Learning Generative Models of Microtubule Distributions

Aabid Shariff

CMU-CB-12-101

March 2012

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Robert F. Murphy, Chair
Gustavo K. Rohde, Co-chair
Russell Schwartz
Philip R. LeDuc
George C. Tseng, University of Pittsburgh

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

To my parents, my brother and my wife.

iv

Abstract

The field of location proteomics seeks to characterize the distributions of all proteins across all cell types and conditions over time. In order to further understand the behavior of proteins in cells (systems biology), we need cell simulations that take into account location information of the proteins (location proteomics). One way this gap can be bridged is by building models in a hierarchical, conditional manner so that models of all cell components can be constructed by automated learning from cell images.

Building on the work done by Zhao and Murphy [2007] where models of cell, nuclear and object-type proteins were described, this thesis focuses on building models of microtubules. Microtubules are dynamic filamentous structures in cells that are important in many cellular processes such as cell division, motility and intracellular transport. Because of their small size and high density in cells, high throughput imaging technologies such as fluorescence microscopy make it harder to trace to extract information such as number and length. Because of this, one of major challenges for building automated methods that "learn" is the availability of limited or no ground truth data of the traces.

I develop a 3D generative model of microtubules and a model parameter estimation approach from confocal fluorescence microscopy images. The estimation approach is an indirect method that compares simulated with real images to estimate model parameters. The chapters in this thesis are organized based on the type of image data (2D vs 3D) and the cell preparation for imaging (fixed vs live cell preparations). Parameters are extracted from images of microtubules in the presence of nocodazole (a microtubule depolymerizing drug), showing the numbers and lengths to decrease over time, and from cell types of different lineages where their numbers and lengths are compared. Continuing on theme of building hierarchical conditional models, I describe a vesicle location model conditioned on a model of microtubules. The final chapter concludes with a summary with its implications and future work.

Acknowledgments

I do not have enough words to sincerely thank my advisors Dr. Robert Murphy (Bob) and Dr. Gustavo Rohde for the very best thing a student could ask for: training me to think "computationally". This has had a profound impact in every aspect my life. Their patience and support have allowed me to learn so many things all through these years.

I would also like to thank my thesis committee members: Drs. Russell Schwartz, Philip LeDuc and George Tseng. Their guidance during early PhD years with fascinating interdisciplinary computational biology courses and later for my thesis research have been invaluable to me.

I would like to sincerely thank my colleagues at Carnegie Mellon: Luis Coelho for being a great Lane Center office-mate and research buddy; Tao Peng, Estelle Glory and Justin Newberg, for making it easy to transition into Bioimage Informatics; Wei Wang, Andy Chen and Anupama Kuruvilla for fascinating Linear Algebra and pattern recognition discussions. I also want to sincerely thank my Murphy lab colleagues: Jieyue Li, Ivan Cao-Berg, Aparna Kumar, Taraz Buck, Armaghan Naik, Josh Kangas, Baek-Hwan Cho, Shannon Quinn and Greg Johnson for the great corridor discussions that almost never end. A special thanks to Lane Fellows Arvind Rao and Le Song who made complex statistics and machine learning concepts easy to understand.

I cannot thank enough my collegues in the PhD program: Justin, Jacob, Ying, Lidio, Suvrajit, Sabah, Om, Anindita, Grace, Guy, Andrej, John Sekar, Ming and Cordelia. They have all made our program bigger and better, and more fun. A special thanks to awesome colleagues from my batch: Arvind Ramanathan and Ahmet Bakan, for making my journey nothing but great all through these years.

Ever since I arrived in Pittsburgh, my friends have been there always for any kind of fun. Starting with Varsha, Sangeetha and Nitin, our group had grown to awesome people like RK, Jhun, Advay, Merve, Kiran, Sammanaz, Anmol, Saumya, and so many other close friends that the list goes on and on. I want to reserve a very special thanks to Varsha for introducing me to this amazing PhD program. Without her special friendship, none of

this work would be possible.

I want specially thank friends from the ECE department. What started as colleagues at an Optimization class, my friendship with Dhruv Batra, Devi Parikh and Kshitiz Kumar has bloomed into a long-lasting one.

I am extremely fortunate to have a great program coordinator: Thom Gulish. He made my transition to Carnegie Mellon smooth, and always took care of anything that I needed. Alongside work, over coffee, we have enjoyed discussions on politics that really gave me those critical breaks needed during work.

I would like to thank my Mom and Dad for the gift of free thinking. They have always encouraged me to do whatever I wanted in life and guided me in every step of the way. My brother, Sajid, has been a major source of fun, feedback, and a voice on the other side of the phone on any random discussion that I want to have. I also have to thank my uncle and aunt in Virginia for being my "escape" to family and to great food that I miss from home. And finally, I have to thank my luck for meeting an amazing person - my wonderful wife, Afreen. Along with her family, her support in every step of the way has been incredible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction[1]

Understanding the many complex cellular and subcellular processes underlying biological phenomena will require approaches for obtaining spatiotemporal information for the tens of thousands of proteins expressed in a typical cell. These measurements (in many cases in the form of statistical estimates) can then be used in modeling and simulation efforts where the goal is to predict and help understand cellular systems. Such measurements can also be used to compare cell types across conditions.

Microtubules are filamentous structures located in the cytoplasm of cells that play critical biological roles in cell division, cell motiliy and intracellular transport. They are present in many different cell types in the both eukaryotic and prokaryotic systems in wide ranging intracellular locations such as the axons fibers of neurons and the hairs of plant roots. Apart from their biological importance, microtubules are fascinating molecular machines to study in vitro because of their self-organizing properties. This chapter

---

[1]Part of this chapter is from [Shariff et al., 2010b,a]

introduces an automated method to build generative models of microtubule distributions in cells and to quantify biologically relevant information from fluorescence microscopy images of microtubules from intact cells.

1.1   Structure, behavior and distribution of Microtubules

Microtubules grow out from the microtubule organizing center (MTOC), such as the centrosome in eukaryotic cells, for example forming a star networked structure. A protofilament is formed by the polymerization of a heterodimer, that is made up of two monomers - the $\alpha - tubulin$ and $\beta - tubulin$ subunits. A microtubule consists of 13 such protofilaments that are held in a parallel fashion forming the wall of a tubular structure Lodish et al. [2007]. See Figure 1.1 from [Li et al., 2002] for a model of a microtubule region derived from cyro-electron microscopy.

Like actin filaments, microtubules have a polarity with a $plus$-end and a $minus$-end. The $minus$-end is attached the centrosome where nucleation occurs, and the $plus$-end is where the growth (and shrinkage) occurs. The nature of the growth and shrinkage is such that they occur in phases: microtubules either grow or shrink and they do so rapidly and appear to function in a stochastic manner. Their behavior is a function of (1) availability of tubulin monomers, (2) forces acting on the microtubules, and (3) the availability of other molecules such as Mg^{2+} and GTP. The concentration of monomers where the polymerization equals the depolymerization is known as the critical concentration. Because of these dependencies, the behaviors of microtubules can be controlled using temperature, pH, drugs, mechanical stress, etc. Also, because of the dynamic nature of microtubules,

the structure of microtubule network consists of varying numbers and lengths of filaments.

Microtubules attached to centrosomes are located in the cytosolic space of mammalian cells where they form two different structures depending on the cell cycle stage. In the interphase stage, the centrosome is located at the close to the nucleus [Malone et al., 2003], and the microtubules grow out and away from the centrosome. The density of microtubules is very high at the centrosome and the density decreases as we move away from the centrosome with an exception at the cell membrane, where microtubules tend to overlap increasing the density. In the mitotic stage, this apparatus dissipates and forms a spindle-like structure [Lodish et al., 2007] shedding light on the highly dynamic nature of microtubules. Figure 1.2 from [Alberts et al., 2002] shows a cartoon of the structure of microtubules in these two stages of the cell cycle.

1.2   Interaction of microtubules with other proteins

Microtubules are known as cellular highways as they allow for vesicles to be transported inside that cells that allow for vesicles to cruise using molecular motors [Bloom and Goldstein, 1998]. Proteins that interact with microtubules are known as microtubule associated proteins or MAPs that include molecular motors such as dynein or kinesin. These proteins are critical for intracellular transport of vesicles (in axons of neurons or in T-cytotoxic cells) and also in the formation of mitotic or meiotic spindle during cell division.
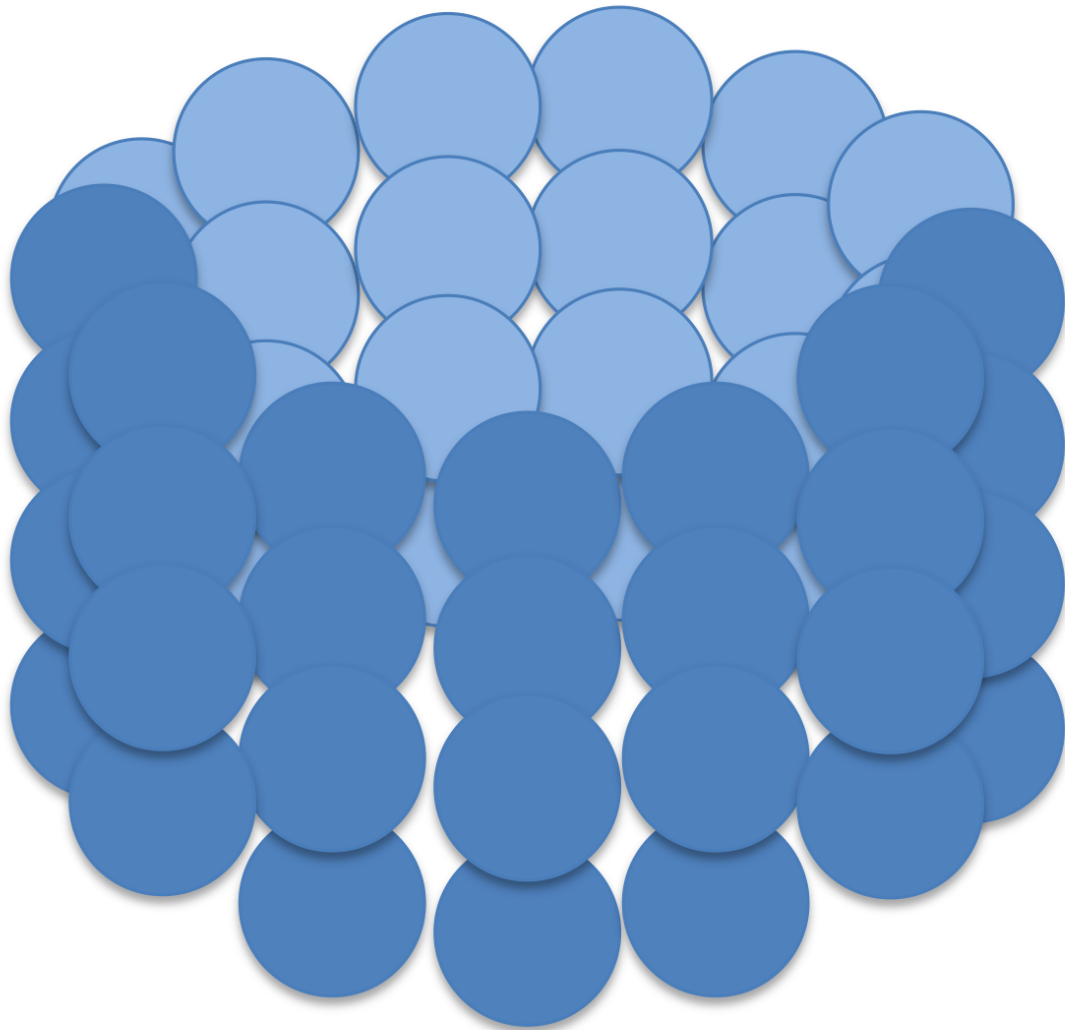
Figure 1.1: A cartoon of the tubular structure of microtubules based on a figure derived using cryo-electron microscopy from [Li et al., 2002]
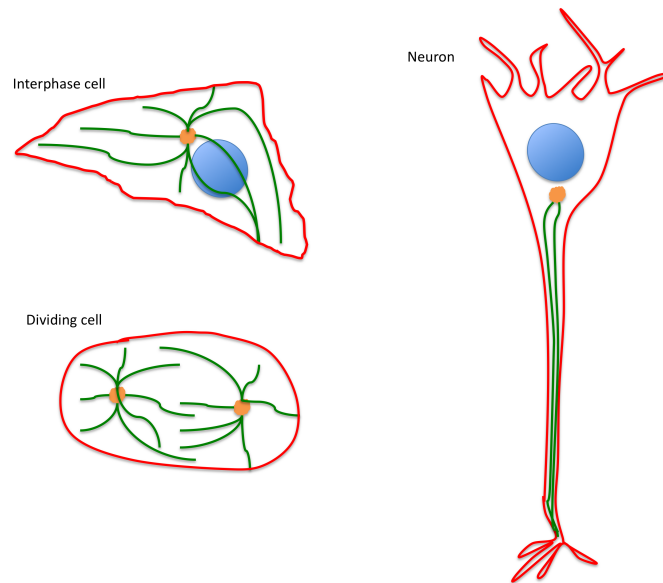
Figure 1.2: Intracellular organization of microtubules based on figure from [Alberts et al., 2002]

1.3 Role of microtubules in disease

Microtubules are known to be a target in plenty of anti-cancer drugs [Jordan and Wilson, 2004]. Because of their importance in cell division, drugs such as taxol act on microtubules by stabilizing them and preventing disassembly thereby inhibiting their ability to further divide.

Malfunction in the microtubule-dependent vesicle transport is directly related to disorders such as kidney disease [Hamm-Alvarez and Sheetz, 1998]. Many neurodegenerative diseases such as Alzheimers and Parkinsons are related to the malfunction of microtubule associated proteins and the microtubule network that leads to the accumulation of protein aggregates in brain cells [Richter-Landsberg, 2008]. Hence, acquiring quantitative information such as number of microtubules or the length distribution of microtubules may help in quantifying the interaction between microtubules and vesicles. Other quantitative information include kinetic parameters reflecting the dynamics of microtubules.

1.4 Simulations of microtubules

Because microtubule organize and interact with other molecules performing many critical cellular functions, cell simulations are necessary to take into account the behavior of the entire microtubule system and the cell. Cells simulations of microtubules have been performed in interphase cells [Tsaneva-Atanasova et al., 2009] and in cells undergoing division [Loughlin et al., 2010] elucidating some of the biophysical interactions between microtubules and its associated proteins within the cell. Automated methods that extract

quantitative information directly from image data for the purposes of cell simulation are critical to increase accuracy and to capture the heterogeneity of protein locations in cells, and that in turn may potentially contribute to heterogeneity in behavior.

1.5 Existing work acquiring such quantitative information

There are several direct methods for estimation of microtubule parameters by tracing described in the current literature. For these, however, the imaging approach is either not suitable for intact cells, or the image resolution is not sufficient to discern individual microtubules throughout the entire extent of the cell [Jiang et al., 2004, Lebbink et al., 2007, Liang et al., 2002, Santamara-Pang et al., 2006, Sargin et al., 2007]. This can be seen in Figure 1.3 in which the high density of microtubules near the centrosomal region makes it impossible to visually or computationally extract individual tracks. Even in regions where individual tracks can be discerned (often near the boundary of the cell), tracing algorithms are invariably hindered by crossing tracks. One solution is to use specialized microscopy methods that greatly enhance estimation of filament like structures: Fluorescence Speckle Microscopy [Ponti et al., 2005], Fluorescence Correlation Spectroscopy [Sisan et al., 2006] and Stimulated Emission Depletion microscopy [Donnert et al., 2006]. However, these methods are not easy to apply on a proteome scale.

Indirect approaches, on the other hand, are more suitable for filament structures since the structures themselves do not have to be matched exactly but rather the pattern they form

in an image is matched instead. A compelling example of such an approach was used to validate models of the mitotic spindle [Sprague et al., 2003]. In that study, however, very limited and simple image features such as mean of fluorescence intensity was used to compare patterns in the images. Another excellent example of an indirect method was analysis of the structure and dynamics of the actin filament network in the lamellipodia of a migrating cell [Schaub et al., 2007]. However, images in this work were cropped to a representative region in the lamellipodia that would not be expected to yield accurate estimate parameters for the entire cell. The method of comparison used only a distribution of correlation lengths from images, which may not be adequate to completely quantify complex patterns in images resulting from overlapping filament structures.

1.6 Approach of this thesis to addressing these biological questions

1.6.1 Bioimage Informatics

This thesis falls under the field of bioimage informatics where automated methods are used to quantify and interpret images of cells and tissues from different modalities [Coelho et al., 2010]. The origins of the field of bioimage informatics in the mid 1990s were in the development of automated microscope systems that included hand-constructed automated analysis algorithms [Giuliano et al., 1997] and the successful application of machine learning methods to recognize subcellular patterns [Boland et al., 1997]. The algorithms behind bioimage informatics are firmly rooted in signal processing, providing a sound theoreti-
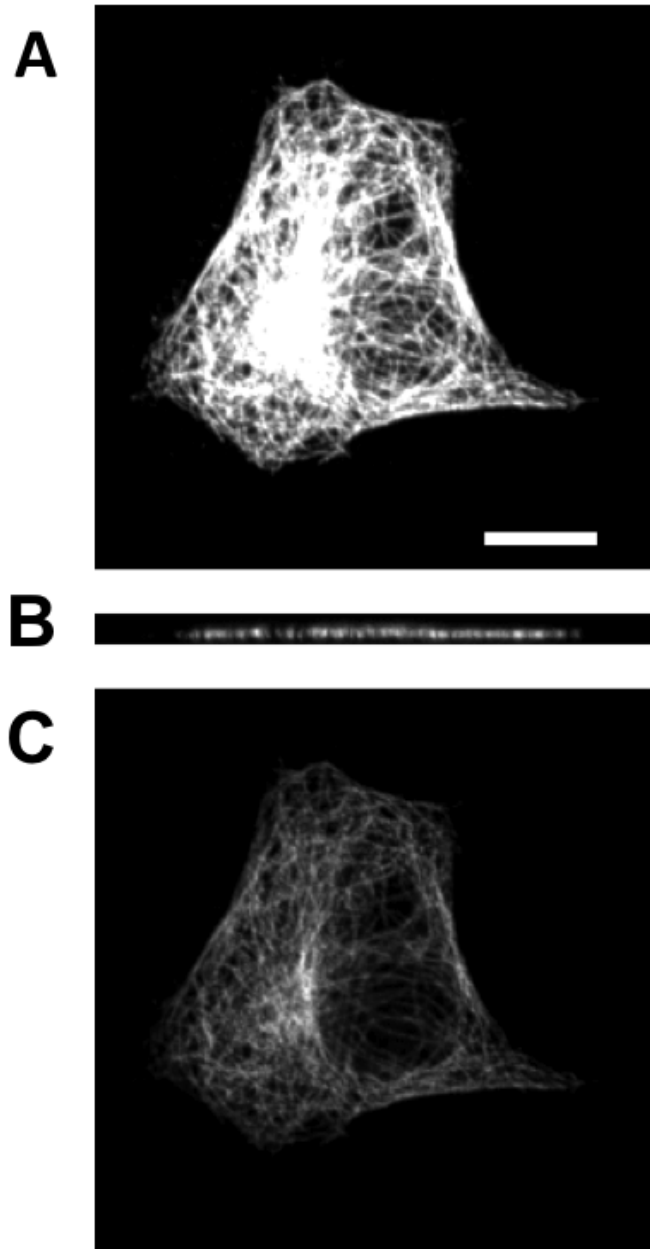
Figure 1.3: Example image from the HeLa image showing the intracellular organization of microtubules

cal foundation for using machine learning techniques to extract meaningful information from large sets of bioimages. Spatio-temporal events within a cell can be captured by microscopy, and quantified through image processing and machine learning methods to produce meaningful conclusions about the data within the experimental context.

For any given task, in order to determine how many images should be acquired, it is best to be able to accurately characterize the variability for a given cell type under a given condition. Hence the more data, the better. In a task where the desire is to assign a label to a protein location pattern, images for one to ten cells can be sufficient to place a protein into a known location class. However, it may require images for as many as 50 to 100 cells to adequately learn a new category.

Depending on the biological domain, image processing techniques such as segmentation/object detection [Coelho et al., 2009], tracing [Al-Kofahi et al., 2008], tracking [Smal et al., 2008] and registration [Al-Kofahi et al., 2003] can be used. See Figure 1.4 for an outline of the steps that is adapted from [Shariff et al., 2010b]. Briefly, image segmentation is used to separate cells in a field, object detection can be used to get the shapes of objects such as cell, nucleus, vesicle and other organelle boundaries, tracing is used to quantify the numbers, lengths and relative sizes of branching or filamentous structures in images, tracking is used to capture the dynamics of movement inside cells from one image frame to the next, and finally image registration is the application of a geometric transformation to align an object in one image to a template object in another image.

For interpretation of the data, machine learning can be used to train an automated method [Bishop, 2006]. Image features, numerical descriptors that can be computed di-

10
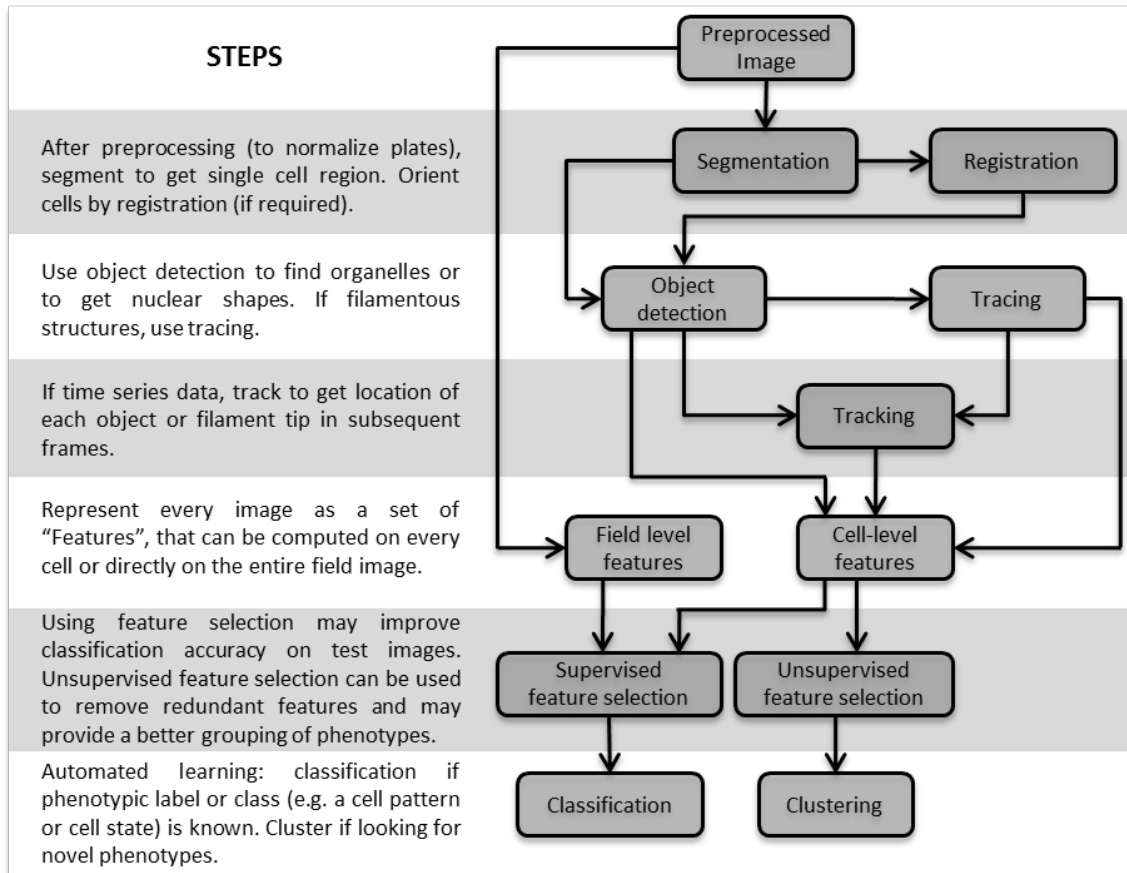
Figure 1.4: Outline of the steps in Bioimage Informatics. Adapted from [Shariff et al., 2010b]

rectly from an image to represent its important aspects or the pixel values can be directly used. Feature selection and extraction methods can be used to select a subset of the features or create new sets of features by recombining the original features that are most informative in discriminating the various classes. Unsupervised methods such as clustering or semi-supervised methods can be used when there are no or limited instances respectively of labeled data.

1.6.2 Fluorescence microscopy images are useful to acquire images of proteins localized in intact cells

Many approaches have been described for obtaining subcellular location data of large numbers of protein distributions [Ross-Macdonald et al., 1999, Hoja et al., 2000, Jarvik et al., 2002, Koroleva et al., 2005, Kumar et al., 2002]. Green fluorescent protein (GFP) tagging has emerged as the most widely used tool for this purpose and has enabled proteome-scale studies (see [Huh et al., 2003] for a prominent example using GFP-fusions in yeast). A notable exception is the work by the Human Protein Atlas project [Barbe et al., 2008, Uhln et al., 2005], which uses antibody-based methods and has generated millions of images for over six thousand antisera against various proteins. In either case, fluorescence microscopy is used to acquire and to interpret the information content in such collections visually, although automated approaches can play an important role in extracting more detailed quantitative information from them [Glory and Murphy, 2007].

1.6.3 The need for generative models

Potential frameworks to characterize protein location patterns from such image data include descriptive techniques and generative models. In short, descriptive techniques seek to describe the content of images using numerical feature vectors, one vector per cell or image. These techniques enable automated subcellular location determination using supervised learning approaches (see [Boland and Murphy, 2001], for an example) but, in the absence of any associated modeling technique, they cannot be used to provide quantitative physical information pertaining to the protein distributions. Generative models, on the other hand, generalize from examples by learning a description of the underlying process believed to give rise to the image [Pece and Larsen, 2007, Murphy, 2007]. We have previously described a framework to learn generative models of multiple subcellular location patterns from cells [Zhao and Murphy, 2007]. Cell membrane, nuclear and protein object models were constructed so that simulated images representing seven different subcellular location patterns could be generated. In short, one way to fully understand the location patterns of individual proteins in a given cell type is to summarize this information in the form of a model that can accurately represent the statistical variation contained in a set of fluorescence microscopy images.

1.7 Microtubule generative models and model parameters

In the context of generative models, this thesis seeks to demonstrate that physically meaningful parameters describing the process by which microtubule distributions are generated can also be learned from fluorescence microscopy images. The work presented in this thesis is builds on the previous modeling framework, which represented protein distributions as a collection of distinct objects [Zhao and Murphy, 2007], to protein distributions such as microtubule networks, that cannot be easily represented as objects.

Specifically, this thesis describes a generative model of microtubules, and an indirect approach for the estimation of model parameters. The publications that resulted in the context of the work in this thesis is described below:

1. L. P. Coelho, A. Shariff, and R. F. Murphy. Nuclei Segmentation In Microscope Cell Images: A Hand-Segmented Dataset And Comparison Of Algorithms. Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging (ISBI 2009), pp. 518-521.

2. A. Shariff, G. K. Rohde, and R. F. Murphy. Indirect learning of generative models for microtubule distribution from fluorescence microscope images. Proceedings of the ICML-UAI-COLT 2009 Workshop on Automated Interpretation and Modeling of Cell Images (Cell-Image Learning 2009).

3. A. Shariff, R. F. Murphy, and G. K. Rohde. A Generative Model of Microtubule Distributions, and Indirect Estimation of its Parameters from Fluorescence Microscopy

Images. Cytometry A. 2010 May;77(5):457-66.

4. L. P. Coelho, E. Glory-Afshar, J. Kangas, S. Quinn, A. Shariff, and R. F. Murphy (2010) Principles of Bioimage Informatics: Focus on machine learning of cell patterns. Lecture Notes in Computer Science 6004:.8-18.

5. A. Shariff, J. Kangas, L. P. Coelho, S. Quinn and R. F. Murphy (2010). Automated Image Analysis for High Content Screening and Analysis. J Biomol Screen. 2010 Aug;15(7):726-34.

6. A. Shariff, R.F. Murphy, and G.K. Rohde (2011) Automated Estimation of Microtubule Model Parameters from 3-D Live Cell Microscopy Images. Proceedings of the 2011 IEEE International Symposium on Biomedical Imaging (ISBI 2011), pp. 1330-1333.

7. J. Li*, A. Shariff*, G.K. Rohde, and R.F. Murphy (2012) Estimating and comparing microtubule distributions from fluorescence microscopy images of different human cell types. Submitted.

Chapter 2 focuses on describing microtubule generative models and estimation of model parameters from 3D fixed cell fluorescence microscopy images where an assumption of no free tubulin is made. Chapter 3 focuses on both polymerized and free tubulin generative models and its respective model parameter estimation from 3D live cell fluorescence microscopy images. Chapter 4 describes a method to extract 3D model parameters from 2D fixed cell fluorescence microscopy images. Chapter 5 describes how to build

models of protein locations conditional on microtubules. Chapter 6 concludes with a summary of the thesis, with implications of the work.

Chapter 2

Microtubule generative models and estimation from 3D fixed cell fluorescence microscopy images[1]

2.1   Background

One of the major challenges in developing a system that can automatically acquire quantitative structural information about microtubules is the availability of image data that shows all the location information describing microtubules. Microtubules are a three-dimensional filamentous structure inside cells that grow out from the centrosome and dynamically occupy the cytosolic space including the volume around the nucleus. In this chapter, 3D

---

[1]Part of this chapter is from [Shariff et al., 2010a]

fluorescence microscopy images of microtubules are used because 3D images capture all the intensity information of all microtubules present in the cell. There are basic two components to the methods in this chapter: generative modeling of simulated images, and model parameter estimation from 3D images.

2.1.1 Generative Modeling of Microtubules

In order to extract biologically relevant information from microtubules, domain knowledge is incorporated into the system by formulating models of proteins from which artificial images are generated (according to initial estimates of the parameters of the model). As described in Chapter 1, since microtubules have a filamentous structure, a Gaussian object representation cannot be used as was the case for generating vesicle-like structures [Zhao and Murphy, 2007]. Although, texture synthesis approach that was previously used for nuclear texture can be applied to microtubules, biologically sensible parameters such as number of microtubules cannot be used for modeling.

2.1.2 Model parameter estimation from 3D images

It is hard to directly estimate structural information about microtubule networks from raw fluorescence microscopy images. This is because there is a high density of microtubules at the centrosomal region making it hard to discern one filament to another. To account for this drawback the model parameters are iteratively modified until a specified similarity measure between the real input images and the simulated ones is maximized. The critical steps in this procedure are shown in Figure 2.1 and include microtubule pattern genera-

tion, image simulation, and comparison with a real microscopic image. These steps are assembled into an optimization procedure to be detailed below.

2.2 Three dimensional Image Data and Preprocessing

Images of 3D HeLa cells previously obtained by three-color confocal immunofluorescence microscopy [Velliste and Murphy, 2002] were used. This collection contains approximately 50 images for each of nine different proteins, including tubulin. Each image consists of three channels, one reflecting the distribution of DNA (as visualized with propidium iodide after RNAse digestion), total protein (as visualized with a non-specific reactive probe), and a specific protein (as visualized with a well-characterized monoclonal antibody). The spacing between voxels in the image is 0.05 microns in the focal plane (the X and Y directions) and 0.2 microns in the axial dimension (the Z direction). The raw images were first downsampled in the X-Y dimension due to memory and computational issues from 0.05 microns to 0.2 microns per voxel. Hence the final voxel spacing is uniform in all three directions; the number of voxels in the X or Y dimension reduced from 1024 to 256.

2.3 Estimation of the point spread functions (PSF)

Three point spread functions were estimated for the cell membrane, nuclear membrane and alpha-tubulin-GFP channels. The point spread functions for the cell membrane and nuclear channels were estimated using the Diffraction PSF 3D ImageJ plugin:

http://www.optinav.com/Diffraction-PSF-3D.htm. The plugin outputs the emission point spread function. The confocal point spread function is approximated as the square of the emission point spread function. The point spread function for alpha-tubulin-GFP channel was directly estimated from the fluorescence microscopy image. Line intensities along the X dimension and along the Z dimension were computed with clearly distinguishable and well separated microtubules wrapped around the nucleus. The line profiles were registered and truncated to size 7, and averaged for the X and Z dimension. A 3D Gaussian was manually fit and was used as the point spread function.

2.4 Segmentation to estimate the cytosolic space

Each channel of each image was corrected for background fluorescence by subtraction of the most common pixel value and deconvolved with a theoretical point spread function for the nuclear channel and the cell membrane channel. The images were segmented into single cell regions using seeded watershed segmentation. The cell boundary and nuclear boundary in each slice was then found using the active contour method on the deconvolved cell membrane channel and nucleus channel respectively [Chan and Vese, 2001].

2.5 Generative Model of Microtubule Patterns

Typically, microtubules grow out from the centrosome and grow within the cytosolic space of the cell. Hence, a generative model of the microtubule pattern must be conditioned (dependent) on a nuclear model and a cell membrane model. In order to build a model from

20

a 3D cell image, the nuclear and cell membrane channels were deconvolved with their respective point spread functions described above and segmented semi-automatically using the Active Contour without Edges approach. The central point from which microtubules grow is the centrosome, and its position can be directly estimated from the tubulin channel. The tubulin image was convolved with a 3x3x3 averaging filter. The location of the centrosome was estimated to be the voxel with the maximum intensity. Figure 2.3 shows the cell boundary, the nucleus boundary and the centrosome location for a slice of the image in Figure 2.2.

The growth model consists of generating different numbers of microtubules (each with a specified length) by extending short segments starting from a single point in the cytosolic space (the centrosome). The model of microtubule distribution was constructed using a growth model conditioned on the centrosome location, cytosolic space and the parameters of the model. The growth model consists of generating microtubules as points on a star network with the hub as the centrosome. Let X denote the location, in three dimensions, of the center of the centrosome of a given cell. Assuming the centrosome a sphere, the diameter of a centrosomal structure was fixed to be approximately 0.4 μm. N random points were generated $X_{i,j} : i \in Z, 1 \leq i \leq N, 1 \leq j \leq n_i$ inside the volume of the sphere where N is the number of microtubules to be generated, and n_i is the number of points for each microtubule i. Each point in the sphere is extended in a random direction to a new point $X_{i,1}$ with steplength γ. These short segments are further extended by picking a point $X_{i,2}$ with step length γ that satisfies two constraints. The stiffness constraint is as follows:

$$cos\alpha \leq v_1 \cdot v_2 \leq 1$$

21

where

$$v_1 = \frac{X_1 - X_0}{\|X_1 - X_0\|}$$

$$v_2 = \frac{X_2 - X_1}{\|X_2 - X_1\|}$$

and α is the angle between $(X_2 - X_1)$ and $(X_1 - X_0)$. In our model, $cos(\alpha)$ is called the collinearity parameter. Points are also constrained to be generated in the cytosolic space using a lookup image that was estimated using segmentation.

The length distribution was modeled as a truncated normal distribution [Johnson et al., 1994]. The normal distribution is truncated such that there can be no negative lengths. This distribution was shown earlier to fit the lengths of microtubules well in the meiotic spindle [Yang et al., 2007]. The random variable $X \sim N(\mu, \sigma^2)$ conditioned on $(0 < X < \infty)$ follows a probability density function:

$$f(x; \mu, \sigma, a, b) = \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{1 - \Phi(\frac{-\mu}{\sigma})}$$

where ϕ is the probability density function of the standard normal distribution and Φ is the cumulative distribution function. This distribution is sampled N times, where N is the number of microtubules. The microtubule elongation procedure is iterated for each of N microtubules, until the sampled lengths of the microtubule polymer is satisfied. The following are thus the model parameters:

1. Diameter of the centrosomal sphere: 0.4 μm (fixed)

2. step length: γ (fixed)

3. number of microtubules: n

22

4. collinearity: $cos\alpha$

5. mean of the normal distribution: μ

6. standard deviation of the normal distribution: σ

2.6 Simulated Pattern to Simulated Image

The microtubule structure model is convolved with the estimated point spread function to simulate a fluorescence microscopy image generation process. The resulting polymerized tubulin image is multiplied by a scalar such that the single microtubule peak intensity from the simulated image matches the mean of the peak single microtubule intensity in the raw image. Given specific values for each parameter, an image can be generated that simulates a microtubule distribution, as it would be imaged under the specified condition. Figure 2.4 shows a model of the microtubule network generated by this method and an image that results from convolving it with a point spread function.

2.7 Grid Generation

The model parameters that are varied are the number of microtubules, the mean and standard deviation of the length distribution of microtubules, and the collinearity. The range of the values for the standard deviation of the length distribution and the collinearity $(cos\alpha)$ did not take all possible values, but was based on how much real variation is believed to be present. The parameters varied took the following values:

- n = 5, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300

- μ = 5, 25, 50, 75, 100, 125, 150 microns

- σ = 1, 5, 10, 15, 20, 25 microns

- $cos\alpha$ = 0.9, 0.95, 0.98

Thus, for a given cell morphology, a total of 1651 images were generated.

2.8   Feature Computation

Image features, numerical descriptors that encode the image content, were then calculated for both real and simulated images. (1) Thirteen 3D Haralick texture features were computed from a single co-occurrence matrix for all the 13 directions for each image [Chen et al., 2003]. Two more sets of these features were computed by downsampling the image by two and by four. (2) The image was discretized in subvolumes radially starting from the centrosome. Radial intensity features were calculated by computing the total intensity in these subvolumes and normalizing by their respective volumes. (3) Histogram features were computed that consist of standard measures such as Mean, Variance, Skewness, Kurtosis, Energy and Entropy. (4) The total intensity was computed as a feature that is the sum total of all graylevels values in the 3D image.

2.9 Model Parameter Estimation

We measured the similarity between the query image and each of the simulated images by computing the Normalized Euclidean distance in feature space. In order to do this, a diagonal matrix D was first computed that contain the variances of the features. This variance matrix was then used to compute the Normalized Euclidean distance between a feature vector x_s computed from a set of simulated microtubules (simulated image) and a feature vector x_r corresponding to the image based on which the microtubule simulation was computed (raw image). In this case, the Normalized Euclidean distance is given by

$$d_{rs} = (x_r - x_s)D^{-1}(x_r - x_s)'$$

For any query image, the Normalized Euclidean distances were computed from each of the large grid of simulated images. The optimization problem estimates best fit parameters by minimizing the Normalized Euclidean distance: $[n, \mu, \sigma, cos\alpha] = argmin\ d_{rs}$

2.10 Evaluating the Matching Procedure using Simulated Data

In order to check the models ability to recover parameters when these are known, images of microtubule patterns were simulated using the methodology described. For each simulation we tested whether the estimation procedure could be used to recover the known parameter values. The cost function is dependent on the choice of features computed from the images and the distance metric in feature space. A plot of the Normalized Euclidean

distance as a function of different parameters around the vicinity of the optimal parameters (Number of microtubules = 150, Mean of length distribution = 75 microns, Standard Deviation of length distribution = 10 microns, Collinearity = 0.95) is shown in Figure 2.5. The cost functions for the parameters show clear minima for the number of microtubules and the mean of the length distribution of microtubules, suggesting that the method of minimization could potentially recover parameters. However, in order to test this, we computed the accuracy of the method on simulated data. To determine how well model parameters can be recovered by matching, 400 parameter sets were randomly selected from the grid. Images were generated based on these parameter sets with different random number generator seeds than those from the images in the image grid. Image matching was done with each of the 400 query images.

The error metric used was the mean absolute percentage error (MAPE).

$$MAPE = \frac{1}{400} \sum_{t=1}^{400} \left| \frac{R_t - S_t}{R_t} \right|$$

where

$$R_t - Query\ Image\ parameters,\ S_t - Estimated\ Image\ parameter$$

Table 2.1 shows the average over four realizations of the mean absolute percent error (MAPE) as a measure of the accuracy of recovering model parameters. Various feature sets were tested in various combinations, and the error was observed to be minimum when all the six sets of features were used in the distance function. All the subsequent analyses were performed using all six feature sets in the distance function. Since, the growth model is stochastic, we also studied the error as a function of number of average realizations

Table 2.1: MAPE - Mean Absolute Percent Error estimates (average over four realizations) for recovery of Number of microtubules (N), Mean (μ) and Standard Deviation (σ) of length of microtubules in microns and Collinearity ($cos\alpha$)

FEATURE SETS	MAPE				
	N	μ	σ	$cos\alpha$	**Total**
tot	70	102	141	3.9	316.9
his	12	17	221	2.5	252.5
har	10	19	226	1.5	256.5
ha2	14	24	230	1.7	269.7
ha4	17	32	233	1.7	283.7
rad	30	53	235	1.7	319.7
har, rad	10	19	233	1.2	263.2
tot, har	10	18	223	1.5	252.5
tot, rad	27	48	232	1.5	308.5
his, har	8	15	231	1.3	255.3
har, ha2, ha4	12	21	223	1.5	257.5
tot, har, ha2, ha4	12	21	222	1.4	256.5
har, ha2, ha4, rad	12	21	235	1.3	269.3
his, har, ha2, ha4, rad	9	15	227	1.2	252.2
tot, his, har, ha2, ha4, rad	9	15	226	1.2	251.2

rad - Radial intensity features

his - histogram features

har - Haralick texture features

ha2 - Haralick texture features downsampled by 2

ha4 - Haralick texture features downsampled by 4

tot - total intensity feature

(computing a distance between a query feature vector and an average feature vector over the number of realizations for each parameter set). The error was only observed to decrease by at most a few percentage points (e.g., from an error of 8.7% to 7% for the number of microtubules) as we increased the number of realizations (data not shown). Hence, in order to reduce computation costs, all subsequent comparisons of query images with synthetic images were performed for only a single realization of the parameter set.

2.11   Estimating Parameters from a 3D HeLa Image Dataset

Using this approach, we next estimated parameters from the images in the 3D HeLa dataset. In these computations we restricted the search to be conducted over parameter values that produced images of similar total tubulin as the input real image. This was done by first estimating the amount of variation in the peak intensity of a single microtubule. We chose one standard deviation of this variation and converted it into a standard deviation of total tubulin using the following formula:

$$Total\ Tubulin_{lim} = \frac{\left(\sum_{pixels} image \right) I_{mean}}{\tau I_{lim}}$$

where I_{mean} is the mean of peak single microtubule intensity estimated, I_{lim} is the upper or lower limit of intensity that is one standard deviation away from I_{mean}, and τ is the total intensity from a simulated microtubule point. The simulated images in the grid were searched over this band of total tubulin.

For the 3D image shown in Figure 2.2, the optimal parameters are: number of mi-

28

crotubules = 175; mean of the length distribution = 25 microns; standard deviation of the length distribution = 15 microns and the collinearity = 0.9. The simulated image corresponding to the optimal parameter set based on the matching is shown in the center column of Figure 2.6. In order to check if a visually reasonable match was picked by the algorithm, variations across the best match are also shown with images of varying number of microtubules (A), mean of the length distribution (B), standard deviation of the length distribution (C), and the collinearity of the microtubules (D). The leftmost image of Figure 2.6A shows an example of a bad parameter set that has very few microtubules. Figure 2.7 shows the estimated images and parameters for three cells in the 3D HeLa dataset. We also present the estimated parameters for 42 images from the dataset as histograms for each of the parameters (Figure 2.8).

2.12 Discussion

A model-based approach is presented here to generate microtubule patterns that mimic some of the aspects of microtubule distributions in live cells. The algorithm generates images and measures similarity between each of the generated images and the query image by computing a Normalized Euclidean distance in feature space. The structural information about the microtubule distribution in a query image is approximated as the parameters of the generative model that generated the simulated image with the smallest Normalized Euclidean distance.

A stochastic path generation algorithm is used to create microtubule distributions. The microtubule segments in our growth model are extended using a persistent random

walk procedure where successive segments are related by a range of correlation coefficients Rudnick and Gaspari [2004]. The collinearity parameter used here is a lower bound on the correlation coefficient (with the upper bound fixed at one) that can be understood as a single stiffness parameter. A related stiffness parameter that is commonly used in persistent random walk methods is the persistence length that can also be estimated from our growth model. The persistent random walk growth model is simple approach but has been used previously to generate microtubule filament patterns [Brangwynne et al., 2007].

The parameter estimation approach is validated using simulated data. Using the same modeling for simulation and recovery, results showed that the average error for recovering the number of microtubules in an image was about 9% while the error in the recovery of the mean length parameter was around 15%. We have also extracted microtubule distribution parameters from real images. In this case results are harder to interpret since the correct values are unknown. Overall, the recovered parameters are able to generate images of similar overall appearance to those of the corresponding real images. Also, the ranges of recovered parameter values (Figure 2.8) are of the appropriate order according to the findings in a study of microtubules in intact cells [Gorbsky and Borisy, 1985].

Although the methods have been validated using simulated data, and have used them to estimate parameters that appear to be reasonable from real data, more can be done to further increase confidence in these estimates. In the next chapter, parameters are estimated from cells under conditions where the number and length of microtubules are expected to change (specifically in the presence of microtubule depolymerizing drug: nocodazole).

In addition, the modeling approach can be easily expanded to incorporate more bio-

logically relevant information. For example, the growth model can be made to include kinetic parameters such as growth and shrinkage rates to model dynamic instability of microtubules, or parameters that capture its interaction with molecular motors [Karsenti et al., 2006]. It may be possible to incorporate some of these parameters by mapping them to the current model parameters (such as length distribution).
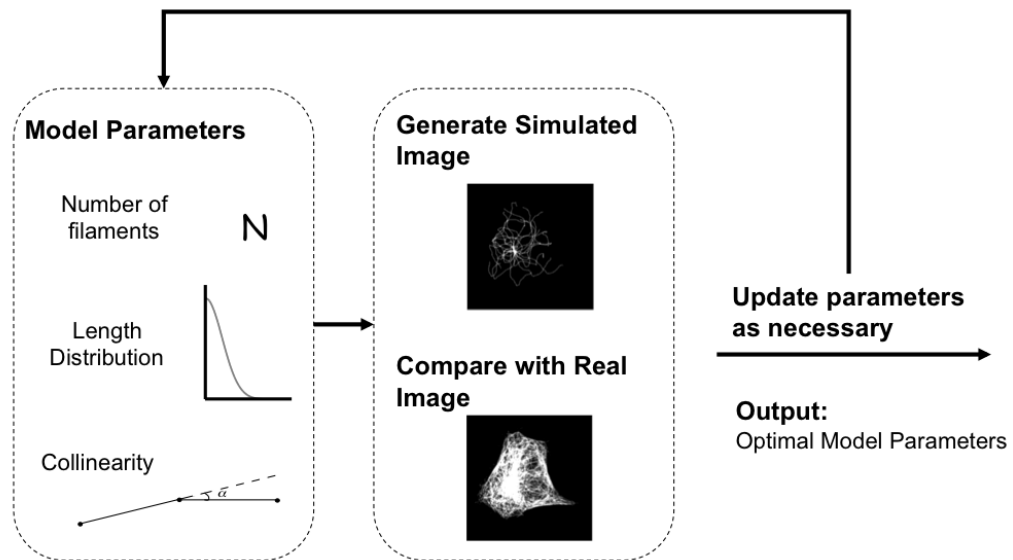
Figure 2.1: Overview of the approach using 3D images
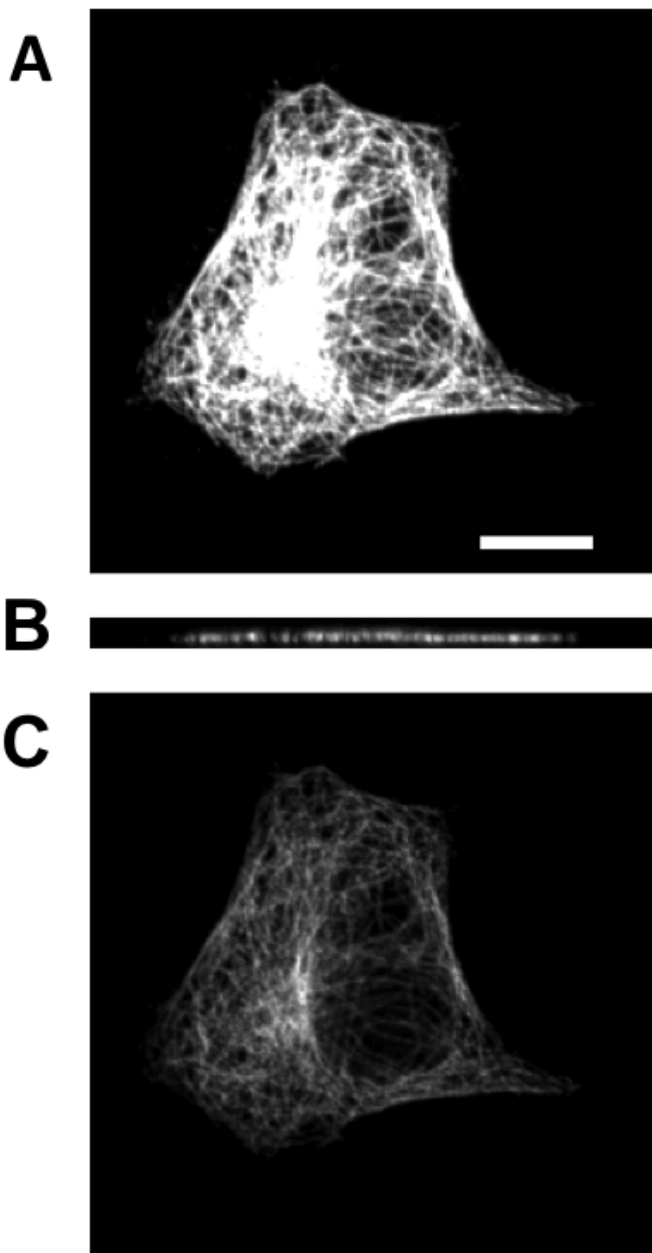
Figure 2.2: Example image from the 3D HeLa dataset. (A) shows the sum X-Y projection of the image (B) shows a slice along the X-Z and (C) shows a slice along the X-Y. The scale bar is $10 \ \mu m$.

Figure 2.3: Cell and nuclear boundaries and centrosome location
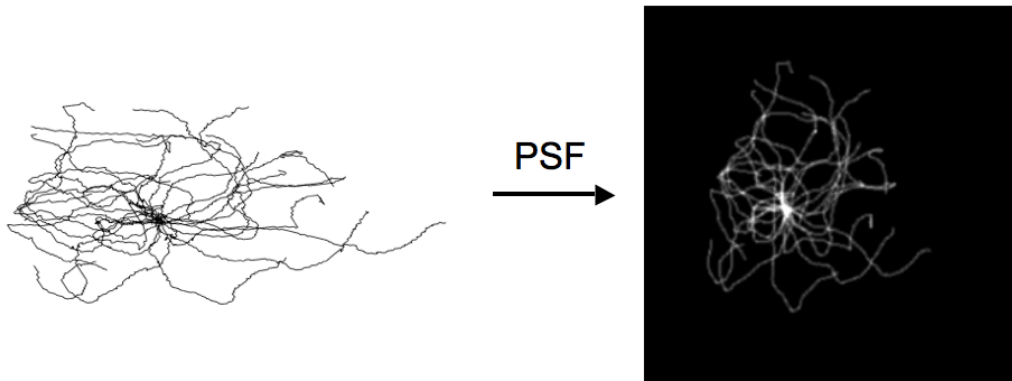
Figure 2.4: An example rendering of a microtubule 3D model (view from a 3D corner) converted to image (sum projected along Z-axis) using a point spread function. The background color is changed to reflect model and image.
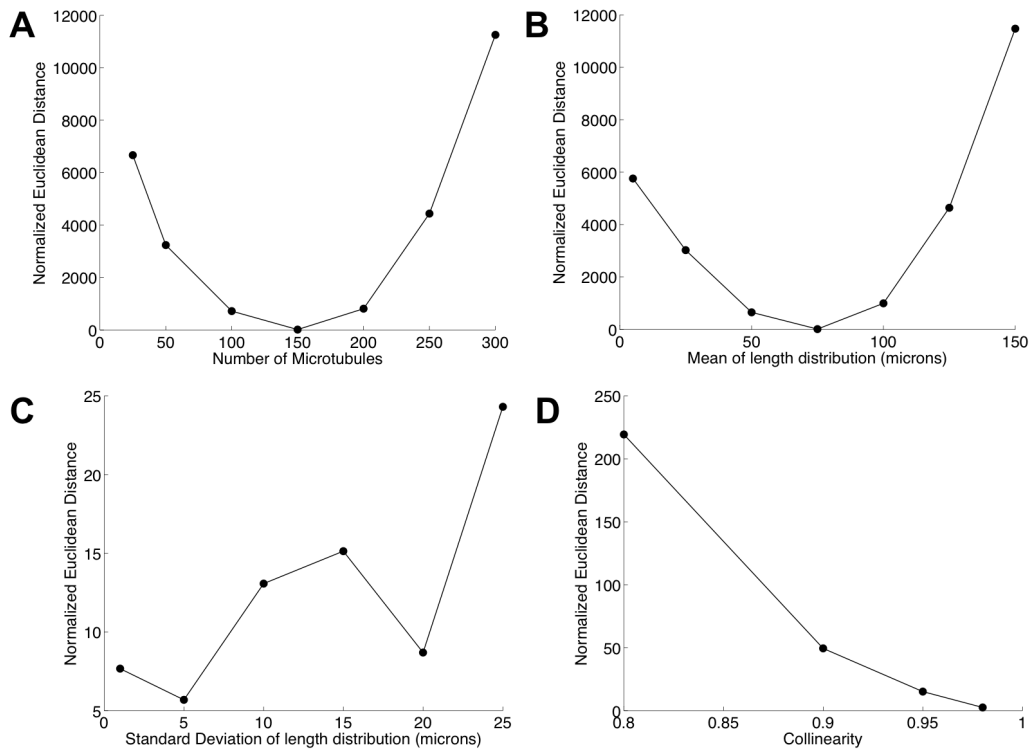
Figure 2.5: Cost function plots for (A) Number of microtubules (B) The mean and standard deviation (C) of the length distribution of microtubules and (D) Collinearity
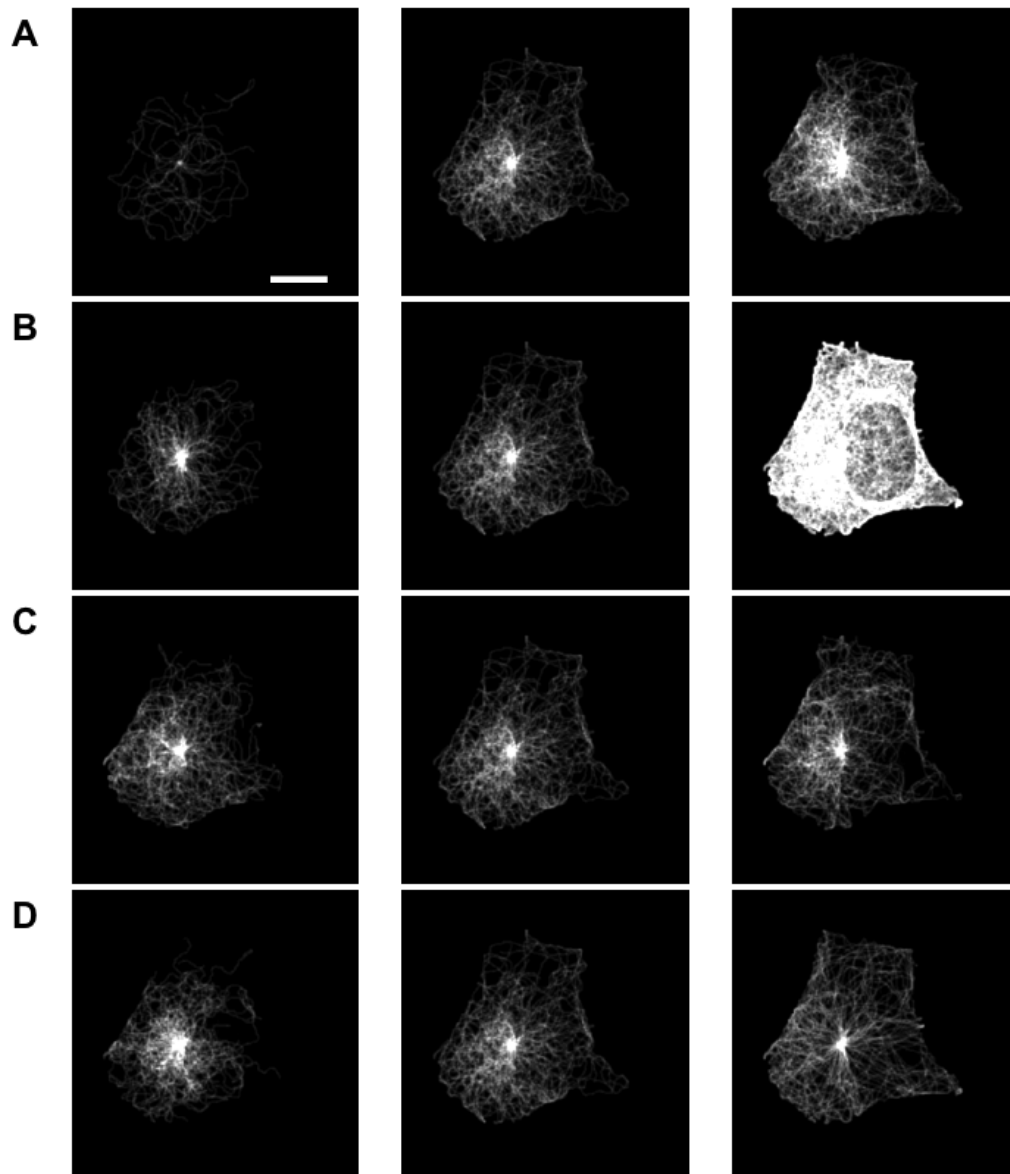
Figure 2.6: Variation on the optimal match for the (A) number of microtubules (B) mean and standard deviation (C) of the length distribution of microtubules and (D) collinearity. The scale bar is 10 μm.

36

Number of Microtubules = 100
Mean of Length Distribution = 50 μm
SD of Length Distribution = 20 μm
Collinearity = 0.9

Number of Microtubules = 50
Mean of Length Distribution = 100 μm
SD of Length Distribution = 10 μm
Collinearity = 0.9

Number of Microtubules = 125
Mean of Length Distribution = 25 μm
SD of Length Distribution = 25 μm
Collinearity = 0.9
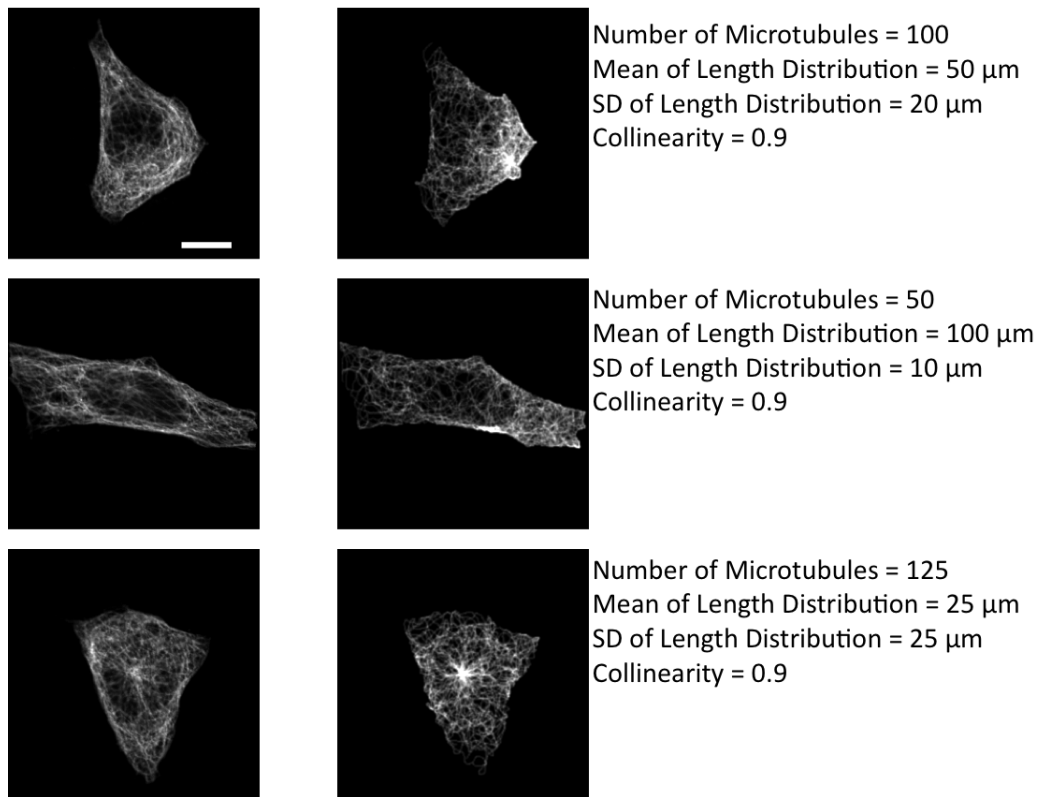
Figure 2.7: Query images (left column) from the 3D HeLa dataset and estimated images (right column) along with the estimated model parameters. The scale bar is 10 μm.
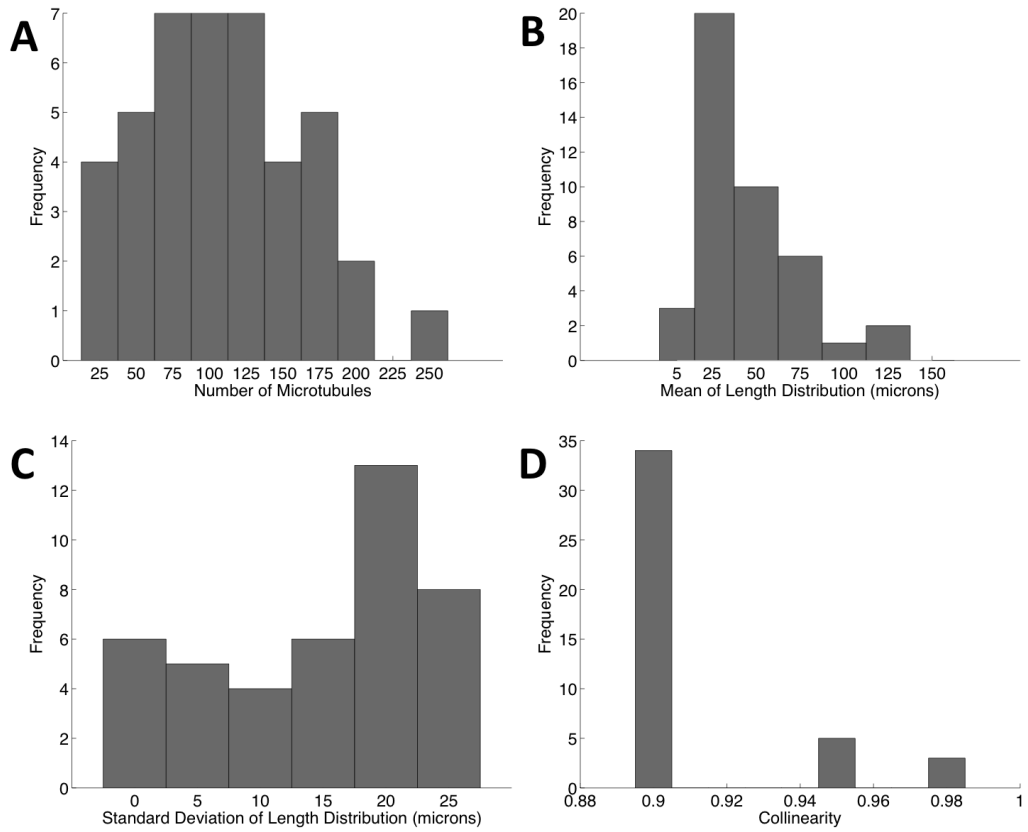
Figure 2.8: Histograms of the parameters estimated for 42 cells of the 3D HeLa dataset. (A) number of microtubules (B) the mean and standard deviation (C) of the length distribution of microtubules and (D) collinearity

Chapter 3

Polymerized and free tubulin generative models and parameter estimation from 3D live cell fluorescence microscopy images[1]

3.1 Background

Fixed cell preparations for imaging requires the use of detergents like Triton-X that permeabilizes the cell membrane that causes the free monomeric tubulin to diffuse away. Since the live cell preparation does not require this, presence of free tubulin must be modeled to

---

[1]Part of this chapter is from [Shariff et al., 2011]

estimate parameters for the indirect estimation approach.

In the previous chapter, a generative model of microtubules was described and an indirect method of estimating its parameters from images that were assumed to not include free tubulin was developed. Also, since whole cell images with known parameters were not available, the ability of the method to accurately estimate model parameters using synthetic images generated using the model was tested. These tests revealed a low error in estimation but estimates for real images could only be described as generally consistent with current knowledge. Here estimation of microtubule model parameters is described from 3D fluorescence microscopy images of live cells under conditions in which changes in those parameters are expected. This was done by acquiring images of living NIH 3T3 cells expressing fluorescently-tagged tubulin in the presence and absence of nocodazole, a drug that is known to depolymerize microtubules [Solomon, 1980].

3.2 Data Acquisition of live cell images of tubulin

3D confocal microscopy images were acquired at five different time points in the presence and absence of nocodazole, keeping all imaging parameters fixed. NIH 3T3 cells expressing EGFP-tagged alpha tubulin were cultured in DMEM supplemented with 10% Fetal Calf Serum and 100 U/ml penicillin and 100 ug/ml streptomycin. The cells were grown to 80% confluency. On the day of imaging, the media was changed to Opti-MEM and a final concentration of 0.5 ug/ml of Hoechst was added to the imaging dish to label the nuclei. The dish was incubated for at least 3 hours in a CO_2 incubator before the image acquisition. The imaging dish was placed in a heated chamber that was maintained at 37 C

throughout the image acquisition. 3D images were acquired using a confocal fluorescence microscope. The spacing between voxels was 0.09 microns in the focal plane and 0.48 microns along the axial dimension. 3D images of five different cells were acquired at 0, 10, 20, 30, 40 min after addition of nocodozale or buffer. Due to photobleaching, full 3D images could not be acquired for the same cell at each time point, and therefore different cells were imaged at each time point (only interphase cells were selected).

Figure 3.1 shows an example set of such images for various times of treatment with nocodazole. Cells treated with nocodazole for 40 min appear to have all of their microtubules depolymerized.

3.3   Fluorescent bead acquisition

As described in Chapter 2, the modeling approach requires a model of the point spread function of the microscope used for acquisition. An empirical estimate of the function was generated using 20 nm fluorescent beads (488 nm absorption). An empirical estimate was used directly instead of a theoretical one since the former tends to be more accurate than the latter. 0.1 ml of a suspension of beads in optiMEM was placed on a clean glass slide and quickly covered by a coverslip. 3D images were acquired as above.

3.4   Generative model of microtubules

The generative model of polymerized tubulin distribution previously described for HeLa cells [Shariff et al., 2010a] was applied to NIH 3T3 with only minor modifications. While

the plasma membrane position for HeLa images was estimated using a fluorescence channel showing total cell protein, this channel was not available in the 3T3 images. The tubulin image itself was therefore used for this purpose since the presence of free tubulin allowed for a reliable estimate of cell boundaries.

3.5   Point spread function

3D images of beads were segmented into individual bead regions using Ridler-Calvard thresholding and registered using the 3D centroid of the bead. The beads were then averaged to estimate the point spread function.

3.6   Free tubulin distribution estimation and generation

As mentioned earlier, the generative model described in the previous chapter only took into account polymerized tubulin because the images were acquired by immunofluorescence staining of fixed cells lacking appreciable free tubulin. This is because permeabilization of cells with detergents like Triton-X to allow antibody penetration causes most of the free tubulin to diffuse away. However, live cell imaging of fluorescently-tagged tubulin detects both free tubulin monomers and polymerized microtubules. Therefore the previous model was extended to account for free tubulin by estimating histograms of free tubulin intensities $h(reg, nz)$ for each nuclear or cytoplasmic region reg and for each 2D slice number nz. Free tubulin regions in each of the 2D slices was estimated by first detecting and removing the polymerized tubulin regions, as follows. The input image was blurred

using a Gaussian filter with standard deviation of 3, and the resulting image was subtracted from the input image. The subtracted image was binarized to separate zero and non-zero pixels. Since the binary image has small clusters of disconnected objects seemingly forming microtubule fibers, the binary image is blurred again to connect objects that are close to each other. This operation was performed using a Gaussian filter with standard deviation of 2. The resulting image was again binarized. This ad hoc approach resulted in a reasonable definition of microtubules (as shown in Figure 3.2). In order to generate free tubulin images for simulations, the histograms $h(reg, nz)$ were sampled to generate the corresponding distribution of free tubulin in all regions of the cell, $f(x)$.

3.7   Tubulin Image Formation

Here, the tubulin fluorescence image formation used for generating simulated images is described. Let $I(x)$ be the tubulin fluorescence image. Let $p(x)$ and $f(x)$ be the polymerized tubulin and free tubulin images respectively. Let $*$ denote a 3D convolution. Then, $I(x) = psf * [p(x) + f(x)]$, where psf is the point spread function of the imaging system (estimated as above). This can be written as:

$$I(x) = [psf * p(x)] + [psf * f(x)]$$

$$psf * p(x) = psf * [\lambda p'(x)] = \lambda[psf * p'(x)]$$

where $p'(x)$ is the model generated in pixel coordinates by the generative model for a given set of parameters and λ is the scaling factor that matches the single polymerized tubulin intensities in the simulated images to the real images (see below). Let $f_2(x) = psf * f(x)$.

43
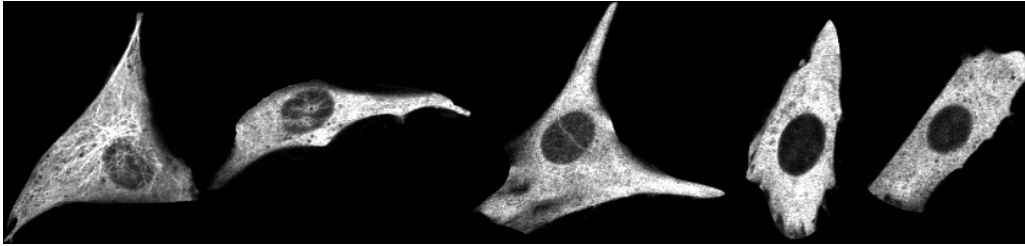
Figure 3.1: Example images of NIH 3T3 cells expressing EGFP-tagged alpha-tubulin at various time points after addition of 20 uM nocodazole (from left to right, 0, 10, 20, 30, and 40 min).
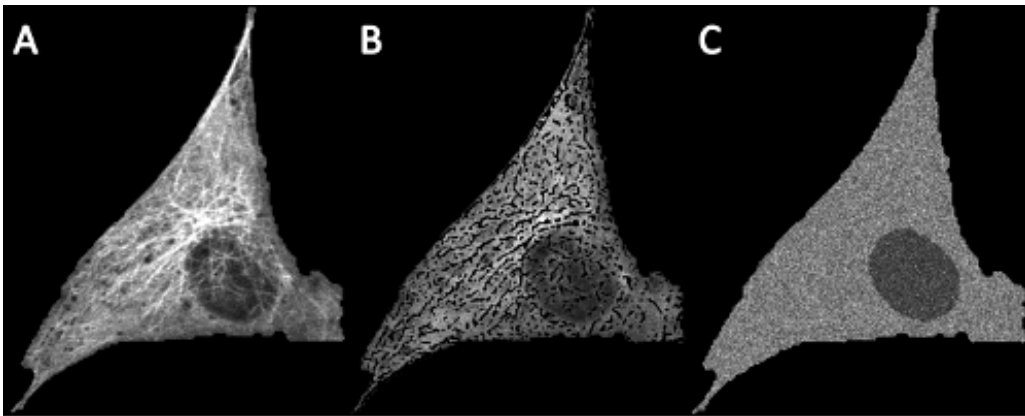
Figure 3.2: (A) 2D slice from a 3D image stack of a cell untreated with nocodazole. (B) Removal of polymerized tubulin (C) Regeneration of free tubulin distribution by sampling from free tubulin intensity histograms estimated from (B).

The above equation then becomes:

$$I(x) = \lambda.[psf * p'(x)] + f_x(x)$$

Hence, for a given set of parameters Θ, $I(x|\Theta)$ can be generated. For a given set of parameters, the amount of free tubulin was adjusted by scaling $f_2(x)$ according to the total amounts (total intensity) available (see Figure 3.3 for an example).

3.8   Single microtubule intensity estimation

The intensity of a single microtubule was estimated from the 2D slice and region just below the nucleus of the cell. The reason for this is that the microtubules (if present) in this region have a very minimal overlap and are generally traceable. λ was defined as:

$$\lambda = \frac{\varphi[p_R(x)]}{\varphi[p_S(x)]}$$

where $\varphi[.]$ is the single microtubule intensity in the real (R) and simulated (S) images. $\varphi[p_R(x)]$ was estimated by averaging tubular pixel values and subtracting out the average free tubulin pixel values. The tubular pixel regions were detected using the method described by Frangi et al. [Frangi et al., 1998] (see Figure 3.4 for an example). The remaining regions were assumed to be free tubulin. $\varphi[p_S(x)]$ was estimated directly from generated polymerized tubulin images $p(x)$. λ was estimated from many images across the dataset and a single average value $\overline{\lambda}$ was used.

45
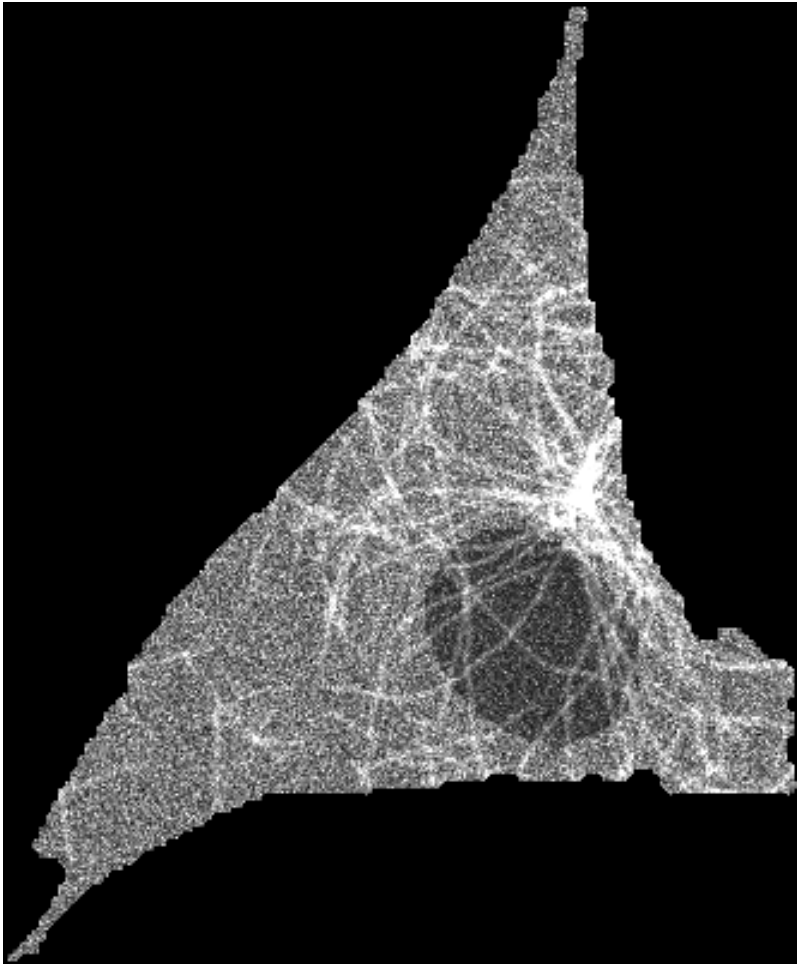
Figure 3.3: A 2D slice in the 3D stack of a simulated image. The image was generated with the number of microtubules set to 100, the mean of the length distribution to 60 microns, the standard deviation of length to 6 microns and the collinearity to 0.9961

3.9   Library generation

As described in Chapter 2, a library of simulated images was generated for all combinations of discrete values of the four parameters:

- Number of microtubules = 0, 5, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220

- Mean of length distribution () = 5, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220 microns

- Coefficient of variation of length = 0, 0.1, 0.2, 0.3

- Collinearity ($cos\alpha$) = 0.97, 0.984, 0.992, 0.996, 1

3.10   Feature selection and Matching

As described in Chapter 2, parameters are indirectly estimated by choosing the synthetic image from the library that is most similar to a given real image. This choice is made using numerical features calculated to describe the fluorescence distributions, and a critical component of this approach is the choice of features and distance function. We describe here a feature selection method to include in the distance function using training data.

All but one of the five images at this time point were therefore used to train the feature selection approach, and the features selected were used to estimate model parameters from all the images except the ones that were used for training. Cells corresponding to the 40-min time point do not appear to have polymerized tubulin.

Therefore features were selected so as to minimize the normalized Euclidean distance in feature space between 4 images of the 40-min time point of nocodazole treated cells and simulated images for 0 microtubules (only free tubulin).

This procedure was repeated by holding out each image in turn (five-fold cross-validation). Figure 3.5 shows the parameter estimates averaged over the five folds and the five replicates per time point. Hence all points are averaged over 25 (5 folds x 5 replicates) except that the last time point is averaged over five folds only. The number and mean of length distribution for nocodazole-treated cells decrease as a function of time, but in the control case, these parameters do not show a decreasing trend. The standard deviation error bars are very large in some of the points. This is because the parameters are averaged over different cells that are likely to have varying numbers and lengths of microtubules because of their varying sizes. However, there is a clear decrease in the number and mean of the length from the first and last time points in the nocodazole treated case as opposed to the untreated case.

3.11   Discussion

A microtubule distribution estimation system was validated by estimating parameters from an image set of live cells. The estimated parameters follow the expected trend: cells treated with nocodazole tend to have less polymerized tubulin. Future work will include improving many of the image processing routines to achieve higher efficiency and robustness, as well as exploring the dependence of the estimates on the accuracy of the point spread function.
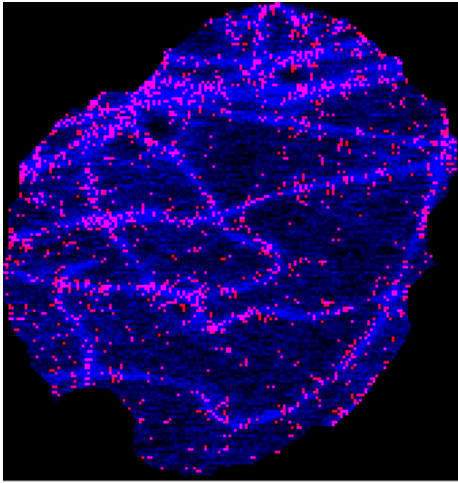
Figure 3.4: Single microtubule intensity detection on microtubules in a slice just below the nucleus. The tubulin image is shown in blue and the points identified as showing a single microtubule are marked in red.
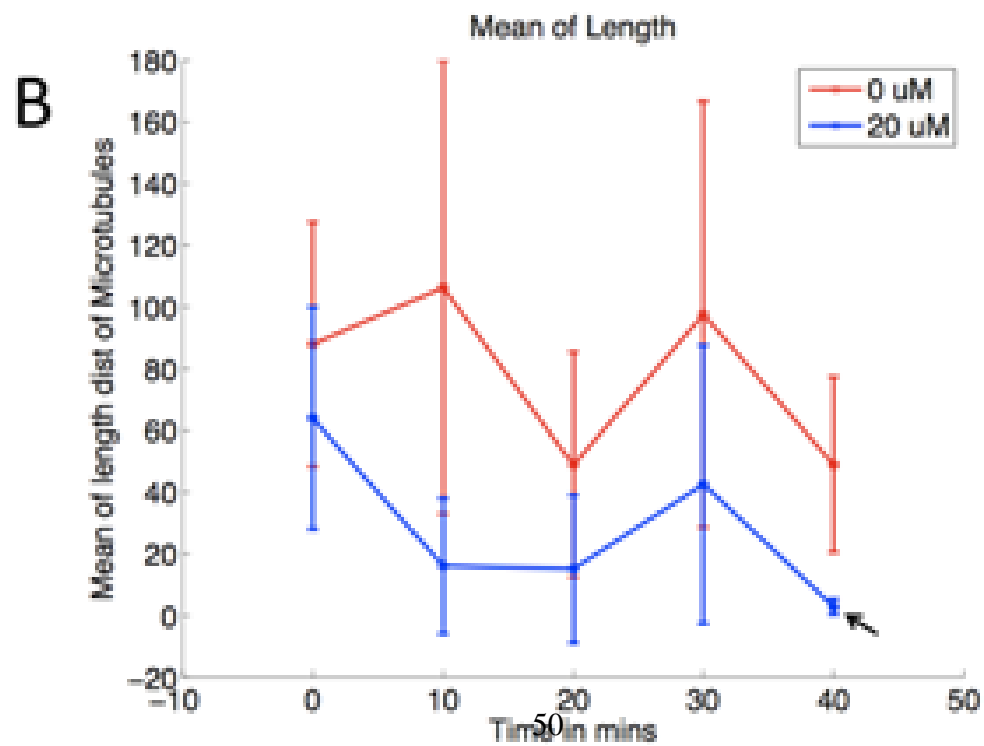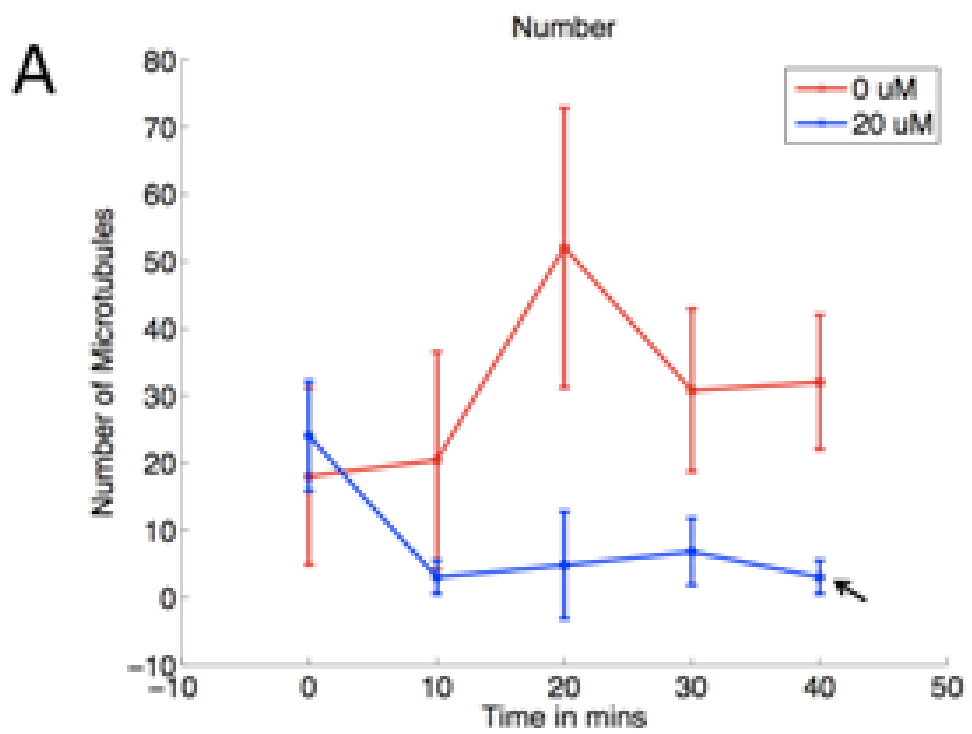
Figure 3.5: Parameter estimates of the number (A) and mean length (B) averaged over different folds and repetitions.

Chapter 4

Microtubule generative models and estimation from 2D fixed cell fluorescence microscopy images

4.1 Background

While limited information is available about microtubule distributions [Brinkley et al., 1981, Reaven, 1982], information on those distributions in intact cells for different cell types has not been readily available. One of the main reasons for this is the difficulty of measuring individual microtubules in whole intact cells. Electron microscopy can be used to trace microtubules, but the specimen preparation for imaging does not allow for intact cells to be imaged. Fluorescence microscopy can be used to image intact cells, but

microtubules typically overlap and are often densely packed inside cells. To accurately trace them, and to make estimation of quantitative parameters, is very hard in intact cells. Hence previous work comparing cell types has often focused on the tips of microtubules where tracing is possible or the comparison is only qualitative [Wolf and Spanel-Borowski, 1995].

In the previous chapters, an indirect method for estimating quantitative parameters such as the number and the mean length of microtubules was developed from 3D fluorescence microscopy images of microtubules [Shariff et al., 2011, 2010a]. 3D image data of intact whole cells, however, can be difficult to obtain in a high throughput (several thousand cells) fashion. 2D images of microtubule structure, on the other hand, are more common. In this chapter, a method of estimating 3D microtubule model parameters of an intact cell from its 2D image fluorescence microscopy data is described.

4.2   Data Acquisition

4.2.1   3D image data of HeLa cells

3D images of HeLa cells previously obtained by three color confocal immunofluorescence microscopy were used to visualize three cell components: the cell membrane, nucleus and microtubules [Velliste and Murphy, 2002].

4.2.2 2D image data of eleven cell types from the Human Protein Atlas

The data used here are confocal fluorescence microscopy images of fixed cells of three different cell types: A-431, U-251MG and U-2OS, from the Human Protein Atlas [Barbe et al., 2008]. The images are analyzed as 8-bit TIFF images, with two files each obtained using a different emission wavelength of fluorescence from a single image field. These two channel files show the locations of (i) microtubules and (ii) nuclei. Each of the field images are of size 1728 x 1728 and the pixel size is 0.08 microns in the sample plane.

4.3 Point Spread Function (PSF) estimation

The confocal PSF was generated theoretically using the SVI PSF calculator for the Zeiss LSM 510 confocal microscope (http://www.svi.nl/NyquistCalculator). The pinhole size was set to 1 Airy Unit. The numerical aperture was 1.4 and the emission-excitation data used to generate the PSF was for the Alexa555 dye (http://probes.invitrogen.com/handbook/boxes/0442.html).

4.4   3D Cell and Nuclear Shape Generation from a 2D Slice of Microtubule Channel and Nucleus Channel

The generative model of microtubules that was conditioned on the shape of the cell and the nucleus (Shariff et al. [2010a]). These shapes were estimated from a 3D confocal stack of images of a total protein stain and a DNA stain respectively. A method for how to generate an approximate 3D shape of a cell and nucleus from a 2D slice (purely for the purpose of being able to generate a synthetic microtubule distribution) is first described. For the 3D shape generation, an assumption was made that the 2D slice contains information about the 3D structure because of the nature of the point spread function of the confocal microscope, which allows some of the out-of-focus light to reveal information about the structure along the Z-dimension.

The field images were first downsampled for computational efficiency from 0.08 microns to 0.2 microns. They were then segmented into single cell regions using a seeded watershed method and the 2D cell boundaries were found by thresholding the single cell regions for above zero pixels. This was used for cell size calculation and for 3D morphology generation (see below). The 3D shape of each cell was estimated by interpolating from the bottom of the cell to a small ellipse whose major axis is aligned with that of the cell. The microtubule channel image acquired at the center of the cell was used, i.e. $z = Z/2$, where Z is the height of the cell.

This image contains information about the cell boundary at the bottom-most region because the out of focus light from the bottom slice is visible in the center slice (as mi-

crotubules of very low intensity). Hence, the boundary of the bottom slice was found by thresholding for above zero intensity pixels. The size (pixel area) of the ellipse is modeled as a fraction of the area of the bottom slice: this is given by $a = 2^{-z}A$, where A is the pixel area of the bottom slice, and z is the distance from the bottom. This equation was estimated from the average area profile of the 2D slices in the 3D HeLa stack (data not shown). The shape of each cell was then estimated by using distance transform based shape interpolation [Luo and Hancok, 1997]. Given the height of the cell and the z-sampling step-size (0.2 microns), a 3D stack of the shape of the cell was generated. The nuclear morphology was generated based on the same procedure above using the nucleus channel image.

Figure 4.1 shows an example of microtubule and nucleus images and the resulting 3D cell and nucleus shape models.

4.5 Centrosome location detection (in 3D)

The 3D coordinate of the centrosome was estimated by breaking the problem into two parts. First, the XY-coordinate was estimated and second, the Z-coordinate. The XY-coordinate was chosen as the pixel with the maximum intensity value in the vicinity of the nucleus after smoothing with an averaging filter of size 25 pixels. For the Z-coordinate, linear regression was used to estimate the location as a function of the following predictor variables: (i) Maximum intensity of the microtubule image, (ii) Mean intensity of the microtubule image, and (iii) pixel intensity of the XY coordinate in the microtubule image. The parameters of the linear regression were estimated from the 3D HeLa images where

the 3D centrosome was estimated using the method described in [Shariff et al., 2010a].

The models and centrosome location were then used to generate microtubules in the cytosolic space.

4.6   Growth model of microtubule patterns:

The growth model of microtubule patterns is similar to the one used in Shariff et al., 2010a, with one modification: if the microtubule is required to make a turn in 3D space such that the 3D angle is greater than 56.6 degrees (this value is chosen manually to account for appearance of real microtubules as well as the generability of the model), the growth procedure for it is terminated. In order to ensure that the input parameters are exactly the same as the output parameters, the Algorithm 1 was used to generate the images.

4.7   Simulated image library generation

The model was convolved with a theoretical point spread function and multiplied it with a scalar to match the microtubule intensity estimated. The single microtubule intensity for each cell type was estimated using the method described in Chapter 3 or in [Shariff et al., 2011]. Using this approach, a library of simulated images was generated for each cell geometry (cell shape and nucleus shape) and contained all combinations of the following parameter values:

1. Number of microtubules = 5, 50, 100, 150, 200, 250, 300, 350, 400, 450;

2D slice of tubulin channel

2D slice of DNA channel

3D model of cell and nucleus shape
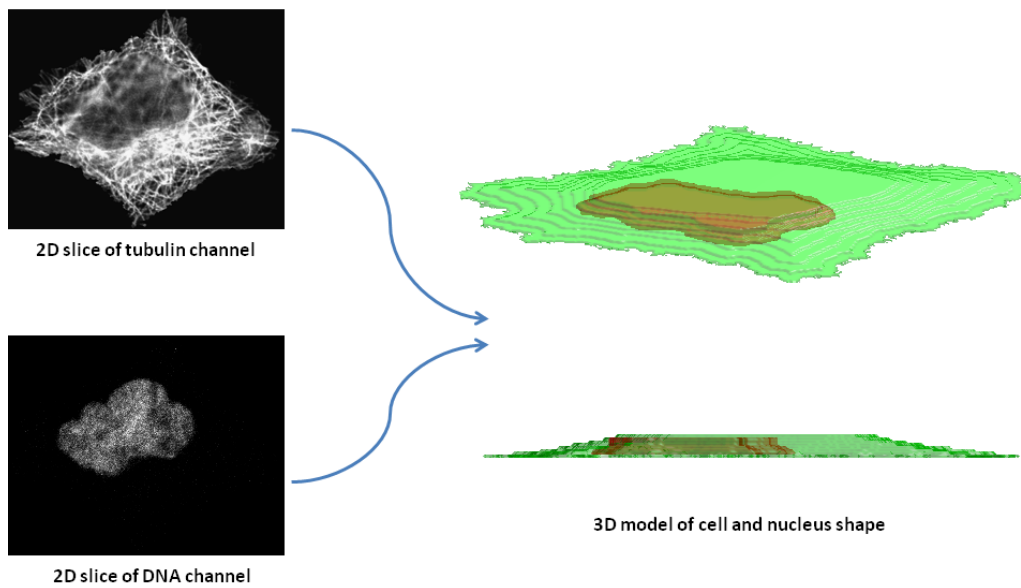
Figure 4.1: Generation of 3D cell geometry from 2D slices of microtubule and nucleus channel

---

Algorithm 1 Microtubule Generative Model

   1. **Input parameters:** number of microtubules (n), mean of the length distribution (mu), coefficient of variation of length, collinearity and cell height;

   2. Sample lengths from Truncated Normal distribution;

   3. Sort lengths from longest to shortest;

   4. Iterate until all lengths are generated, starting with the longest microtubule:

for $i = 1$ to n **do**

   if storage has microtubule of desired length generated **then**

      (1) use the generated microtubule length;

      (2) remove chosen microtubule from storage;

      (3) **continue**, to the next microtubule.

   end if

   loop

      (1) Generate the microtubule using the method in Shariff et al., 2010a and two modifications.

      if the microtubule length cannot be generated **then**

         (a) add to storage and re-generate the microtubule.

         if repeating 100 times does not generate the microtubule of desired length **then**

            return declare "input parameters cannot be generated".

         end if

      end if

   end loop

end for

---

2. Mean of length distribution = 5, 10, 15, 20, 25, 30, 35, 40, 45 microns;

3. Coefficient of variation of length = 0, 0.1, 0.2, 0.3;

4. Collinearity ($cos\alpha$) = 0.97, 0.98466, 0.9961;

5. Cell Height = 1.2, 1.4, 1.6 microns.

4.8   Features and matching

For all the 3D simulated images in the library, the central 2D slice was used to compute 2D versions of the features that were used in Shariff et al., 2010a. More details about the implementations of the 2D version of the features can be found in [Boland and Murphy, 2001], and in addition we appended the feature vector with edge features. Following feature computation, the normalized euclidean distance in feature space was minimized for matching to estimate the parameters of distribution of microtubules in real 2D HPA images [Shariff et al., 2010a].

4.9   Recovering 3D Microtubule Generative Model Parameters from 2D Images: comparisons with real 3D estimates

In Chapter 2, the method of parameter estimation involves computing features from 3D stacks of fluorescence microscopy images of microtubules from HeLa cells. Since esti-

mates need to be made from 2D images, 2D versions of the features were computed and applied the 2D method of matching on the central slice (at half height of the cell) of the 3D HeLa image data, and compared it with the 3D method. The half height was chosen as the preferred slice because the 2D images from the HPA were also acquired at half the height of the cell. The mean absolute percentage error (MAPE) was computed.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} |(P_i - \hat{P}_i)/P_i| * 100 \tag{4.1}$$

where P_i is the true parameter and $\hat{P}$ is the estimate, between the estimates from 2D image data and the estimates obtained using the 3D generative method over 42 cells. The error was 43% for the number of microtubules, and 49% for the mean of length distribution. The estimates from a single 2D slice are reasonably close to those from the entire 3D image.

4.10 Comparing the model parameters from the three cell types shows differences

3D Microtubule model parameters were estimated from 2D fluorescence microscopy images of three different cell types from the HPA [Barbe et al., 2008], with the application of the whole framework including library generation, feature calculation and matching. Figure 4.2 shows examples of query images and corresponding images synthesized using the parameters estimated from them. A t-test comparing the estimates of the number, mean of length of microtubules, and their product, from 100 cells of each of the three cells types suggests that there is a significant difference in the means of the distributions across the

60

three cell types.

4.11 Discussion

4.11.1 Summary

An automated method to estimate 3D microtubule model parameters from 2D images is developed here. The method is dependent on the 3D structure of the cell and the nucleus, and the centrosome location. An automated approach is described to generate the same using only the 2D microtubule image and 2D nucleus image acquired at the center (half height) of the cell. This method was applied to compare model parameter estimates from over 400 images of cells and 11 cell types from a set of cells obtained from the HPA.

4.11.2 Comparison with Existing Methods.

To my knowledge, this study is the first attempt to quantify the number and mean of the length distribution of microtubules in intact cells across different cell types. Methods such as electron microscopy can image intact cells, but have interference from other cell components [Osborn et al., 1978]. More invasive methods of preparation such as extraction of microtubule network can allow electron microscopy to generate traceable images, but are no longer representative of the intact cells [Letourneau, 1982]. Fluorescence microscopy, on the other hand, can be used to obtain information about proteins at monomer-level resolution of localization without interference from other cell components in intact cells.
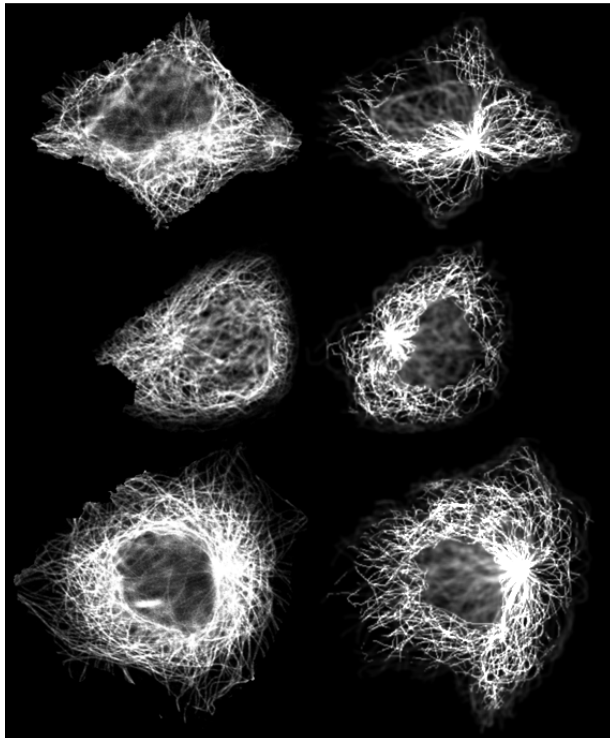
Number of Microtubules = 250
Mean of Length Distribution = 25 µm
SD of Length Distribution = 5 µm
Collinearity = 0.9847

Number of Microtubules = 200
Mean of Length Distribution = 30 µm
SD of Length Distribution = 6 µm
Collinearity = 0.97

Number of Microtubules = 350
Mean of Length Distribution = 35 µm
SD of Length Distribution = 3.5 µm
Collinearity = 0.9847

Figure 4.2: Best match simulated images center slice (right) for the real images (on the left), and estimates of parameters

One reason for studying microtubule distributions across cell types is to understand the correlation between subcellular localization patterns of microtubule associated proteins (MAPs) and the microtubule network. There is evidence of varying levels of proteins across cell types [Duerr et al., 1981], and also cell-specific proteins regulating microtubules [Shestakova et al., 1998]. In this paper, the cell types chosen for the HPA are from varying lineages such as mesenchymal, epithelial and glial tumors, that are hence expected to have different localization patterns of microtubule associated proteins (MAPs). Analysis done here shows that some cell types have significant differences in the number and the mean of the length distribution of microtubules. Although analysis of the images acquired from the HPA reveals that about half (49%) of the antibodies analyzed showed identical subcellular localization in all three cell lines, and over 82% in two cell lines [Barbe et al., 2008], it is unclear what proportion of these are microtubule associated proteins. It is possible that the proteins involved in regulating the number and length are not identical in their distributions across cell types.

There is evidence that the number and lengths of microtubules are correlated with the size of the cell [Brinkley et al., 1981, Goniakowska-Witalinska and Witalinski, 1976]. Therefore, the area of the bottom-most slice (sum of pixel values of the binary image) was computed as the value reflecting the size of the cell, for each of the cell types. To quantify the correlation, the correlation coefficient was computed between the cell size and the product estimated. The values were 0.81 for A431 cells (red), 0.64 for U2OS cells (green), and 0.80 for U251-MG cells (blue). Figure 4.3 shows a plot of the all the cells and a best fit line indicating the relationship between the cell size and total tubulin content indicated by the product estimated. The correlation coefficient for all the cell
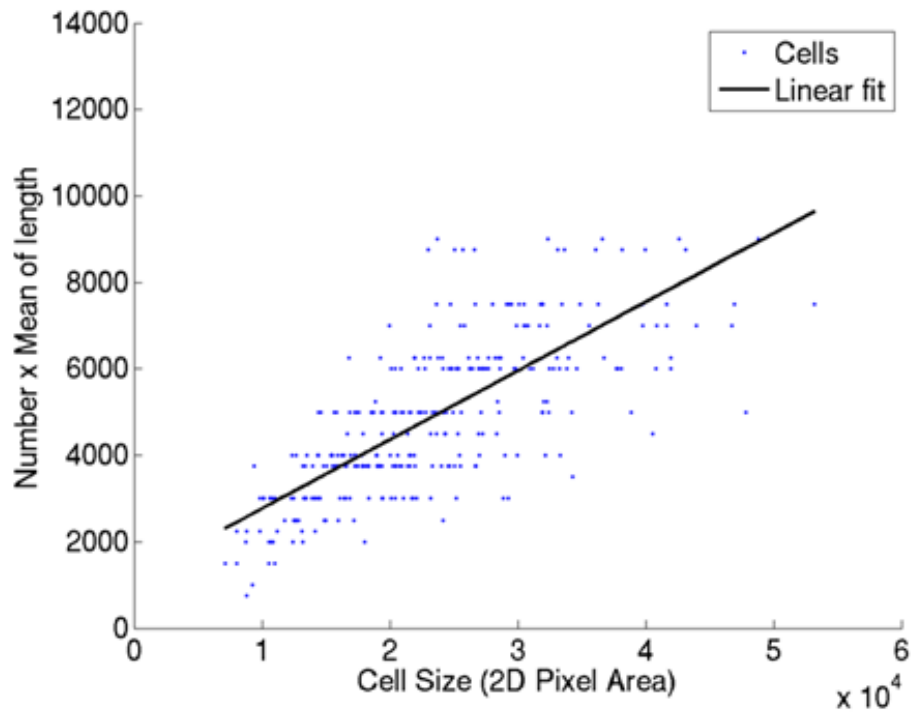
Figure 4.3: Correlation between the cell size and the product of number and mean length estimated.

types together was 0.76. This adds more confidence to the estimates of the automated approach and further confirms the existing hypothesis using the work done by alternative approaches.

In future, reducing the error in the comparison between the parameter estimates from the 2D and 3D methods will be required. A potential improvement would be to identify optimal focal plane (or planes) for acquisition (z-location) of the microtubule image that better approximates the microtubule in the whole 3D cell. Another improvement would be to perform the MAPE comparison of estimates using the 2D and 3D methods on a larger number of 3D images instead of just 42 cells.

Another future work can be to estimate from more cells and cell types from both 2D and 3D datasets across different cell lines and across species. This would give more insight into the distribution of microtubules across cells from different lineages, cell morphologies, etc. We also plan to estimate parameters from time series data sets. Acquisition of 3D time series data sets for microtubules has been difficult because of issues with photobleaching and phototoxicity. Since a method of 2D analysis is now possible, the problem can be minimized allowing the acquisition of time series. The generative model and estimation approach can also be extended to estimate microtubule dynamics parameters such as growth and shrinkage kinetics in live intact cells.

Chapter 5

Building models conditional on microtubules: A vesicle location model

5.1 Background

Cytoplasmic vesicles are membrane-covered organelles that are localized inside the cell and are important for various cellular tasks such as endocytosis and exocytosis. For example, synaptic vesicles are released from the axon terminal to release neurotransmitters that are required for propagating nerve impulses. Depending on the contents of vesicles and markers present on the vesicle membrane, they are targeted towards specific locations inside the cell. There are least five major types of vesicles, endosomes, peroxisomes, lysosomes, phagosomes, and exosomes. They are also categorized according to their contents (e.g.. digestive enzymes in lysosomes) and functionality (e.g.. endosome for endocytosis)

A critical component of vesicle targeting is the association of vesicles and microtubules for intracellular transport. Although microtubules are not necessary for short-range transport, they are required for rapid transport of vesicles [Bloom and Goldstein, 1998]. Figure 5.1A shows a two-color fluorescence microscopy image of an A431 cell that is fluorescently tagged to shown the distributions of a vesicle and microtubule protein. The image suggests that vesicles are spatially localized in a close proximity to microtubules.

Cell modeling and simulations are important tools for understanding the various roles vesicles play inside cells. As described in Chapter 1, cell simulations, for example, are often used to model behavior of vesicle transport on microtubules [Tsaneva-Atanasova et al., 2009]. Such simulations invariably require vesicle location information that is representative of that cell type and condition. Modeling the spatial distribution of vesicles inside cells taking into account information from image data remains an important unsolved problem that could have a significant impact in measuring and understanding related biological processes. Although descriptive parameters can be extracted that are useful for tasks such as classification of subcellular patterns [Zhao et al., 2005], only generative models can generate instances that can be used in cell simulations.

In previous chapters, a generative model was described for microtubules conditional on the cell membrane and nucleus along with a method to estimate parameters describing the spatial distribution from fluorescence microscopy images of HeLa cells. Previously, a vesicle location generative model that is conditional on the nucleus and cell membrane was described [Zhao and Murphy, 2007]. However, such models for vesicle distribution should be conditional on microtubules. In this chapter, an approach for modeling vesicular distributions conditioned on models of microtubule distributions is described. An approach for

68
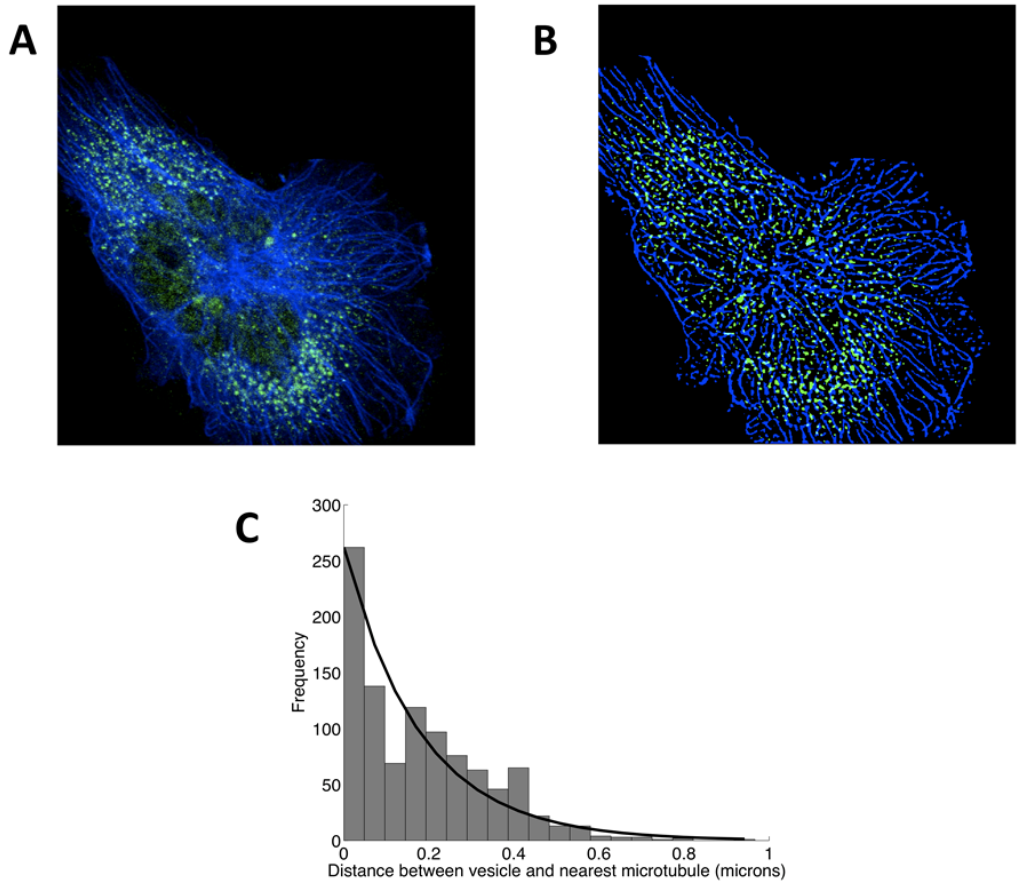
Figure 5.1: (A) A two-color image of a vesicle protein (green) and microtubules (blue) (B) segmentation of the two channels (C) distributions of the distances between vesicles and nearest microtubules

how to estimate model parameters is also described from confocal microscopy images of microtubules and vesicle proteins using examples from the Human Protein Atlas [Barbe et al., 2008]. Cluster analysis is performed to analyze the variation in microtubule association among a group of vesicle proteins. Using the parameters estimated from images of lysosomes and endosomes, a model of vesicles is constructed in a hierarchical framework.

5.2   Image Collections

The data used here are 392 confocal fluorescence microscopy images of fixed A431 cells from the Human Protein Atlas [Barbe et al., 2008]. The images are analyzed as 8-bit TIFF images with three files each obtained using a different emission wavelength of fluorescence from a single image field. The three images show the locations of (1) vesicle-dependent proteins, (2) nuclei and (3) microtubules. The vesicle-related protein images were acquired using 128 fluorescently labeled antibodies and the other two using fluorescent stains [Barbe et al., 2008]. Each of the field images are of size 1728 x 1728 and the pixel size is 0.08 microns in the sample plane.

5.3   Dependence of vesicle location on microtubules

The goal of this work is to be able to generate a model of vesicle distribution given quantifiable parameters conditional on microtubules. Since microtubules and vesicles are known to interact, this affinity is quantified by computing the distances between vesicles and the nearest microtubule. Using confocal microscopy image data of the microtubule and vesicle

channel, the spatial distribution of microtubules and vesicles are estimated by segmentation. The input image was blurred using a Gaussian filter with standard deviation of 3, and the resulting image was subtracted from the input image. The subtracted image was binarized to separate zero and non-zero pixels. Since the binary image has small clusters of disconnected objects seemingly forming microtubule fibers or vesicle-like structures, the binary image is blurred again to connect objects that are close to each other. This operation was performed using a Gaussian filter with standard deviation of 2. The resulting image was again binarized. This ad hoc approach resulted in reasonable definition of both vesicles and microtubules.

Figure 5.1B shows the segmentation of microtubules (blue) and vesicles (green) of the cell in Figure 5.1A. Next, the distance between a vesicle and its nearest microtubule was computed. First, the centroids of all vesicles were computed using the segmented binary image. Then, the distance between each vesicle and its nearest microtubule was found using a distance transform of the binarized microtubule image. Figure 5.1C shows the distribution of distances between the centroid of vesicles and the nearest microtubule for the same cell. The distribution was fit with an exponential (solid line). The mean of the exponential is used in our analysis as measure of the distance between a specific vesicle localized protein and microtubules.

5.4   Identification of multiple populations in vesicle data

The data from the Human Protein Atlas are images of labels tagged as proteins and the corresponding microtubule channel. Since, vesicles can be classified either as lysosomes,

endosomes, peroxixomes or phagosomes, cluster analysis was performed to identify potential multiple populations where one of parameters is based on affinity of vesicles and microtubules: the mean of the exponential distributed microtubule-vesicle distances. In addition to the proteins being localized in vesicles, there is also non-vesicle fluorescence. Hence, two more features were computed: the fraction of fluorescence in the vesicle and the protein fluorescence fraction overlapping with nucleus. The feature vectors were clustered by k-means clustering using Normalized Euclidean as the distance metric. For each value of k-clusters, the clustering was repeated using 100 different random starts. The number of clusters was chosen as the one that minimized the Akaike Information Criterion. Figure 5.2 shows the distributions of the three features across the clusters. The green cluster had two points that are outliers because of high noise levels in the field images where some regions get incorrectly detected as vesicular objects. From Figure 5.2A, it was concluded that the vesicle objects could not be clustered into different classes on the basis of microtubule-vesicle affinity. However, clusters appeared to be significant across fluorescence fractions in vesicles and those overlapping with the nucleus.

5.5 Generative model of vesicles conditioned on microtubules

I now describe how vesicle locations are simulated taking into account the affinity (estimated as described above) as well as a given microtubule distribution. The method utilizing a lysosomal protein pattern is illustrated. The mean of the exponential distribution
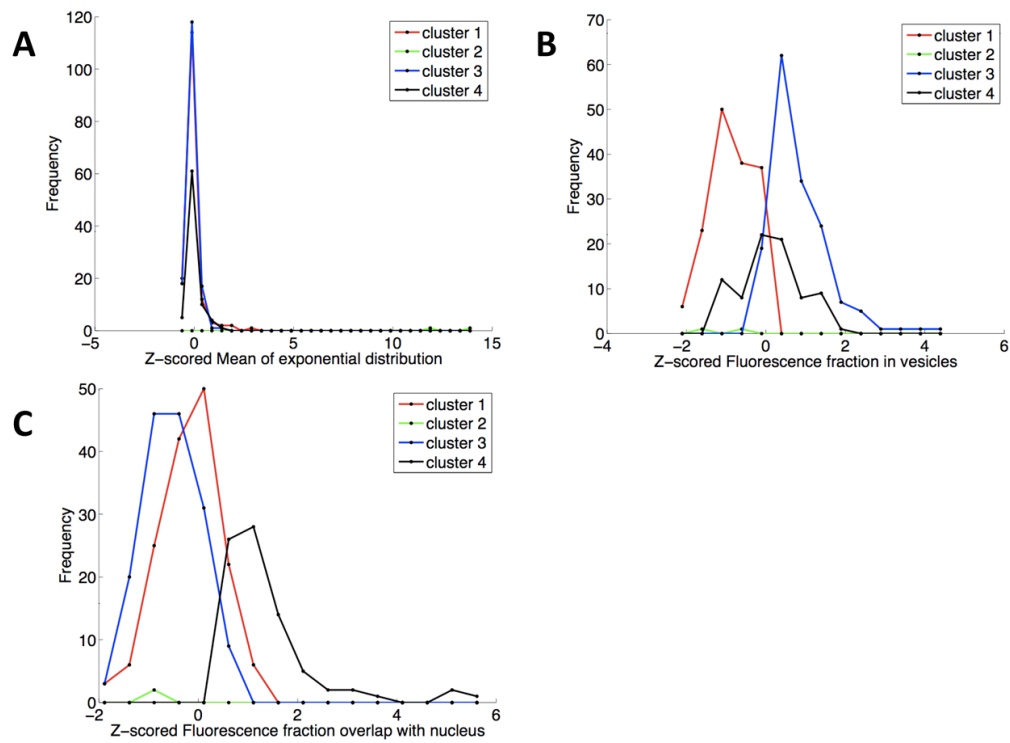
Figure 5.2: Clustering of vesicle image data. (A) Distributions across clusters of mean exponential parameter, (B) fluorescence fractions in vesicles and (C) fluorescence fraction overlapping with the nucleus. Max. bin size = 10.

of distances was estimated from fluorescence microscopy images of a lysosomal protein: LAMP-2. Next, a model for microtubules was generated using the procedure described in [Shariff et al., 2010a]. The model parameters are the number of microtubules, the mean and standard deviation of the length distribution, and collinearity. These parameters were sampled from the distributions that were estimated from HeLa cells [Shariff et al., 2010a].

Figure 5.3A shows a three-dimensional rendering of microtubules generated (and its nucleus) that was used for our vesicle model of LAMP-2. Next the sampled values of the number of vesicles per cell, size and fluorescence intensities of the vesicles distribution from HeLa data were used to generate the vesicles using Gaussian object based generative models. The parameters for generating vesicles conditioned on microtubules are: (1) mean fluorescence intensity, (2) size of the vesicle (3) number of vesicles per cell, and (4) the mean of the exponential distribution fit using the distances between vesicles and the nearest microtubule and object centers. All the parameters for the vesicles except for the mean of the exponential parameter were estimated from HeLa cells as described in [Zhao and Murphy, 2007].

For the vesicle locations, the estimates from the HPA data were used to compute a single mean for the exponential distribution. The mean was converted into a 3D spatial probability distribution by convolving the generated 3D microtubule model with a 3D Gaussian filter whose standard deviation is approximated to have the decay of a two tailed exponential distribution with the mean parameter as input. This empirical density is sampled to generate the spatial locations of the vesicle points. In order to generate the distribution of vesicles (X_{VC}), we require the locations of the vesicles that are generated by sampling from a 3D spatial probability density function (P) that is a function of the

locations of the microtubules X_{MT} and the affinity parameter.

$$X_{VC} \sim P(X_{MT}, \mu_{LAMP2})$$

A two dimensional slice from P is shown in Figure 5.3B. Figure 5.3C is a two dimensional slice from a 2-color three dimensional simulated microscopy image of the vesicles (green) and the corresponding microtubules (blue). This procedure was repeated using estimates from an endosomal protein TfR and an example image generated is shown in Figure 5.3D.

5.6 Discussion

This article describes a method to build a model of vesicle distribution that is conditional on microtubules. The model outputs locations of the vesicles based on (1) microtubule location information and (2) the estimate of the mean of the exponentially distributed distances between vesicles and the nearest microtubule. We used estimates of two vesicle proteins: Transferrin receptor (endosomal) and LAMP-2 (lysosomal) and generated images of vesicles distributions conditioned on microtubules.

The work described here represents an important step towards bridging detailed models learned from large collections of images for proteins contained in discrete objects with models of microtubule growth learned by inverse modeling. We plan to extend this work by merging it with models of vesicle movement obtained by automated tracking. It is hoped that approaches like this will enable the construction of models that capture essential cell behaviors without requiring the simultaneous measurement of the thousands of
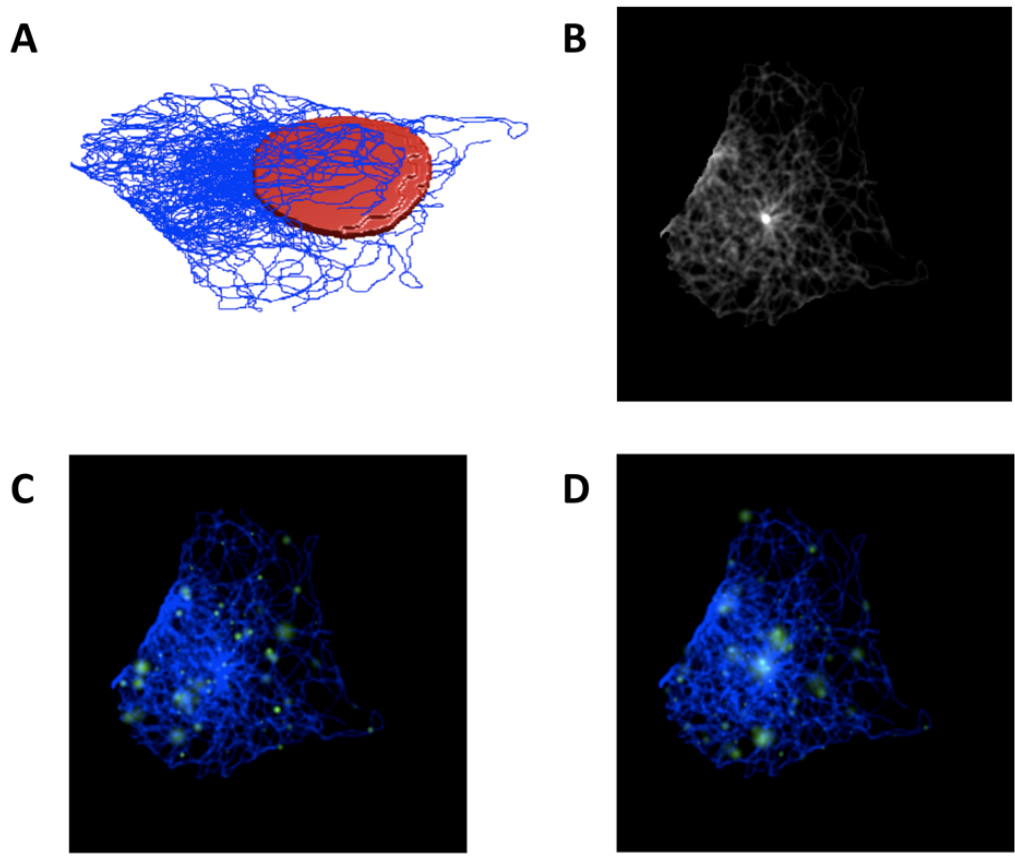
Figure 5.3: Generative Model of vesicles. (A) A 3D rendering of a generated microtubule distribution (B) a spatial probability distribution of vesicle locations of a 2D slice (C) a simulated image of vesicle locations and microtubules of LAMP-2 and (D) Transferrin receptor

proteins in the same living cell, something that is not possible with current technology.

Chapter 6

Conclusion

This chapter summarizes the contribution of this thesis in the field of systems biology. Specifically, this chapter outlines how the tools developed allow the extraction of biologically relevant parameters from images of microtubules in the framework of generative modeling. The chapters in this thesis can be classified based on the type of image data (2D vs 3D image acquisition, appearance of free vs polymerized tubulin) that are used to extract quantitative parameters (see below). Implications of each of the chapters and future work based on this thesis is also outlined.

6.1 Contribution of this thesis - Generative model of microtubules

This thesis uses the framework of generative modeling for models of microtubules as it allows for generating patterns that reflect the development of cellular components that is very intuitive in biology. This further allows the parameters of the generative model to be realistic - number, length distribution and stiffness of microtubules. The models are also generated in a conditional manner. Specifically, the model of microtubules are described here as dependent on cell and nuclear shape models.

The growth model used is a persistent random walk for extending microtubule segments where successive segments are related by a range of correlation coefficients. The collinearity parameter used here is a lower bound on the correlation coefficient (with the upper bound fixed at one) that can be understood as a single stiffness parameter that has been frequently used to explain growth models of microtubules.

6.2 Contribution of this thesis - Model parameter estimation from confocal fluorescence microscopy images

The model parameters for the generative model of microtubules are the following:

1. number of microtubules

2. collinearity

3. mean of the normal distribution of length

4. standard deviation of the normal distribution of length

These model parameters are biological relevant quantitative information that is hard to extract from fluorescence microscopy images. Features are numerical values computed as a function of an image or pixel intensity values that are required to be mapped to these biological relevant parameters. Using the generative model of images of microtubules for a given set of model parameters, features can be computed. In order to map the features back to model parameters, an estimation approach needs to be described. As mentioned earlier, this thesis describes extraction of these parameters from different classes of images: 2D vs 3D images and free vs polymerized tubulin.

6.3   3D fixed cell preparations

In fixed cell preparations, the free tubulin monomers escape the cell leaving only the polymerized filaments in the cell to appear intact. Chapter 2 describes how to estimate these parameters from 3D images. Using a non-parametric approach of minimizing the normalized euclidean distance between the query image and a library of simulated images, it is possible to recover the parameters with low errors: 9% for the number of microtubules and 15% for the mean of the length distribution.

6.4 2D fixed cell preparations

Chapter 4 describes how to estimate 3D model parameters from 2D images. In this chapter, a method for how to generate the 3D shape of the cell using a center 2D slice of the tubulin channel is described. A method for estimating the 3D coordinates of the centrosome is also described. Estimating the model parameters from the three different cell types from the Human Protein Atlas reveals significant differences in the means of the parameters taking into account their variances. Estimating parameters from different cell types allows for comparing cell types in terms of their microtubule distributions.

6.5 3D live cell preparations

In chapter 3, fluorescence microscopy images of live GFP-tagged cells to alpha-tubulin were acquired revealing the presence of free tubulin monomers and the polymerized microtubules. A microtubule depolymerizing drug nocodazole was used and images of cells at different time points was acquired. A feature selection approach is described to select for features that reliably estimate the model parameters. The results show that the number and the mean of the length of microtubules decrease in the presence of nocodazole.

6.6 Models conditioned on microtubules

Chapter 5 shows an example of how to build models conditioned on microtubules. Specifically, the chapter descibes a vesicle location model by sampling from a 3D spatial dis-

tribution where the higher probability indicates the location of a microtubule. The work showed that it was possible to estimate vesicle location model parameters from images of lysosomal protein LAMP2 and an endosomal protein TfR.

6.7 How does this work change with improving technologies such as resolution of the images with image acquisition?

Simulations have been used across various scientific fields in order to test models and understand behavior. The work presented here allows parameters of these models to be estimated automatically from biological images. Specifically, the work here shows that when the resolution of images is such that filamentous structures are not traceable, an indirect approach can be used where simulated images are compared with real images to estimate model parameters. Direct approaches such as tracing can used when future technologies that allow for clear distinction of microtubules in a high throughput manner are available.

6.8 Implications and Future work

6.8.1 Conditional Models

Ultimately, we seek to build models in a hierarchical, conditional manner so that models of all cell components can be constructed by automated learning from cell images. In the future, we anticipate that this model can be merged with generative models of other protein patterns. Microtubules are critical for intracellular transport and vesicles that are transported by molecular motors along microtubules. There are numerous other protein pattern images of microtubule associated proteins (MAPs), such as the microtubule end binding protein (mEB1), that are dependent on microtubule network.

6.8.2 Other filamentous structures

The current approach can be used to learn models for other structures that have a network/filamentous appearance. Particularly, patterns that make up the cell cytoskeleton, such as actin and intermediate filaments, or proteins that make up the connective tissue, such as collagen fibers, may be quantified by this method. For example, in an analysis of actin images, there are quantitative measures that can be used as features for the indirect approach described in this thesis [Weichsel et al., 2010].

6.8.3 Comparing microtubules

In future, biologically relevant parameters from different cell types across species and conditions, such as temperature, pH and drugs, can be estimated so that comparisons based on their microtubule distributions can be made.

6.8.4 Regression to predict parameters of distribution of microtubules

From a computational point of view, using a brute-force search where many simulated images need to be generated, demands a large amount of computational effort, and the resulting values of parameters can be only chosen from the grid of the library generated. Therefore, in future the library of images can be adaptively expanded to generate more detail where it is needed, or a regression model can be trained to create a mapping function between model parameters and the features of the synthetic or real images.

6.8.5 Are we estimating all quantitative information required to answer biological questions regarding microtubules?

This thesis describes building static representations of microtubule distributions, but in order to simulate the behavior of microtubules, critical information such as the kinetic parametes of microtubules need to be estimated. Data acquisition of 3D time series images of microtubules has been difficult because of issues with photobleaching and phototoxicity. Since the 2D method of parameter estimation from images is now possible, analysis of time series is possible. However, the generative model and estimation approach needs to

be extended to estimate microtubule dynamics parameters such as growth and shrinkage kinetics in live intact cells.

6.9   Availability

All the code for the work is available on the Murphy lab software website `http://murphylab.cbi.cmu.edu/software/software_from_papers.html`. These models and training data will also be available in PSLID [Huang et al., 2002], and in future will be integrated with cell simulations software such as V-cell [Moraru et al., 2002].

Bibliography

O. Al-Kofahi, A. Can, S. Lasek, D. H. Szarowski, J. N. Turner, and B. Roysam. Algorithms for accurate 3d registration of neuronal images acquired by confocal scanning laser microscopy. *J Microsc*, 211(Pt 1):8–18, Jul 2003. 1.6.1

Yousef Al-Kofahi, Natalie Dowell-Mesfin, Christopher Pace, William Shain, James N. Turner, and Badrinath Roysam. Improved detection of branching points in algorithms for automated neuron tracing from 3d confocal images. *Cytometry A*, 73(1):36–43, Jan 2008. doi: 10.1002/cyto.a.20499. URL `http://dx.doi.org/10.1002/cyto.a.20499`. 1.6.1

Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular biology of the cell*. Garland Science Taylor & Francis Group, 4 edition, 2002. ISBN 0815332181. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0815332181`. (document), 1.1, 1.2

L. Barbe, E. Lundberg, P. Oksvold, A. Stenius, E. Lewin, E. Bjorling, A. Asplund, F. Ponten, H. Brismar, M. Uhlen, and H. Andersson-Svahn. Toward a confocal subcellular

atlas of the human proteome. *Mol Cell Proteomics*, 7:499–508, 2008. 1.6.2, 4.2.2, 4.10, 4.11.2, 5.1, 5.2

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738. 1.6.1

G. S. Bloom and L. S. Goldstein. Cruising along microtubule highways: how membranes move through the secretory pathway. *J Cell Biol*, 140(6):1277–1280, Mar 1998. 1.2, 5.1

M. V. Boland, M. K. Markey, and R. F. Murphy. Classification of protein localization patterns obtained via fluorescence light microscopy. In *Proc. 19th Annual Int Engineering in Medicine and Biology Society Conf. of the IEEE*, volume 2, pages 594–597, 1997. doi: 10.1109/IEMBS.1997.757680. 1.6.1

M.V. Boland and R.F. Murphy. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells. *Bioinformatics*, 17:1213–1223, 2001. 1.6.3, 4.8

Clifford P. Brangwynne, F. C. MacKintosh, and David A. Weitz. Force fluctuations and polymerization dynamics of intracellular microtubules. *Proc Natl Acad Sci U S A*, 104 (41):16128–16133, Oct 2007. doi: 10.1073/pnas.0703094104. URL http://dx. doi.org/10.1073/pnas.0703094104. 2.12

B.R. Brinkley, S.M. Cox, D.A. Pepper, L. Wible, S.L. Brenner, and R.L. Pardue. Tubulin

assembly sites and the organization of cytoplasmic microtubules in cultured mammalian cells. *J Cell Biol*, 90:554–562, 1981. 4.1, 4.11.2

T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans Image Process*, 10(2):266–277, 2001. doi: 10.1109/83.902291. URL `http://dx.doi.org/10.1109/83.902291`. 2.4

Xiang Chen, Meel Velliste, Shmuel Weinstein, Jonathan W. Jarvik, and Robert F. Murphy. Location proteomics: building subcellular location trees from high-resolution 3d fluorescence microscope images of randomly tagged proteins. volume 4962, pages 298–306. SPIE, 2003. doi: 10.1117/12.477899. URL `http://link.aip.org/link/?PSI/4962/298/1`. 2.8

L. P. Coelho, A. Shariff, and R. F. Murphy. Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. In *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro ISBI '09*, pages 518–521, 2009. doi: 10.1109/ISBI.2009.5193098. 1.6.1

Luis Pedro Coelho, Estelle Glory-afshar, Joshua Kangas, Shannon Quinn, Aabid Shariff, and Robert F. Murphy. Principles of bioimage informatics: Focus on machine learning of cell patterns. *Lecture Notes in Bioinformatics*, pages 8–18, 2010. 1.6.1

Gerald Donnert, Jan Keller, Rebecca Medda, M Alexandra Andrei, Silvio O. Rizzoli, Reinhard Lhrmann, Reinhard Jahn, Christian Eggeling, and Stefan W. Hell. Macromolecular-scale resolution in biological fluorescence microscopy. *Proc Natl Acad*

Sci U S A, 103(31):11440–11445, Aug 2006. doi: 10.1073/pnas.0604965103. URL `http://dx.doi.org/10.1073/pnas.0604965103`. 1.5

A. Duerr, D. Pallas, and F. Solomon. Molecular analysis of cytoplasmic microtubules in situ: identification of both widespread and specific proteins. *Cell*, 24:203–211, 1981. 4.11.2

Ro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. pages 130–137. Springer-Verlag, 1998. 3.8

K. A. Giuliano, R. L. DeBiasio, R. T. Dunlay, A. Gough, J. M. Volosky, J. Zock, G. N. Pavlakis, and D. L. Taylor. High-content screening: A new approach to easing key bottlenecks in the drug discovery process. 2(4):249–259+, 1997. ISSN 1087-0571. 1.6.1

Estelle Glory and Robert F. Murphy. Automated subcellular location determination and high-throughput microscopy. *Dev Cell*, 12(1):7–16, Jan 2007. doi: 10.1016/j. devcel.2006.12.007. URL `http://dx.doi.org/10.1016/j.devcel.2006.12.007`. 1.6.2

L. Goniakowska-Witalinska and W. Witalinski. Evidence for a correlation between the number of marginal band microtubules and the size of vertebrate erthrocytes. *J Cell Sci*, 22:397–401, 1976. 4.11.2

G. Gorbsky and G. G. Borisy. Microtubule distribution in cultured cells and intact tissues: improved immunolabeling resolution through the use of reversible embedment cytochemistry. *Proc Natl Acad Sci U S A*, 82(20):6889–6893, Oct 1985. 2.12

S. F. Hamm-Alvarez and M. P. Sheetz. Microtubule-dependent vesicle transport: modulation of channel and transporter activity in liver and kidney. *Physiol Rev*, 78(4):1109–1129, Oct 1998. 1.3

M. R. Hoja, C. Wahlestedt, and C. Hg. A visual intracellular classification strategy for uncharacterized human proteins. *Exp Cell Res*, 259(1):239–246, Aug 2000. doi: 10.1006/excr.2000.4948. URL `http://dx.doi.org/10.1006/excr.2000.4948`. 1.6.2

Kai Huang, J. Lin, J. A. Gajnak, and R. F. Murphy. Image content-based retrieval and automated interpretation of fluorescence microscope images via the protein subcellular location image database. In *Proc. IEEE Int Biomedical Imaging Symp*, pages 325–328, 2002. doi: 10.1109/ISBI.2002.1029259. 6.9

Won-Ki Huh, James V. Falvo, Luke C. Gerke, Adam S. Carroll, Russell W. Howson, Jonathan S. Weissman, and Erin K. O'Shea. Global analysis of protein localization in budding yeast. *Nature*, 425(6959):686–691, Oct 2003. doi: 10.1038/nature02026. URL `http://dx.doi.org/10.1038/nature02026`. 1.6.2

J. W. Jarvik, G. W. Fisher, C. Shi, L. Hennen, C. Hauser, S. Adler, and P. B. Berget. In vivo functional proteomics: mammalian genome annotation using cd-tagging. *Biotechniques*, 33(4):852–4, 856, 858–60 passim, Oct 2002. 1.6.2

Ming Jiang, Qiang Ji, and Bruce McEwen. Model-based automated segmentation of kinetochore microtubule from electron tomography. *Conf Proc IEEE Eng Med Biol Soc*,

3:1656–1659, 2004. doi: 10.1109/IEMBS.2004.1403500. URL `http://dx.doi.org/10.1109/IEMBS.2004.1403500`. 1.5

Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Vol. 1 (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2 edition, 1994. ISBN 0471584959. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0471584959`. 2.5

Mary Ann Jordan and Leslie Wilson. Microtubules as a target for anticancer drugs. *Nat Rev Cancer*, 4(4):253–265, Apr 2004. doi: 10.1038/nrc1317. URL `http://dx.doi.org/10.1038/nrc1317`. 1.3

Eric Karsenti, Franois Ndlec, and Thomas Surrey. Modelling microtubule patterns. *Nat Cell Biol*, 8(11):1204–1211, Nov 2006. doi: 10.1038/ncb1498. URL `http://dx.doi.org/10.1038/ncb1498`. 2.12

Olga A. Koroleva, Matthew L. Tomlinson, David Leader, Peter Shaw, and John H. Doonan. High-throughput protein localization in arabidopsis using agrobacterium-mediated transient expression of gfp-orf fusions. *Plant J*, 41(1):162–174, Jan 2005. doi: 10.1111/j.1365-313X.2004.02281.x. URL `http://dx.doi.org/10.1111/j.1365-313X.2004.02281.x`. 1.6.2

Anuj Kumar, Seema Agarwal, John A. Heyman, Sandra Matson, Matthew Heidtman, Stacy Piccirillo, Lara Umansky, Amar Drawid, Ronald Jansen, Yang Liu, Kei-Hoi Cheung, Perry Miller, Mark Gerstein, G Shirleen Roeder, and Michael Snyder. Subcellular localization of the yeast proteome. *Genes Dev*, 16(6):707–719, Mar 2002.

doi: 10.1101/gad.970902. URL `http://dx.doi.org/10.1101/gad.970902`. 1.6.2

Misjal N. Lebbink, Willie J C. Geerts, Theo P. van der Krift, Maurice Bouwhuis, Louis O. Hertzberger, Arie J. Verkleij, and Abraham J. Koster. Template matching as a tool for annotation of tomograms of stained biological structures. *J Struct Biol*, 158(3):327–335, Jun 2007. doi: 10.1016/j.jsb.2006.12.001. URL `http://dx.doi.org/10.1016/j.jsb.2006.12.001`. 1.5

P.C. Letourneau. Analysis of microtubule number and length in cytoskeletons of cultured chick sensory neurons. *J Neuro Sci*, 2:806–814, 1982. 4.11.2

Huilin Li, David J. DeRosier, William V. Nicholson, Eva Nogales, and Kenneth H. Downing. Microtubule structure at 8 a resolution. *Structure*, 10(10):1317–1328, Oct 2002. (document), 1.1, 1.1

Lichen Liang, Qiang Ji, and B. F. McEwen. Extraction of 3d microtubules axes from cellular electron tomography images. In *Proc. 16th Int Pattern Recognition Conf*, volume 1, pages 804–807, 2002. doi: 10.1109/ICPR.2002.1044881. 1.5

Harvey Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, Matthew P. Scott, Anthony Bretscher, Hidde Ploegh, and Paul Matsudaira. *Molecular Cell Biology*. W. H. Freeman, 6th edition, June 2007. ISBN 0716776014. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0716776014`. 1.1, 1.1

Rose Loughlin, Rebecca Heald, and Franois Ndlec. A computational model predicts xeno-

93

pus meiotic spindle organization. *J Cell Biol*, 191(7):1239–1249, Dec 2010. doi: 10.
1083/jcb.201006076. URL `http://dx.doi.org/10.1083/jcb.201006076`.
1.4

B. Luo and E.R. Hancok. Slice interpolation using the distance transform and morph-
ing. *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International
Conference on*, pages 1083–1086, 1997. 4.4

Christian J. Malone, Lisa Misner, Nathalie Le Bot, Miao-Chih Tsai, Jay M. Campbell,
Julie Ahringer, and John G. White. The c. elegans hook protein, zyg-12, mediates the
essential attachment between the centrosome and nucleus. *Cell*, 115(7):825–836, Dec
2003. 1.1

Ion I. Moraru, James C. Schaff, Boris M. Slepchenko, and Leslie M. Loew. The virtual
cell: an integrated modeling environment for experimental and computational cell biol-
ogy. *Ann N Y Acad Sci*, 971:595–596, Oct 2002. 6.9

R. F. Murphy. Systematic description of subcellular location for integration with pro-
teomics databases and systems biology modeling. In *Proc. 4th IEEE Int. Symp.
Biomedical Imaging: From Nano to Macro ISBI 2007*, pages 1052–1055, 2007. doi:
10.1109/ISBI.2007.357036. 1.6.3

M. Osborn, R.E. Webster, and K. Weber. Individual microtubules viewed by immunoflu-
orescence and electron microscopy in the same ptk2 cell. *J Cell Biol*, 77:27–34, 1978.
4.11.2

Arthur E. C. Pece and Rasmus Larsen. Guest editorial: Generative model based vision.

Comput. Vis. Image Underst., 106:3–4, April 2007. ISSN 1077-3142. doi: 10.1016/j.
cviu.2006.10.006. URL http://dl.acm.org/citation.cfm?id=1235884.
1235964. 1.6.3

A. Ponti, A. Matov, M. Adams, S. Gupton, C. M. Waterman-Storer, and G. Danuser. Peri-
odic patterns of actin turnover in lamellipodia and lamellae of migrating epithelial cells
analyzed by quantitative fluorescent speckle microscopy. *Biophys J*, 89(5):3456–3469,
Nov 2005. doi: 10.1529/biophysj.104.058701. URL http://dx.doi.org/10.
1529/biophysj.104.058701. 1.5

E. Reaven. Stereological analysis of microtubules in cells with special reference to their
possible role in secretion. *Methods Cell Biol*, 25:273–283, 1982. 4.1

Christiane Richter-Landsberg. The cytoskeleton in oligodendrocytes. microtubule
dynamics in health and disease. *J Mol Neurosci*, 35(1):55–63, May 2008.
doi: 10.1007/s12031-007-9017-7. URL http://dx.doi.org/10.1007/
s12031-007-9017-7. 1.3

P. Ross-Macdonald, P. S. Coelho, T. Roemer, S. Agarwal, A. Kumar, R. Jansen, K. H. Che-
ung, A. Sheehan, D. Symoniatis, L. Umansky, M. Heidtman, F. K. Nelson, H. Iwasaki,
K. Hager, M. Gerstein, P. Miller, G. S. Roeder, and M. Snyder. Large-scale analysis
of the yeast genome by transposon tagging and gene disruption. *Nature*, 402(6760):
413–418, Nov 1999. doi: 10.1038/46558. URL http://dx.doi.org/10.1038/
46558. 1.6.2

Joseph Rudnick and George Gaspari. *Elements of the Random Walk: An introduc-*

tion for Advanced Students and Researchers. Cambridge University Press, May 2004. ISBN 0521828910. URL `http://www.amazon.com/exec/obidos/ redirect?tag=citeulike07-20&path=ASIN/0521828910`. 2.12

A. Santamara-Pang, T.S. Bldea, C.M. Colbert, P. Saggau, and I.A. Kakadiaris. Towards segmentation of irregular tubular structures in 3d confocal microscope images. In *Proc. MICCAI Workshop in Microscopic Image Analysis and Applications in Biology (MI-AAB), Copenhagen, Denmark, Oct. 1-6, 2006*, 2006. 1.5

M. E. Sargin, A. Altinok, E. Kiris, S. C. Feinstein, L. Wilson, K. Rose, and B. S. Manjunath. Tracing microtubules in live cell images. In *Proc. 4th IEEE Int. Symp. Biomedical Imaging: From Nano to Macro ISBI 2007*, pages 296–299, 2007. doi: 10.1109/ISBI.2007.356847. 1.5

Sbastien Schaub, Jean-Jacques Meister, and Alexander B. Verkhovsky. Analysis of actin filament network organization in lamellipodia by comparing experimental and simulated images. *J Cell Sci*, 120(Pt 8):1491–1500, Apr 2007. doi: 10.1242/jcs.03379. URL `http://dx.doi.org/10.1242/jcs.03379`. 1.5

A. Shariff, R.F. Murphy, and G.K. Rohde. A generative model of microtubule distributions, and indirect estimation of its parameters from fluorescence microscopy images. *Cytometry A*, 77:457–66, 2010a. 1, 1, 3.4, 4.1, 4.4, 4.5, 4.6, 12, 4.8, 5.5

A. Shariff, R.F. Murphy, and G.K. Rohde. Automated estimation of microtubule model parameters from 3-d live cell microscopy images. *Proceedings of the 2011 IEEE Inter-*

national Symposium on Biomedical Imaging (ISBI 2011), pages 1330–1333, 2011. 1, 4.1, 4.7

Aabid Shariff, Joshua Kangas, Luis Pedro Coelho, Shannon Quinn, and Robert F. Murphy. Automated image analysis for high-content screening and analysis. *J Biomol Screen*, 15(7):726–734, Aug 2010b. doi: 10.1177/1087057110370894. URL `http://dx.doi.org/10.1177/1087057110370894`. (document), 1, 1.6.1, 1.4

E. Shestakova, J. Vandekerckhove, and J.R. De Mey. Epithelial and fibroblastoid cells contain numerous cell-type specific putative microtubule-regulating proteins, among which are ezrin and fodrin. *Eur J Cell Biol*, 75:309–320, 1998. 4.11.2

Daniel R. Sisan, Richard Arevalo, Catherine Graves, Ryan McAllister, and Jeffrey S. Urbach. Spatially resolved fluorescence correlation spectroscopy using a spinning disk confocal microscope. *Biophys J*, 91(11):4241–4252, Dec 2006. doi: 10.1529/biophysj.106.084251. URL `http://dx.doi.org/10.1529/biophysj.106.084251`. 1.5

I. Smal, K. Draegestein, N. Galjart, W. Niessen, and E. Meijering. Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: Application to microtubule growth analysis. 27(6):789–804, 2008. doi: 10.1109/TMI.2008.916964. 1.6.1

F. Solomon. Neuroblastoma cells recapitulate their detailed neurite morphologies after reversible microtubule disassembly. *Cell*, 21(2):333–338, Sep 1980. 3.1

Brian L. Sprague, Chad G. Pearson, Paul S. Maddox, Kerry S. Bloom, E. D.

Salmon, and David J. Odde. Mechanisms of microtubule-based kinetochore positioning in the yeast metaphase spindle. *Biophys J*, 84(6):3529–3546, Jun 2003. doi: 10.1016/S0006-3495(03)75087-5. URL `http://dx.doi.org/10.1016/S0006-3495(03)75087-5`. 1.5

Krasimira Tsaneva-Atanasova, Andrea Burgo, Thierry Galli, and David Holcman. Quantifying neurite growth mediated by interactions among secretory vesicles, microtubules, and actin networks. *Biophys J*, 96(3):840–857, Feb 2009. doi: 10.1016/j.bpj.2008.10.036. URL `http://dx.doi.org/10.1016/j.bpj.2008.10.036`. 1.4, 5.1

Mathias Uhln, Erik Bjrling, Charlotta Agaton, Cristina Al-Khalili Szigyarto, Bahram Amini, Elisabet Andersen, Ann-Catrin Andersson, Pia Angelidou, Anna Asplund, Caroline Asplund, Lisa Berglund, Kristina Bergstrm, Harry Brumer, Dijana Cerjan, Marica Ekstrm, Adila Elobeid, Cecilia Eriksson, Linn Fagerberg, Ronny Falk, Jenny Fall, Mattias Forsberg, Marcus Gry Bjrklund, Kristoffer Gumbel, Asif Halimi, Inga Hallin, Carl Hamsten, Marianne Hansson, My Hedhammar, Grel Hercules, Caroline Kampf, Karin Larsson, Mats Lindskog, Wald Lodewyckx, Jan Lund, Joakim Lundeberg, Kristina Magnusson, Erik Malm, Peter Nilsson, Jenny Odling, Per Oksvold, Ingmarie Olsson, Emma Oster, Jenny Ottosson, Linda Paavilainen, Anja Persson, Rebecca Rimini, Johan Rockberg, Marcus Runeson, Asa Sivertsson, Anna Skllermo, Johanna Steen, Maria Stenvall, Fredrik Sterky, Sara Strmberg, Mrten Sundberg, Hanna Tegel, Samuel Tourle, Eva Wahlund, Annelie Waldn, Jinghong Wan, Henrik Wernrus, Joakim Westberg, Kenneth Wester, Ulla Wrethagen, Lan Lan Xu, Sophia Hober, and Fredrik Pontn. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell*

Proteomics, 4(12):1920–1932, Dec 2005. doi: 10.1074/mcp.M500279-MCP200. URL `http://dx.doi.org/10.1074/mcp.M500279-MCP200`. 1.6.2

M. Velliste and R. F. Murphy. Automated determination of protein subcellular locations from 3d fluorescence microscope images. In *Proc. IEEE Int Biomedical Imaging Symp*, pages 867–870, 2002. doi: 10.1109/ISBI.2002.1029397. 2.2, 4.2.1

Julian Weichsel, Nikolas Herold, Maik J. Lehmann, Hans-Georg Krusslich, and Ulrich S. Schwarz. A quantitative measure for alterations in the actin cytoskeleton investigated with automated high-throughput microscopy. *Cytometry A*, 77(1):52–63, Jan 2010. doi: 10.1002/cyto.a.20818. URL `http://dx.doi.org/10.1002/cyto.a.20818`. 6.8.2

K.W. Wolf and K. Spanel-Borowski. Acetylation of alpha-tubilin in different bovine cell types: implications for microtubule dynamics in interphase and mitosis. *Cell Biol Int*, 19:43–52, 1995. 4.1

Ge Yang, Benjamin R. Houghtaling, Jedidiah Gaetz, Jenny Z. Liu, Gaudenz Danuser, and Tarun M. Kapoor. Architectural dynamics of the meiotic spindle revealed by single-fluorophore imaging. *Nat Cell Biol*, 9(11):1233–1242, Nov 2007. doi: 10.1038/ncb1643. URL `http://dx.doi.org/10.1038/ncb1643`. 2.5

Ting Zhao and Robert F. Murphy. Automated learning of generative models for subcellular location: building blocks for systems biology. *Cytometry A*, 71(12):978–990, Dec 2007. doi: 10.1002/cyto.a.20487. URL `http://dx.doi.org/10.1002/cyto.a.20487`. (document), 1.6.3, 1.7, 2.1.1, 5.1, 5.5

Ting Zhao, Meel Velliste, Michael V. Boland, and Robert F. Murphy. Object type recognition for automated analysis of protein subcellular location. *IEEE Trans Image Process*, 14(9):1351–1359, Sep 2005. 5.1