window generated using this procedure as $f_{s,\theta,a}(\vec{x})$, where $\vec{x}$ is a 2D coordinate within the sub window of size s .

Given an input image $I(\vec{x})$, with $\vec{x}$ a 2D image coordinate, the likelihood that a particular pixel location $\vec{x}$ contains an underlying filament is estimated by:

$$\mathcal{L}(\vec{x}) = I(\vec{x}) \max_{s,\theta,a} \{I \star f_{s,\theta,a}(\vec{x})\} \quad (1)$$

where the $\star$ operation denotes the normalized cross correlation (Haralick and Shapiro, 1992) between the image I and $f_{s,\theta,a}$. The maximization procedure above is performed exhaustively, for a pre determined set of parameters (s, θ, a) . In this work we have used the following parameters while searching for the solution for the maximization problem stated in 1: $s = \{5, 7, \dots, 15\}$, $a = \{0, 1, 2\}$ and θ ranges from 0 to 360 degrees in steps of 5 degrees.

Finally, we note that the maximization problem stated above can also yield the angle θ that best describes the orientation of any underlying filament bundle at location $\vec{x}$. We denote this orientation vector, at location $\vec{x}$, by

$$O(\vec{x}) = [\sin \bar{\theta} \cos \bar{\theta}]^T. \quad (2)$$

Smoothing the Orientation Field—The orientation vector given by eqn. (2), as can be expected, suffers from estimation noise (primarily contributed by the deviation from ideal filamentous structures in the real images) and lack of meaningful orientations in non-filamentous areas such as the background and the cytoplasm. However, a simple local smoothing filter applied to the individual components of the orientation field gives incorrect results in our formulation since the space of angles is highly redundant. For example, angles θ_1 and $\pi + \theta_2$, where $\theta_1 \sim \theta_2$, although very similar in filament orientation, might be construed as mathematically vastly different while performing a simple smoothing.

Therefore, instead of locally smoothing the components of the orientation vectors, we locally smooth the respective components of the 2×2 *Orientation Matrix* $\mathbf{O}$ defined by

$$\mathbf{O}(\vec{x}) = \mathcal{L}(\vec{x})[O(\vec{x})O(\vec{x})^T]. \quad (3)$$

The local smoothing of $\mathbf{O}(\vec{x})$ is carried out component by component. In other words, we separate the 4 components of $\mathbf{O}(\vec{x})$ at the pixel locations and smooth them individually over the pixel locations using a chosen smoothing filter. Note that, in equation (3), if $O(\vec{x}_1) = [\sin \bar{\theta} \cos \bar{\theta}]^T$ and $O(\vec{x}_2) = [\sin(\pi + \bar{\theta}) \cos(\pi + \bar{\theta})]^T$, then $\mathbf{O}(\vec{x}_1) = \mathbf{O}(\vec{x}_2)$ (assuming $\mathcal{L}(\vec{x}_1) = \mathcal{L}(\vec{x}_2)$). Also note that the orientation matrix is weighed by the likelihood image so that we do not incorporate meaningless orientation vectors from regions which have low possibility of having filaments.

Suppose the (component-wise) smoothed matrix is called $\mathbf{O}_s(\vec{x})$, then, the smoothed orientation vector (We also call it $O(\vec{x})$ by an abuse of notation) can be recovered from $\mathbf{O}_s(\vec{x})$ by setting it to the eigenvector of $\mathbf{O}_s(\vec{x})$ with the higher eigenvalue. For the component-wise smoothing of $\mathbf{O}(\vec{x})$, we have used a 5×5 Gaussian filter with 0 mean and standard deviation equal to 1 pixel length.

2.2 Filament Centerline Localization by Reverse Diffusion

The map $\mathcal{L}(\vec{x})$ provides an estimate of the likely presence of a filament at pixel location $\vec{x}$. A logical next step is to select locations $\vec{x}$ for which $\mathcal{L}(\vec{x})$ is greater than some threshold α as potential candidates for filament positions. This procedure, however, can be inaccurate and the standard choice is to follow the thresholding by the standard skeletonization procedure described in Chang et al. (2001). We here describe a different skeletonization procedure that is able to produce robust estimates of filament locations by utilizing more information from the likelihood and smoothed orientation functions described earlier. The idea is to define filament centerline locations as the ‘crests’ of the likelihood function $\mathcal{L}(\vec{x})$, where ‘crests’ are defined as local maxima in the direction perpendicular to the local orientation map.

We start with a coarse binary segmentation of the likelihood image $\mathcal{L}(\vec{x}) > \alpha$. The set of coordinates that satisfy this inequality are denoted as $\vec{X} = [x_1, x_2, \dots, x_N]$ and serve as starting ‘particles’ to which we apply our reverse diffusion process. However, we constrain the movement of the particles to be along the direction perpendicular to $O(\vec{x})$, denoted by $\hat{O}(\vec{x})$. The differential equation describing this process is given by:

$$\frac{d\vec{x}_i(t_i)}{dt} = \hat{O}(\vec{x}_i(t_i)). \quad (4)$$

The time t_i here is an artificial parameter allowing us to specify the iterative maximization of the following objective function:

$$E(t_1, t_2, \dots, t_n) = \lambda \sum_{i=1}^n \mathcal{L}(\vec{x}_i(t_i)) - (1 - \lambda) \sum_{j=1}^n \sum_{k=1}^n \mathcal{K}(\vec{x}_k(t_k), \vec{x}_j(t_j)) (\vec{x}_k(t_k) - \vec{x}_j(t_j))^2, \quad (5)$$

where $\lambda \in [0, 1]$ is a parameter that weights the two terms in the equation above, and

$\mathcal{K}(x_i, x_j) = 1/\sqrt{2\pi\sigma^2} e^{-(x_i - x_j)^2/2\sigma^2}$. In the results shown below, we have set $\lambda = 0.5$. The first term $\sum_{i=1}^n \mathcal{L}(\vec{x}_i(t_i))$ gives preference for positions where the likelihood function is highest. The second term aims to make particles move towards each other so that they do not move independently of each other. The goal is then to find variables $t_1, t_2, \dots, t_n$ that, through equation (4), maximize equation (5).

To that end, we utilize the standard gradient ascent method with fixed step size.

Differentiating $E(t_1, t_2, \dots, t_n)$ in equation (5) with respect to t_i we obtain the gradient descent step for the maximization of E as

$$\begin{aligned} \Delta t_i = \tau \frac{\partial E}{\partial t_i} = & \tau \lambda (\nabla \mathcal{L}(\vec{x}_i) \bullet \hat{O}(\vec{x}_i)) - \tau (1 - \lambda) \left(\sum_{k=1}^n (4\mathcal{K}(\vec{x}_i, \vec{x}_k) (\vec{x}_i - \vec{x}_k) \right. \\ & \left. + 2 \frac{\partial \mathcal{K}(\vec{x}_i, \vec{x}_k)}{\partial t_i} (\vec{x}_i - \vec{x}_k)^2) \right) \end{aligned} \quad (6)$$

where we imply $\vec{x}_i(t_i)$ by $\vec{x}_i$, and where τ is an iteration time step that can be chosen arbitrarily or by a suitable line-search method (Box et al., 1969). We note that $\bullet$ denotes the vector dot product. Thus the gradient ascent equation can be written as

$$\vec{x}_i(t_i + \Delta t_i) = \vec{x}_i(t_i) + \Delta t_i \nabla \hat{O}(\vec{x}_i(t_i)), \quad (7)$$

and the explicit evolution equation for the particles $\vec{x}_i \in \vec{X}$ is given by combining equations (4), (6), and (7):

$$\begin{aligned} \vec{x}_i(t_i + \Delta t_i) = & \vec{x}_i(t_i) + (\tau \lambda (\nabla \mathcal{L}(\vec{x}_i) \bullet \hat{O}(\vec{x}_i)) - \tau (1 - \lambda) \left(\sum_{k=1}^n (4\mathcal{K}(\vec{x}_i, \vec{x}_k) (\vec{x}_i - \vec{x}_k) \right. \\ & \left. + 2 \frac{\partial \mathcal{K}(\vec{x}_i, \vec{x}_k)}{\partial t_i} (\vec{x}_i - \vec{x}_k)^2) \right) \hat{O}(\vec{x}_i)). \end{aligned} \quad (8)$$

The third panel in Figure 1 shows the output of this procedure when applied to the likelihood image shown in the second panel of the same figure.

Convergence—Anytime a new iterative approach is presented, it is important to consider whether the method is guaranteed to converge or if the iterative method could go on forever. To understand that the method described above converges, in this case, it suffices to show that 1) the optimization function is bounded, and 2) each successive iteration augments the value of the objective function being maximized. In this case, it is clear that the objective function (5) is bounded above by $\lambda n \max_i \mathcal{L}(y_i)$, where y_i correspond to all pixels in the input image. Now we note that the method we utilize to maximize this procedure is based on the steepest ascent procedure which, given a small enough choice of τ or the use of an appropriate line search method, guarantees that each step in the procedure increases the value of the cost function. It is important to note, however, that this function is nonlinear and non convex. Therefore any such gradient-based procedure is only guaranteed to converge to a local optimum, which may not correspond to the globally optimal solution.

2.3 Local Filament Centerline Extraction by a Minimum Spanning Tree

The final part of our algorithm infers the local centerlines by connecting the converged centerline particles from the previous step through a minimum spanning tree (MST) (Graham and Hell, 1985), where each converged particle is treated as a graph node and the edge distance between any two nodes is the simple Euclidean distance separating them. We also can reasonably assume that for a dense set of nodes such as the output from the reverse diffusion step, two nodes on one filament centerline and connected by the MST cannot be too far apart, therefore we disconnect edges in the MST that are longer than a prescribed length. The output of this procedure is shown in the last panel of Figure 1. The local filament elements output by the MST can be utilized to calculate local curvatures of the filaments at non-bifurcating points. This is an important measurement in cytometry where perturbation of the cytoskeleton by external agents frequently express themselves as a change of curvature distribution of the network.

3 Experiments

Our filament localization algorithm has been tested on a database of real and simulated images to test for both accuracy and applicability. To generate our filter bank, we used a filament width $w = 0.32 \mu\text{m}$ and a fine sampling of curvatures, orientations and scales in our filter bank. In all results shown, the threshold parameter for the likelihood function was set to $\alpha = 0.1$. In all experiments, the width of the Gaussian function $\mathcal{N}(\vec{x}_i, \vec{x}_j)$ was set to $\sigma = 1.6 \mu\text{m}$.

3.1 Validation on simulated images

We have tested the filament localization step on 100 simulated images of size 128×128 pixels consisting of filament networks that vary in complexity and background noise. The lowest filament count per image is eighty and the highest is one hundred and thirty, with each filament count having two independently generated images. The length of individual filaments are drawn randomly from a Gaussian distribution of mean 40 pixel lengths ($2.5 \mu\text{m}$) and a standard deviation of 5 pixel lengths ($0.32 \mu\text{m}$ in this simulation). The starting point of each filament in each image is chosen randomly, and the filament is allowed to grow to the sampled length incrementally, with the direction of each incremental selected

randomly from a uniform distribution ranging from -10 to 10 degrees. Panels (a), (b), and (c) of Figure 3 show three example images simulated using this approach. After the filament traces were simulated, the resulting images were convolved with a Gaussian PSF to simulate optical blurring. Note that since the individual traces were simulated, they are known exactly and can be used as ground truth to evaluate any tracing result.

Figure 4 shows the result of our filament localization step vis-a-vis two other popular filament detection methods from real or binary images, the standard binary thinning procedure (Chang et al., 2001) and Beamlet decomposition (Donoho et al., 2001). In Figure 4(e), the Y axis shows the percentage of correctly localized centerline points with respect to the starting points; here we call the ‘correctly localized centerline points’ as the proportion of pixels that are correctly localized on real filament centers with respect to the total number of pixels output by the algorithm as filament centers. Any output pixel that is within a unit pixel distance from an actual filament centerline pixel in the ground truth is deemed a ‘correct’ localization. In Figure 4(f), the Y axis shows the percentage of correctly covered centerline points in the ground-truth with respect to the points output by the three competing algorithms; here by a ‘correctly covered’ centerline point, we mean that at least one point output by the respective algorithm is within one pixel distance of the true centerline point in question. The X axis in both Figures 4 (e) and (f) are plotted with increasing filament density (starting with eighty and ending at one hundred and thirty randomly generated filaments per image), and hence complexity of the image. Figures 4(a) shows an example of an original simulated image. Panels (b), (c) and (d) of the same figure show the result of applying binary thinning (Chang et al., 2001) method, the Beamlet extraction ((Donoho et al., 2001)), and our method respectively. The panels 4(b) and 4(c) show comparable performance. Panel 4(c) is evidently not a good choice for filament localization. Figure 4(e) and (f) shows the result of applying our method, binary thinning and the Beamlet extraction method to the database of 94 images. (The results of the two independent realizations of each filament count has been averaged and reported). For simulated images generated with a Gaussian PSF, binary thinning gives comparable performances to our method, except in congested areas where a thresholding of the likelihood image produces confusing boundaries. Our method slightly outperforms binary thinning in this simulation, which contained no image noise. The differences, however, are statistically significant. The mean accuracy for the method we propose is 0.96 and 0.84 for the data in portions (e) and (f) of this figure, respectively. The mean accuracy for the binary thinning method for the same data was 9.6 and 0.80. The p -value for the paired t -Test comparison is zero (within numerical precision error). We note this plot also shows that the constrained reverse diffusion-based approach nearly always (with the exception of two data points) outperforms the other methods tested. In addition to simulations without image noise, we’ve also tested the robustness of the reverse diffusion and binary thinning methods to recover filament centerlines in noisy images. To that end, we’ve repeated the experiment described above by adding zero mean Gaussian noise to each simulated image. Figure 5 displays the overall average accuracy for each experiment which was repeated between signal to noise ratios (SNR) 6.62 and -2.23 dB. Results show a more significant difference between the methods when noise was present.

3.2 Real Filament Images

We have applied the filament localization step to image datasets of cells. To that end, we have utilized the LIPS-III database containing maximum intensity projections of spinning disk confocal microscopy images of stomata guard cells ($0.064 \mu\text{m}$ per pixel, 119×304 pixels). The data is fully described in (Higaki et al., 2013). Parts (d) and (e) of Figure 3 show two examples of real cell images stained for actin. Figure 6 shows the result of applying both the method we describe here (second column) and the binary thinning method (third column from the left) on the raw images. We note that both the reverse diffusion-based approach and the binary thinning method utilized the same likelihood function, as well as the same initial threshold.

In order to further quantify the differences between the method we propose and the standard binary thinning algorithm, in Figure 7 we demonstrate the application of both methods to an image sub-region, where the ground truth has been manually delineated. Part (a) shows the unaltered image, part (b) shows the estimated likelihood function, and part (c) shows the manual delineation. The second row of this figure shows our method as well as the standard binary thinning method being applied after the likelihood image was thresholded using $\alpha = 0.1$. Finally the third row both methods applied when $\alpha = 0.05$. As shown here, the reverse diffusion-based approach is significantly more robust to changes in the initial threshold.

In addition, we have quantified how well both methods match the ground truth displayed in part (c) of Figure 7. We have manually delineated the ground-truth in a set of real filament images and performed a similar comparison of our method vis-a-vis the standard binary thinning as in Figure 7. The resultant average pixel distance error for our method after the likelihood image was thresholded using $\alpha = 0.21$ and $\alpha = 0.17$ were 0.29 and 0.42 respectively, while the same errors with the standard binary thinning and the respective thresholds were 0.34 and 0.73. This shows that while our method performs better on dense real filaments it is also considerably robust to the threshold α , whereas binary thinning is heavily dependent on the threshold and the error increases by over 100% for a slight change in the initial threshold.

The tree of centerline pixels output by our method can be subsequently used for calculating many important biological properties of filament networks in cells. As an example, we consider the problem of estimating the centerline location of DNA fragments deposited on mica using AFM images of 538 bp DNA fragments, imaged as described in (Lysetska et al., 2002), prior to and after UV light exposure. Figure 8 shows DNA fragments prior to UV exposure (part A), as well as after 20 (B) and 40 (C) minutes of exposure (Lysetska et al., 2002) (by permission of Oxford University). Parts (D) and (E) show the centerlines extracted from (A) and (B), respectively. We utilized our software to extract the centerlines corresponding to DNA fragments, from these compute the normalized curvature distribution for each image. Part (F) shows a histogram of the normalized curvature computed from the DNA fragments before and after 20 minutes of radiation (data from 40 minutes of radiation were omitted since the filaments were no longer visible). That is, the curvature of each centerline pixel was computed for each image, and the results (of all pixels) were aggregated into the histograms shown in this figure. Results show a slight increase in curvature with UV radiation. The mean normalized curvature was 0.36 ($\sigma = 0.31$) for data prior to UV radiation,

and 0.37 ($\sigma = 0.31$) after 20 minutes of radiation. The difference, however, is not statistically significant as far as can be concluded from the available images. Our goal is to show the technique can be used to quantify differences between filament measurements obtained using different conditions. From this limited sample, the technique we described allows us to conclude that no statistically significant difference in curvature is detected.

4 Summary and Discussion

We have described an algorithm to extract the location of curvilinear structures in microscopy images. We demonstrate the feasibility of our approach by applying our tool to estimate actin filament networks in confocal microscope images of filament distributions in several cells. We show that quantitative measurement of filament centerlines is possible even in the case of dense networks with complicated intersections and bifurcations. Although in this work we focus on extracting information related to actin network from microscopy images, we note that the constrained diffusion algorithm can be used as a general linear network estimator, as well as a line thinning algorithm. Experiments using real and simulated data showed that our method outperformed two other existing algorithms (Donoho et al., 2001; Chang et al., 2001). In addition, we have demonstrated how our method can be used to extract meaningful information about actin networks in complicated geometric configurations.

We note that in the simulated data experiments our method slightly outperforms the standard binary thinning method. Qualitative and quantitative results on real data, however, show a more pronounced improvement both in accuracy of the centerline locations, the ability to deal with complicated topologies, as well as robustness with respect to changes in the initial threshold. We stipulate that this is due to the fact that it is difficult to simulate the intricate complexity of real images. Real images have several artifacts which we have not implemented in our simulation. Some of which are varying depth of focus issues, combinations of very long and very short filaments, as well as blurring from out of focus fluorescence, to name a few. We believe that if these were included our simulation, the improvements over binary thinning in the simulated data results could have been greater. In addition we note that our algorithm showed greater robustness to slight variations in the threshold used in the filter bank detection step, as well as greater robustness to noise. This represents an important advantage since it translates to reduced effort in ‘calibrating’ the method to different types of data.

Finally, we believe it is useful to discuss a few of the inherent drawbacks of the procedure we describe here. The first obvious one is that our current implementation is for 2D images. This is not a severe impediment, however, since the same method could be applied to 3D images by extending the methodology to 3D as well, given that all operations defined here (filtering, reverse diffusion, etc.) can be performed in 3D. Naturally, this would come at a significant increase in computational cost. In addition, we note that computation time for this process is significantly higher than for the binary thinning method for example. For instance, our method, when applied to Figure 6(d) with size 304×119 , took 383 seconds to complete, while the corresponding times for binary thinning was only 0.05 seconds. We note, however, that our algorithm was implemented in the Matlab (Mathworks, 2012)

programming language, with extensive use of 'FOR' loops, which are notoriously slow. We believe the time of computation, however could be significantly improved by implementing the code in a compiled (as opposed to interpreted) language. Lastly, we note that the method can be applied to localize filaments where the complexity of the network (number of filaments and their proximity) is not so high as to, combined with optical blurring due to diffraction, eliminate the 'crests' that our algorithm depends on. In such cases, we believe an inverse modeling approach such as the one discussed in Shariff et al. (2010) would be a better estimation method.

Acknowledgments

The authors would like to thank Dr. Kris Dahl, from Carnegie Mellon University, for informative discussions. SB, CL, and GKR acknowledge support from grant NIH GM090033.

References

- Bitplane. Bitplane Scientific Software. 2012.
- Box, MJ., Davies, D., Swann, WH. Non-linear optimisation techniques. Oliver & Boyd; 1969.
- Chang S, Kulikowski CA, Dunn SM, Levy S. Biomedical image skeletonization: a novel method applied to fibrin network structures. Medinfo. 2001; 84(2):901–906.
- Disanza A, Steffen A, Hertzog M, Frittoli E, Rottner K, Scita G. Actin polymerization machinery: the finish line of signaling networks, the starting point of cellular movement. Cell Mol Life Sci. 2005; 62(3):955–970. [PubMed: 15868099]
- Donoho, DL., Huo, X., Jermyn, I., Jones, P., Lerman, G., Levi, O., Natterer, F. Multiscale and Multiresolution Methods. Springer; 2001. Beamlets and multiscale image analysis; p. 149-196.
- Fleischer F, Ananthakrishnan R, Eckel S, Schmidt H, Käs J, Svitkina T, Schmidt V, Beil M. Actin network architecture and elasticity in lamellipodia of melanoma cells. New Journal of Physics. 2007; 9(11):420.
- Frangi, RF., Niessen, WJ., Vincken, KL., Viergever, MA. Proc MICCAI'98 LNCS. Springer-Verlag; 1998. Multiscale vessel enhancement filtering; p. 130-137.
- Gonzales, R., Woods, R. Digital Image Processing. Addison-Wesley Publishing Company; 1992.
- Goodman, J. Introduction to Fourier Optics. McGraw-Hill; 1996.
- Graham RL, Hell P. On the history of the minimum spanning tree problem. IEEE Ann Hist Comput. 1985; 7(1):43–57.
- Haralick, RM., Shapiro, LG. Computer and Robot Vision. Vol. II. Addison-Wesley; 1992.
- Higaki T, Kutsuna N, Hasezawa S. Lips database with lipservice: a microscopic image database of intracellular structures in arabidopsis guard cells. BMC Plant Biol. 2013; 13:81. [PubMed: 23679342]
- Huang B, Babcock H, Zhuang X. Breaking the diffraction barrier: super-resolution imaging of cells. Cell. 2010; 143(7):1047–1058. [PubMed: 21168201]
- Karlson WJ, Hsu PP, Li S, Chien S, McCulloch AD, Omens JH. Measurement of orientation and distribution of cellular alignment and cytoskeletal organization. Annals of Biomedical Engineering. 1999; 27:712–720. 10.1114/1.226. [PubMed: 10625144]
- Lichtenstein N, Geiger B, Kam Z. Quantitative analysis of cytoskeletal organization by digital fluorescent microscopy. Cytometry A. 2003; 54(1):8–18. [PubMed: 12820116]
- Longair M. Simple Neurite Tracer. 2010
- Loss L, Bebis G, Parvin B. Iterative tensor voting for perceptual grouping of ill-defined curvilinear structures. Medical Imaging, IEEE Transactions on. 2011; 30(8):1503–1513.
- Lück S, Sailer M, Schmidt V, Walther P. Three-dimensional analysis of intermediate filament networks using sem tomography. Journal of Microscopy. 2010; 239(1):1–16. [PubMed: 20579265]

- Lysetska M, Knoll A, Boehringer D, Hey T, Krauss G, Krausch G. Uv light-damaged dna and its interaction with human replication protein α : an atomic force microscopy study. *Nucleic Acids Res.* 2002; 30(12):2686–91. [PubMed: 12060686]
- Mathworks. MATLAB. 2012
- Ott A, Magnasco M, Simon A, Libchaber A. Measurement of the persistence length of polymerized actin using fluorescence microscopy. *Phys Rev E.* 1993; 48:R1642–R1645.
- Petroll WM, Cavanagh HD, Barry P, Andrews P, Jester JV. Quantitative analysis of cell fiber orientation during corneal wound contraction. *J Cell Sci.* 1993; 104(2):353–363. [PubMed: 8505365]
- Selvin PR, Syed R, Sobh N. Illinois tool: Fiona (fluorescence imaging with one nanometer accuracy). 2010
- Shariff A, Murphy RF, Rohde GK. A generative model of microtubule distributions, and indirect estimation of its parameters from fluorescence microscopy images. *Cytometry Part A.* 2010; 77A(5):457–466.
- Thomason DB, Anderson O, Menon V. Fractal analysis of cytoskeleton rearrangement in cardiac muscle during head-down tilt. *Journal of Applied Physiology.* 1996; 81(4):1522–1527. [PubMed: 8904563]
- Wang YL. Noise-induced systematic errors in ratio imaging: serious artefacts and correction with multi-resolution denoising. *Journal of Microscopy.* 2007; 228(2):123–131. [PubMed: 17970912]
- Wearne SL, Rodriguez A, Ehlenberger DB, Rocher AB, Henderson SC, Hof PR. New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience.* 2005; 136(3):661–680. [PubMed: 16344143]
- Weichsel J, Herold N, Lehmann MJ, Kräusslich HG, Schwarz US. A quantitative measure for alterations in the actin cytoskeleton investigated with automated high-throughput microscopy. *Cytometry Part A.* 2010; 77A(1):52–63.
- Xu T, Li H, Shen T, Ojic N, Vavylonis D, Huang X. Extraction and analysis of actin networks based on open active contour models. *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on.* 2011:1334–1340.

Figure 1.

Overview of filament localization procedure. The input image (leftmost panel, data from (Wang, 2007)) is first analyzed to determine the likelihood that a filament is present at any given pixel (second panel). The likelihood function is then thresholded to obtain an initial estimate of the filament locations. The initial estimate is ‘evolved’ with the constrained reverse diffusion-based algorithm described in the text so as to estimate the centerlines of each likely filament (third panel from the left). The estimated filament centerlines are then analyzed so as to determine their local connectivity and topology (right most panel).

Figure 2.
Examples of filters used to measure evidence of filaments in sub-portions of a filament image. In this case we have shown straight line and curved line elements with varying degrees of rotation.

Figure 3.

(a), (b) and (c) show three images from a simulated database of artificial filaments with filament counts of 95, 100 and 105 respectively. (d) and (e) show two maximum intensity projections of actin filaments in stomata guard cells from the LIPS-III database (see text for details).

Figure 4.

Part (a) shows a simulated filament image having 94 filaments from our database. Parts (b), (c) and (d) show the filament localized images after application of binary thinning ((Chang et al., 2001)), Beamlet extraction ((Donoho et al., 2001)) and our method respectively. (e) shows the result of applying our method, binary thinning and the Beamlet extraction method to the database of 94 images. The Y axis shows the percentage of correctly localized centerline points with respect to the starting points (see text for more details). (f) shows the proportion of the true centerline pixels that are within one pixel distance of any output particle by our method, binary thinning and the Beamlet extraction method to the database of 94 images. There are two images for each setting of the filament count and random parameters for both (e) and (f), the results of the two cases have been averaged.

Figure 5. Average filament localization accuracy (proportion of pixels which are correctly localized) for reverse diffusion and binary thinning-based methods as a function of SNR.

Figure 6.

(a), (d) and (g) show original confocal microscope images of three HeLa (g) and plant stomata guard cells (a,d) cells, where the actin filaments have been labeled. (b), (e) and (h) show the corresponding localized centerline points output from our localization algorithm in red, describing the local network structure. (c), (f) and (j) show results of the binary thinning (Chang et al., 2001) on (a), (d) and (g), demonstrating that (Chang et al., 2001) is not an ideal way to localize dense filaments with irregular profile geometry.

Figure 7.

(a) shows a close-up of a zoomed in area of Figure 6(d). 7(b) shows the enhanced filament image from (a). (c) shows a manual delineation of the strongly oriented filaments in (b) in red. (d) and (f) show the result of application of our algorithm (in green) to (b) with threshold $\alpha = 0.21$ and $\alpha = 0.17$ respectively. Similarly, (e) and (g) show the application of (Chang et al., 2001) to (b) with the same respective values of α (in magenta).

Figure 8. Parts (A), (B), and (C) show DNA fragments imaged using AFM prior to, after 20 minutes, and after 40 minutes, respectively, of UV radiation (Lysetska et al., 2002). Parts (D) and (E) show the extracted filaments. Part (F) shows a histogram of normalized curvatures extracted from these images (see text).