Generative Models for Super-Resolution Single Molecule Microscopy Images of Biological Structures

Suvrajit Maji

CMU-CB-12-104
August 2012

Lane Center for Computational Biology
School of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213

Thesis Committee:
Marcel P. Bruchez, Advisor
Gustavo K. Rohde
Frederick Lanni
Daniel M. Zuckerman, University of Pittsburgh

Submitted in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy

This research was supported by the National Institutes of Health grant 5R01-GM086237 and the Human Frontier Science Program grant RGP0063/2009. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, or the U.S. Government, or any other entity.

Keywords: Single molecule, super resolution, STORM imaging, Hough Transform, generative models, probabilistic graphical model, Markov random field, nonparametric belief propagation

To my parents and my sisters.

Abstract

The ultimate objective of any biological imaging method is to understand the underlying biology especially through observing, analyzing and understanding the structures. Biological imaging using conventional light microscopy is limited in resolution due to the diffraction barrier of light and hence obtaining the detailed structural information is difficult. Recently, there has been advancement in this field to break the resolution limit with several super resolutions microscopy methods based on localization of single molecules such as STORM, PALM, etc. which are evolving into important tools for structural biology. A catalog of molecular positions provides insight into underlying structures potentially at molecular length scales and demands computational approaches that utilize the inherent positional information to extract meaningful structural biology–scale information about cellular structures. These methods still suffer from localization of single molecules in 3-d and we show that there are imaging methods which can improve the localization of multiple distinct molecules in a cell in 3-D. Moreover, there are limitations for dynamic imaging with these methods, since dynamic structures require information from fewer positions (i.e. shorter time window) to minimize underlying motion. Hence, the datasets are inherently incomplete. Our aim is to provide an information bridge between super-resolution microscopy and structural biology by using generative models to get a "molecular length-scale" picture of cellular structures. We hypothesize that generative models can accurately reconstruct biological structures using less data and with better resolution and infer useful biological information such as characteristic lengths, orientation of filamentous structure, molecular distributions for proteins inside a cell.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor and mentor Dr. Marcel Bruchez, for his invaluable support, guidance and encouragement for the past four years. It has been a privilege and a wonderful experience pursuing my research projects under his supervision and this thesis would not have been possible without his support. I would also like to thank my thesis committee members Dr. Gustavo Rohde, Dr. Frederick Lanni and Dr. Daniel Zuckerman for their helpful comments and discussions on my thesis projects.

I would like to sincerely thank my lab colleagues Qi Yan, Rowena Mittal, Yi Wang, Richa Verma, Anmol Grover, Saumya Saurabh, Christopher Pratt, Dmytro Yuschenko, Cheryl Telmer for creating such a vibrant environment in the research group and never let me feel as the odd one, in spite of me being the only non-experimental student in the lab. I can never thank them enough for being such great colleagues and friends and making the journey such a pleasant one. I will always cherish those fun moments and the wonderful time that I have spent with them in the lab.

I would like to thank Dr. Gregory Fisher who was my office mate for majority of my graduate studies here at Carnegie Mellon for keeping my interest in mathematics continuing with wonderful discussions and for his encouragement and motivation. I would like to thank other members of the Molecular Biosensor and Imaging Center, Sue Andreko, Dr. Chris Szent-Gyorgyi, Dr. Byron Ballou, Dr. Lauren Ernst, Dr. Haibing Teng, Alison Dempsey, Mingrui Zhang, Nina Senutovitch, Elvira Highley, Michael Patrick, Siddhesh Angle, William Clafshenkel, Margaret Fuhrman and Sally Adler for their encouragement. Thanks to Donna Smith and Robert Bordnar for taking care of all my administrative and logistics needs at the Center.

Thanks to all my colleagues and friends in the CPCB PhD program for making my stay in the program an enjoyable time. I would also like to thank Thom

Gulish for helping me with all my administrative requirements in the PhD program.

A very special thanks goes to my friends Sourav Chatterjee, Debaditya Dutta, Yujie Ying, Hitashyam Maka, Professor Sourav Bhattacharya, Mrs. Sudarshana Bhattacharya for their love and support during my stay in Pittsburgh.

I will forever be indebted to my parents and my two sisters for their unconditional love and support in everything I have pursued in my life, without them I would not be here.

CONTENTS

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION TO SUPER–RESOLUTION SINGLE MOLECULE IMAGING

Learning about most biological processes begins at the very basic level of the molecular structures of the components involved. In order to obtain the relevant information and be able to study them, there must be a way to visualize the structure. It will not be an exaggeration if we say that the development of the entire field of biology has been possible due to our capability to image the cells, which began with Antonie Van Leeuwenhoek's prototype of modern day's microscope more than 300 years ago when he discovered bacteria, protozoans, muscle cells, etc. Before, that Zacharias Jansen and his father Hans Jansen (1595) of Holland invented a compound light microscope and later Robert Hooke (1665) from England further refined it and actually coined the term 'cell' by looking at walls in cork tissue (plant). The major improvement in the microscope optics was achieved in 19th century due to the effort of Carl Zeiss and Ernst Abbe. It is only recently that we have seen methods such as X-ray diffraction, Nuclear Magnetic Resonance (NMR), Electron microscopy (EM), cryoEM and so on, besides light microscopy, which can capture the structural information at various scales and various components with amazing level of details. However, for learning about many biological processes and structures, often the most practical way is to image them using light microscopy methods due to the capability of staining and tagging individual cells and individual molecules of interest. Now to obtain a

detailed knowledge about the underlying biology, it is imperative that we should be able to observe the biological structures at the best possible resolution, which is defined as the ability of an optical system to resolve between closely spaced objects. We will discuss this concept in a later section. However, for many problems the resolution of the images obtained are not sufficient to reveal the intricate structural details due to the diffraction barrier of light, which was first introduced by Ernst Abbe in 1873. Super-resolution (SR) imaging has led to various important studies in biology, which could not have been achieved with conventional microscopy. There are varieties of recent methods that achieve resolution far beyond the diffraction limit. These belong to two categories of methods, which works on the principle of optical switching of the fluorophores by manipulating the neighboring molecules in different states of activation so that it is easier for them to be optically resolved (Hell and Wichmann, 1994) . The first one is patterned illumination to spatially modulate the fluorescence of the molecules so that a subset of molecules are emitting at any given time and thus the effective point spread function of the molecules are reduced within the diffraction limited region causing an improvement of resolution. This is an ensemble imaging method and some of the most popular methods are stimulated emission and depletion (STED) microscopy (Hell and Wichmann, 1994) , ground state depletion (GSD) (Folling et al., 2008; Kroug, 1995), saturated structured illumination microscopy (SSIM) (Gustafsson, 2000) and so on. The other category comprises of the methods where the idea is to stochastically activate individual molecules at different times. This makes it easier to determine the position of the molecules (localization) and then the structure can be reconstructed based on the measured position of the molecules. The methods that are based on this principle are called stochastic optical reconstruction microscopy (STORM) (Rust et al., 2006),

photo-activated localization microscopy (PALM) (Betzig et al., 2006), fluorescence PALM (fPALM) (Hess et al., 2006). These methods allow us to study the dynamics of complex heterogeneous systems such as living cells at the single molecule level. The ability to obtain information of individual molecules using single particle tracking (SPT) has opened up new avenues that were previously not possible using ensemble averaging techniques. However, meaningful biological studies using SPT and Super-Resolution imaging require an extremely precise localization of single molecules in three dimensions.

There are various optical fluorescence imaging techniques available in literature, which are used in the studies of cellular structures and biological processes. In particular super-resolution imaging methods such as localization microscopy can achieve extremely good lateral localization accuracy, but the axial localization is not that good. There are several methods which allow some determination of z position from the localization data sets e.g. defocused (Juette et al., 2008; Speidel et al., 2003; Toprak et al., 2007; Zhang and Menq, 2008) and distortion (astigmatism) approaches (Holtzer et al., 2007; Huang et al., 2008; Kao and Verkman, 1994) , but compromises with photon efficiency and axial localization accuracy. These approaches work best for quantum dots and other bright probes, but are still limited with typical fluorescence probes in living cells. In addition, to implement these with multiple colors as currently designed would require an exceptionally split (and inefficient) collection path, running multiple colors through multiple filters and distortion or focus-shifting optics. Also, conventional microscopes image only one focal plane at any time. Therefore, to study three-dimensional dynamics inside living cell one has to move the objective in sequential steps, which limits the localization and events of interest

due to slower acquisition speed with respect to the timescale of cellular events. Recently, a dual plane (Ram et al., 2008; Toprak et al., 2007; Watanabe et al., 2007) and dual objective (Ram et al., 2009a) imaging system that eliminates such imaging limitations has reported impressive 3-dimensional localization with around 10-20 nm axial localization accuracy. However, such imaging setup is not ideal for efficient multicolor localization. In addition, post-processing and analysis of localization data would still be computationally expensive. In order to perform super-resolution imaging on multiple colors, in 3-dimensions, we need to have a more efficient optical design, and novel and faster algorithms that allow determination of the object position from such a collection system, in spite of different optical transfer functions for each of the collected colors.

1.1 Breaking the diffraction barrier of light

Spatial resolution for observing fluorescence molecules under light microscopy is limited due to the diffraction barrier of light. In general, any point object in a microscope generates a diffraction pattern (Fraunhofer). The size of the diffraction-limited spot depends on the wavelength of the light and the angle of the objective. The separation between two objects then is limited by the interference pattern. If the first minima of one object either coincide with principle maxima of the second object then that provides the limit of the resolution. Ernst Abbe first calculated the resolution limit and it is called as the Abbe Limit (Abbe, 1873) which is discussed a little later.

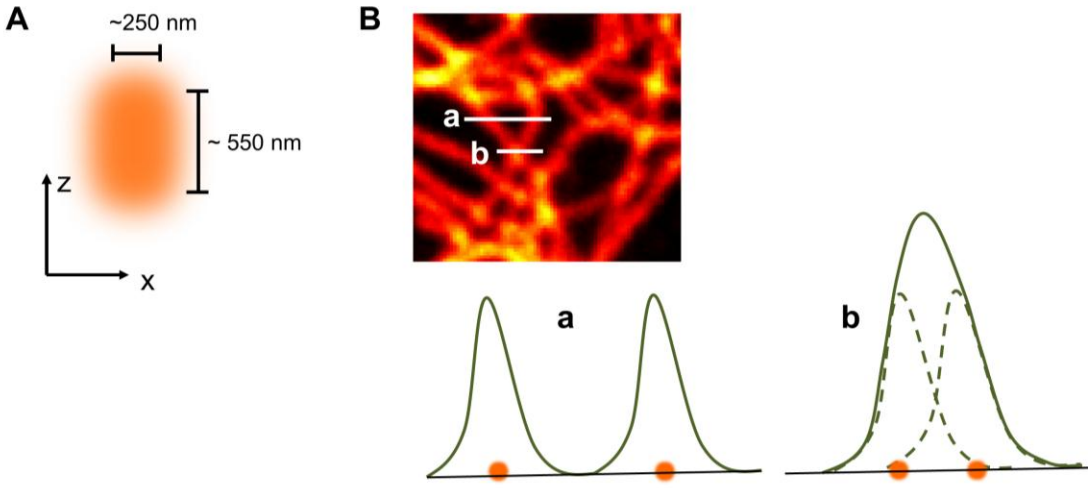


Figure 1.1 Diffraction limited resolution of conventional microscopy

For a conventional microscope the focal spot of a point emitter is shown in Figure 1.1A . The width of the point emitter as imaged through the objective is the actual diffraction-limited resolution. When these optical setup is used for imaging biological structures whose features are smaller than the diffraction limited spot size then we see the image shown in Figure 1.1B. This forced scientists to think about ways to resolve this barrier of biological imaging.

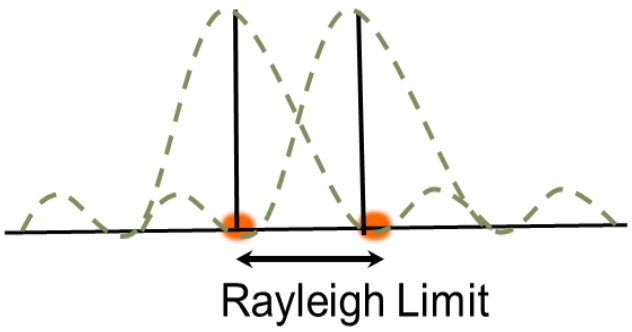


Figure 1.2 Resolution limit for overlapping fluorophores

Ernst Abbe provided a quantitative estimate of the resolution limit given by the following:

$$\Delta x, y = \frac{\lambda}{2\,NA}, \qquad \Delta z = \frac{2\lambda}{NA^2} \tag{1.1}$$

where the lateral resolution is given by $\Delta x, \Delta y$ and the axial resolution is given by Δz. λ is the wavelength of the light, n is the refractive index of the medium and NA is the numerical aperture.

Later Rayleigh provided a more appropriate limit for the diffraction limited resolution as shown in Figure 1.2. For noise free images the resolution limit according to Rayleigh criteria is given by:

$$\Delta x, y = 0.61\frac{\lambda}{NA}, \qquad \Delta z = \frac{2\lambda n}{NA^2} \tag{1.2}$$

When two point emitters are farther than this resolution limit, they appear as two separate objects as shown in part (a) of Figure 1.1B and are easily resolvable , whereas if they are less than this limit, then they appear as a single object and unresolvable as shown in part (b) of Figure 1.1B.

For typical microscope setups, the lateral resolution is around 200-250 nm and the axial resolution is around 500 nm. As we can see in Figure 1, that when objects are separated by a distance larger than the resolution limit they are seen as separate objects otherwise they will appear as a single unresolvable object. As a consequence, obtaining detailed knowledge and visualization of sub-cellular structures such as vesicles, microtubules, mitochondria, etc. which are sub-resolution (< 10 - 100 nm) sizes, are not possible as they appear as blurred spots

when imaged under light microscopes (Abbe, 1873). To overcome this limitation, new optical imaging systems have been designed along with computational approaches that now enable us to observe those structures far beyond the diffraction limit and thus these methods are collectively called Super-resolution imaging.

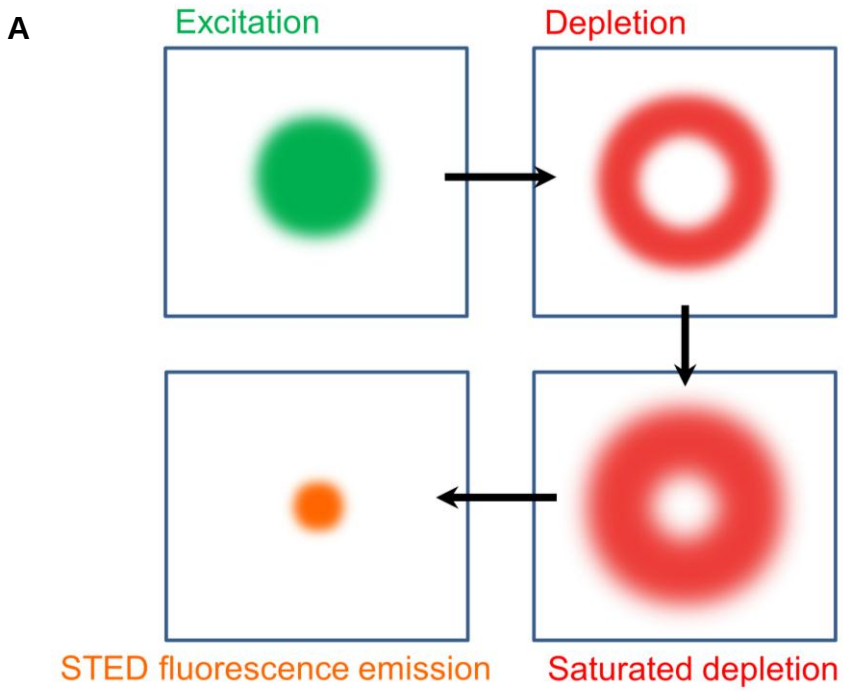
1.2 Super-Resolution methods

Here we describe the various super-resolution methods currently used in literature in little more details. The first categories of methods are based on non-linear optical imaging with a deterministic activation of fluorophores and the second category is based on computational approach post image acquisition, mostly through the localization of single molecules, which are activated stochastically.

1.2.1 Optical methods for super-resolution

There are non-linear optical approaches such as Stimulated Emission and Depletion (STED), Ground State Depletion (GSD), Saturated Structured Illumination Microscopy (SSIM), Vertico Spatially Modulated Illumination (Vertico-SMI) (Reymann et al., 2008) which effectively reduces the Point spread function (PSF) through optical manipulation technique. The idea is to use patterned illumination to spatially modulate the fluorescence behavior of a subset of molecules and thus achieve sub-diffraction resolution. These methods fall under the category of far-field microscopy with light waves showing the properties of Fraunhofer diffraction. The other class of optical methods uses the near-field properties of light where the phenomenon of diffraction of light is no

longer true. Some of the notable methods are near-field scanning optical microscopy (NSOM) (Betzig et al., 1986) which places a detector very near to the specimen surface, the distance being less than the wavelength of light and apertureless near-field scanning optical microscopy (ANSOM). Within this near field the evanescent waves is not diffraction limited and hence nanometer spatial resolution is possible. The limitation of these methods is that specimens have to be placed at immediate proximity to the optical probes and it can be used for only imaging the surface structures. The concept of achieving resolution far beyond the diffraction barrier was first introduced by Stefan Hell through a family of techniques collectively called as reversible saturable (switchable) optical fluorescence transitions or RESOLFT (Hell and Wichmann, 1994). The underlying principle of this concept is to reversibly and deterministically switch between two distinct states A and B or a fluorescent 'on' and dark 'off' state which is determined by the probability of molecules in each state. The most common examples are STED, GSD, and SSIM.



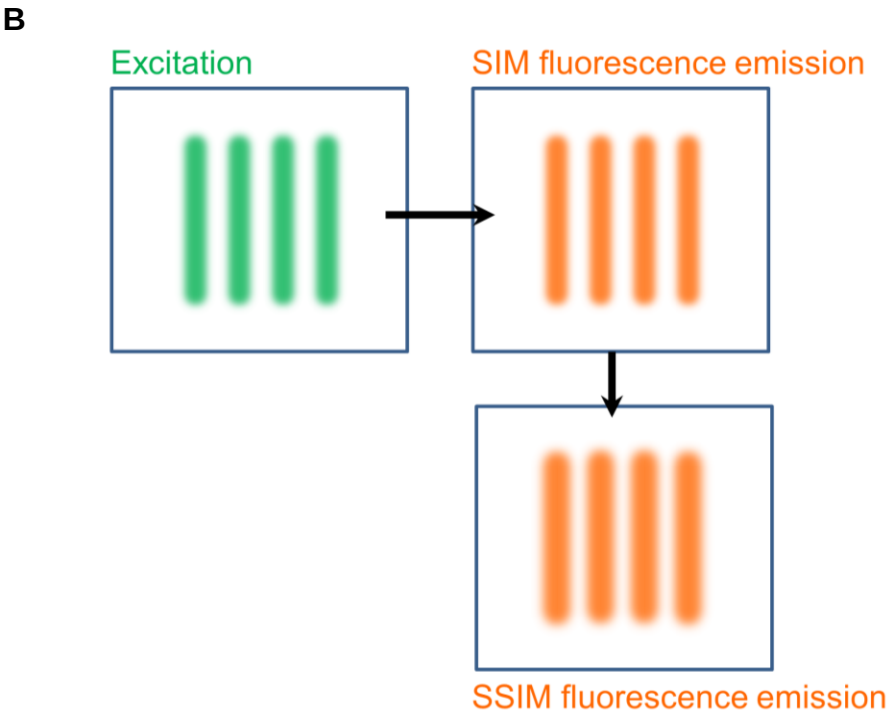


Figure 1.3. Working principle of (A) STED (B) (S)SSIM

The working principle of STED is shown in Figure 1.3A (Huang et al., 2010). At first, the fluorophores are excited to a higher energy state by a focused light beam shown in the green and followed up with depletion beam shown in red to bring the molecules back to the ground state through a process called stimulated emission. The intensity profile of the depletion beam is usually doughnut shaped so that when the depletion happens it is usually on the outer sides of the focal spot leaving the molecules towards the central region to be still in the excited state. This produces a considerable decrease in the overall size of the fluorescent spot and thus effectively improves the image resolution to a sub-diffraction level. Although the depletion intensity pattern is produced by the

diffraction-limited optics, since the molecules respond in a non-linear manner, it can achieve the enhancement.

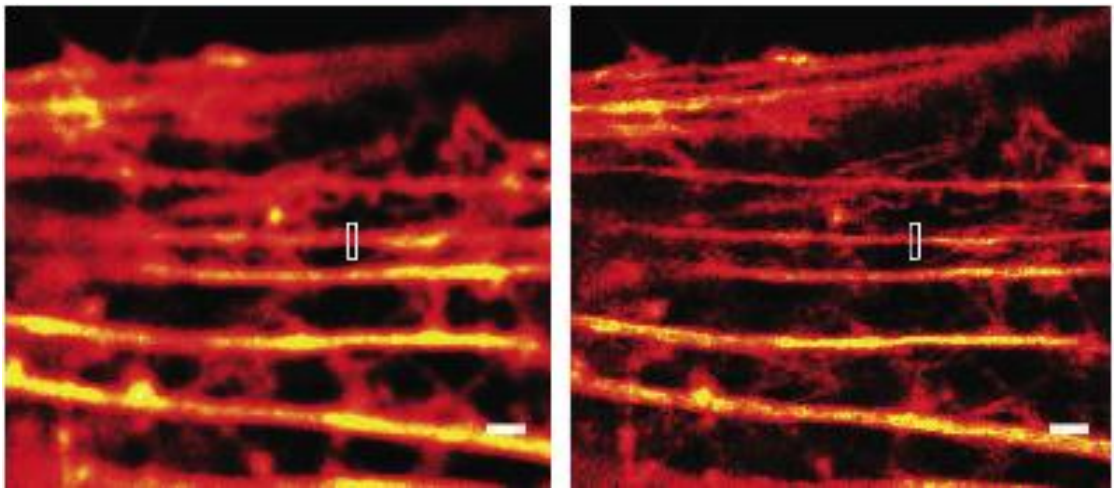


Figure 1.4. Resolution enhancement using STED (Fitzpatrick et al., 2009)

The resolution (δ) of the microscope is inversely proportional to the root of the intensity of the depletion beam:

$$\delta = \frac{\Delta}{\sqrt{1 + I_{dep}/I_{sat}}} \tag{1.3}$$

where Δ is the diffraction-limited focal spot size measured as the full width at half maxima of the intensity, the I_{dep} is the intensity of the depletion beam and I_{sat} is the intensity of the saturated beam

In case of the structured illumination microscopy (SIM) and saturated SIM (SSIM) as shown in the Figure 1.3B, the illumination is sinusoidal producing the pattern shown in green. This in turn generates a similar pattern, shown in orange, when the molecules respond linearly to the excitation beam. When the excitation beam intensity strength is increased, the fluorescence pattern becomes saturated and then it generates the pattern shown in SSIM fluorescence emission with narrower unexcited regions, which effectively improves the resolution.

So theoretically, these methods are limited by the amount of depletion light source, but there are some practical limitations such as optical aberrations, photo stability of the fluorophores, etc. The optical resolution achieved by STED is about 20 nm for organic dyes and 50-70 nm for fluorescent proteins. The resolution obtained with SIM is 100 nm and 50 nm with SSIM in lateral dimension. These methods have also been used for 3-D imaging, for example, isoSTED with a z depletion pattern can achieve resolution of about 50nm in all three dimensions. In 3D SIM, three beams of patterned illumination are projected onto the samples and it creates an interference pattern known as moiré fringes in the lateral and the axial directions. The resolution achieved by 3D SIM is ~100 nm in lateral and ~300nm in the axial directions. A detailed review can be found in (Huang et al., 2010).

1.2.2 Computation based super-resolution methods

Any biological structure can be thought of as a collection of individual components (molecules). Therefore, the resolution can also be improved through localization of the centroid of the single molecule, which are fluorescently activated over the biological structure in a stochastic manner. We are going to mainly focus on the far-field approaches with Fraunhofer diffraction. The notable amongst those methods are Photoactivated localization microscopy (PALM), Fluorescence PALM (FPALM), Stochastic Optical Reconstruction Microscopy (STORM), Fluorescence Imaging with One Nanometer Accuracy (FIONA) (Yildiz et al., 2003; Yildiz and Selvin, 2005), Super High Resolution Imaging with Photobleaching (SHRImP) (Gordon et al., 2004), Single molecule High REsolution Colocalization (SHREC) (Churchman et al., 2005) , point accumulation for imaging in nanoscale topography (PAINT) (Sharonov and Hochstrasser, 2006) and so on. One of most interesting non-localization based computational method is Super-resolution Optical Fluctuation Imaging (SOFI) (Dertinger et al., 2009). There is a recent method for modeling single molecule data called Bayesian analysis of blinking and bleaching (3B) (Cox et al., 2012) which provides a very powerful and interesting approach on studying biological problems with single molecule imaging . Some of these notable techniques are described below.

1.2.2.1 STORM / (F)PALM

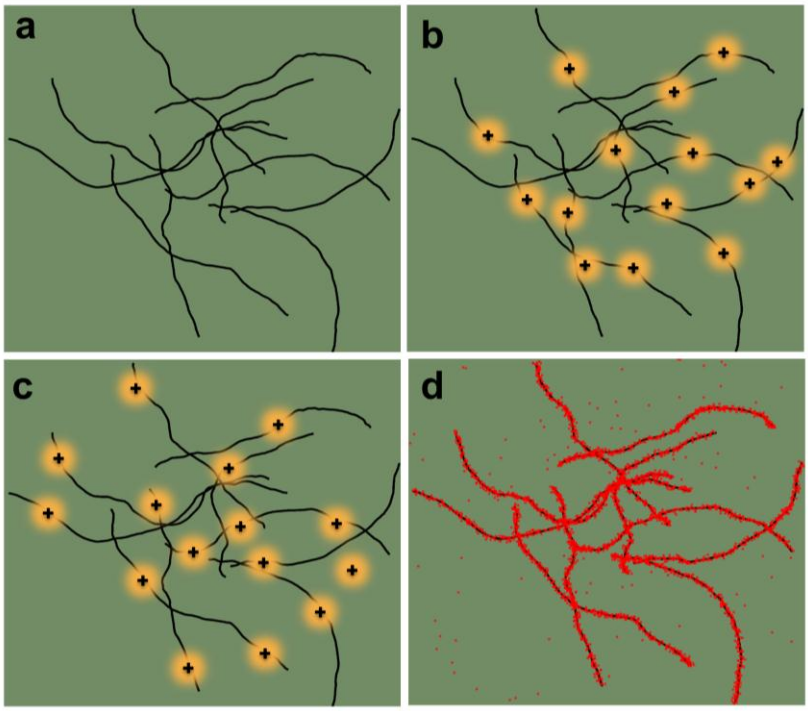


Figure 1.5. STORM , (F)PALM imaging method for biological structures

Figure 1.5 shows how STORM and PALM imaging reconstructs the biological structure. Figure 1.5a shows the ground truth for the structure. Figure 1.5b and Figure 1.5c shows stochastic activation of the two different subsets of molecules at different time points. Figure 1.5d shows the reconstruction of the biological structure after localization of the single molecules from all such time points.

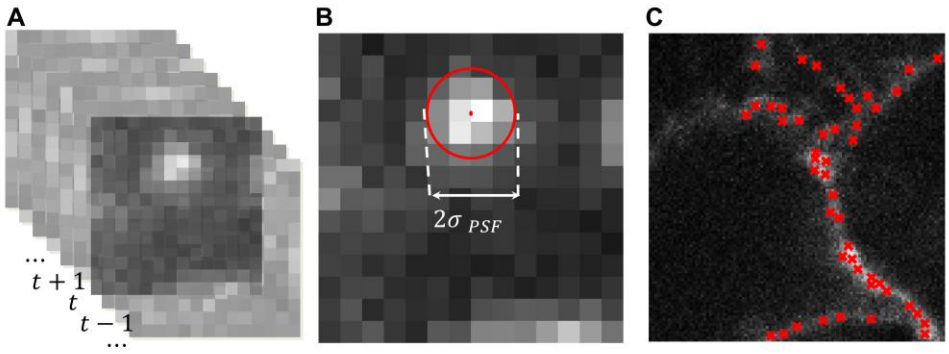


Figure 1.6. STORM , (F)PALM Imaging principle (A) Image acquisition (B) Single Molecule Localization (C) Localized position mapping

STORM uses photo-switchable probes and is reversible whereas (F)PALM is an irreversible process since after a single step of photobleaching , the molecules do not reappear. The basic imaging principle of STORM and PALM microscopy is shown in Figure 1.6. The images are acquired and then analyzed frame by frame. The diffraction limited spots are then segmented and fitted with a parameterized theoretical point spread function usually a Gaussian or an Airy function described in chapter 2. The estimation of the centroid of the Gaussian determines the position of the single molecules. This is performed for all the objects in all the frames and then all the positions are accumulated in a composite image with a higher-level pixel sampling.

The structural resolution is dependent on the localization accuracy σ_{xy} of the single molecules and the density of the molecules that are detected. The localization accuracy (Thompson et al., 2002) is given by :

$$\sigma_{xy} = \sqrt{\frac{\sigma_{PSF}^2}{N} + \frac{a^2/12}{N} + \frac{8\pi\sigma_{PSF}^4 b^2}{a^2 N^2}} \tag{1.4}$$

where σ_{PSF} is the standard deviation of the point spread function, a is the pixel size, b is the background noise and N is the number of photons.

If $\bar{\sigma}_{xy}$ is the mean localization accuracy for all the molecules in the image and $\bar{d}_{NN}$ is the mean of the pairwise nearest neighbor distance of the molecules , providing the sampling density information, then the structural resolution R is given by (Kaufmann et al., 2012):

$$R = \sqrt{(2.35\ \bar{\sigma}_{xy})^2 + (2\ \bar{d}_{NN})^2} \tag{1.5}$$

Although, we see a substantial improvement in the spatial resolution with these imaging methods, this still lacks the capability of high-speed image acquisition, which is critical for studying dynamic imaging. Recent demonstrations using very high laser power improved the frame-capture timescale by an order-of-magnitude by accelerating the localization and deactivation cycle time (Jones et al., 2011). While this approach achieved 0.5–2 second acquisition speeds, this still poses a challenging limit for many biological processes with timescales of milliseconds or even lower. Other methods such as FIONA, SHRImP and SHREC are all based on the same principle with different conditions and applicability.

1.2.2.2 SOFI

This method uses the fluorescence blinking of fluorophores to improve the spatial resolution. The working principle of SOFI (Dertinger et al., 2009) is based on the assumption that blinking behavior of the neighboring fluorescent molecules is statistically independent, whereas the single molecule spatio-

temporally correlates with itself. As a result the temporal correlation of each pixel through the time series can generate a higher resolution image due to the reduced effective point spread function size.

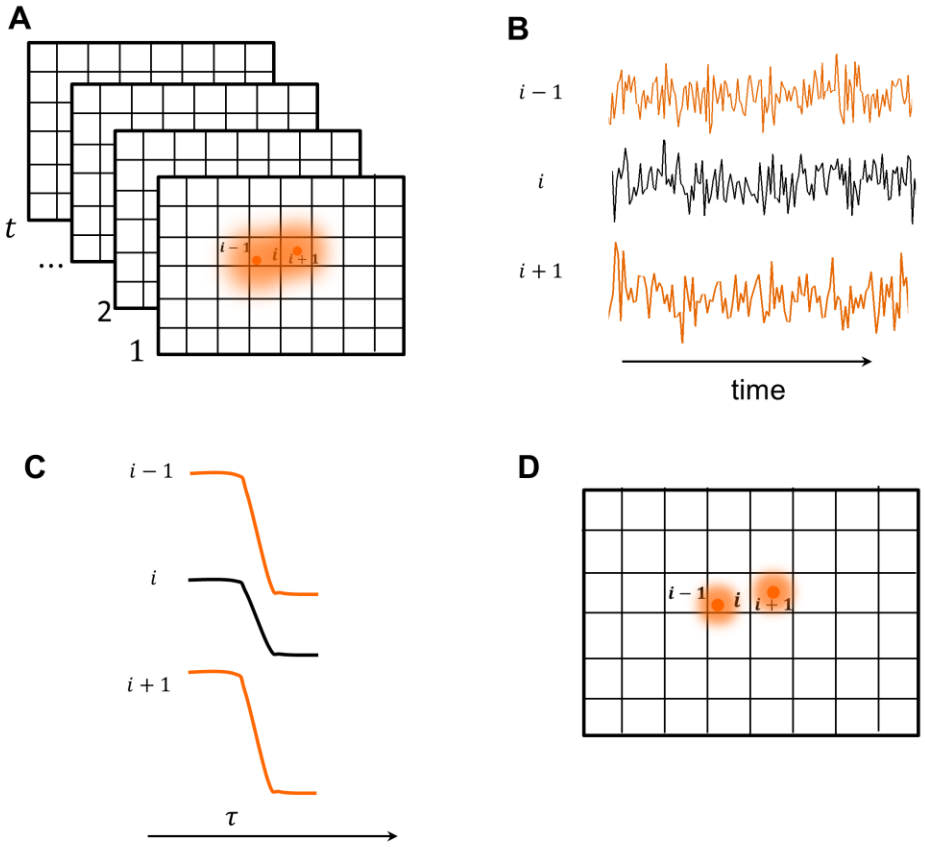


Figure 1.7 Basic principle of SOFI analysis. (A) Emitter distribution (B) Time series fluctuation of the pixels (C) second-order correlation function calculated from the fluctuations in (B). (D) SOFI intensity for the corresponding pixels.

The emitter fluorescence distribution is shown for two overlapping fluorophores in Figure 1.7A. The first step in SOFI is to collect the signal from the emitter fluorescence distribution and convolve the signals with the PSF of the optical system. Then the convolved values are recorded on sub-diffraction pixels so that

the information of the diffraction-limited spot is spread over multiple pixels. As a result, each pixel now contains the time series information, consisting of the sum of signals (Figure 1.7B) from the emitters whose PSFs are part of the pixel. The next step is to calculate the second-order correlation function from the fluctuations recorded in each pixel as shown in Figure 1.7C. Then the higher order statistical cumulant, given by the integral over the second-order correlation function is computed for each pixel. This results in the SOFI image as shown in Figure 1.7D. The expected spatial resolution enhancement that can be achieved by SOFI imaging is a factor of $\sqrt{2}$, since the emitter signal is processed by a second-order correlation function that is proportional to the squared PSF. If we take even higher order correlations then it can reduce the noise further. There is a variant of SOFI called variance imaging for super-resolution (VISION) (Watanabe et al., 2010) which has achieved 80 ms temporal resolution , although the spatial resolution enhancement is limited.

There are some advantages of SOFI over localization-based methods. Since the fluorophores are uncorrelated, it can automatically distinguish between overlapping fluorophores and it can remove the background automatically. It requires dark state lifetime of the fluorophores to be on the order of the frame rate and the acquisition is usually faster than localization microscopy. It can be used alongside with most wide-field microscopy. There are some limitations to this method such as, the assumption that the positions of emitters are unchanged during the image acquisition, though this problem is fixable. The short acquisition time can generate noise in the correlation values and the fluorophore on-off switching rate will limit the acquisition speed for the method. SOFI has no single molecule information and so can be used only for super-resolution image reconstruction and not for single molecule studies.

1.2.2.3 Multiple Fluorophore fitting

In super-resolution microscopy, the usual approach is to localize the single molecules and perform the reconstruct the structures or do single particle tracking. However in many cases the density of single molecules high and for that reason the localization suffers due to overlapping fluorophores, since the model that is used to fit the single molecule intensity profiles is assuming there is only a single emitter in that space. Therefor a significant amount of data is either not properly localized or is discarded. In order to retain the valuable data approaches have been developed such as (Huang et al., 2011).This method uses a Bayesian maximum likelihood estimation method to localize multiple fluorophores in a given region of interest.

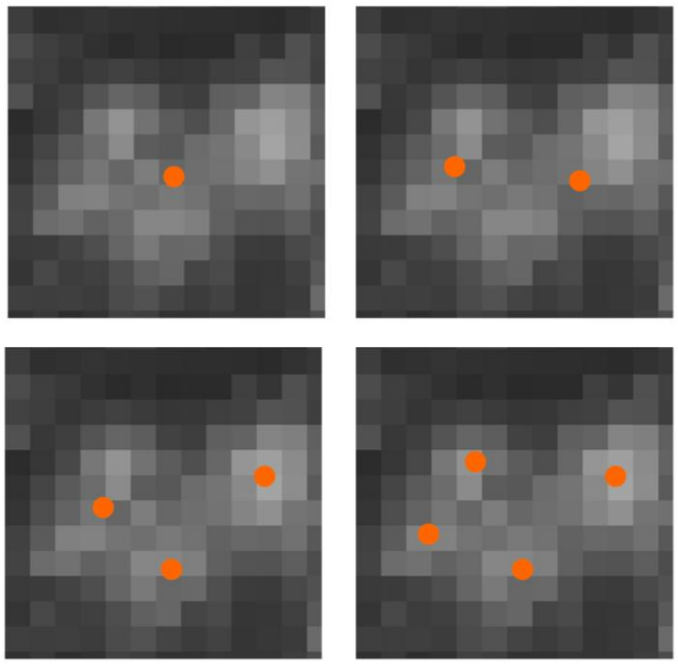


Figure 1.8 Example of multi-fluorophore fitting. Top left is 1 emitter fitting. Top right is 2 emitter fitting. Bottom left is 3 emitter fitting. Bottom right is 4 emitter fitting

Figure 1.8 shows an example of multiple fluorophore fitting where we see that with 4 emitter fitting , all the molecules are correctly localized compared to the lower order fittings.

1.2.2.4 DAOSTORM

This algorithm was originally developed as DAOPHOT (Stetson, 1987) for studying images of crowded stars in astronomy, has been adopted for resolving high-density super-resolution images. The basic idea is to find the fluorophores in the image on the first pass with single emitter fitting approach. Then subtract the fit from the original image and perform the fitting on the residual image in a iterative manner until no further emitters are left in the residual. Standard DAOSTORM (Holden et al., 2011) uses a fixed-shape model PSF although it has been extended to variable shape in the 3D-DAOSTORM (Babcock, 2012). The 3d version is significantly more efficient and faster than the 2d version.

1.2.2.5 3B Localization Microscopy

The Bayesian analysis of blinking and bleaching (3B) method models the blinking and bleaching mechanisms of multiple fluorophores using a Markov Chain Monte Carlo (MCMC) approach. The 3B method works by modeling over the full time series and it generates several possible models and a weighted average of all those models generates a probability map of the location of the fluorophores. It factors the prior information on the blinking and bleaching behavior of the fluorophores, their numbers, location and the temporal dynamics.

The emitting fluorophores are modeled using a Gaussian profile with a state space (x, y, r, b), where x, y is the position, r is the spot radius and b is the brightness of the emitter. The state transition for the Markov chain model is shown below:

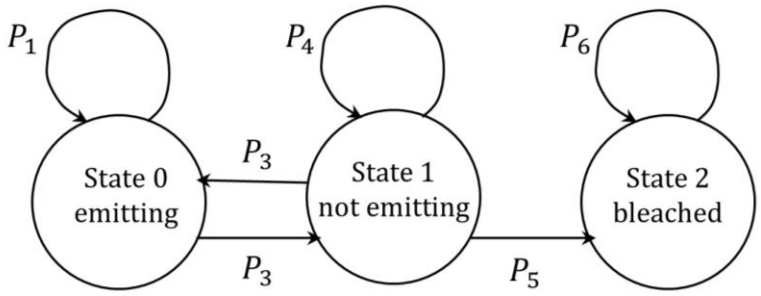


Figure 1.9 State transition diagram for the fluorophores in 3B Method

The next step is to compute the different probability distributions, with the goal of finding the maximum a posteriori (MAP) location estimates. The marginal integrals are calculated using a hybrid of MCMC and forward Hidden Markov Model (HMM) algorithm. To build the final image from the MAP estimates of the locations , the positions are mapped to the pixel in the higher resolution pixel grid and the intensity values are calculated from the accumulated fluorophores weighted with the MAP intensity values.

In principle, the 3B methodology is similar to just point estimates for sparse emitter locations. However, it's actual effectiveness can be seen for high-density case where simple point estimates are not sufficient to provide enough information, since there will be a lot of ambiguous estimates, in which case a

multi fluorophore fitting approach, as discussed earlier, is required. Since the 3B method averages over several models, it automatically accounts for the ambiguous cases as well. The different situations is shown in Figure 1.10 below (Lidke, 2012) .

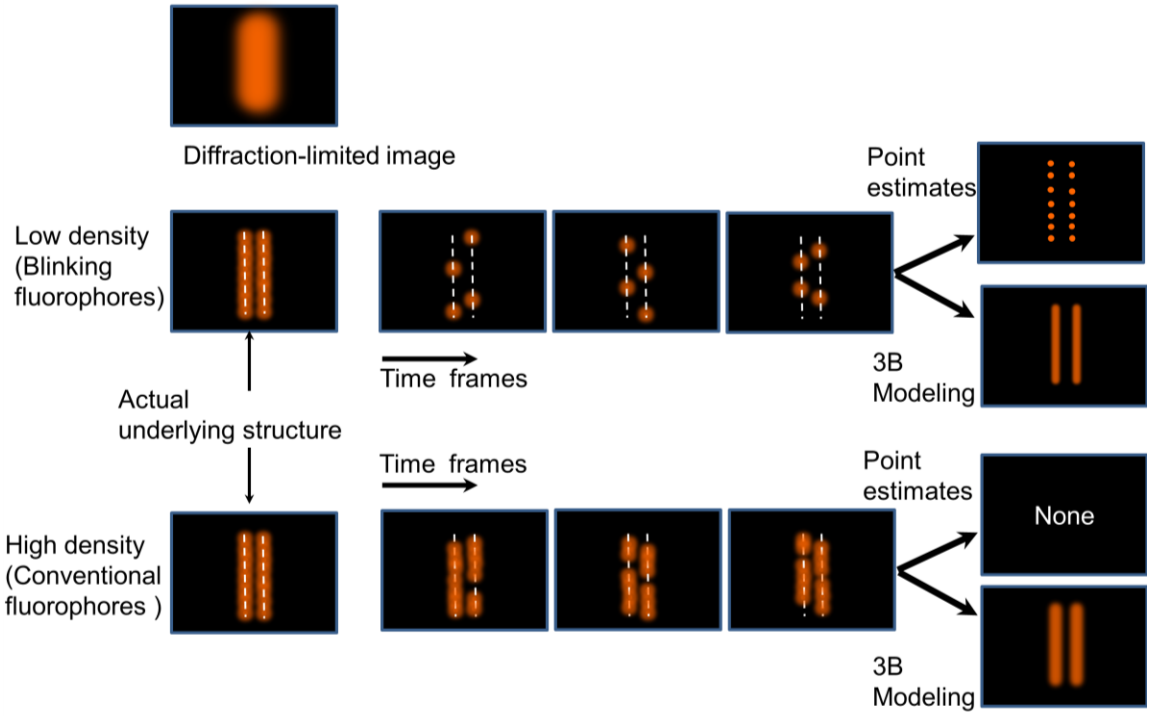


Figure 1.10 3B Approach for single molecule super-resolution imaging

The 3B method has can be used with most common fluorescence microscopy experiments. If we compare with SOFI, which also can deal with overlapping fluorophores, the 3B method has some advantages. For reconstructing the structures with similar details, SOFI requires more data than the 3B method, although the reconstruction is much faster for SOFI. In addition, the resolution enhancement with SOFI is limited whereas the 3B method can achieve resolution

up to 50 nm with relatively short acquisition time of few seconds. For 3B method, with more data the spatial resolution is expected to improve at the expense of temporal resolution and vice-versa. The computational time complexity varies linearly with the number of emitters times the number of pixels, for the 3B method. This actually is quite computationally intensive when there is a large model to analyze. In short, the 3B method is an exciting and powerful concept for analyzing and reconstructing dynamic images of biological structures using super-resolution microcopy.

1.3 Spatial and Temporal resolution information of super-resolution methods

The imaging methods used in fluorescence microscopy cover spatial resolution from 5 mm to ~ 10 nm (Fernandez-Suarez and Ting, 2008; Huang et al., 2010). Widefield and Total Internal Reflection Fluorescence (TIRF) microscopy generally covers the range ~ 5 mm to 400 nm and milliseconds temporal resolution. Confocal has a spatial range of 5 mm to 200 nm and a temporal resolution of milliseconds. GSD and SSIM has a spatial range of 5 mm to ~ 80 nm. STED has a spatial range of 5 mm to 10 nm and a temporal resolution of seconds. PALM and STORM has a spatial range from 5 mm to 10nm with a temporal resolution of seconds. NSOM has a spatial range from 5 mm to 10nm. EM and cryoEM has a spatial range from less than 100 μm to less than 1 nm. NMR has a spatial range of 100 μm to less than 1nm and temporal resolution of less than seconds.

1.4 Modeling of Single Molecule Super Resolution Images of Biological Structures

So far we have seen all the different methods which can either image biological structures at sub-diffraction resolution using improved optical setups, fluorescent probes or performing post processing of the single molecule images acquired to provide the super-resolution images or the single molecule information the biological problem. Some methods are better than the other in respect of providing improved spatial resolution and some are superior to the other in providing better temporal resolution for imaging dynamic biological processes. All these methods may be extremely good at providing the raw structural image, but what none of these methods are able to provide is the capability of understanding the images automatically with minimal human supervision or interpretation. We can take the super-resolution imaging concept one-step further with the novel idea of actually reconstructing the images using generative models for the biological structures.

The 3B method describe earlier is extremely powerful in that sense of providing the spatial and temporal dynamics of the biological structures and other information about the actual physical process of blinking and bleaching behavior of the fluorescent emitters. What we are aiming is that we would tackle this problem from the view point of structural model rather than the actual physical process that generates the fluorescence over the structure to make it absolutely independent of any physical process of the probes.

Various computational methods from the branch of statistical machine learning and computer vision (Berlemont et al., 2008; Li et al., 2009a; Schaub et al., 2007; Stoitsis et al., 2008; Taylor et al., 2011; Thomann et al., 2003) have been applied to study biological and biophysical structures and processes. These applications are

mostly on confocal or other conventional images of biological structures like actin filaments with parametric feature extraction techniques such as Hough Transform in (Berlemont et al., 2008; Li et al., 2009a; Schaub et al., 2007; Stoitsis et al., 2008; Taylor et al., 2011; Thomann et al., 2003) , Radon Transform and Beamlet Transform in (Berlemont et al., 2008; Li et al., 2009a; Schaub et al., 2007; Stoitsis et al., 2008; Taylor et al., 2011; Thomann et al., 2003). Some are applications to electron microscopy (EM) images such as in (Berlemont et al., 2008; Li et al., 2009a; Schaub et al., 2007; Stoitsis et al., 2008; Taylor et al., 2011; Thomann et al., 2003) . Again some of the methods are for studying dynamic trafficking and tracking of single molecules such as in (Berlemont et al., 2008; Li et al., 2009a; Schaub et al., 2007; Stoitsis et al., 2008; Taylor et al., 2011; Thomann et al., 2003). Methods such as active contour models from computer vision and method like particle filters with Monte Carlo Sampling methods from machine learning has been employed for studying actin filaments (Berlemont et al., 2008; Li et al., 2009a; Schaub et al., 2007; Stoitsis et al., 2008; Taylor et al., 2011; Thomann et al., 2003) from conventional microscopy images. Again, various generative models (Fudenberg and Paninski, 2009; Svoboda et al., 2009; Zhao and Murphy, 2007) have been used to model biological structures from conventional microscopy images. These models generally are parametric models based on several structural components, that is used to learn and build the generalized cellular and sub-cellular structures with several instances of the similar structure. The structure of localization-based SR imaging data is different from that of conventional microscopy and hence we need to develop methods from first principles or apply approaches already existing in the computational fields to address the needs of super-resolution and single molecule imaging field.

1.5 Thesis outline

The essence of this thesis is about inferring biological structure from single molecule images and improving the localization image data, which already has a sub-diffraction resolution. The idea is to be able to reconstruct the structures at much lower spatial sampling which will enabling us to improve the temporal resolution for dynamic imaging. Since our primary focus is on localization microscopy, one of the objectives is to determine methods of improving the localization capability. With that idea, we start with a method for 3D localization of single molecule, using multi-focal plane imaging (first introduced in (Ram et al., 2008; Ram et al., 2009a), in chapter 2. The motivation of course is to point towards the fact that given an appropriate optical setup we can still push the boundary for localization microscopy, which would help in ultimately reconstructing the structures or track single molecules more accurately. Since our ultimate goal is to determine the biological structures, we take the concept of localization microscopy to a different level with completely different perspective of studying single molecule super-resolution data using generative models, starting in chapter 3. We present a proof of principle using a parametric feature extraction technique called Hough Transform for showing that we can establish the underlying biological structures with primitive shapes from sparse single molecule data and thereby potentially improving on the temporal resolution. This chapter is mostly adapted from the publication (Maji and Bruchez, 2012). In chapter 4 we present another approach of modeling arbitrary biological structures using single molecule super-resolution data with a generative probabilistic graphical model framework. The manuscript for this chapter is in preparation. The motivation of this work is to start with a subset of data and improve the model gradually in a iterative manner using the biologically

relevant model constraints for the structure of interest. In chapter 5 we discuss some clustering and manifold learning algorithms for 2d point data sets which could be useful for recovering some straightforward super-resolution structures and provide additional information. In chapter 6 we discuss a method on Monte Carlo based data association algorithm, which is usually used for tracking multiple objects. We wanted to test if a single framework can both perform single particle tracking and recover structures from static data. These methods pave the way for future development of sophisticated modeling of biological structures from super-resolution microscopy images.

CHAPTER 2

DUAL PLANE THREE-DIMENSIONAL LOCALIZATION OF SINGLE MOLECULES

Conventional microscopes use serially stepped single focal plane imaging to study biological process and that limits the capability to look at fast processes in 3-dimensions. Localization of single object with single plane intensity information often leads to a poor estimate of axial positions especially near the focus where the change in point spread function (PSF) with Z displacement is minimal. One method to make this change of PSF with Z positions more pronounced for better estimation of the axial position is by distortion of PSF or astigmatism as in (Holtzer et al., 2007; Huang et al., 2008), but this technique is limited in its spatial range to about 1 μm or less and still shows minimal distortion near the true focal plane. In order to do single particle localization and tracking with better spatial and temporal precision; we can use a dual plane imaging setup (Ram et al., 2008). Dual plane information (in-focus and out-of-focus) would allow us to estimate the 3D location more accurately than classical approaches, because the in-focus data would provide us with Z-information for object positions away from the focus better and out-of-focus data would provide better Z-information for object positions near focus. Together they would provide improved Z-displacement sensitivity for all the axial positions. The axial range for the dual plane approach is more than 2 μm, which is a significant advantage over astigmatism approach as far as biological processes are

concerned. The objective is also to track multiple objects for analyzing the interactions of different biological molecules.

2.1 3D Image model

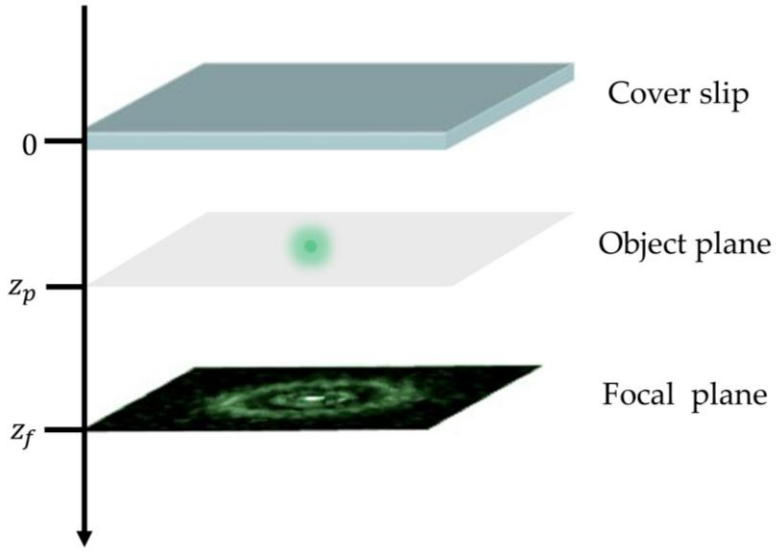


Figure 2.1. Object representation

Here $\xi = (x, y, z_f)$ is a point on the image plane, $\xi_p = (x_p, y_p, z)$ is the actual object coordinate and z_f is the focal distance.

For three-dimensional localization, the object intensity profile could be fitted to a three-dimensional Gaussian PSF given by:

$$I(x,y,z) = I_B + I_0 \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2} - \frac{(z-\mu_z)^2}{2\sigma_z^2}\right] \quad (2.1)$$

Here (μ_x, μ_y, μ_z) are the coordinates of the centroid and $(\sigma_x, \sigma_y, \sigma_z)$ are the standard deviation of the point spread function in x, y and z direction respectively, I_0 is the peak intensity and I_B is the offset for the intensity profile.

However, a more correct way to localize single molecules is to use a 3D Airy function as the point spread function (Aguet et al., 2005) based on (Born and Wolf, 1999) and described in (Gibson and Lanni, 1989; Gibson and Lanni, 1991) is given by:

$$PSF(\xi, \xi_p|\beta) = \left| A \int_0^1 J_0\left(K \frac{NA}{z_f} \rho r\right) \exp[-iW(\rho, \xi, \xi_p|\beta)] \rho d\rho \right|^2 \tag{2.2}$$

where A is a constant complex amplitude, J_0 is the zeroth order Bessel function of the first kind. k is the wavenumber, $r = \sqrt{(x-\mu_x)^2 + (y-\mu_y)^2}$ is the radial distance from the centroid of the object and NA is the numerical aperture of the microscope, $W(\rho, \xi, \xi_p|\beta)$ is the phase aberration term defined as $W(\rho, \xi, \xi_p|\beta) = k.OPD$, β is the set of optical parameters and OPD is the optical path difference between the object plane and the detector plane and is given by:

$$OPD = \Delta z \frac{\pi NA^2}{\lambda n_i} \rho^2 \tag{2.3}$$

Below is a model representation of the 3D point spread function

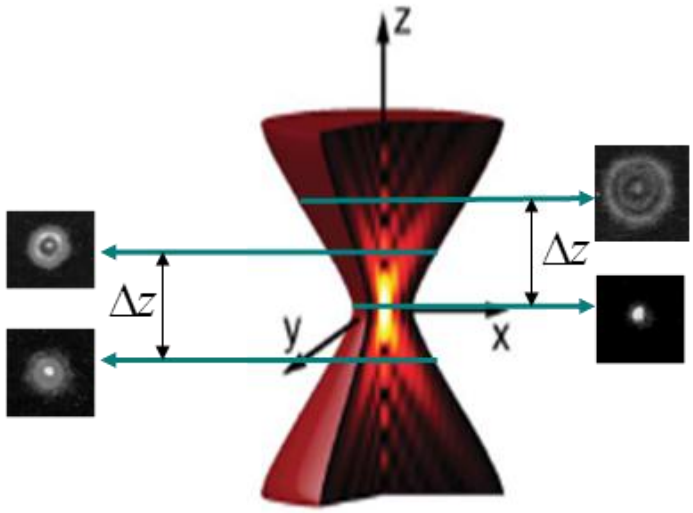


Figure 2.2 3D Point spread Function with varying airy disk patterns at different axial positions

A single focal plane is symmetric to the positive and negative Z displacement whereas dual plane is inherently asymmetric for the displacement and may contain more information about the 3d position of a molecule.

2.2 Global fitting using Dual-plane information

The z-position estimates for fluorescent objects close to focus is usually very difficult, so a good way to resolve this problem is to use information from more than one plane. Localization using multifocal plane data has been shown in (Ram et al., 2008; Ram et al., 2009a). Here we use a similar approach to address the problem to show the working principle of dual plane method. Suppose the

intensity distribution of an object in any two planes separated by a distance Δz (as shown in Figure 2.2) is given by:

$$I_1(\xi, \xi_p|\beta) = B_{1,xy} + PSF(\xi, \xi_p|\beta)$$

$$I_2(\xi, \xi_p|\beta) = B_{2,xy} + PSF(\xi + (0,0,\Delta z), \xi_p|\beta)$$

(2.4)

where $B_{1,xy}$ and $B_{2,xy}$ are the background noise for plane 1 and plane 2.

Then the actual object intensity profiles in two planes can be fitted simultaneously (global fitting) to their corresponding theoretical forms given in Equation 2.4 Global fitting can be achieved by minimizing the objective function for least square error:

$$\left|I_1(\xi, \xi_p|\beta) - F_1\right|^2 + \left|I_2(\xi, \xi_p|\beta) - F_2\right|^2 \quad \text{or}$$

$$Z_p^* = \arg\min_{z_p} \sum_{k=1}^{2} \left\|I_k(\xi, \xi_p|\beta) - F_k\right\|^2$$

(2.5)

where F_1 and F_2 are fluorescence image intensity profile for the object in plane 1 and plane 2 respectively. The fitting results from the two focal planes should provide a better estimate of the actual centroid due to higher information content than if we had just one plane.

2.3 Simulation results

The simulated dual plane data is shown in Figure 2.3. The data is generated with the parameter values described in Table 2.1. The goal is to estimate the z position

,using the information from both the planes, of all the images given its true position by which the data was generated. The global fitting results for the two cases are shown in Figure 2.4.

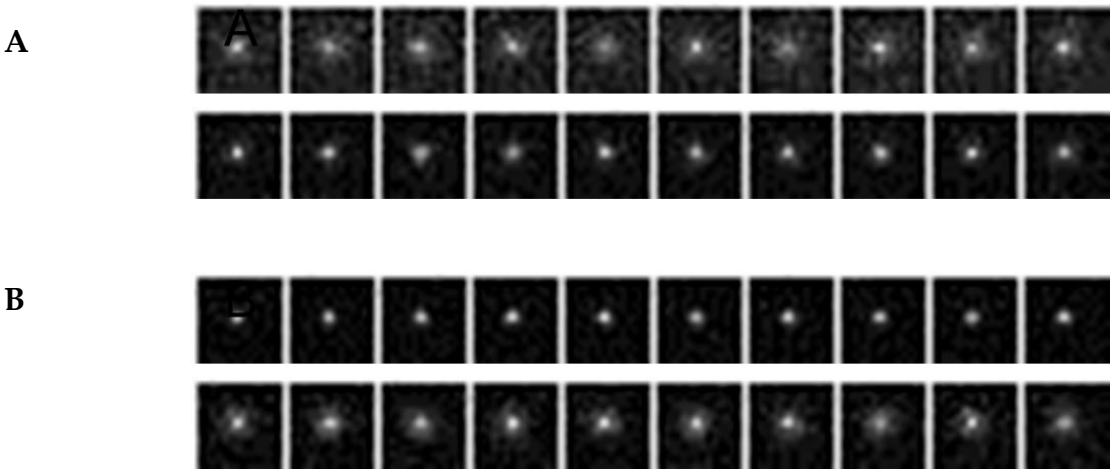
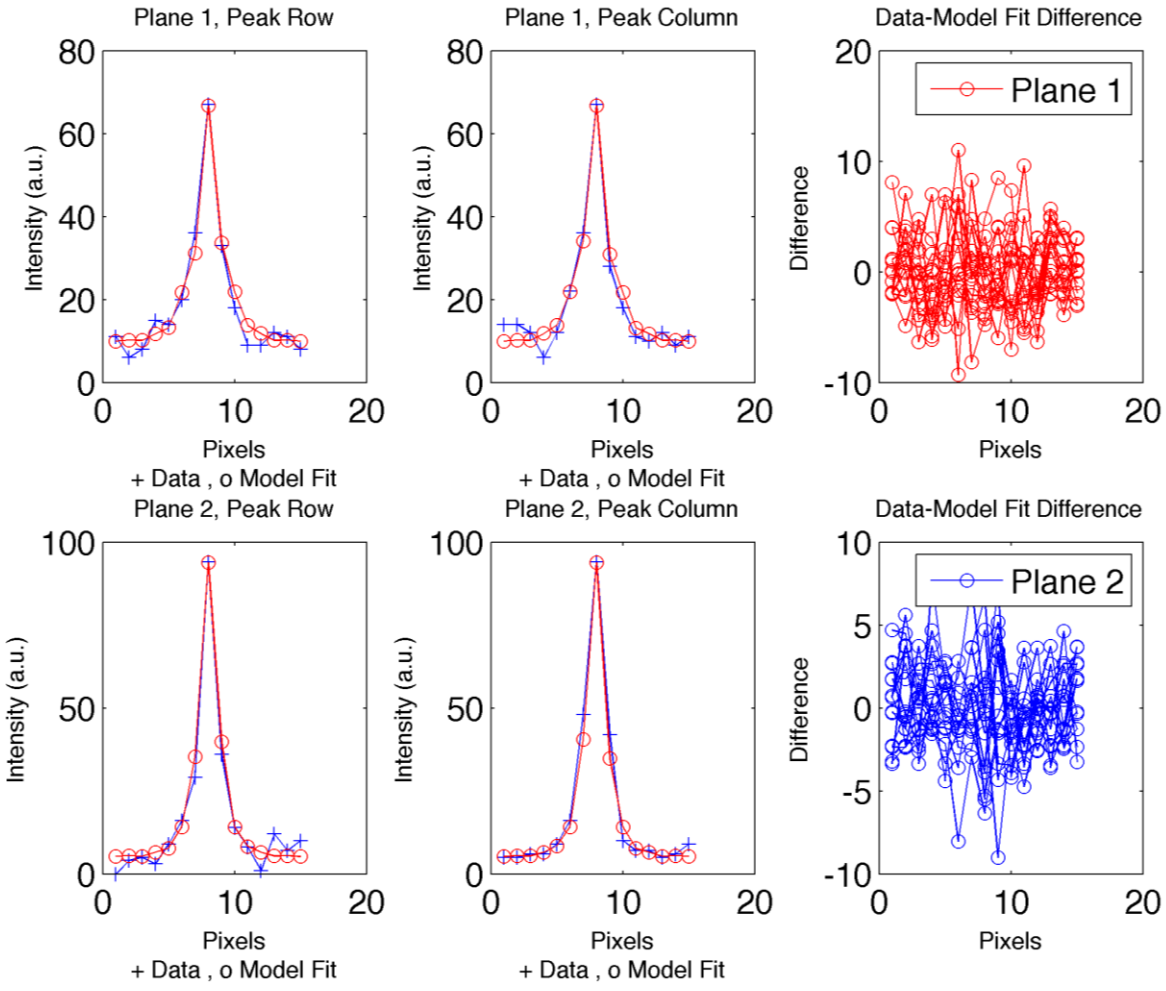


Figure 2.3. Simulated single molecule data in a dual plane setup. (A) True Z-position of the object centroid is 400nm**.** Top row is focal plane 1 (out-of-focus) and bottom row is focal plane 2 (in-focus). (B) True Z-position of the object centroid is 100nm**.** Top row is focal plane 1 (in-focus) and bottom row is focal plane 2 (out-of-focus).

The simulation details are provided in the Table 2.1. In the first two columns of Figure 2.4A and B, the blue curves are the simulated data and the red curves are from the dual plane fitting model given by equation 2.4 and 2.5. The third column shows the difference between the data and model fits. The fitting estimate of the Z positions are quite close to the theoretical position values for all

the positions. Figure 2.5 shows the variability of the estimates for the same Z positions for different instances of the z position data. The simulation shows very good improvement of the dual plane method over single plane localization, where the accuracy is generally around 40-50 nm (Ram et al., 2008) for a photon count of about 1000. The localization is even worse in single plane when objects are close to the focus. There are however limitations to the dula-plane method, mostly in the physical setup.

A



B

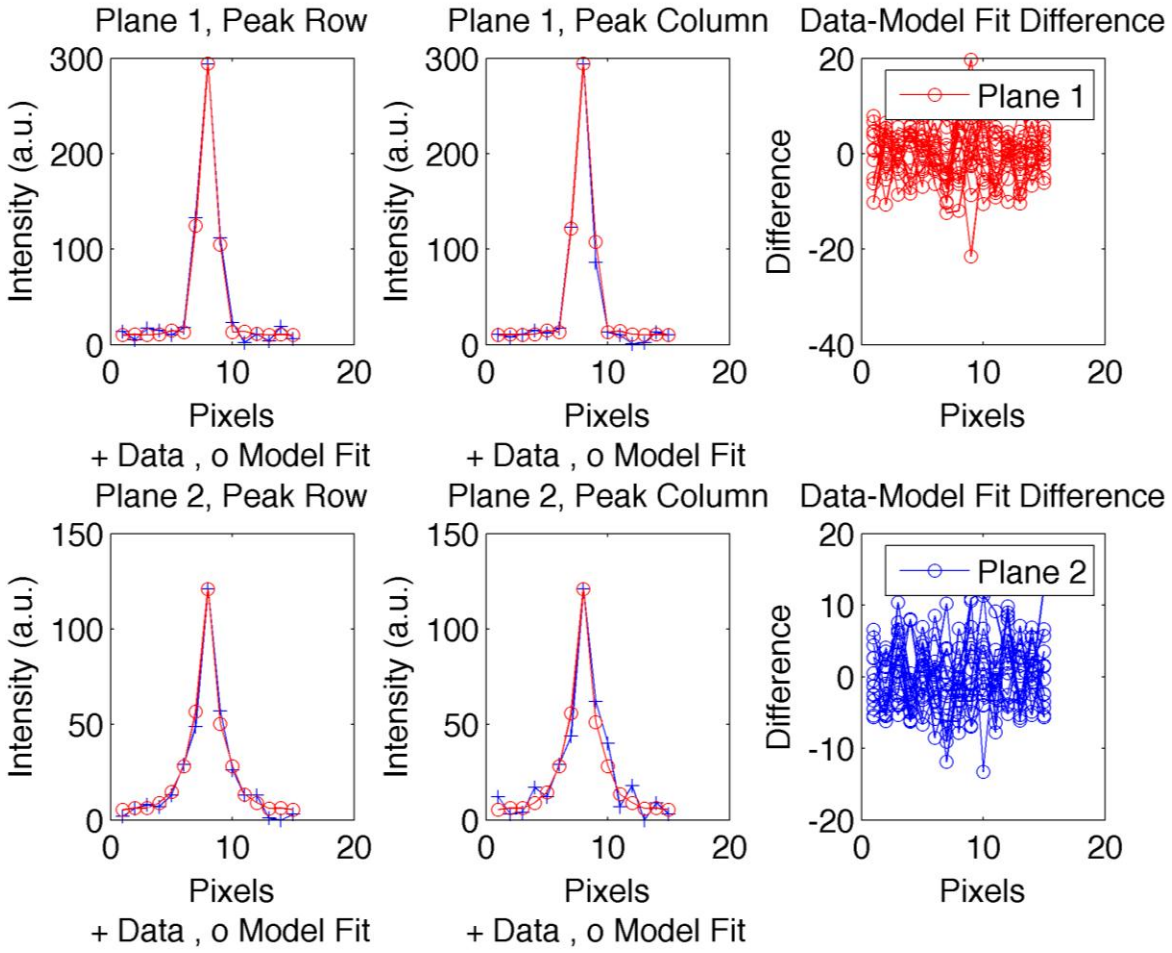
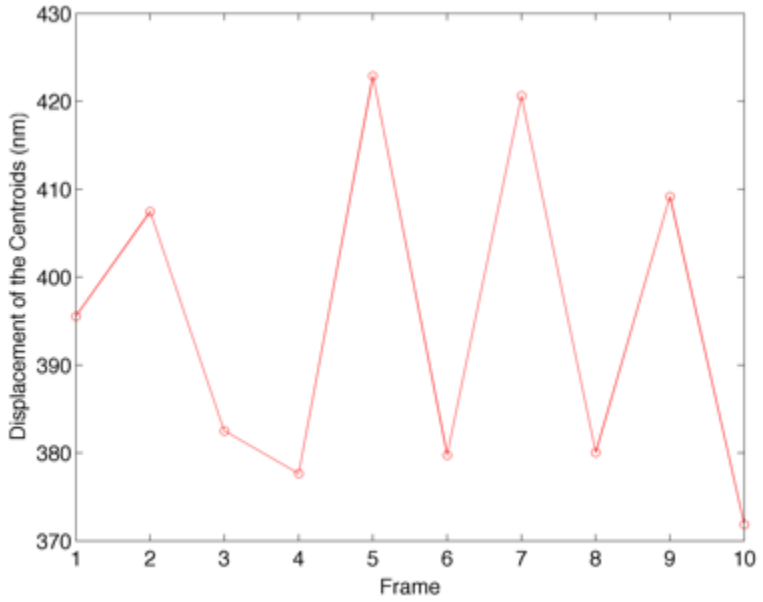


Figure 2.4 Global fitting of dual plane single molecule image. (A) true Z = 400 nm. (B) true Z = 100nm. The blue cross is the data and the red circle is the dual plane Airy model fit. The first column is Intensity vs x plot and the second column is Intensity vs y plot.

Model Parameters	True	Estimated (1 fit)	True	Estimated (1 fit)
X_0 *(μm)*	0.975	0.9796	0.975	0.9761
Y_0 *(μm)*	0.975	0.9691	0.975	0.971
Z_0 *(μm)*	0.400	0.3955	0.100	0.11386
α $(1/\mu m)$	17.52	15.33	17.52	14.8
Photon detection rate (Photons/second)	1000	633	1000	1188.76
Photon detection rate ratio between focal planes	1(equally)	-	1(equally)	-
Plane separation *(μm)*	0.7	-	0.5	-
Numerical Aperture	1.45	-	1.45	-
Wavelength *(μm)*	0.52	-	0.52	-
Refractive index:	1.515	-	1.515	-
Background for focal plane 1 (Photons/pixel/second)	10	9.83	10	9.58
Background for focal plane 2 (Photons/pixel/second)	5	5.21	5	5.37
Exposure time (second)	1	-	1	-
Magnification	100x	-	100x	-
Pixel size *(μm^2)*	13x13	-	13x13	-
Gaussian noise mean (e-)	0	-	0	-
Gaussian noise variance (e-)	16	-	16	-

Table 2.1. The details of model parameters and dual plane fitting for two simulated data samples shown in the Fig. 2.3

A



B

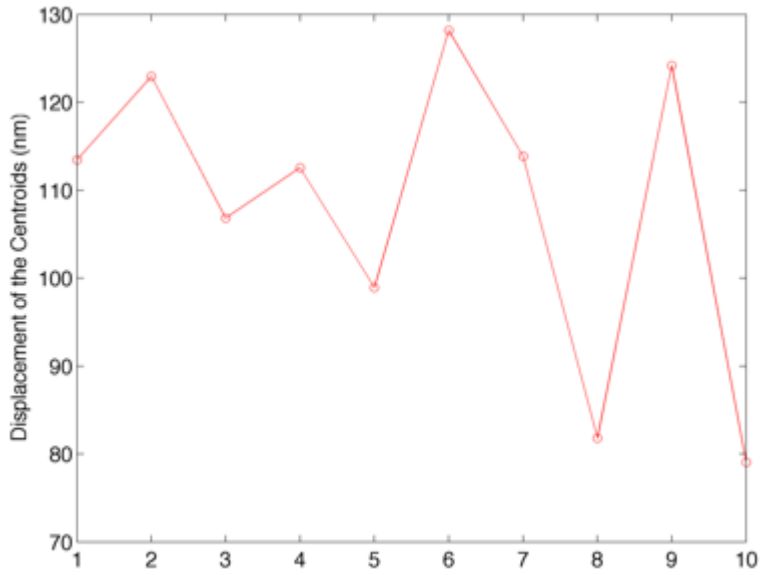


Figure 2.5. Localization for 10 frames shown in Fig. 3 . A) The Mean Z-localization accuracy when true Z_0 = 400 nm is 17.27 and the standard deviation of the 10 Z-position estimates is 18.96. B) The Mean Z-localization accuracy when true Z_0 = 100 nm is 16.22 and the standard deviation of the 10 Z-position estimates is 16.94.

2.4 Discussion

For simulated dual plane single molecule data, the localization accuracy is less than 20 nm, which is comparable to astigmatism based axial localization (but dual plane has other advantages) and is a significant improvement over single plane localization accuracy that usually around 40 - 50 nm. Although, it would be more meaningful to test the performance on real data, we have demonstrated here the effectiveness of multi-focal plane imaging for improved localization of single molecules in 3d with simulated data. The dual plane localization approach may be further improved by using maximum likelihood estimation method by incorporating the dual plane information in the model. Although here we have discussed one approach of improving the 3d localization accuracy of single molecule, there are other approaches involving modified PSFs (Pavani et al., 2009; Thompson et al., 2011) or the optical setups including astigmatism as already mentioned before or develop newer methods. The idea here is to either improve the localization as much as possible, so that ultimately it can help more accurate SPT studies by reducing the ambiguities in the trajectories or achieve better resolution structures from localization microscopy data. The field of super-resolution microscopy is rapidly progressing with newer microscopy techniques and therefore it is an opportunity to introduce novel computational approaches for single molecule data analysis for biophysical studies. We have introduced one such concept in Chapter 3.

CHAPTER 3

GENERATIVE MODELS FOR SUPER-RESOLUTION SINGLE MOLECULE IMAGING

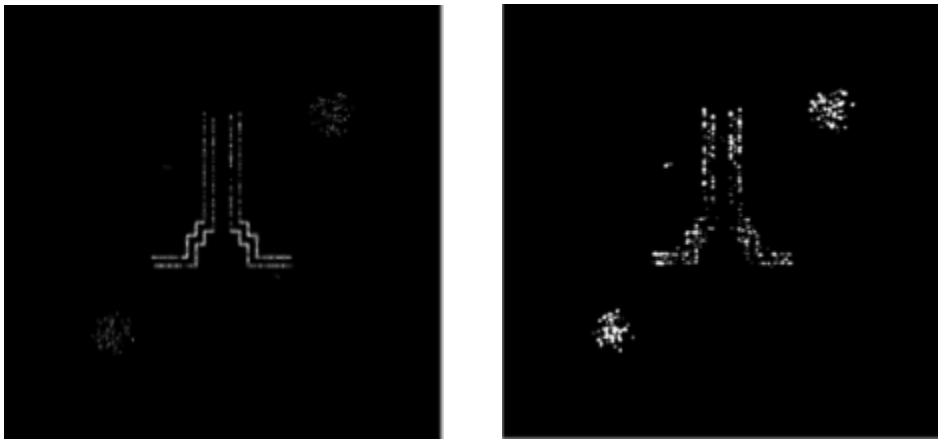
Super-resolution (SR) imaging has recently led to a number of important insights in biology that could not have been achieved with conventional microscopy due to optical resolution limitations (Abbe, 1873; Hell and Wichmann, 1994). A variety of approaches now achieve resolution far beyond the diffraction limit. Localization based approaches such as STORM (Rust et al., 2006), PALM (Betzig et al., 2006), FPALM (Hess et al., 2006) and related methods have been employed effectively for static and slowly-moving structures. These approaches require sequential acquisition of positions of individually resolved fluorescent molecules, which are then assembled into a high-resolution image. The resolution in these images is related to the localization accuracy and the sampling density, with high-resolution images requiring comprehensive sampling of the molecular positions. Because of these requirements, localization microscopies still struggle to provide high spatial and temporal resolution images, primarily due to the time-scale mismatch between acquisition and biological motion. Recent demonstrations using very high laser power improved the frame-capture timescale by an order-of-magnitude by accelerating the localization and deactivation cycle time (Jones et al., 2011). While this approach achieved 0.5–2 second acquisition speeds, this still poses a challenging limit for many biological processes.

Recently, computational methods from the branch of statistical machine learning and computer vision (Berlemont et al., 2008; Li et al., 2009a; Schaub et al., 2007; Stoitsis et al., 2008; Taylor et al., 2011; Thomann et al., 2003) have been applied to biological structures and biophysical processes. Various generative models (Fudenberg and Paninski, 2009; Svoboda et al., 2009; Zhao and Murphy, 2007) have been used to facilitate analysis of conventional microscopy images. The structure of localization-based SR imaging data is different than that of conventional microscopy. The catalog of molecular positions provided by this approach provides information about the underlying structures at molecular length scales. Such data requires computational approaches that utilize the inherent positional information to extract meaningful structural biology–scale information about those cellular structures. Because localization microscopy relies on sequential acquisition of molecular positions, a shorter acquisition window results in identification of fewer molecular positions from the underlying structure. Dynamic localization datasets are inherently incomplete; yet represent a statistical sampling of the complete underlying structure. Our aim is to provide an information bridge between super-resolution microscopy and structural biology by using generative models to build a top-down molecular scale picture of cellular structures.

We hypothesized that generative models can accurately identify underlying biological structures at high resolution using significantly less data than the full localization based reconstruction at Nyquist sampling. Such models can be used to extract useful biological information such as characteristic lengths and inclination angles of filamentous structures, organelle size and shape and other representative characteristics of the underlying structures.

3.1. Why do we want to use Generative models in Super-resolution Imaging?

Let us consider the toy example in Figure 3.1A. It is an image based on the exact location of the actual objects where we can clearly see the structural details. Now through localization microscopy we would most likely not be able to see this image, rather we would see a much blurrier image in the context of molecular proportions. Now Figure 3.1B is the single molecule localization based reconstruction of the same image. Although, we are able to see some details, others are obviously lost such as the "step" like portion of the structure.



A **B**

Figure 3.1. (A) actual location based image. (B) Localization (MLE) based reconstruction

However if we have a generative model for the same example we can reconstruct the original image without much of a problem and with full structural details. This is a very simple toy example, so it is easy for us to think of a generative model that could generate the straight lines or parametric curves easily, but as

real world biological examples are not that simple, we have to devise a modeling framework, which can accommodate such complexities.

3.2. Structural Examples

Many Biological structures can be roughly characterized as simple geometric shapes like lines, circles, ellipse and so on.

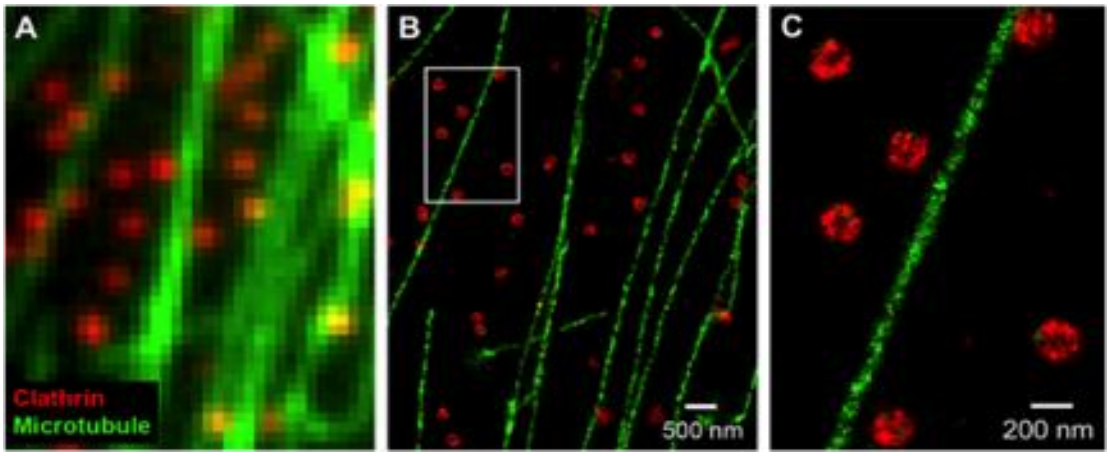


Figure 3.2. Conventional fluorescence image (A) and STORM reconstruction (B,C) of annular clathrin-coated pits and linear microtubules in a cell (adapted from (Bates et al., 2007))

We can clearly see that localization based methods enhances the structural details and resolution in comparison to conventional methods. We want take this one step further by using generative methods, which according to our hypothesis can improve the resolution and provide with more useful information about the biological structure.

3.3. Inferring Biological Structures from Super-Resolution Single Molecule Images using Generative Models: Simple Parametric Shapes

Here we apply a parametric feature extraction method known as the Hough Transform(Duda and Hart, 1972) to identify basic structures using sparse single molecule (SM) data in 2-d. As a proof of concept, we have chosen HT as our test method. We would like to emphasize here that we chose HT as our test method only due to its simplicity and popularity in robust detection of parametric curves. This approach is robust to noise sources common in localization datasets. In addition, it is robust to occlusion and the presence of features unrelated to the parameterized features of interest. As implemented here, the Hough Transform efficiently infers underlying structures in spite of substantially reduced molecular sampling density and recovers quantitatively useful information about the sample set based on the parametric definitions of the objects. This computational framework lays the groundwork for extension to more generalized parametric objects in 2-d and 3-d.

The Hough Transform (HT) and its close relative Radon Transform has been previously used to study biological features from images (Maly and Borisy, 2001; Stoitsis et al., 2008; Verkhovsky et al., 2003; Zhou and Zheng, 2008). We extend the method to the analysis of localization based super resolution image datasets. Although we evaluate only the parametric case, the generalized Hough Transform (GHT) and variants can be extended to non-parametric cases. In case of the standard HT applied here, the parameter space for lines is 2-d and for circles is 3-d, both remaining computationally tractable for typical SR datasets(Duda and Hart, 1972). In contrast, GHT variants usually involve a 4-d parameter space with position, orientation and scale (Ballard, 1981), and are

substantially more computationally expensive. An efficient extension of GHT called displacement vector GHT (DV-GHT) is proposed in (Kassim et al., 1999). Some other improved and faster variants have been proposed for 2-d (Fung et al., 1996; Galamhos et al., 1999; Illingworth and Kittler, 1987; Kimura and Watanabe, 2000; Kimura and Watanabe, 2004; Olson, 1998; Suetake et al., 2006; Xu et al., 1990) and 3-d (Khoshelham, 2007). HT and GHT are inherently parallelizable, so large-scale computation can be managed by performing hardware-based parallel processing using the latest GPUs (Gómez-Luna et al., 2011) or field programmable gating arrays (FPGA)(Geninatti et al., 2009), potentially making some of these generalized methods computationally approachable.

3.4 Hough Transform

The Hough Transform (HT) (Duda and Hart, 1972) is a standard computer vision tool for recognition of global patterns in an image space by recognition of local patterns such as points or peaks in a transformed parameter space. The basic idea of HT method is to identify parametrizable curves such as lines, polynomials, circles, ellipsoids, and others using a voting procedure on the parameter space based on features in the image. It is worth noting that HT is generally applicable to images with Gaussian noise and may not provide optimal solution for images with Poisson noise from the perspective of detection theory. However, here we have set of localized points from images with Poisson noise and the points are transformed into image pixel features. Each input feature contributes to a global consensus shape that most likely generated the image point. Localization datasets produce discrete features, namely the set of found molecular positions.

Since each point is treated independently, outlier noise pixels will add small peaks and occluded points will just alter the peak intensities in the parameter space without changing the actual structure. In addition, points from other shapes will not significantly contribute to the peaks for the consensus shape in the transformed parameter space. These traits make the HT robust to noise, partial occlusion and the presence of other shapes, common problems encountered with localization microscopy. HT does not require any prior information about the number of solution classes and can find multiple instances of the shape at once. We have applied the classical HT to extract linear and circular structures from SR biological datasets. HT implicitly generates the observable structural data from a probability density function through a Bayesian process (Toronto et al., 2007). Hence, HT is an implicit generative model of parameterized shapes.

3.5 Hough Transform as a Generative Model for Biological Structures Using Single Molecule Data

In classical machine learning a generative model is defined as a model that can randomly generate observable data with a parameter set defined by a full joint probability distribution with priors. The working principle of the Hough Transform (HT) is essentially a voting process. Investigated from a Bayesian perspective, if the votes follow a probability distribution, the joint probability distribution of all the input feature points is, in effect, the voting process. The mathematical proof has been shown elsewhere (Toronto et al., 2007) for conventional images and edge points found through edge detection. In the current application, the features are localized single molecules from labeled

biological structures that can be represented as parametric objects. The proof can be straight forwardly extended to this situation.

Parameterization of a structure is based on a function that defines the structure in terms of a set of variables. The parametric normal form of a line is:

$\rho = \mathrm{x}\cos\theta + \mathrm{y}\sin\theta$, with a 2-d parameter space $\ominus_l = (\rho, \theta)$

(3.1)

The parameter ρ is the distance of the line from the origin of the coordinate space and θ is the angle formed by the line relative to the horizontal axis.

The parametric equation for a circle is:

$(x-a)^2 + (y-b)^2 = r^2$ with a 3-d Hough parameter space $\ominus_c = (a, b, r)$

(3.2)

The parameter (a , b) is the center of the circle and r is the radius of the circle.

The working principle for a classical HT is explained below.

Fig. 3.3A represents the parametric normal form line $\rho = \mathrm{x}\cos\theta + \mathrm{y}\sin\theta$, drawn in solid blue color, passing through a point (50, 50), with $\theta = 45^o$ and $\rho = 69.3$. Here the origin is (1,1). Fig. 6B shows a sinusoidal curve in the Hough parameter space (ρ, θ), corresponding to the point (50,50) in the real space. When we have three points (Fig. 3.3C), the Hough parameter space has three sinusoidal curves (Fig. 3.3D) corresponding to the three points in real space and they have an intersection point corresponding to a particular pair of (ρ, θ) values indicating that the three points are collinear in the real space (Fig. 3.3C). The individual curves are accumulated in a matrix (the Hough matrix), and consensus lines are identifiable as peaks within this accumulation matrix (in this case, a single point

with a value of 3). When there are multiple lines in the image space, there will be several intersections of the sinusoidal lines (peaks) for the group of points falling on the corresponding lines in the image space. Line end points are determined based on votes and a pre-defined maximum gap allowed between two points. If the distance between points exceeds a threshold, the line is terminated at the previous point generating an end point.

The detection of circles works on the same voting principle as that of lines, only the Hough parameter space is 3-d (a, b, r).For each input point on the original circle (Fig. 3.4B) there will be a range of circles (depending on the discretization of the parameter space) in the Hough accumulator space with the input point as the center. The intersection of those circles will define the center of the circle in the original image space. For the above example with 5 and 20 points, the intersection of the circles in Hough space (Fig. 3.4C) is around (100,100) as the original circle (Fig. 3.4B). This example also shows how more input points, produces more votes for a particular circle increasing the probability of locating the center of the circles. The Hough space for multiple objects is shown in below in Figure 3.5**.** The accumulator slices are of the same size as the image space and the stack length is the total length of the radius range that has to be searched. So the accumulator array has a dimension of Image Width x Image Height x Length of Radius discretization. For objects with a known radius, the search space is 2-d and calculations are much faster.

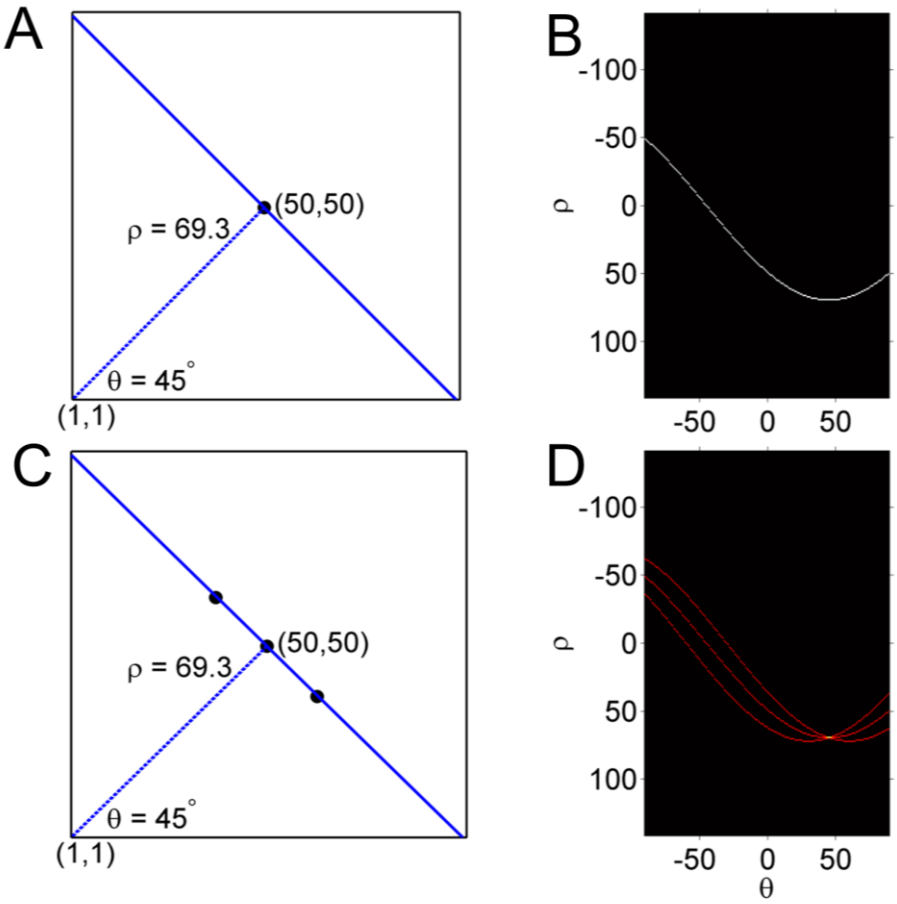


Figure 3.3. Illustration of working principle of the Hough Transform for lines

(A) Parametric normal form line passing through a point (50, 50). (B) Hough matrix parameter space with sinusoidal line corresponding to (50, 50). (C) 2 additional points added to (A). (D) Sinusoidal curves intersect for the three collinear points, one peak in the Hough space corresponds to one line.

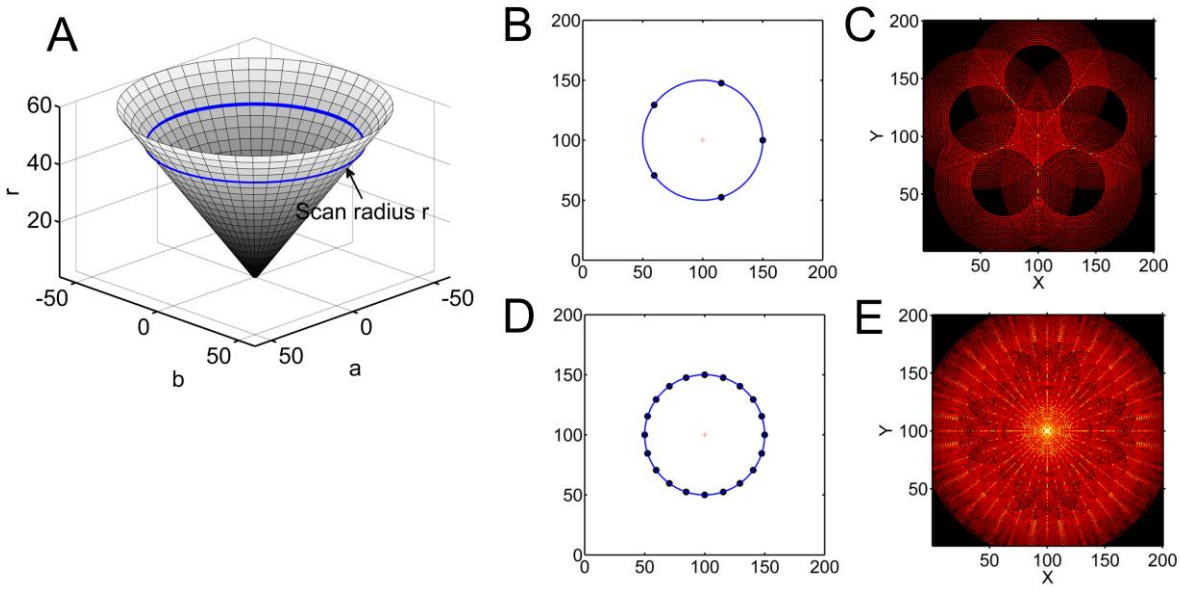


Figure 3.4. Illustration of working principle of the Hough Transform for Circles. (A) Hough accumulator space for a circle *(a,b,r)* when the radius r is unknown. The scanning circles in the parameter space are on the cone surface in the 3D space. (B) 5 points on a circle (100, 100, 50). (C) Circles in the Hough accumulator space corresponding to each of the input points in (B). (D) 20 points on a circle (100, 100, 50). (E) Circles in the Hough accumulator space corresponding to each of the input points in (D). The intersecting peak represents the center of the circle we are searching.

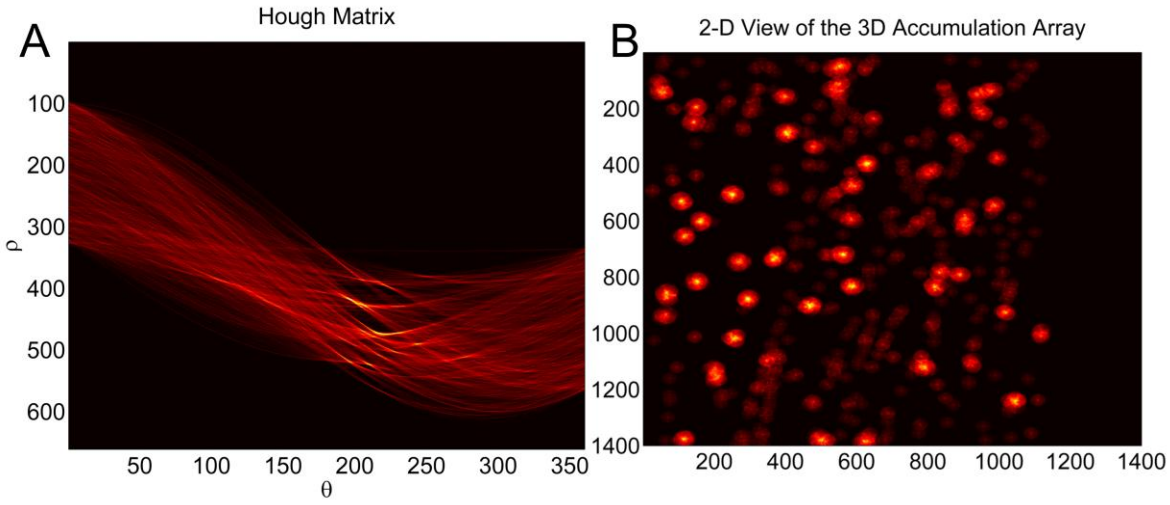


Figure 3.5. Example of Hough space for multiple lines and circles in the real data (Fig. 3.11). (A) Hough Matrix for the lines (microtubules) at 5% data density (B) Hough accumulator space for circles (CCPs) at 5% data density

Figure 3.5A shows a sinusoidal curve in the Hough parameter space (ρ, θ), corresponding to the each line passing through a point in the real space. When there are multiple lines in the real space, there will be several intersections of the sinusoidal lines for the group of points falling on the corresponding lines in the real space. Therefore using this method we can identify the line segments in the real space. For each input point on the original circle, there will be a range of circles (depending on the discretization of the parameter space) in the Hough accumulator space with the input point as the center. The intersection of those circles will define the center of the circle in the original image space.

3.6 Simulated data analysis

The detection of lines and circles using HT was performed for 100 random samplings of the data points on the structures at each data density. To remove spurious feature peaks in the Hough parameter space, which could be a result of outlier points, we have used a 2-d median filter for lines and a discrete filter with a Laplacian of Gaussian kernel in order to smooth the 3-d Hough accumulator matrix for circles. To quantify the reconstruction, the structural similarity score was calculated for each random sample using a Complex Wavelet Structural Similarity Measure (Sampat et al., 2009) (CW-SSIM) and the mean of those scores was calculated for each data density. These calculations were performed for different position noise and at different outlier noise densities as described above.

The details of the discretization of parameter space, feature range constraints for the real data and values for other parameters are provided in the section below.

3.7 Parameter Information for Hough Transform reconstruction of real data

For Hough line parameter space we set the orientation parameter $\theta = [-90^o, 89.5^o]$ with steps of 0.5^o and ρ was discretized with a resolution of 6 at the scale =25x from data space. A 2D median filtering was applied to Hough Matrix H with sliding window $= [length\,(row);\ length\,(column)]\,/75$. For finding the peaks in H, the maximum peak was set to 5000, the peak separation window = [15, 19] and the peak threshold was set to 0.2 – 0.3 of the maximum of H. For finding the actual lines from Hough peaks, the minimum line length was set to increase from 126 to 168 and Hough bin gap filling parameter value was set to decrease from 47.5 to 33.6 for 5% to 100% data density.

For circles, the radius range $=scale \times [10, 120]/pixel\ size$. The pixel size is 158 nm, scale = 25x and the resolution of radii= 0.4. The minimum spatial separation between peaks was 55 pixels and the minimum separation in radius between peaks was 73-75 pixels from the centroid. A Laplacian of Gaussian filter (LoG) and unsharp mask filter with parameter value of 0.2 was applied to the 3D accumulation Hough matrix CH .The discrete Laplacian of Gaussian filter that we have used is given by:

$$LoG = \begin{bmatrix} 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -3 & -3 & -3 & -1 & -1 & 0 \\ 0 & -1 & -3 & -3 & -1 & -3 & -3 & -1 & 0 \\ -1 & -3 & -3 & 6 & 13 & 6 & -3 & -3 & -1 \\ -1 & -3 & -1 & 13 & 24 & 13 & -1 & -3 & -1 \\ -1 & -3 & -3 & 6 & 13 & 6 & -3 & -3 & -1 \\ 0 & -1 & -3 & -3 & -1 & -3 & -3 & -1 & 0 \\ 0 & -1 & -1 & -3 & -3 & -3 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \end{bmatrix}$$

The threshold of peaks was chosen between 0.53 – 0.63 of the maximum of the CH for the data densities between 5% -100%. The maximum number of peaks was set to 200 for the region of interest.

3.8 Structural Similarity Index Measure (SSIM)

To measure the reconstructed structure with the original mask we have used an image similarity method called the Complex Wavelet Structural Similarity Index (CW-SSIM), which is an extension of spatial domain SSIM and inspired by the concept of the human visual system (HVS). It works on the principle that small geometric distortions produce a consistent phase shift in the local wavelet coefficients and that a consistent phase shift of the coefficients does not change

the structural information content of the images. The advantage of this method is that it does not require explicit correspondences between pixels being compared and is insensitive to small geometric distortions hence it is robust to structural distortions, this approach may overestimate the similarity index by a small factor even if there is slight dissimilarity between the images being compared.

CW-SSIM Index is defined by:

$$\tilde{S}(c_x, c_y) = \frac{2\left|\sum_{i=1}^{N} c_{x,i} c_{y,i}^{*}\right| + K}{\sum_{i=1}^{N}\left|c_{x,i}\right|^2 + \sum_{i=1}^{N}\left|c_{y,i}\right|^2 + K}$$

where $c_x = \{c_{x,i} \,|i = 1, \ldots, N\}$ and $c_y = \{c_{y,i} \,|i = 1, \ldots, N\}$ are two sets wavelet coefficients obtained at the same spatial location in the same wavelet sub bands of the two images being compared and K is a small positive scalar constant which improves the robustness of CW-SSIM measure at low local signal to noise ratio. We have used this index for a simpler case since we compared binary images. We have tried to use other measures such as Multi-scale Structural SIM (MS-SIM), Information content Weighted SIM (IW-SSIM), and the standard SSIM, but on visual inspection they seem to underestimate the similarity score by a significant factor and we found CW-SSIM to be more appropriate.

3.9 Results

3.9.1 Simulated data generation

The basic structural elements in biology are often simple geometric shapes such as lines, circles and ellipsoids (Blum, 1973). To mimic filamentous structures such as actin fibers or microtubules and circular shaped structures such as clathrin-coated pits or endosomes we have generated artificial data consisting of binary lines and circles in distinct channels (Figure 3.6**).**

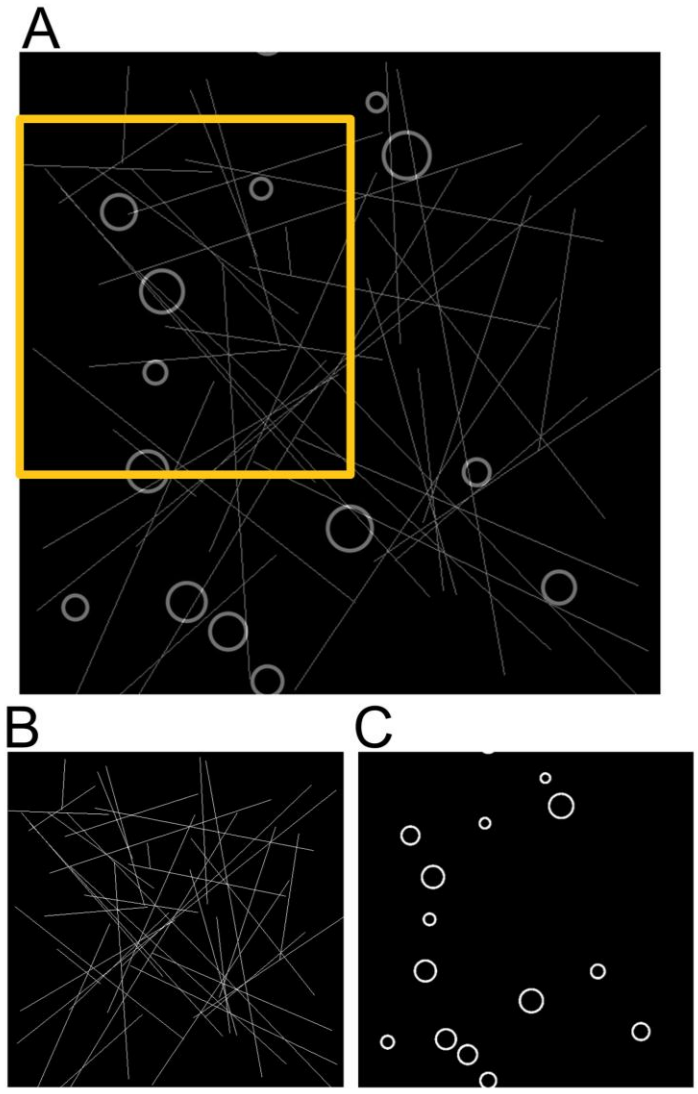


Figure 3.6. Structural mask for simulated data (A) Lines and Circles, cropped image in the yellow rectangle box is shown in Figure 3.8 (B) lines only (C) circles only

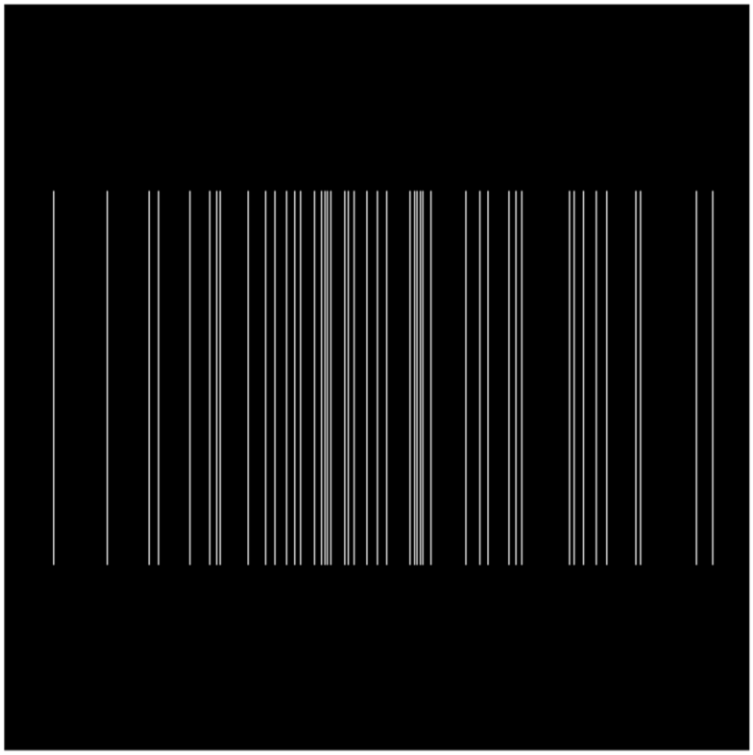


Figure 3.7. Parallel line mask

The density of lines in the example mask corresponds to real biological structures such as lamellipodial actin networks (Resch et al., 2002) if the mask area represents a 640 nm x 640 nm region of a cell (a 1 pixel = 1 nm^2 scale). Active pixel points from the mask structures are randomly selected to simulate stochastic activation of fluorescent molecules, analogous to PALM and STORM imaging. This reduces the selection bias of molecules from a certain region of the structures and retains the relative density of the molecules for all regions. For all simulated and real datasets, the found or simulated molecular positions were the input to the HT calculations. A number of papers have reviewed robust approaches for identifying molecular positions from localization datasets (Ram et al., 2009b; Smith et al., 2010).

3.9.2 Noise Sources

The two basic noise sources in localization-based SR imaging are position noise (localization accuracy) and outlier noise (background signal)(Bates et al., 2007; Betzig et al., 2006). The position noise represents the limitations inherent in finding the true position of a molecular emitter, while the outlier noise represents spurious localizations and nonspecific fluorophore binding sites typical of real datasets. Outlier noise was generated as 'Salt and Pepper' noise in MATLAB although any type of noise can be considered. The position noise of 0, 5 and 10 pixels represent the FWHM of the Gaussian spread of position relative to the true active-pixel location in the mask. Outlier noise densities tested were 0, 0.002, 0.005, 0.01, 0.02 and 0.05 expressed as the fraction off-mask pixels considered as a found molecular position. Outlier noise densities above 0.002 are extremely high for single molecule datasets and unrealistic, but were included to assess the robustness of the reconstruction method to high degrees of noise. Additional simulations were performed at other intermediate position noises. While only three cases are shown here all are available.

Simulations of the linear and circular masks at different outlier noise, position noise and sampling density demonstrated that the HT is able to reconstruct the linear and circular structures robustly and accurately at high outlier noise levels and position noise levels similar to those seen in real single molecule localization data (Betzig et al., 2006; Gordon et al., 2004; Rust et al., 2006). The reconstructed lines and circles are shown in Figure 3.8 and the reconstruction performance is quantified using a complex wavelet structural similarity index measure (CW-SSIM) is shown in Figure 3.9

3.9.3 Reconstruction from simulated data

Figure 3.8 shows a cropped section of the reconstructed lines (top row) and reconstructed circles (bottom row) overlaid on the point datasets for the mask shown in Fig. 3.6A at different outlier noise densities, a position noise of 5 pixels and a data density of 15% (fraction of total number of possible points that constitutes the structure). The full reconstruction for lines and circles at all the position noise and outlier densities has also been performed.

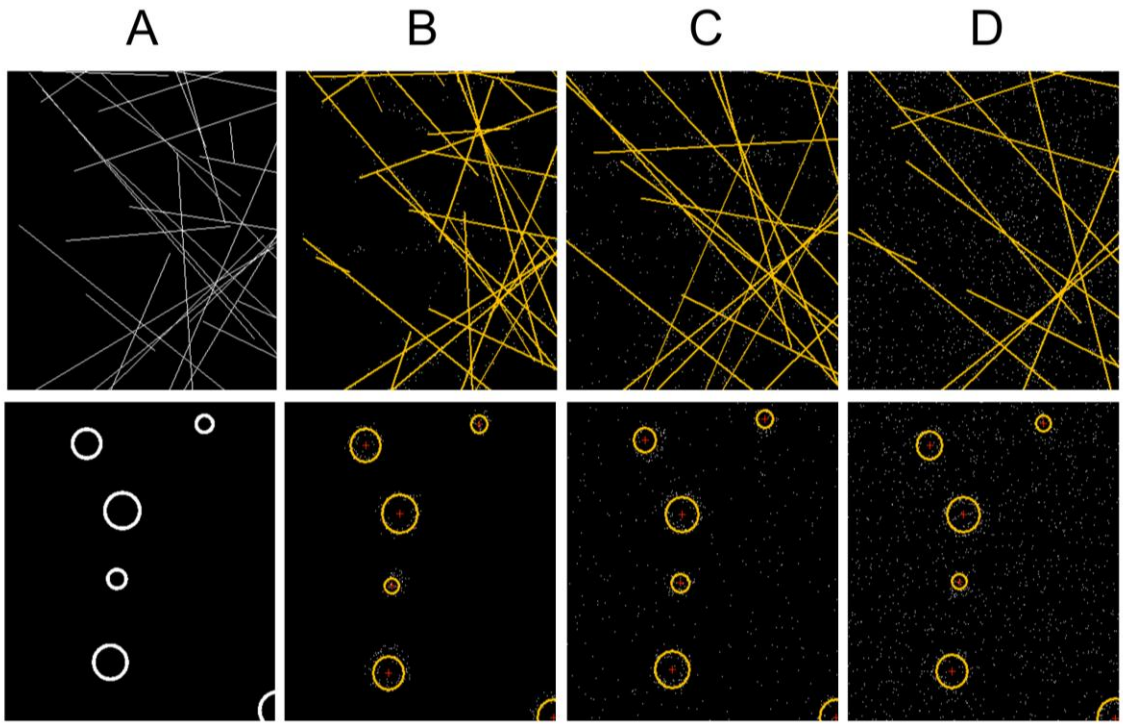


Figure 3.8. Representative linear and circular structure reconstruction. Column (A) Mask (B) outlier noise density 0 (C) outlier noise density 0.005 (D) outlier noise 0.02. Position noise is 5 pixels with data density of 15% for all cases here.

The plots shown in the top row of Fig. 3.9A, B, C for lines reveals that at lower position noise cases, the reconstruction measure is close for different outlier noise densities; although as expected, it is better at low outlier noise. In general, the reconstruction gets better with increased sample density but beyond a data density of 10–15% (low position noise) and 15–20% (high position noise), more data does not provide more information about the structure and the CW-SSIM measure reaches a plateau. This indicates that collection of SM-SR data has an optimum value for dynamic experiments. It should be noted here that the reconstruction measure does not reaches 100% even with no position and outlier noise. This is due to the complexity of the structural features, both density and configuration wise. With better estimators than HT, we should have better results. The plots shown in the bottom row Fig. 3.9A, B, C for circles reveal a similar trend at various position noise and outlier noise to that of line reconstruction. The reconstruction for circles is significantly better than the lines, an improvement expected due to the 3-d parametric space for circles.

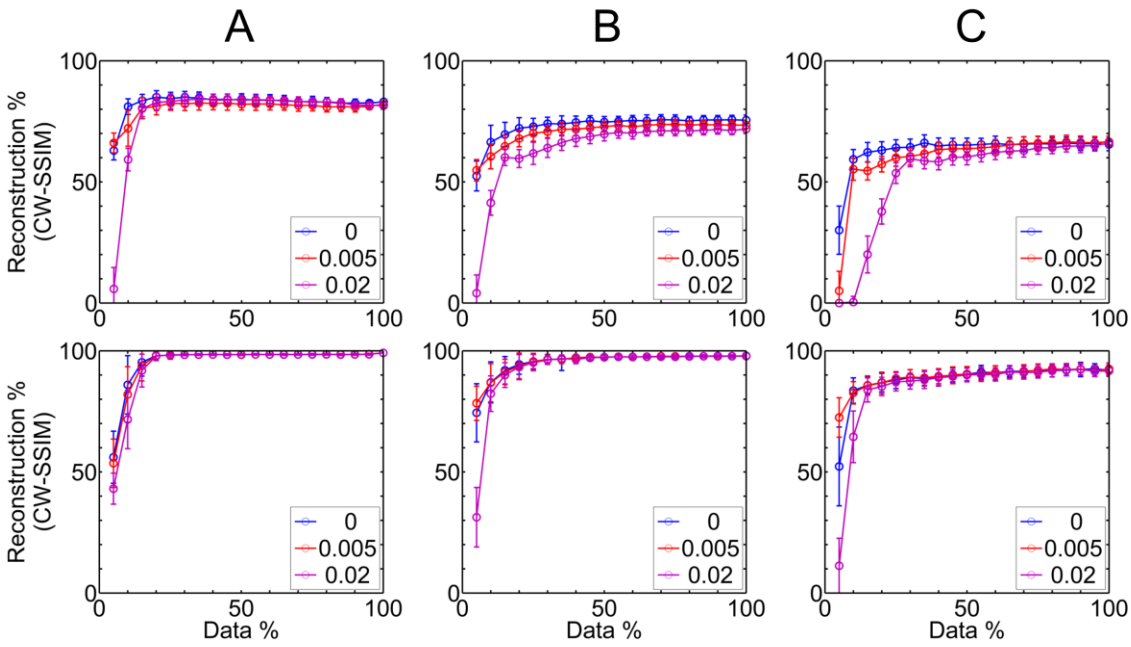


Figure 3.9. Reconstruction measure using Structural Similarity Index CW-SSIM. A total of 100 random simulations were performed at each data density and at outlier noise densities of 0, 0.005 and 0.02. Top row is for lines and bottom row is for circles. Column (A) Position noise of 0 pixel. (B) Position noise of 5 pixels. (C) Position noise of 10 pixels.

Reconstruction measure for all the noise densities are also performed and shown in the Supporting Figure S2 in (Maji and Bruchez, 2012)

The HT is more robust to outlier noise than to position noise in these simulations. This is likely a result of the Hough accumulator, which scores votes for objects that are coincident with a feature and does not account explicitly for localization uncertainties (objects that are near to a feature). Improvements to the algorithm could incorporate localization uncertainty directly.

Overall, Fig. 3.9 demonstrates that most of the structural information can be recovered with only a fraction of the single molecule data for analysis of lines and circles. In this example mask, for lines, about 15% data identifies 80–85% or more of the input structures while for circles around 10% of the data identifies more than 90–95% of the input structures. Sampling beyond these levels only modestly increased the information recovery. For lines and circles, inclusion of additional data density beyond these levels only resulted in modest additional feature identification (< 10%). With an improved HT we would likely improve the performance in recovering the dense linear structures, for example using Monte Carlo optimization over parameter space or maximum likelihood shape reconstruction (Zelniker and Clarkson, 2006).

We also performed similar analysis for parallel sets of lines to determine the resolution, calculated as the smallest pairwise distances between all the lines, at different data densities. The reconstruction result is shown in Figure 3.10 for the mask in Figure 3.7, and we found that the highest resolution is obtained at 10–15% of the input data. This is a marked contrast to the spatial sampling requirements according to the Nyquist theorem, requiring a measured molecular density at half the length scale of the smallest feature size in the data.

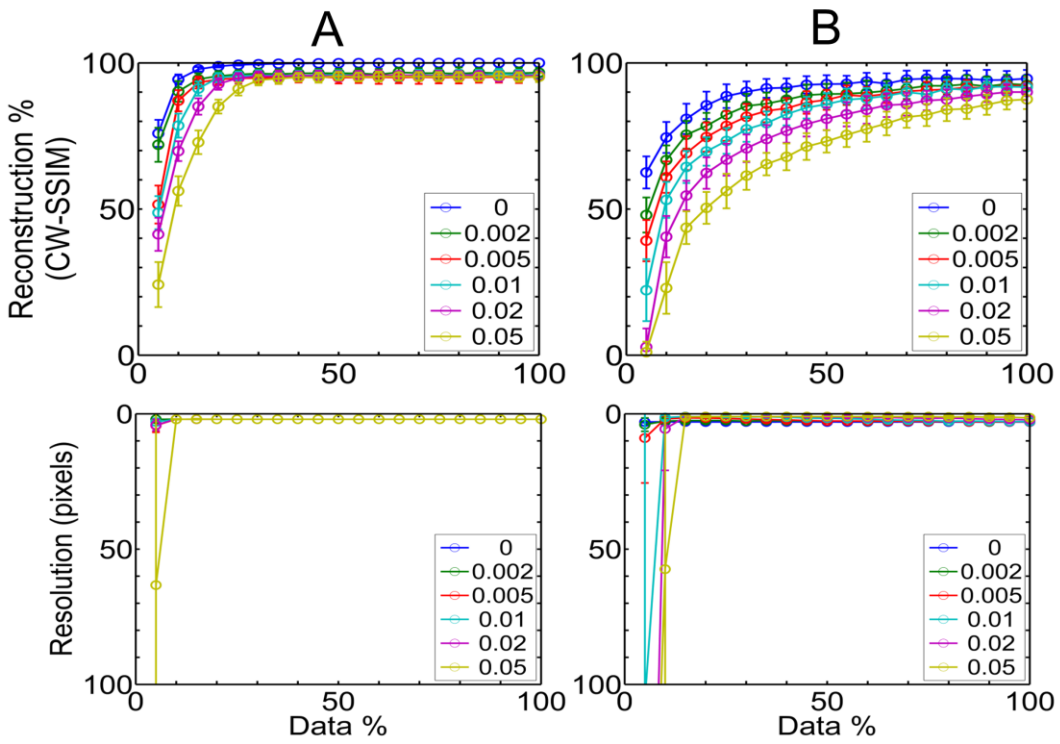


Figure 3.10. **Parallel line reconstruction.** Reconstruction measure using Structural Similarity Index CW-SSIM (top row) and resolution, calculated as the minimum inter line distance (bottom row) at indicated outlier noise densities. A total of 100 random simulations were performed at each data density. Column (A) Position noise of 0 pixel. Column (B) Position noise of 2 pixels.

3.9.4 Reconstruction from real data

We obtained the molecular position table from the previously published two-color STORM datasets (Bates et al., 2007) that labeled clathrin (red) and tubulin (green) in BS-C-1 cells. We applied the Hough Transform reconstruction for lines and circles independently on the two channels. The reconstruction is shown in Fig. 3.11 and the full reconstruction at more data densities has also been performed. It is not possible to determine the CW-SSIM without the actual structure, so the performance is gauged visually and with quantitative feature analysis. We have validated the robustness of the HT on the real data by performing the feature extraction and analysis with 100 random samplings at each of three data densities. The statistics from these analyses are shown in Table 3.1. The parameter extraction and distribution properties from the 100 random samplings are very consistent, evidenced by the negligible standard deviations in the mean and median parameter values. It should be noted that at 100% density the data remains the same for each sampling and hence the feature extraction is exactly the same for all the sampling instances with standard deviation of practically zero for all the parameter values. This method is robust to cross-talk (Fig. 3.13) (as explained in the methods section) of the multicolor channels and so it was not necessary to perform density filtering (Bates et al., 2007) prior to analysis.

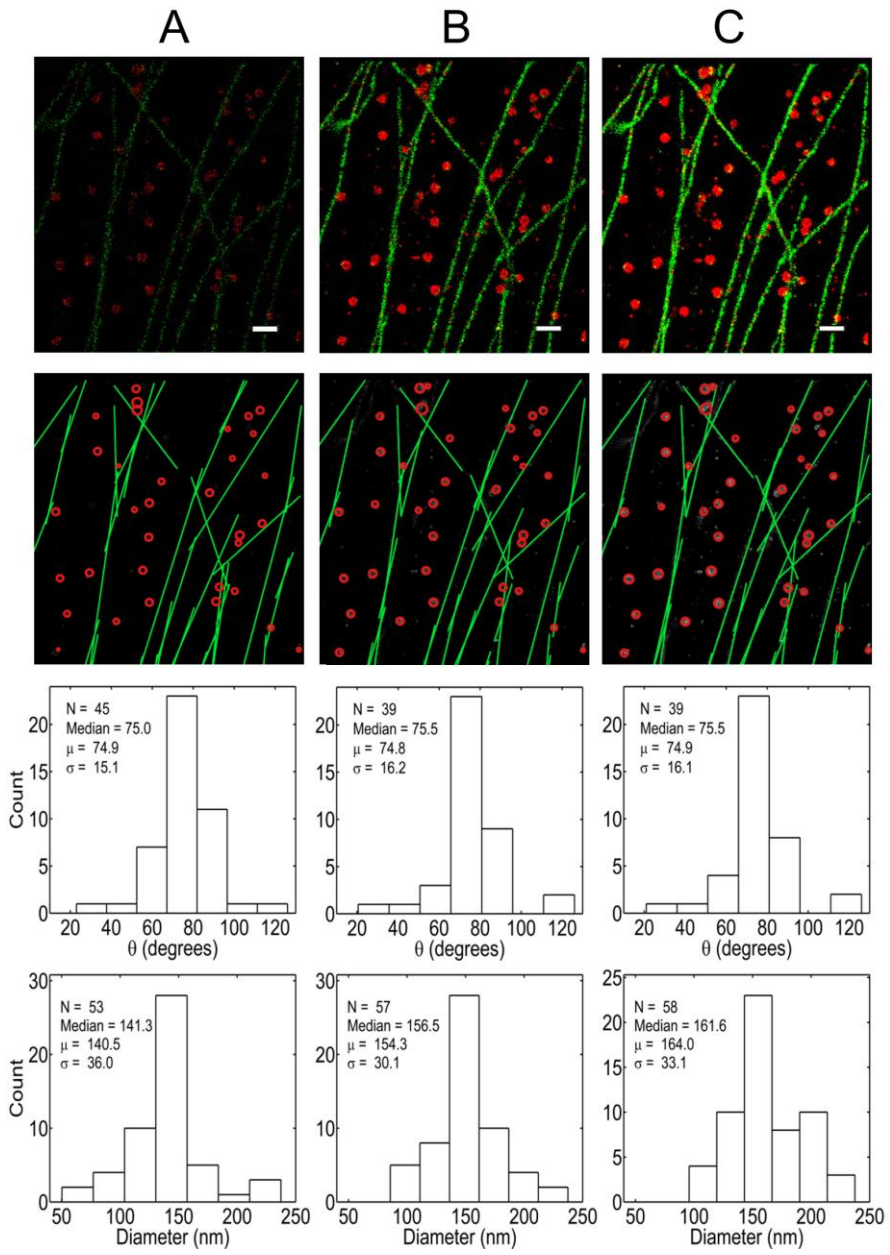


Figure 3.11. Single molecule localized data of clathrin (red) and tubulin (green) . Top row is the plotted positions from both channels. Scale bar is 500 nm. Second row is the representative reconstructed structures from both channels, overlaid on the data. (A) 10% data (B) 50% data. (C) 100% data. Third row is the histogram of orientation angle of the reconstructed line segments and the bottom row is the histogram of the diameters of the reconstructed circles.

The original data provided was in camera pixel coordinate space. We have performed the reconstruction at 25x scaling from the original coordinate space (~6 nm x 6 nm pixel-size). This scale retains most close points without being binned into the same pixel when we discretize the coordinates for analysis. Most of the structural information is obtained at just 10% of the single molecule localization data (Fig. 3.11) and very little additional information is recovered at higher data densities. This holds true for both the image reconstruction and the extracted distributions of quantitative traits from the objects. The quantitative information extracted from the HT parameters for objects identified in the tubulin and clathrin localization data is shown in the histograms of Fig. 3.11 (third row–tubulin, fourth row–clathrin). The histograms of tubulin orientation are practically identical with a mean and median of about 75° for the three data densities shown here. The distribution of clathrin vesicle diameters is also similar for the three data densities. The mean and median values of the distributions of clathrin diameters are slightly higher with increasing data density, increasing from 140 nm (10%) to 160 nm (100%), a likely consequence of the increased data density providing more votes from localizations at the periphery of the circular objects. As with any automated analysis, there are some missed structures and some spurious structures in the reconstruction. These represent ~10% of the distinct features identified by manual inspection. The choice of parameters could be optimized iteratively to achieve the best possible solution.

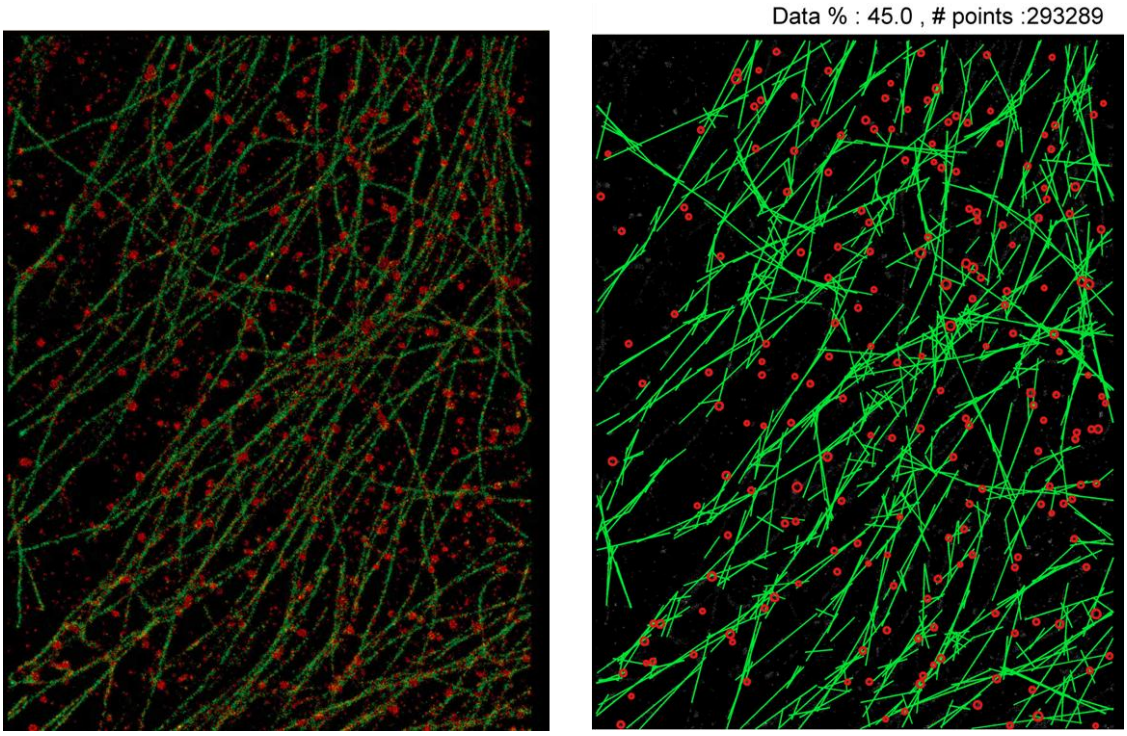


Figure 3.12. Single molecule localized data of clathrin (red) and tubulin (green) with high density of the filaments in the region of interest. The figure on the left is the localized data and the figure on the right is the HT reconstruction at 45% of the total data density.

We have compared this HT approach to an alternative feature extraction method called Blob detection (Hinz, 2005) with the Laplacian of Gaussian (LoG) as the kernel, is an established method for object detection generally applied to intensity images. We wanted to see if Blob detection can extract the circular features in a more efficient and informative way, since this method is also very popular in computer vision.

We applied blob detection to datasets with 10% and 50% of the clathrin localizations included, and attempted to extract quantitative parameters from the blob analysis (Fig. 3.13). This approach generated multiple blob circles of different radii at multiple scales for the same feature, so we had to filter out the smaller circles with an aggressive size filter, eliminating some circles of a biologically relevant length scale. While this approach correctly locates the possible features, it tends to overestimate the circle size as can be seen from Fig. 3.13C and G and the diameter histograms D and H. Moreover, since it does not discriminate between different feature types, it is not robust to cross talk from the other channel. For quantitative analysis of sparse localization data, the HT is significantly more robust than the blob detection.

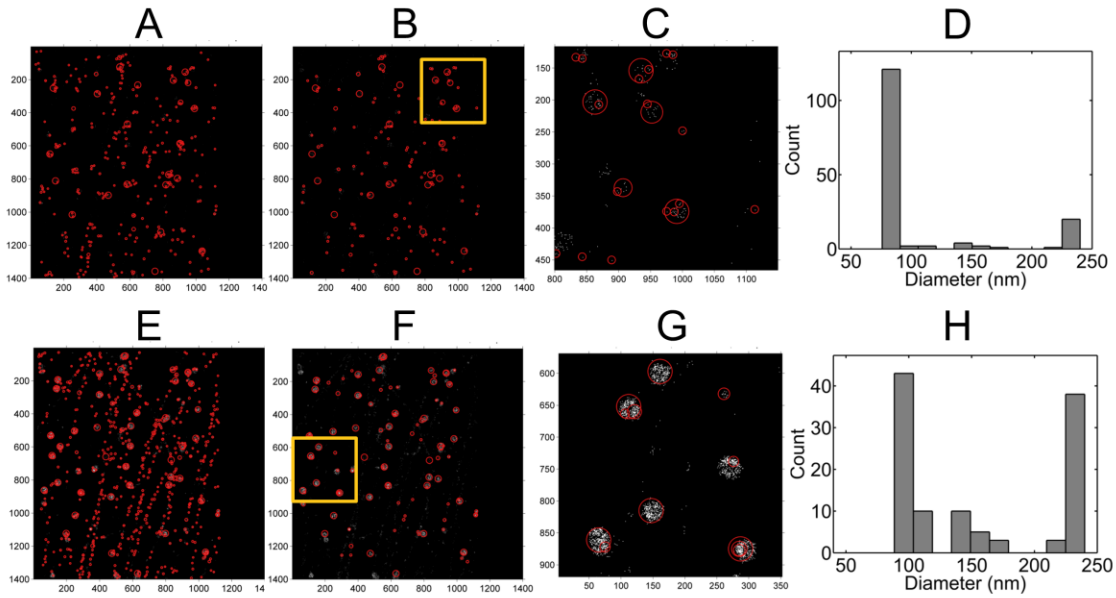


Figure 3.13. Laplacian of Gaussian (LoG) blob detection of circular features. Multi-scale kernel size range is set to 1.0% - 10% of the image size (1400x1400) and radius search range of 1.6 – 19 pixels which corresponds to ~ 10 to 120nm.It is a multiscale detection hence there are more than one circles with different radius for a detected blob. (A) Detection at 10% data density. (B) Same as (A), circles with radius less than 6 pixels (~38 nm) are removed. (C) Close up view of the yellow region in (B). (D) Histogram of the detected blob radii in (B) (E) Detection at 50% data density. (F) Same as (E), circles with radius less than 6.5(~41nm) pixels are removed. (G) Close up view of the yellow region in (F). (H) Histogram of the detected blob radii in (F)

Hough Transform is robust to outlier noise, as we have seen from the reconstruction of lines and circles in the simulated data at various outlier noises as seen in Figure 3.9. In practice, most of the multicolor fluorescent single molecule data will have cross-talk from the different channels making the data much noisier as can be seen in the Figure 3.14. This method is robust to cross-talk

of the multicolor channels and so it was not necessary to perform density filtering prior to analysis.

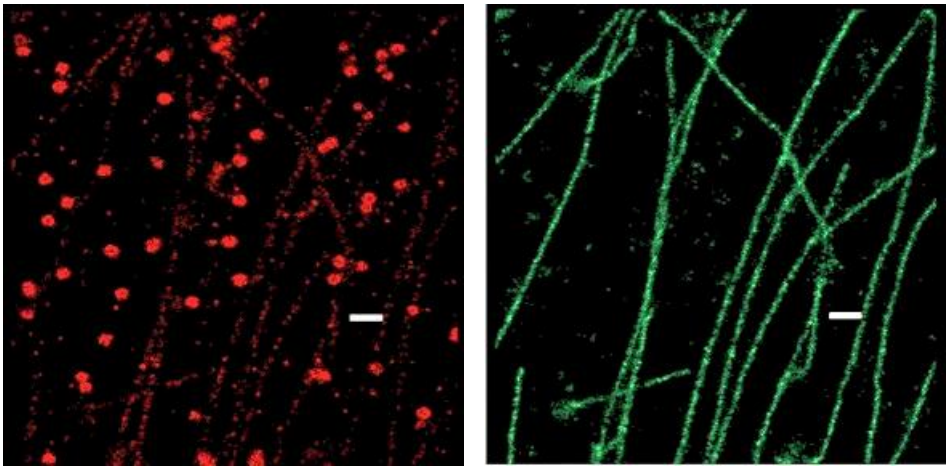


Figure 3.14. CCP(left) and Tubulin(green) data showing cross-talk from the green and red channel. Scalebar is 500 nm

This approach can be potentially used with dynamic SM-SR imaging of structural components in cells to improve the temporal resolution by a factor of 5 to 10. Since the parameters of the method represent physical traits such as radius of circles or orientation angle of lines, we were able to extract meaningful distributions of object properties with this approach.

The quantitative information for the real data is shown in the histograms of Fig. 3.11. The histograms of tubulin orientation are practically identical with a mean and median of about 75° for the three data densities shown here. The

distribution of clathrin vesicle diameters is also similar for the three data densities; however, the actual mean and median values of the distributions are slightly higher for higher data density cases, increasing from 140 nm (10%) to 160 nm (100%).

Data density		10%	50%	100%
$\theta_{average}$ (degree)	μ	76.04	75.6	74.9
	σ	1.4	0.64	1.6×10^{-13}
θ_{median} (degree)	μ	75.8	76.4	75.5
	σ	1.02	1.1	0.0
$D_{average}$ (nm)	μ	137.2	152.8	164.0
	σ	3.5	1.5	2.6×10^{-13}
D_{median} (nm)	μ	138.1	156.5	161.6
	σ	6.4	2.6×10^{-13}	2.6×10^{-13}

Table 3.1. HT extracted feature parameter values for the real data over 100 random sampling at 10 , 50 and 100 % data density. $\theta_{average}$ and θ_{median} are the mean and median values of the orientation angle of all the lines for a particular sampling.$D_{average}$ and D_{median} are the mean and median values of the diameter of all the circles. μ and σ are respectively the mean and standard deviation over the 100 random sampling for the average and the median values of the distributions at each random sampling.

Parameter	Value, Range
Lines	
θ (degree)	0.5, [-90, 89.5]
ρ (pixels)	6
maximum peaks	5000
peak separation (2d) (pixels)	[15,19] or [17,19]
peak threshold	(0.23-0.26) × max(H)
minimum line length (pixels)	(126-168)
H bin gap filling (pixels)	(47.5-33.6)
Circles	
r (pixels)	0.4, [10, 120] × scale/pixelsize
maximum peaks	200
minimum spatial separation between peaks (pixels)	55
minimum radius separation between peaks (pixels)	(71-75)
peak threshold	(0.53-0.63) × max(cH)

Table 3.2. HT Parameter information for the HT reconstruction of the real dataset**.** [,] indicates fixed range values for all conditions and those in (-) are values that vary from 5–100% data density. The single values listed for the parameters θ, ρ and r are the discretization steps. *scale =25* and *pixelsize =158 nm.* A detailed list of parameter values for all data densities (5% steps) are provided in the Supporting Table S1 of (Maji and Bruchez, 2012) .

3.10 Discussion

Generative models allow efficient reconstruction of underlying parametric objects in both simulated and real localization microscopy datasets at data densities between 10 - 20%. These approaches substantially improve the efficiency of SM–SR imaging to generate quantitative biological and structural information. This approach can be potentially used with dynamic SM-SR imaging of structural components in cells to improve the temporal resolution by a factor of 5 to 10. Since the parameters of the method represent physical traits such as radius of circles or orientation angle of lines, we were able to extract meaningful distributions of object properties with this approach. Moreover, the method appears to be fairly robust in extracting the features from the real dataset as evident from Table 3.1. More careful quantification of the parameter space could be used to extract, for example, the underlying molecular density for a feature, since the classical HT method is based on implicit Bayesian voting of the localized points in the datasets. It is also possible to obtain the persistence length of the tubulin from the obtained coordinates of the lines with further analysis.

The difference of estimated clathrin diameter histograms in Figure 3.11 could be due to the fact that at lower data density the edge points are most likely underrepresented in the vote counts relative to high data densities. To overcome this issue we can apply weighted voting for circle detection so that even a small number of points towards the outer edge of the circles can receive enough votes to be considered as a valid shape. We have tested this correction, but found that the full normalization appeared to overestimate the boundary. The correct level of voting normalization could be estimated through a statistical learning of several such objects at low data density. There is always systematic bias in

estimating the structure at consistent data densities typical of biological experiments. In spite of this, quantitative comparisons across treatment conditions with similar data densities remain informative in assessing differences in biological datasets.

As seen from the results section, the classical HT for line detection was limited to narrow filamentous structures since it has no accommodation for the uncertainty of the molecular position. Methods do exist for such purposes (Zhang and Couloigner, 2007). In this study, we have shown that given sparse molecular positions we can generate the corresponding biological structures with high efficiency using simple shape primitives. Variants of the HT and (Cootes et al., 1998; Cootes et al., 1995; Davies et al., 2002; Mokhtarian and Mackworth, 1992; Simoncelli et al., 1992; Staib and Duncan, 1992) can detect arbitrary shaped structures. Here we have applied only the classical form of the HT for inferring basic parameterizable biological shapes; but the next step is to perform parameter optimization by estimation using Monte Carlo sampling and arbitrary shape extension using variants of the classical HT such as the Generalized HT (Ballard, 1981), which can be used for shapes without a parametric form, Randomized or Probabilistic HT (Fung et al., 1996), or the Progressive Probabilistic HT (Galamhos et al., 1999). The extension of generative models to single molecule super-resolution image analysis may be particularly useful for dynamic imaging of cellular components at high spatial and temporal resolution.

CHAPTER 4

GENERATIVE PROBABILISTIC GRAPHICAL MODEL

Modeling arbitrary shaped biological structures using super-resolution image data is a non-trivial task. The information that we directly have is the position of the molecules and the diffraction-limited image of the entire structure. However, we also have another set of information, which is the set of inter-molecular distances, and if the structure is already known then the molecular arrangement can be used. This arrangement can provide the local orientation of the molecules. Now since we are dealing with sub-diffraction structures with unknown molecular density, we have a problem with incomplete information and it will require estimation of missing components. This type of problem is well studied with generative probabilistic graphical models.

In machine learning, generative probabilistic models are models which can tell us how the observed data was generated given some the hidden (latent) variable. This method works on the principle of how cause (variables) generates effect (observable data); hence, they are called generative models. Generative models attempts to estimate the joint probability distribution over all hidden and observed variables. Probabilistic graphical models (Bishop, 2006; Jordan, 2004; Murphy, 2010; Yedidia et al., 2003) are the class of machine learning algorithms, which deals with probabilistic inference on graphical networks and has been widely used for image analysis in computer vision and also in medical imaging.

When the graphical model is undirected it is called Markov Random Field (MRF) (Bishop, 2006). It can represent cyclic dependencies, whereas a directed graphical model is called Bayesian network, which can represent induced dependencies. The nodes (could be image pixels or single molecules) represent random variables and the edges represent the conditional independence assumptions. Thus, they can represent joint probability distributions in a compact manner. Another advantage in using graphical model is that it can be extended for structural data beyond two dimensions and also for temporal tracking, by using an even larger graphical framework and linking nodes in adjacent time steps (Isard, 2003). We are going to review the basic theory of graphical model and non-parametric belief propagation and then setup for the application to localization data sets of biological structures.

4.1 Basics of Graphical Model

A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ **c**onsists of a finite set of nodes (or vertices) $i \in \mathcal{V}$ connected by a set of edges$(i, j) \in \mathcal{E}$. Each node i is associated with a random variable X_i. The edges of the graph represent some meaningful relationships or physical constraints between the random variables through a joint probability distribution over the random variables. In MRFs, these distributions are characterized through a set of conditional independencies.

4.2 Markov Random Field

Markov Random Fields are undirected graphical models where are the random variables representing the nodes of the graph have Markov property. In simple notation, the Markov property can be represented as:

$$P(X_i = x_i | X_{i-1} = x_{i-1}, \ldots, X_0 = x_0) = P(X_i = x_i | X_{i-1}).$$

Then we can define various types of conditional dependencies on the nodes such as with pairwise, local or global Markov property. We are particularly interested in the pairwise Markov Random Field and we will discuss it in a little more detail.

4.2.1 Pairwise Markov Random Fields

Suppose we have some random variables x_i which are unobserved (hidden or latent) and y_i which are known (observables), then we can define a relationship between the two variables by $\phi_i(x_i, y_i)$, or, since y_i is known, simply $\phi_i(x_i)$, called as the evidence for the node x_i. For example if we have an image then each pixel *i* can be represented by a node variable x_i and the pixel intensity value (observable) can be represented by y_i. The conditional distribution for nodes x_i and x_j is given by $\psi_{ij}(x_i, x_j)$. This is also known as the adjacency potential or conformational potential $\psi_{ij}(x_i, x_j)$ and is defined such that the message passing is bi-directional but the potential itself is not symmetric. Therefore, we will have

$$\psi_{ij}(x_i, x_j) = \psi_{ji}(x_j, x_i)$$

but $\psi_{ij}(x_i, x_j) = \psi_{ij}(x_j, x_i)$ is not necessarily true

The joint probability distribution over all the nodes is given by the following equation:

$$P(x) = \prod_{i \in \mathcal{V}} \phi_i(x_i) \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) \tag{4.1}$$

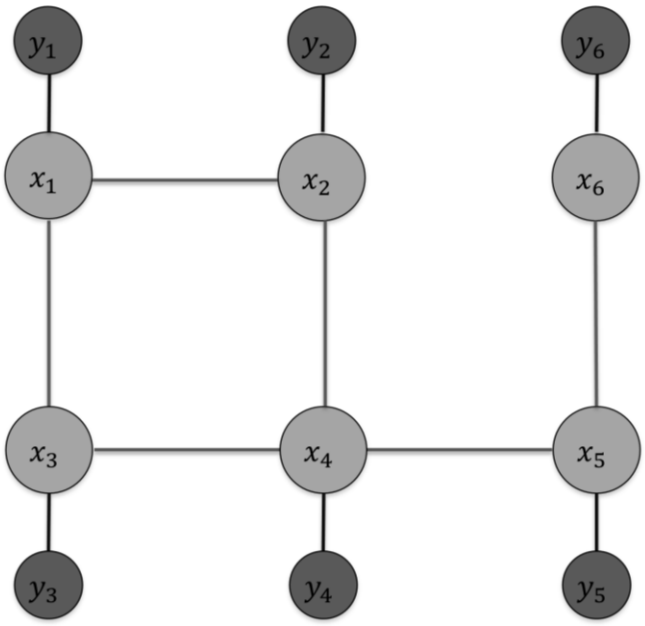


Figure 4.1. Simple example of an undirected graphical model (MRF). The nodes are in light gray and the circle in dark gray represents the observation for the corresponding nodes.

For the model in Figure 4.1, the joint probability distribution is given by:

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$
$$= \frac{1}{Z}\phi_1(x_1)\psi_{12}(x_1, x_2)\phi_2(x_2)\psi_{13}(x_1, x_3)\ \phi_3(x_3)\psi_{24}(x_2, x_4)\psi_{34}(x_3, x_4)\phi_4(x_4)$$
$$\times \psi_{45}(x_4, x_5)\phi_5(x_5)\psi_{56}(x_5, x_6)\phi_6(x_6)$$

(4.2)

where Z is the normalization constant known as the partition function, which is in fact the limiting factor for the computational tractability of any graphical model approach. For smaller graphical models the inference can be performed exactly but it becomes increasingly difficult and intractable, so approximate inference is usually performed using methods such as belief propagation, which works quite efficiently in most practical cases.

4.3 Inference in Graphical Models

In probabilistic graphical model, usually the objective is to compute the marginal probabilities and posterior distributions and also the maximum a posteriori (MAP) probabilities or the mean values of the hidden variables given some observation or evidence. For instance for finding the posterior mean value, we have to first find the marginal probability at each node. A specific example is shown later in example Figure 4.3, with the probability replaced with messages, as the inference method used here is belief propagation described in the next section.

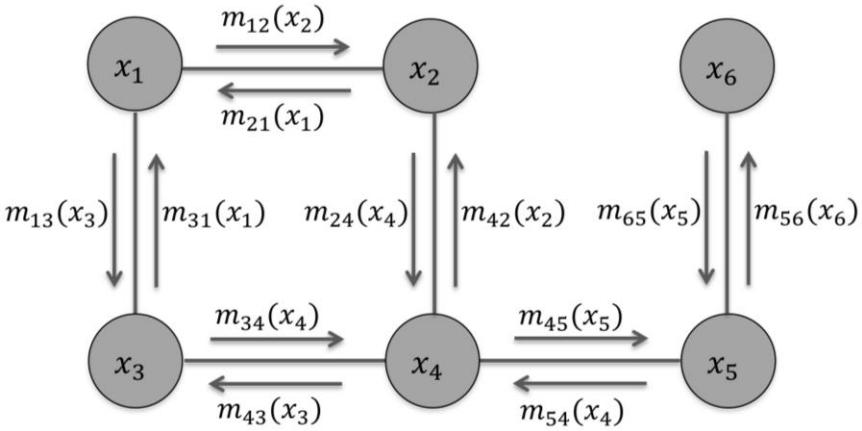
4.4 Belief Propagation

Belief propagation (BP) (Jordan, 2004; Sudderth et al., 2010; Weiss and Pearl, 2010; Yedidia et al., 2003) is a widely used algorithm for performing inference on Bayesian networks through a message passing method. It can be applied to both directed and undirected graphical models. The inference for a very large graphical model becomes computationally intractable so approximate inference methods such as loopy belief propagation are used, although it is not guaranteed to converge or be accurate, but it does in most practical cases of computer vision and imaging. It can handle missing values so it can be useful in our modeling with single molecule imaging data.

BP is a method used for inference in a graphical model for computing the marginal probability distribution of the hidden nodes conditional on the observed nodes. The basic principle of this method is passing and updating the 'messages' about the 'states' between the adjacent hidden nodes of the graph in an iterative process until convergence. Then through a consensus, the marginal probabilities or 'belief' for each node is computed. For discrete random variables representing the nodes, the messages can be passed as matrices, and for the continuous case, the messages can be passed using a mean and covariance matrix of the multivariate Gaussian for the jointly distributed random variables.

The way BP is performed is through passing a message $m_{ij}(x_j)$ from node i to node j. This message can be perceived as node i's belief about node j being in certain states with various likelihoods. The messages from node i to j is updated with all the messages flowing into i (Figure 4.2 A).

A



B

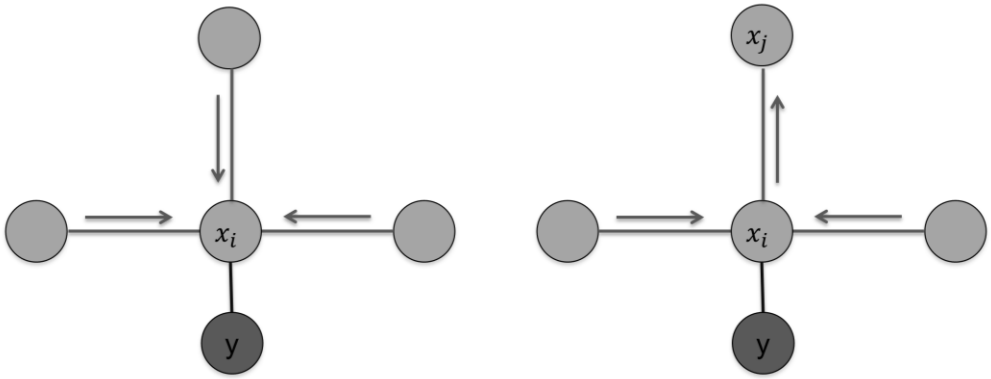


Figure 4.2. Belief propagation in a graphical model**.** (A) An illustration of messaging passing in pairwise MRF. (B) An illustration of message update and belief equation. It is a kind of distributed way of computing the marginal for the nodes.

The iterative message updating scheme can be written as:

$$m_{ij}(x_j) = \sum_{x_i} \phi_i(x_i)\, \psi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \tag{4.3}$$

The approximate posterior marginal distributions, which are called as the node's belief, is given by

$$b_i(x_i) = \alpha\, \phi_i(x_i) \prod_{k \in N(i)} m_{ki}(x_i) \tag{4.4}$$

where α is the normalization constant to ensure the beliefs sum to 1 and $N(i)$ is defined as the neighborhood set for the node i $N(i) \backslash j$ is the neighbor of i other than j in equation 4.3. m_{ji}'s the messages that the node i receives from its neighbors $N(i)$.

4.4.1 Example

Let us first consider the example in Figure 4.3, which is similar to the example in Figure 4.2A, but with the connection between node 1 and node 2 removed, to make it a MRF without loop. Following the equations 4.3 and 4.4, let us try to compute the belief for node 6.

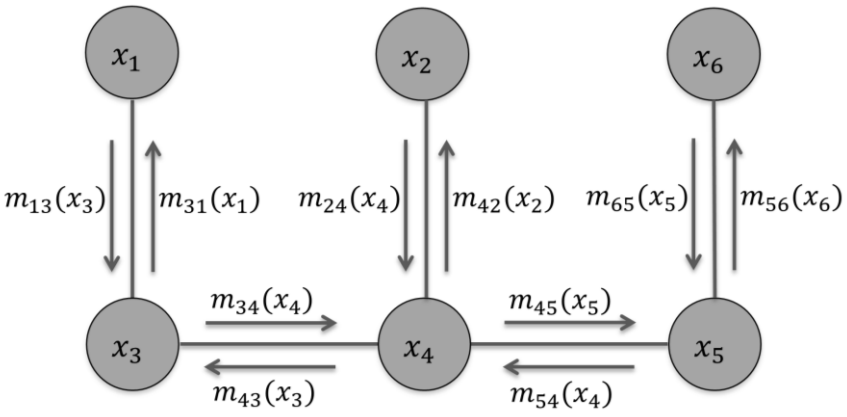


Figure 4.3. Belief propagation in a graphical model without loop**.** An illustration of messaging passing in pairwise MRF without loop.

The belief for node 6 is computed as:

$$b_6(x_6) = \alpha\, \phi_6(x_6)\, m_{56}(x_6)$$

$$m_{56}(x_6) = \sum_{x_5} \phi_5(x_5)\, \psi_{56}(x_5, x_6)\; m_{45}(x_5)$$

$$m_{45}(x_5) = \sum_{x_4} \phi_4(x_4)\, \psi_{45}(x_4, x_5)\; m_{24}(x_4)\, m_{34}(x_4)$$

$$m_{34}(x_4) = \sum_{x_3} \phi_3(x_3)\, \psi_{34}(x_3, x_4)\; m_{13}(x_3)$$

$$m_{24}(x_4) = \sum_{x_2} \phi_2(x_2)\, \psi_{24}(x_2, x_4)$$

$$m_{13}(x_3) = \sum_{x_1} \phi_1(x_1)\, \psi_{13}(x_1, x_3)$$

Now, substituting we get,

$$b_6(x_6)$$

$$= \alpha\, \phi_6(x_6) \sum_{x_5}\sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} \phi_5(x_5)\, \psi_{56}(x_5, x_6)\, \phi_4(x_4)\, \psi_{45}(x_4, x_5) \phi_3(x_3) \psi_{34}(x_3, x_4)$$

$$\times\, \phi_2(x_2)\, \psi_{24}(x_2, x_4) \phi_1(x_1)\, \psi_{13}(x_1, x_3)$$

$$= \sum_{x_5}\sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} \alpha\, \phi_6(x_6) \phi_5(x_5)\, \psi_{56}(x_5, x_6)\, \phi_4(x_4)\, \psi_{45}(x_4, x_5) \phi_3(x_3) \psi_{34}(x_3, x_4)$$

$$\times\, \phi_2(x_2)\, \psi_{24}(x_2, x_4) \phi_1(x_1)\, \psi_{13}(x_1, x_3)$$

$$= \sum_{x_5}\sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p_6(x_6)$$

Therefore, if the graph has no loops, then the message passing converges in finite steps and BP provides the exact marginal for each node. The time complexity for the algorithm is proportional to the number of edges in the graphical model.

Now if we try to compute the belief $b_6(x_6)$ for node 6 in example 4.2A, which has a loop, we will have the same expressions for $m_{56}(x_6)$ and $m_{45}(x_5)$ as above and the following:

$$m_{24}(x_4) = \sum_{x_2} \phi_2(x_2)\, \psi_{24}(x_2, x_4)\; m_{12}(x_2)$$

$$m_{12}(x_2) = \sum_{x_1} \phi_1(x_1)\, \psi_{12}(x_1, x_2)\; m_{31}(x_1)$$

$$m_{13}(x_3) = \sum_{x_1} \phi_1(x_1)\, \psi_{13}(x_1, x_3)\; m_{21}(x_1)$$

and so on.

We can now see that the messages are not converging and rather bouncing from one node to its adjacent node and so on. So for undirected graph with loops the belief propagation is not guaranteed to converge and even if it does it may not be accurate. Therefor approximate estimates of marginal are computed for graphs with loops using what is called as loopy belief propagation (LBP) (Ihler et al., 2005; Murphy et al., 1999; Yanover and Weiss, 2003). In practice, it has been observed that the beliefs converge to rather good approximations. Another approach for approximate inference is by using particle filtering.

4.4.2 Nonparametric Density Estimates

When the likelihoods of density function are known then we can model the data directly. If the true distribution is unknown but at least their parametric form is known then we can model the data using parameter estimation by assuming a particular form of the density, for instance, say a Gaussian distribution. In this case, we need to estimate only the parameters such as mean and variance. However, when there is no information about the likelihoods or their parametric form, ,which is in fact the most common case in real data, then we have to resort to determine the density function directly from the data, without assuming a particular form for the underlying distribution. The usual way is to approximate a probability density function (PDF) by a set of smoothed data samples(Sudderth et al., 2010) .

$$\hat{p}(x) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sigma}K\left(\frac{x - X_i}{\sigma}\right)$$

where $\{X_i\}$ are M independent samples from p(x) and $K(.)$ is the kernel and σ is the bandwidth also called as the smoothing parameter.

If we have a Gaussian kernel and set $\sigma = w(i)$ and the covariance Λ_i , then we have:

$$\hat{p}(x) = \frac{1}{M}\sum_{i=1}^{M} w(i)\,\mathcal{N}(x,\, X_i\,,\, \Lambda_i)$$

(4.5)

Here the input messages are Gaussian kernel density estimates.

The simplest form of nonparametric density estimate is the histogram.

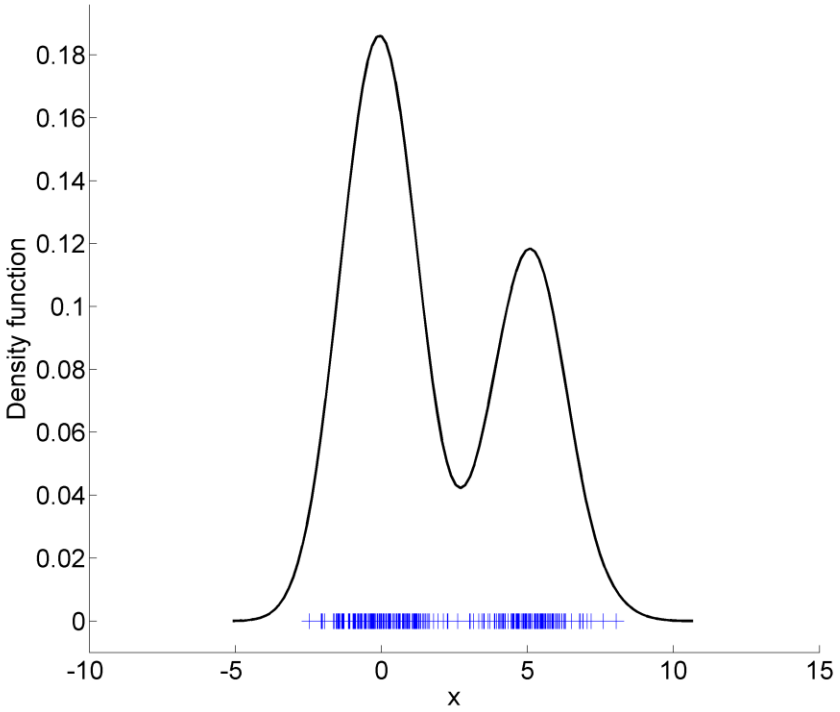


Figure 4.4 Nonparametric kernel density Estimate.

4.4.3 Nonparametric Belief Propagation

Belief propagation can produce exact conditional marginal for discrete valued random variables, whereas for continuous valued random variables, the message update integral as defined in 4.5 becomes intractable. In order to make the method tractable the messages are represented as kernel density estimate (Ihler et al., 2005; Murphy et al., 1999; Yanover and Weiss, 2003) as defined above.

A reasonable assumption for the density is to be a normalized Gaussian (x, μ, Λ), then the message with mixture can be approximated using the form of (4.5)

$$m_{ij}(x_j) = \sum_{x_j} w_j^{(k)} \mathcal{N}(x_j, \mu_j^{(k)}, \Lambda_j)$$

(4.6)

The pairwise potential is formulated over the cliques using the rules form graph theory and this clique potential to be meaningful should be finitely integrable, which means the following conditions should hold:

$$\int \psi_{ij}(x_i, x_j)\, dx_i < \infty \;, \int \psi_{ij}(x_i, x_j)\, dx_j < \infty \qquad \forall\, (i,j) \in \mathcal{E}$$

$$\int \phi_i(x_i)\, dx_i < \infty \qquad \forall\, i \in \mathcal{V}$$

(4.7)

One of the important features of the potentials here is that they are informative (Ihler et al., 2005; Murphy et al., 1999; Yanover and Weiss, 2003) , which means constraining the value for one variable has a restrictive effect on other variables too. This property can be useful, when we model actual structures, since the constraint state space may be interrelated. In graphical models, in order for the inference algorithms to be applicable, the above normalization constraints should hold which ensures the convergence of the algorithms, but in most practical cases, the variables are bounded within a finitely large range.

We have developed our method upon the base implementation (Sigal et al., 2004) of the Non-parametric Belief propagation described in (Isard, 2003) called Particle Filter Message Passing (PAMPAS).

4.4.4 Particle Filter Message Passing

Inference on graphical model using belief propagation has been proposed in various works such as Non-parametric Belief Propagation (NBP), Belief propagation Monte Carlo (BPMC), Particle Message Passing (PAMPAS), etc. (Isard, 2003; Wei-Kai and Cohen, 2006). For discrete random variables we have equations 4.3 and 4.4 and for continuous case we have:

$$m_{ij}(x_j) = \int_{\mathbb{R}^D} \phi_i(x_i)\, \psi_{ij}(x_j, x_i) \prod_{k \in N(i) \setminus j} m_{ki}(x_i)\, dx_i \tag{4.8}$$

When the random variables are continuous, the potential functions ϕ_i and ψ_{ij} can be considered as probability density functions and hence m_{ij} in equation 4.8 is also a probability density function. When the marginal distribution at the nodes is Gaussian then the integration is exact otherwise, we have to perform numerical approximation usually through Monte Carlo.

Now the above integral can be approximated to $\widetilde{m}_{ij}$ using a Monte Carlo sampling (Wei-Kai and Cohen, 2006) as follows :

$$\widetilde{m}_{ij}(x_j) = \int_{\mathbb{R}^D} \psi_{ij}(x_j, x_i)\, g(x_i) dx_i$$

$$= E_{g(x)}\left[\psi_{ij}(x_j, x_i)\right]$$

$$= \frac{1}{N} \sum_{n=1}^{N} \psi_{ij}(x_j, s_{ij}^n) \quad , \qquad s_{ij}^n \sim g(x)$$

where $g(x) = \phi_i(x_i) \prod_{k \in N(i) \setminus j} \widetilde{m}_{ki}(x_i)$

(4.9)

Now the sampling method proposed in PAMPAS algorithm is that the messages or beliefs are accumulated from the neighboring nodes and then taking the product weighted by the local evidence of the node in consideration. The basic idea of particle filtering is also discussed in section 6.3.2 of chapter 6. In the case of graphical model, the probability beliefs are just propagated along after being compute by the particle filters. The message m_{ij} can be updated using a weighted particle set $\{ s_{ij}^n, w_{ij}^n \}_{n=1}^N$ sampling (importance sampling) as follows:

$$s_{ij}^n \sim \prod_{k \in N(i) \setminus j} \widetilde{m}_{ki}(x_i)$$

(4.10)

The importance weights can be obtained using the current belief estimate:

$$w_{ij}^n = \phi_i(s_{ij}^n) \quad (4.11)$$

Alternatively, the samples can be drawn from the observation or evidence model and weighted by the product of incoming messages from the neighboring nodes and the effective belief propagation will still remain the same. The samples can be generated from the joint distribution over the graphical model using a Gibbs sampler. The potentials $\psi_{ij}(x_i, x_j)$ in many cases can be expressed as a mixture of Gaussians, which makes the message propagation more tractable. The basic principle of computing the updates for the marginal (Sudderth et al., 2010) is shown in Figure 6.5 The detailed description of the algorithm can be found in (Isard, 2003).

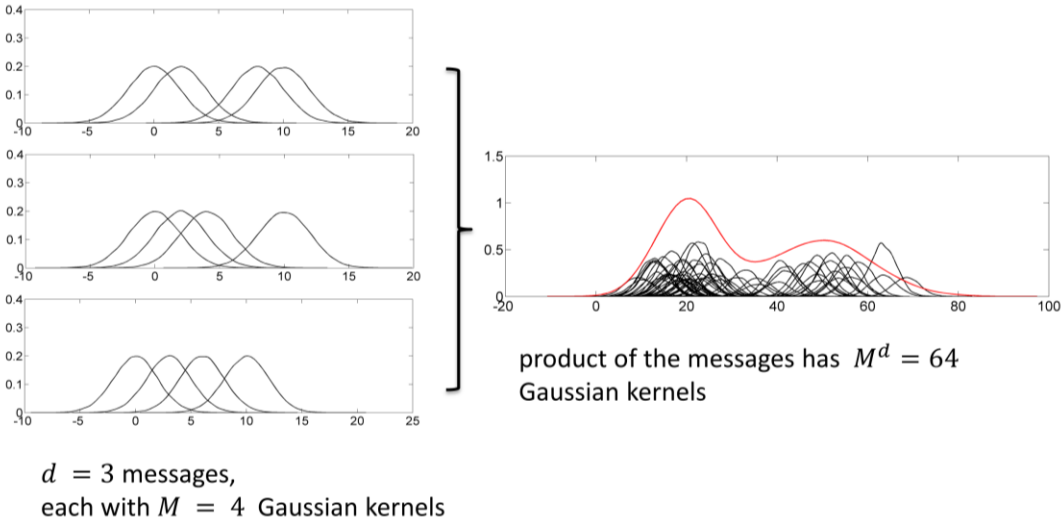


Figure 4.5 Nonparametric belief propagation marginal update schematic (Sudderth et al., 2010).

There is one more important feature about this graphical model approach in general, which is its inherent ability to estimate the state of unobserved nodes even when sufficient information is not present. This is very important for cases where the structural data is partially missing or has been corrupted due to noise or any other reason.

4.5 Formulation of the probabilistic graphical model for super-resolution

The overall idea is to start with diffraction limited image and an initial model of interest and try to learn a generative model from them with the help of partial single molecule data if necessary to constrain the model search space at the

beginning itself. We generate the set of possible models (based on molecular properties) that fit with the diffraction –limited prior by calculating the posterior distributions of the probabilities of the molecular locations based on rational biological rules (e.g. persistence length, molecular dimensions). From there, we can then use more single molecule images to select the best of these possible models, and eventually collecting enough molecular positions that a unique model can be selected. This is similar modeling approach to determine an all-atom protein structure model from electron density maps (DiMaio et al., 2007). We are using nonparametric belief propagation (NBP) with informative constraints and constraining the data points in an iterative manner, so we name the method as **I**teratively **C**onstrained and **E**nhanced STORM (ICE-STORM). The key question is how to reconstruct biological structures in far more details (greater resolution) than we see in normal localization microscopy using far less imaging data. There is recent work using compressive sensing theory in single molecule image where sparse signal is encoded during acquisition and then the full signal is reconstructed using only a fraction of the data (Zhu et al., 2012).

4.5.1 Evidence, Proposal and Conditional distribution

Evidence and proposal distributions for the nodes in the graphical model for the structures should be formed from the given diffraction limited structures. The conditional distribution should be formulated based on the known biological information about the structures such as intermolecular distance, etc. Following the schematic approach in (DiMaio et al., 2007) we can formulate the structure estimation problem as shown in Figure 4.6.

4.5.2 State space

The state space governing the generative PGM is $\vec{u} = [x, y, \theta_T]$ where (x, y) is the coordinate of the molecule, θ_T is the local gradient field orientation of a molecule relative to the filament.

θ_{rel} is the relative orientation of the neighboring molecules and θ_{rel}, here, is however only allowed to take values of 90^o or 0^o and 30^o or 60^o (it depends on how the configuration of the molecules in the tubulin structure is defined) relative to the local gradient since the molecules are arranged on the filament in a particular manner. In case of tubulins, the neighboring molecules are either on the same filament in a linear way or on the adjacent filament but with a shift of one position (alternating alpha and beta units). On the other hand, the local gradient field orientation will vary according to the curvature of the filaments. Therefore the initial evidence for θ_T will be derived from the gradient field of either the diffraction limited image or the initial set of molecular positions.

4.5.3 Orientation Tensor Field

Given any image or localization dataset for a biological structure, the local curvature of the structures can be computed from the gradient of the orientation tensor field (Hill, 2011b). A tensor is a generalization of a vector or matrix. First, we will consider the localization datasets given by the 2d positions $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$. We can construct a tensor at any point say $\vec{x}_k$, which provides information about the orientation for the cluster to which that point belongs and it is based on the distance and direction to the other points in the cluster.

Suppose we have a set of points $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$. Then for defining a tensor at a point $\vec{x}_k$ we take the point $\vec{x}_k$ and another point $\vec{x}_i$ in the neighborhood and define the vectors $X^{ki} = (X_1^{ki}, X_2^{ki}) = \overrightarrow{x_k x_i}$, and apply a non-linear transformation so that the vectors X^{ki} is mapped to $\tilde{X}^{ki}$ as follows:

$$\tilde{X}^{ki} = \left(\tilde{X}_1^{ki}, \tilde{X}_2^{ki}\right) = exp(-((X_1^{ki})^2 + (X_2^{ki})^2)/4\sigma^2) \frac{(X_1^{ki}, X_2^{ki})}{\sqrt{(X_1^{ki})^2 + (X_2^{ki})^2}} \tag{4.12}$$

where σ is a scaling parameter and the mapping function is based on the distance and direction between $\vec{x}_k$ and $\vec{x}_i$. The distance measure is given by the exponential term and the direction given by the unit vector $\frac{(X_1^{ki}, X_2^{ki})}{\sqrt{(X_1^{ki})^2 + (X_2^{ki})^2}}$ in equation 4.12.

Now the initial tensor at $\vec{x}_k$ is then represented as:

$$T(\vec{x}_k) = \sum_{i \neq k} \left(\tilde{X}_1^{ki}, \tilde{X}_2^{ki}\right)^T \left(\tilde{X}_1^{ki}, \tilde{X}_2^{ki}\right) \tag{4.13}$$

These are a 2×2 tensor matrices constructed at the sparse molecular locations $\vec{x}_k$. The field of tensor is calculated by kernel smoothing interpolation of the tensors $T(\vec{x}_k)$. Then if we estimate the orientation of the principle eigenvector of tensors at each point, then it will result into the field of orientations. The details of the entire method can be found in (Hill, 2011a; Su, 2008).

4.5.4 Setting up the probabilistic graphical model for biological structures

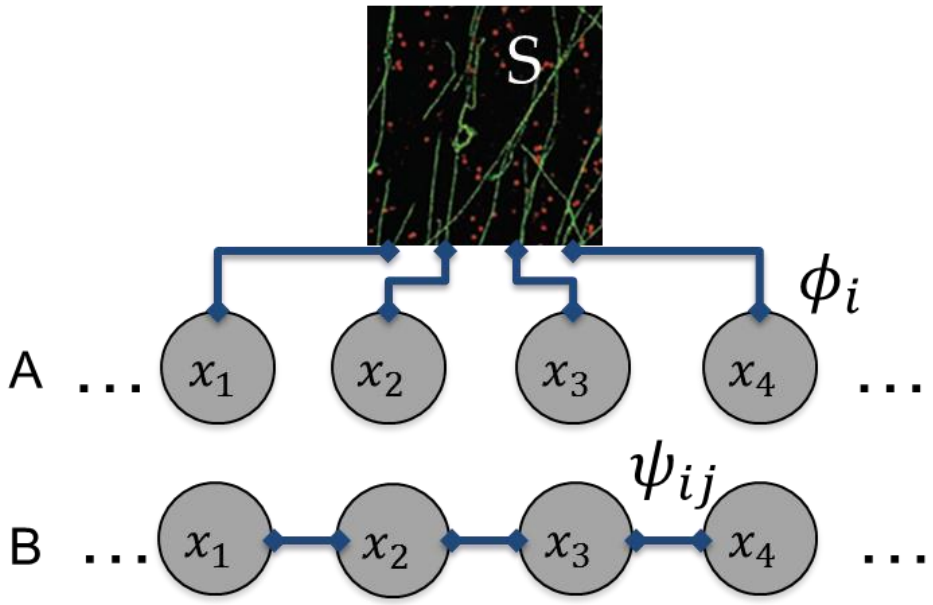


Figure 4.6 (A) Local observation potential at each node (B) Adjacency potential or conditional distribution between adjacent nodes.

Suppose, $\overline{U} = \{\vec{u}_i\},\ i = 1, \dots, N$ defines the vector of state space where

$\vec{u}_i = (\vec{x}_i, \vec{\theta}_i)\,, \qquad \vec{x}_i = (x_i\,, y_i), \qquad i = 1, \dots, N$

$\vec{\boldsymbol{x}}_{\boldsymbol{i}}$ is the position vector of node i and $\vec{\theta}_i$ is the parameter vector.

Now if we define a probabilistic graphical model over this state space then we will have the following posterior distribution:

$$P(\overline{U}|S) = \prod_i \phi_i(\vec{x}_i|S) \prod_{ij} \psi_{ij}(\vec{u}_i, \vec{u}_j)$$

(4.14)

where $\phi_i(\vec{x}_i|S)$ is the local observation potential for node i given a diffraction limited image S and $\psi_{ij}(\vec{u}_i, \vec{u}_j)$ is the adjacency potential or global conformational potential between node i and node j. In this example node i and node j are considered to be neighbors if they are within 50 nm of each other. This conditional distribution models the probability of a particular conformation of the molecules on the structure.

The objective is to maximize the probability given in equation (4.14) for an optimal configuration set $\overline{U}^* = \{\vec{x}_i^*\}$, derived as

$$\overline{U}^* = \underset{\overline{U}}{argmax}\, P(\overline{U}|S)$$

In general, for certain biological structures which have been studied in great detail with high-resolution imaging methods and electron microscopy, we potentially have some extra information for modeling purpose. For example, if we have structures such as tubulins, we can know the inter-molecular distance, which is essentially the alpha and beta protein spacing as the tubulins are labeled selectively. In addition, the structural composition of the alpha and beta tubulin dimer can provide some evidence for the relative orientation of the protein molecules along the tubulin (Figure 4.7).

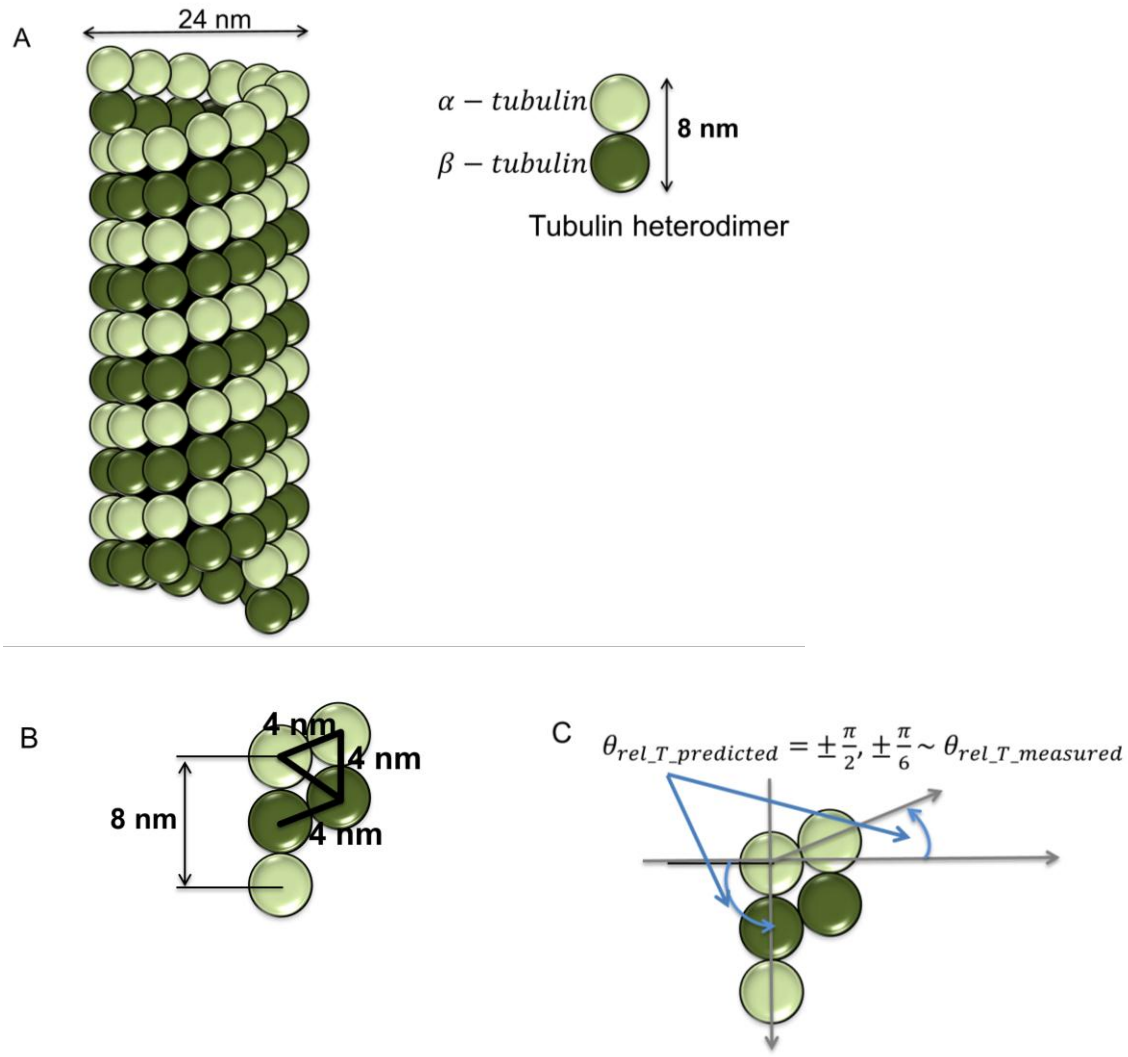


Figure 4.7 Tubulin dimer**.** (A) Microtubule structure (B) Molecular distance constraints (C) Relative orientation from the local orientation constraint as described in Figure 4.8 and the illustrative example in section 4.5.5.

Both of this information can be used to properly guide the graphical model in generating feasible configurations of the structure by incorporating the information in the conditional distribution $\psi_{ij}(\vec{u}_i, \vec{u}_j)$ which can thus be further factorized into two basic components (DiMaio et al., 2007), one for maintaining the proper inter-molecular distance (say 8 nm for the alpha and beta tubulin

distance) and the other with proper relative orientation. Therefore, the adjacency potential can be written as:

$$\psi_{ij}(\vec{u}_i, \vec{u}_j) = \psi_x(\|\vec{x}_i - \vec{x}_j\|) \times \psi_\theta(\|\vec{u}_i - \vec{u}_j\|) \tag{4.15}$$

ψ_x is the distance constraint function between the nodes, which will ensure the inter nodal distance follow a certain distribution and limits and ψ_θ is the orientation constraint between the nodes, for two dimensional case this can be represented by orientation angles. This is similar to the case for filamentous molecule in cells, which are ordered, in a linear fashion (Nogales et al., 1998). Considering the molecules as nodes of the graphical model, we can introduce constraining function described in equation 4.15 as probability distribution over inter-molecular spacing and persistence length of the filaments.

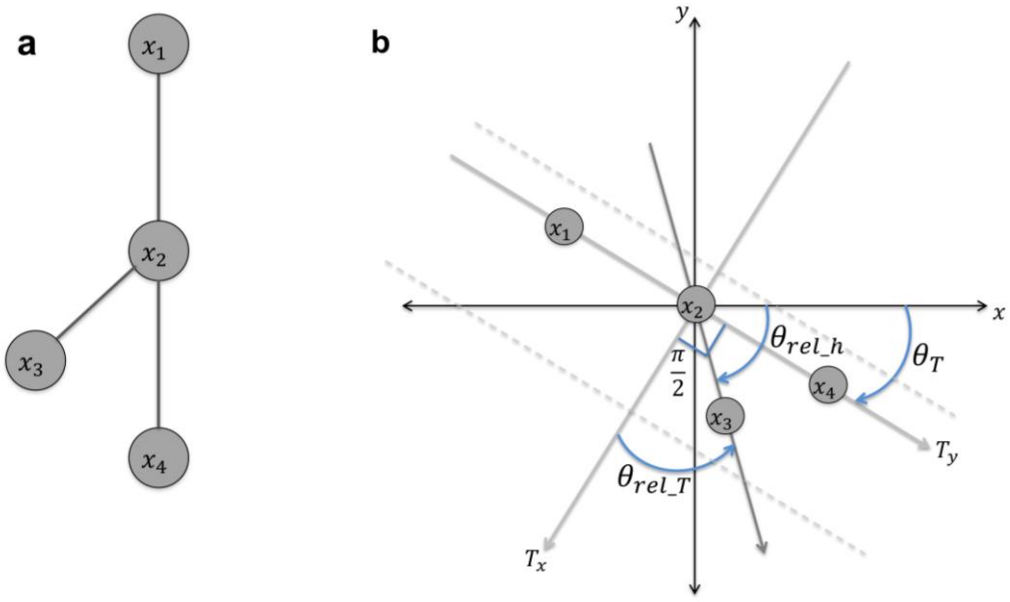


Figure 4.8 (a) A graphical model showing the node connections. (b) Orientation estimation of the vector connecting two nodes (e.g. x_2 and x_3).

The basic idea of the orientation estimation relative to the neighboring molecules is shown in Figure. 4.8b. The relative orientation is also dependent on local alignment vectors of the points on the filaments. Here T_y-T_x is the axis system corresponding to the local orientation field calculated as the gradient of the tensor field ((Su, 2008). A tensor here is defined as a 2×2 symmetric postive definite matrix and it is constructed for each point. The inertial tensor of transformations of the vectors $\overrightarrow{x_i} - \overrightarrow{x_k}$ as seen in equation 4.12 is given by :

$$T(\overrightarrow{x_k}) = \frac{\overrightarrow{x_i} - \overrightarrow{x_k}}{|\overrightarrow{x_i} - \overrightarrow{x_k}|} \exp(-\frac{1}{2\sigma^2} |\overrightarrow{x_i} - \overrightarrow{x_k}|^2) \tag{4.16}$$

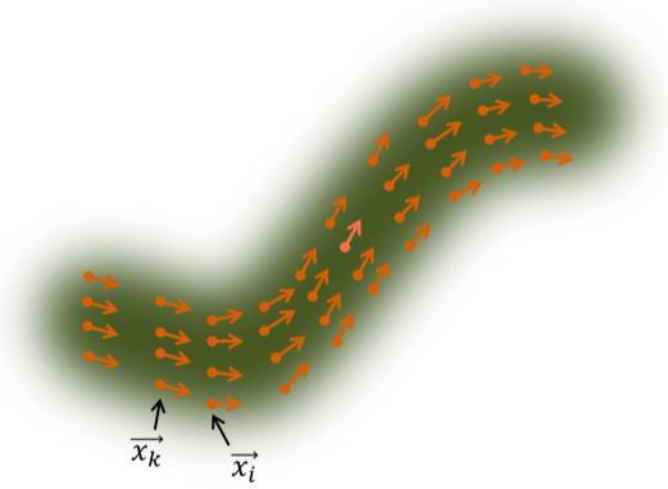


Figure 4.9 Local Orientation Tensor Field.

Then the orientation of eigenvector for largest eigenvalue is used. The orientation field is calculated using an interpolated tensor field for the sparse points based on a kernel smoothing method using weighted log-Euclidean metric. Close

neighbors will have similar orientations and based on that orientation the neighboring molecules will be positioned in a particular configuration (Figure 4.9).

Suppose θ_{rel_T} and θ_{rel_h} are the angles of the vector $\overrightarrow{x_2x_3}$ realtive to the axis T_x and the x-axis respectively. Let θ_T be the local orientation of the points in the neighborhood. It can be considered as the average of the local orientation estimated for the two nodes in consideration.

From Figure 4.8b we have the following relation

$$\theta_{rel\_T} = \theta_{rel\_h} + (\frac{\pi}{2} - \theta_T) \tag{4.17}$$

A simple example for the PGM with all the constraints is illustrated in Figure 4.10. The state space governing the generative PGM is $\vec{u} = [x, y, \theta_T]$ where (x, y) is the coordinate of the molecule, θ_T is the local gradient field orientation of a molecule relative to the filament. θ_{rel_T} is the relative orientation of the neighboring molecules and θ_{rel_T}, here, is however only allowed to take values of 90^o or 0^o and 45^o (30^o if helical) relative to the local gradient since the molecules are arranged on the filament in a particular manner. In case of tubulins, the neighboring molecules are either on the same filament in a linear way or on the adjacent filament (alternating alpha and beta units). On the other hand, the local gradient field orientation will vary according to the curvature of the filaments. Therefore the initial evidence for θ_T will be derived from either the gradient field of diffraction limited image or from the initial set of molecular positions.

4.5.5 Illustrative example

Suppose we have 4 nodes and the initial evidence is given as follows:

Evidence:

$x_1 = (0, 20),\ x_2 = (0, 0),\ x_3 = (-15, -10),\ x_4 = (15, -20)$

Conditional:

Case 1: Fig. 4.9a and 4.9b**.** The hypothetical local orientation for the 4 points is $\theta_T = -\frac{\pi}{2}$.

x_1 should be at a position, which is $\theta_{rel\_T} = -\frac{\pi}{2}\,(w.r.t\ T_x)$ relative to x_2 and 20 units from it.

x_3 should be at a position, which is $\theta_{rel\_T} = \frac{\pi}{4}\,(w.r.t\ T_x)$ relative to x_2 and $10\sqrt{2}$ units from it.

x_4 should be at a position, which is $\theta_{rel\_T} = \frac{\pi}{2}\,(w.r.t\ T_x)$ relative to x_2 and 20 units from it.

Case 2: Fig. 4.9c and 4.9d**.** The hypothetical local orientation for the 4 points is $\theta_T = -\frac{\pi}{3}$

x_1 should be at a position, which is $\theta_{rel\_T} = -\frac{\pi}{2}\,(w.r.t\ T_x)$ relative to x_2 and 20 units from it.

x_3 should be at a position, which is $\theta_{rel\_T} = \frac{\pi}{4}\,(w.r.t\ T_x)$ relative to x_2 and $10\sqrt{2}$ units from it.

x_4 should be at a position, which is $\theta_{rel\_T} = \frac{\pi}{2}(w.r.t\ T_x)$ relative to x_2 and 20 units from it.

Then we calculate

$$\theta_{rel\_h} = \theta_{rel\_T} - (\frac{\pi}{2} - \theta_T)$$

$$\Delta x = r.cos(\theta_{rel\_h})$$

$$\Delta y = r.sin(\theta_{rel\_h})$$

(4.18)

For real data, we first calculate the angle for the pair of points $\overrightarrow{x_i} = (x_i, y_i)$ and $\overrightarrow{x_j} = (x_j, y_j$from their coordinates as:

$$\theta_{rel\_h\_measured} = tan^{-1}(\frac{y_j - y_i}{x_j - x_i})$$

(4.19)

Next, we calculate the θ_{rel_T} using equation 1 and then approximate this value to the closest theoretical value $\theta_{rel\_T\_predicted}$ that the molecules can assume example $\pm\frac{\pi}{2}, \pm\frac{\pi}{4}$. Then we re-calculate $\theta_{rel\_h\_conditional}$ again using equation 4.17 with $\theta_{rel\_T} = \theta_{rel\_T\_predicted}$ and $\theta_{rel\_h} = \theta_{rel\_h\_conditional}$ and the conditional input is then calculated using equation 4.18.

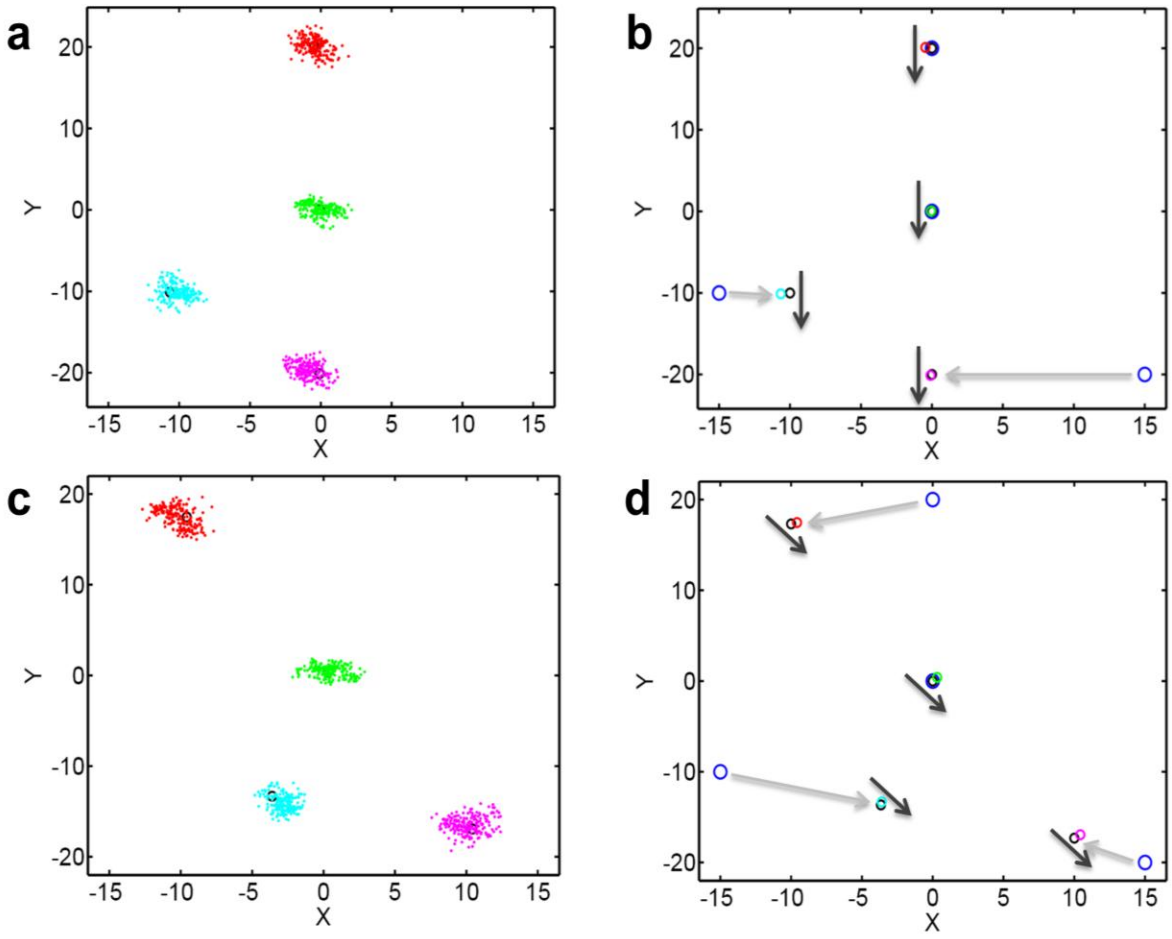


Figure 4.10 Illustration of probabilistic graphical model. (**a**) The local orientation for each of the points is -$\frac{\pi}{2}$. The particle filter sampling for the estimated locations. (**b**) Blue circles are the evidence position and the black circles are the conditional positions. The red, green, cyan and magenta circles are the estimated positions. (**c**) The local orientation for each of the points is -$\frac{\pi}{3}$.

Therefore, we can see that when we have proper relative orientation, the four nodes retain their relative configurations even with different local orientation angles θ_T.

4.6 Iteratively Constrained and Enhanced STORM (ICE-STORM) Imaging

The non-parametric belief propagation (NBP) method provides a framework for estimating the state space values for all the nodes in the PGM of the entire structure. It can even estimate the values for nodes, which does not have prior information, due to missing structure or occlusion or due to noise. Now once since we have the starting information for a set of molecules, we can use the known structural constraints and then apply the NBP to reach a convergence of state space estimates to provide better estimate of the positions, which is our primary objective. Then we can use these refined molecules, include further points in the set, and perform NBP again in an iteratively increasing set of molecules until the model reaches a sufficient. The basic idea for the algorithm is shown by the schematic in Figure 4.11.

4.6.1 ICE-STORM Algorithm

Step 1

a) Start with a random subset of molecules from the original STORM data
b) Calculate local tensor field θ_T for those set of points

Step 2

a) Evidence $\phi_i : [x\ ,\ y,\ \theta_T, \sigma_x, \sigma_y, \sigma_{\theta_T}]$
b) Conditional $\psi_{ij} : [\Delta x, \Delta y, \Delta\theta_T]$
c) Inference using Nonparametric Belief Propagation with particle filters

Step 3

a) Update the evidence and conditional values
b) Impose an increasingly lower uncertainty of position for the points which are already refined from the previous iterations
c) Re-calculate orientation field based on refined position estimates

Step 4

Repeat step 2 to step 3 and stop after n iterations

4.6.2 Steps to parallelize for large super-resolution dataset

a) After n iterations, suppose we have a set of N refined points whose uncertainty in position is assumed to be low ~zero, we mark those N molecules as enhanced set N_{ice}
b) Add another set N points from the original data set and refine them with respect to the N_{ice} points.
c) Perform the task (b) in parallel with a total of $N \times P$ molecules, P being the number of parallel processes.

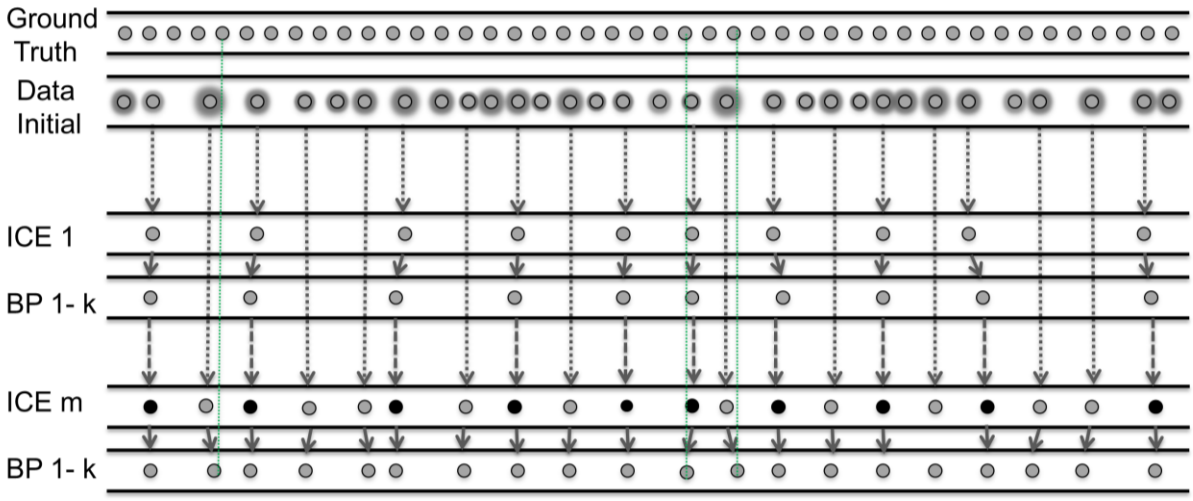


Figure 4.11 Schematic of the ICE-STORM with Belief Propagation. The green dashed line shows the final alignment of the ICE points with the ground truth.

The black circles are improved from the previous iteration whereas the gray ones are from the starting configurations. It is however enough to just start with sufficient molecules in the starting and proceed with the belief propagation iterations.

The starting molecular distance between neighboring points are discretized into multiples of 8nm or 4 nm since in the actual model the molecules are 8nm if aligned linearly or 4 nm apart for molecules in the neighboring filament. If the neighboring filament molecule arrangement in the model are assumed to be helical then there will be a factor multiplied to 4nm for the molecules in the neighboring filament. This neighboring distance discretization is assumed, since the actual density of the molecules is not known; hence, the distance constraint cannot be estimated at arbitrary density.

Here we are using non-parametric belief propagation approach since we are not aware of the probability distributions of our parameters describing the graphical model for the biological structure and hence can accommodate the unknown complexities or underlying patterns of the structures more appropriately.

4.6.3 Nonparametric Belief Propagation with ICE

The objective is to use the graphical model message passing method to update the location of the initial positions represented by red points in Figure 4.13 and the local tensor orientation for all the points using the evidence distribution potential function $\phi_i(\vec{x}_i|S)$. We have used the adjacent position and orientation information for the red dots in the potential function $\psi_{ij}(\vec{u}_i, \vec{u}_j)$ representing the conditional distribution for the state space of the nodes. The starting criteria are that localized points should have stronger beliefs and they should have lower variance and higher weights as we progress through the belief propagation iteration and also the ICE iteration. This means with each ICE iteration we should find the set of points with greater beliefs.

4.6.4 Application to real super resolution data

The idea is to use the localized data for building the model, however if for some reason a significant amount of single molecule for a part or whole of the structure cannot be used or is not available , the we can also use as evidence from the diffracted limited image. This will be then the first level of approximation of the structure.

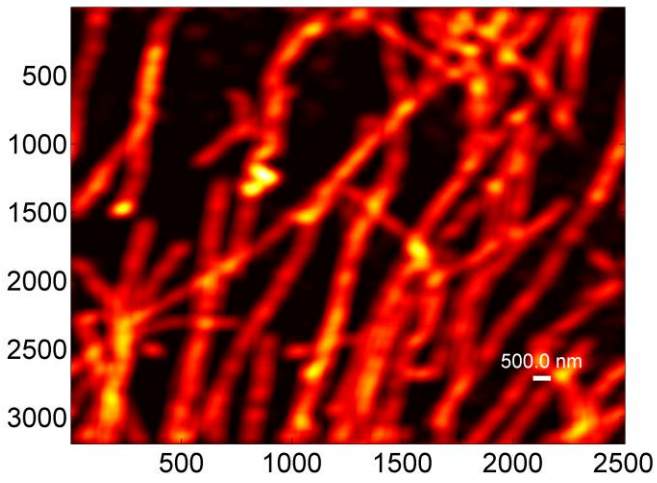


Figure 4.12 Example of a diffraction limited image of Tubulin (generated through Gaussian blurring of the actual super-resolution data)

We can use a subset of the localized single molecule data and generate points that are randomly distributed over the entire structure based on the diffraction-limited image. The Markov random graphical model can be setup using both the data type. If the modeling is performed appropriately, then it will not matter what the data type as long as there are sufficient evidence from the actual localization data and the rest can be fitted using any data type with the correct model constraints. In this case, we have focused on using just the actual localization data. Figure 4.13 shows the position of the starting points in red and the green points, which have been estimated by the PGM.

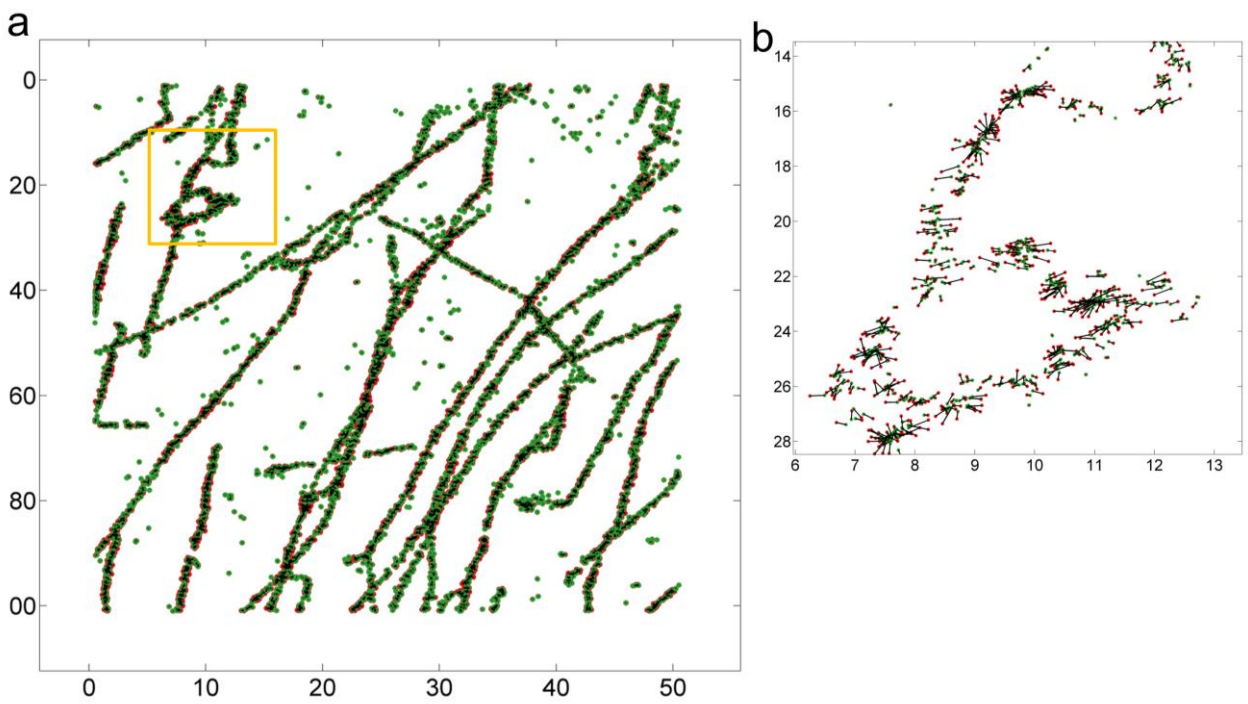


Figure 4.13 Application on real single molecule localization data. (a) Red points are starting positions and green points are after 3 ICE iterations with 2 belief propagation iterations each. (b) Zoomed in view of the displacement of the nodes.

The black lines connecting the red and the green points indicate the magnitude and direction of the displacement. Figure 4.13 shows the improved positions, which effectively increases the spatial density and as a result potentially improves the resolution of the structures. The improved structural features are non-distorted with reduced positional uncertainty and statistically consistent.

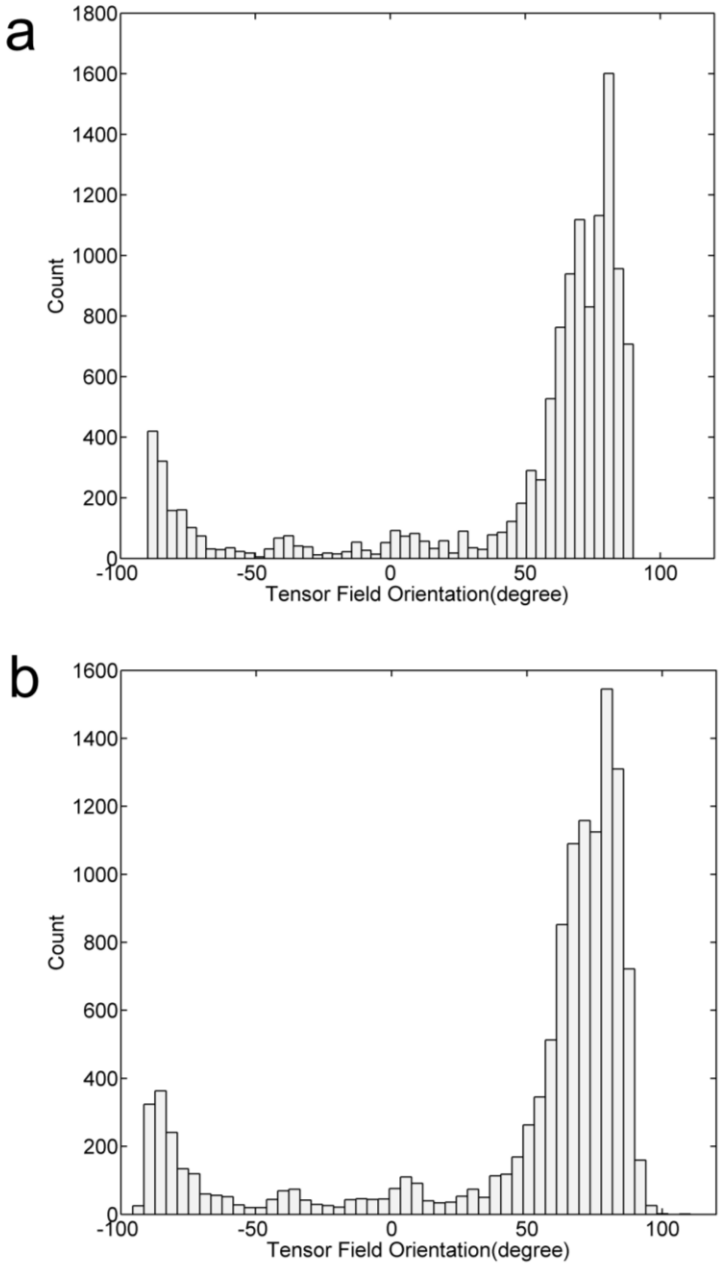


Figure 4.14. Local Orientation Estimation after 3 ICE iterations. (a) Starting orientation estimates for all the molecules. (b) Orientation estimates after 3 ICE iterations with 2 belief propagation iterations each.

The orientation field for the structure does not change very significantly, which is expected since the molecular displacements are small.

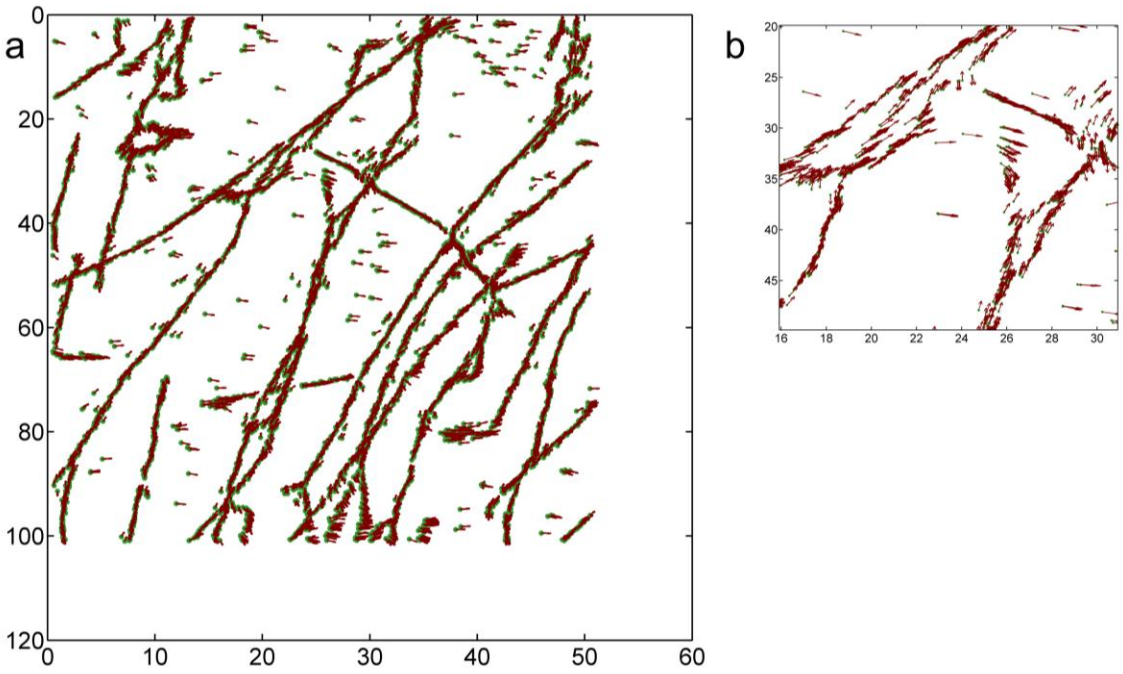


Figure 4.15. Local orientation field for the real data**.** (a) Green points are the final positions and dark red arrows are the local orientation for the points (b) Zoomed in view of a ROI.

Figure 4.15 shows the orientation for the individual molecules after the ICE iterations. The orientation vectors for the points close to each other within filaments are consistent but for points which are sparse are not, since the local orientation requires at least some points to get a proper estimate. Although using a proper prior, this can be solved by using the orientation of the diffraction-limited image, in case where the data points are sparse. Usually with close points, even if the state values (in this case the orientation) are not so consistent

initially, using belief propagation, they can be made more consistent, but since sparse points are far away, they may not be even neighbors and hence no message passing occurs between distant points. Probably separate criteria have to be developed for such points.

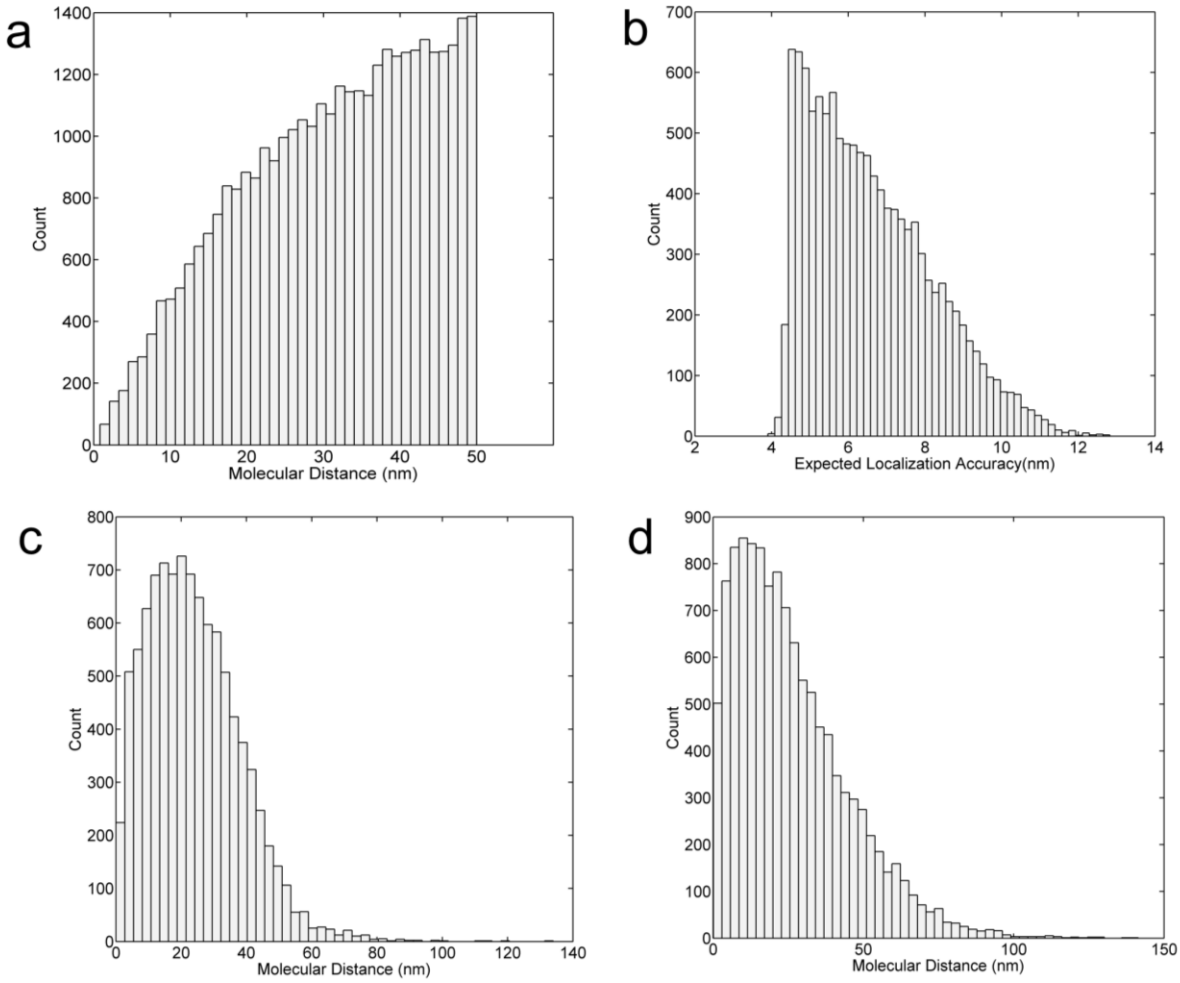


Figure 4.16 Displacement of the molecules after 3 ICE iterations and 2 BP iterations each. (a) Starting pairwise intermolecular distance of neighboring molecules. (b) Expected localization accuracy of the molecules obtained from the individual photon counts. (c) The inter-molecular distance between the neighboring molecules after ICE iteration. (d) Displacement of the individual nodes after 3 ICE iterations with 2 belief propagation iterations each.

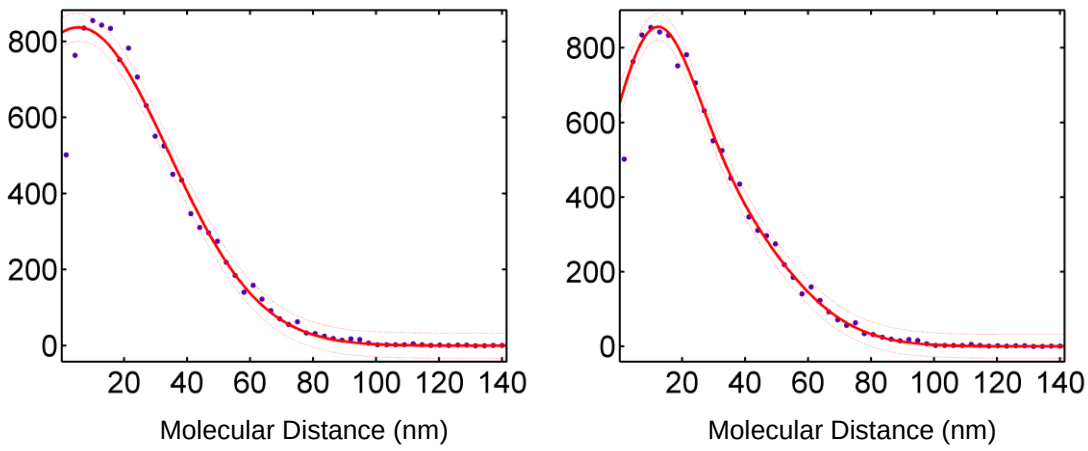


Figure 4.17 Displacement distributions (same as in Figure 4.16) of the molecules after 3 ICE iterations and 2 BP iterations each. The histogram is fitted with single peak (left) and double peak Gaussian (right) for finding the population mean of the displacements.

The sample mean localization uncertainty in Figure 4.16b is 6.6 nm for the used data points whereas the sample mean and median displacements of the nodes after 3 ice iterations as shown in Figure 4.16d is about 26 nm and 22 nm respectively, which is within the 4 standard deviations of the localization uncertainty. The population mean of the displacement is around 5 – 10 nm as calculated in Figure 4.17.

There is an improvement of the molecular position with or without the enforcement of linearization constraint on the orientation. It is slightly more for if the linearization is enforced. The molecules are re-positioned based on the distance and orientation constraints. This improvement is presently judged just on the molecular displacement relative to its starting position and we assuming

the model constraints are being enforced appropriately, which we have verified with simple simulated examples in Figure 4.10.

4.7 Discussion

In this chapter, we have introduced a nonparametric framework to model biological structures using super-resolution localization dataset. We have used a graphical model approach with initial positions as evidence and informative biological constraints such as inter-molecular distance and local orientation as conditionals to generate an improved structure with statistically consistent way. The evidence potential function formulation can have either the prior evidence from the diffraction-limited image from a normal microscope or it can be totally from the super-resolution data depending on the situation and modeling needs. The rules that we want to introduce might be hard to implement due to the nature and quality of the diffraction-limited image and also they might be too simplistic in context of a biological structure to capture the realistic patterns. Again, if there is too much ambiguity in a certain region of the structure and if the single molecule data cannot provide sufficient information then it could be difficult for the generative model to make a correct estimate of the molecular positions in that region. Therefore, the generative model has to be sophisticated enough for it to be able to generate structural models with high degree of information.

Here we have developed a generative probabilistic graphical model with state space consisting of molecular position and orientation (relative to the filament). So, the adjacency potential contains both distance and orientation constraints and we have applied it to real biological datasets. This results in unbiased, but

biologically informed tightening of the data points, which can enhance the efficiency of achieving Nyquist levels of labeling density. There are limitations of antibody labeling for achieving structural biology level labeling densities. The limitations are related to 3-d molecular distributions, which are easily rectified by extending into 3-d, labeling density and repeated sampling, issues that are being addressed experimentally (Lau et al., 2012). These problems and limitations in super-resolution imaging can be handled using a generative modeling approach such as the one we have discussed here. The advantage of using a nonparametric approach such as a PGM is that new information can be incorporated into the state-space without changing the overall modeling approach whereas if we use a parametric approach, we have to re-formulate the whole computational framework with new parametric information.

The only issue right now is the number of molecules to incorporate in the model building, as the intermolecular distance and orientation estimate for neighboring molecules is more accurate if they are closer neighbors. With a sparse dataset with the uncertainty of positions close to or more than the distance between the immediate neighboring molecules, it is difficult to model it accurately. Since, the ultimate goal is to use hundreds of thousands of localized points for modeling the structures; we can propose to parallelize the algorithm as described in the methods section 4.6.

CHAPTER 5

CLUSTERING AND MANIFOLD ANALYSIS

In exploratory data analysis of any type, especially dealing with pattern recognition and shape retrieval, it is perhaps incomplete without the discussion of clustering algorithms. Localization microscopy produces unordered spatial data such for reconstructing structures, but can produce ordered temporal data such as for single particle tracking. There are many "pathfinding" analyses (Li et al., 2009b; Maskell et al., 2006; Oh et al., 2009; Rezaii and Tinati, 2011; Sarkka et al., 2007; Smal et al., 2008b; Yu and Medioni, 2009) which work with temporally ordered data very efficiently. These methods are applicable to localization datasets only if they have an order either spatially or temporally. For that, we have to devise strategies to order the localization data for structures and hence we are exploring methods for clustering and manifold learning. Various clustering approaches can provide initial knowledge about the distribution of feature points. These methods can be applied as separate methods or can be used as a complementary method for the purpose of structure recovery. Here we are only discussing some of the clustering and manifold analysis methods, which have the property of providing a sense of a spatial ordering of the points along with the cluster. This may provide us a unified framework for tracking of single molecules and for shape retrieval from super-resolution data, further discussed in chapter 6.

5.1 Clustering

Here we are going to mention some of the clustering approaches relevant to our purpose.

5.1.1 Density Based Clustering

Clustering of points in $\mathbb{R}^2$ and $\mathbb{R}^3$ provides a meaningful information of the distribution and pattern in the data, and thus can provide a sense of the underlying structures even if no further analysis is performed. There are various methods such DBSCAN (Density Based Spatial Clustering of Application to Noise)(Ester, 1996; Sander et al., 1998) and OPTICS (Ordering Points To Identify the Clustering Structure)(Ankerst, 1999) ,LSDBC(Locally Scaled Density Based Clustering) (Bicici and Yuret, 2007) and of course the very popular k-NN (k-nearest neighbor) and k-means , although kNN or k-means do not provide spatial ordering as such.

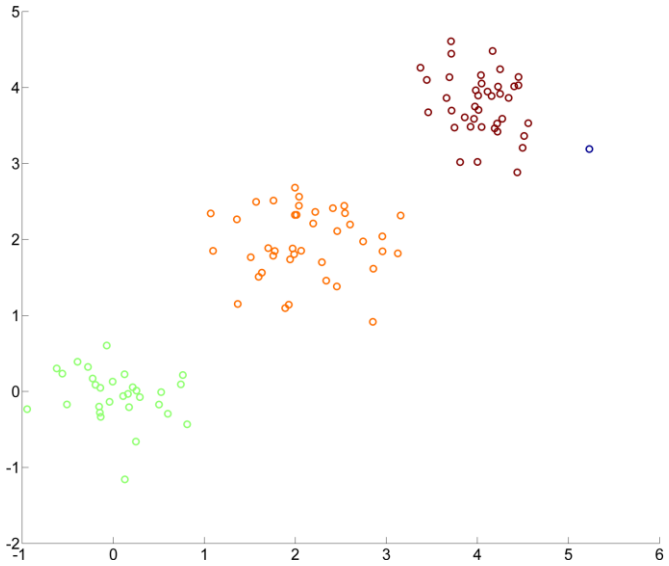


Figure 5.1 Density Based clustering using DBSCAN showing 3 clusters.

DBSCAN works on the principle of density reachability using a ϵ-neighborhood and minimum points to form a cluster as the parameters. It finds the clusters from the estimated density distribution of the points of the arbitrary shaped clusters. OPTICS is a generalized version of DBSCAN and it works with a varying search radius and as such can be used to find clusters with variable density unlike DBSCAN.

Some of these clustering methods such as DBSCAN and OPTICS, can obtain clusters from non-linearly separable clusters which are not separable by simple methods such k-means or even sophisticated clustering such as EM clustering. All these methods have their own advantage and disadvantages, so the proper choice of a clustering method will depend on the type of data and the analysis requirements. As a side note, there are other methods such as graph clustering and support vector machines (SVM) with say radial basis function kernel, which can cluster non-linearly separable clusters. There are numerous other clustering methods from the branch of machine learning, which one can use, but we are discussing here only few of them, which have the property of spatially ordering the points along with finding the clusters.

5.2 Graph Based Clustering

Here we will discuss some of the clustering approaches, which utilize the property of the graph structures of the clusters.

5.2.1 Spectral Clustering

This is a class of unsupervised clustering techniques (Shi and Malik, 2000) where the idea is to perform the spectral (eigenvalue) analysis of the graph. Based on the same principle there is a method for structure preservation and spatial ordering of points (graph embedding later in this chapter).

5.2.2 Similarity Graphs

Similarity graphs (Knappe et al., 2003) are based on local neighborhood relationships either pairwise similarities or pairwise distance between the data points$\{x_1, \dots, x_n\}$.

1. The ε-neighborhood graph: We generate the neighborhood graph with points whose pairwise distances or dissimilarities are smaller than ε. These kind of graphs are undirected and unweighted.

2. k-nearest neighbor graph: Here we generate the neighborhood graph by connecting vertex v_i with vertex v_j if v_j is among the k-nearest neighbors of v_i. This connection makes the graph asymmetric and hence directed but we can have an undirected graph by ignoring the directions, which is called as k-nearest neighbor graph. Otherwise, we can connect vertex v_i with vertexv_j, if

v_i is among the k-nearest neighbors of v_j which the mutual k-nearest neighbor graph.

3. The fully connected graph: Here we generate the neighborhood graph by connecting all points with pairwise positive similarity.

The spectral theory and similarity graph concept is used later in this chapter on manifold learning in section 5.4.

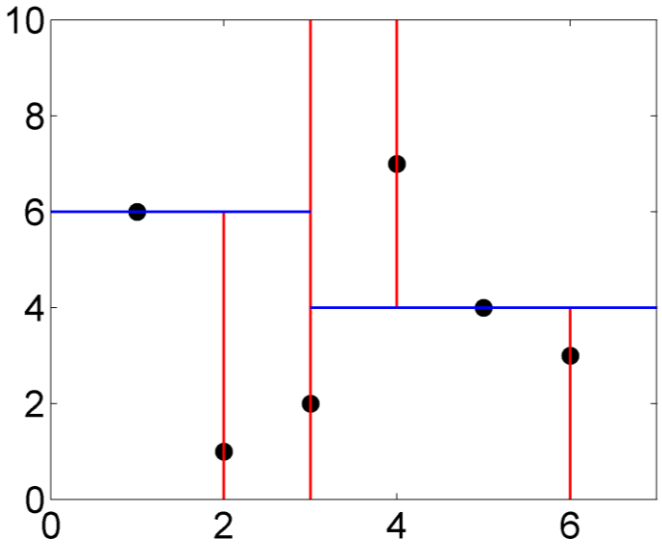
5.3 Computational Geometry

In order to provide an order to the structure data to make it look like trajectories, we have employed some methods from computational geometry, computational topology and graph theory. We have applied a method called Alpha Shapes. It performs clustering on 2D point clouds, using Delaunay tessellation although it is applicable to 3D point clouds too. An alpha shape has been previously applied to protein structures (Albou et al., 2009; Edelsbrunner, 1995; Liang et al., 1998a; Liang et al., 1998b; Liang et al., 1996) This particular approach can provide us with a fairly informative approximation of the overall structure and can be useful on its own without any further analysis. It may also allow us to perform the ordering on individual patches since we have that information.

5.3.1 kd-tree

This is a generalized version of binary tree data structure (Bentley, 1975). It can cluster the data points based on their distance in a hierarchical manner from the connectivity graph. Every node of a k-d tree is a k-dimensional point. Every node acts as a separation plane for the space and it can split the point sets alternatingly by x-coordinate and y-coordinate. Figure 5.2A shows a k-d tree decomposition for the black dots that are spatially ordered by x and y coordinates.

A



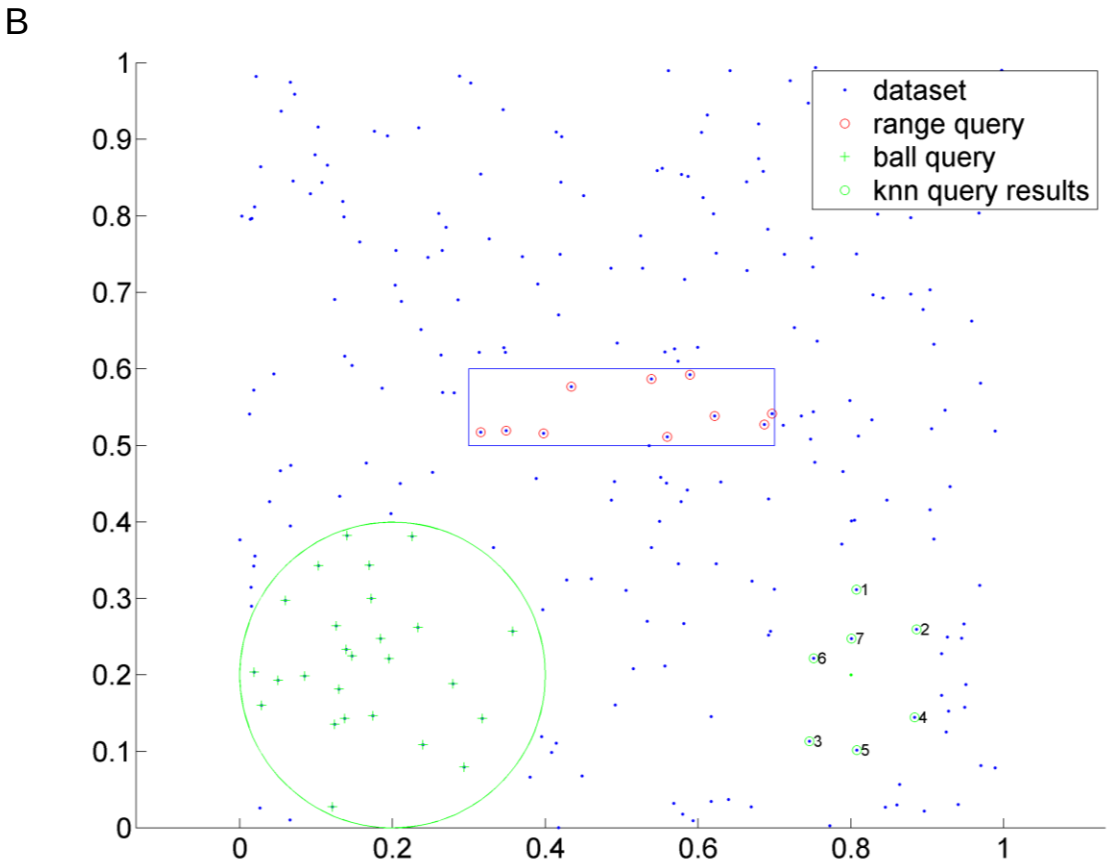


Figure 5.2 (A) Kd-tree decomposition (B) Kd-tree clustering example with different type of search queries

Figure 5.2B shows how all the initial blue points are ordered by k-d tree and then upon different search queries returns the results. K-d tree is also used in kernel density estimation, which we have discussed in chapter 4 in the context of belief propagation in probabilistic graphical model.

5.3.2 Delaunay Triangulation

Given a set of N points P = {x_1, x_2,…, x_N} ∈ R^d, where d = 2 or 3 , if the triangulation DT is such that no point in P lies within the circumcircle for any triangle in DT, then DT is called a Delaunay triangulation. The triangulation can be incrementally modified by addition or removal of points in P. In the case of 2-D triangulations, we can apply constraints on the connecting edges of DT, say based on distances, we can perform topological and other geometric operations such as range queries, nearest neighbor search, Voronoi tessellation and convex hull.

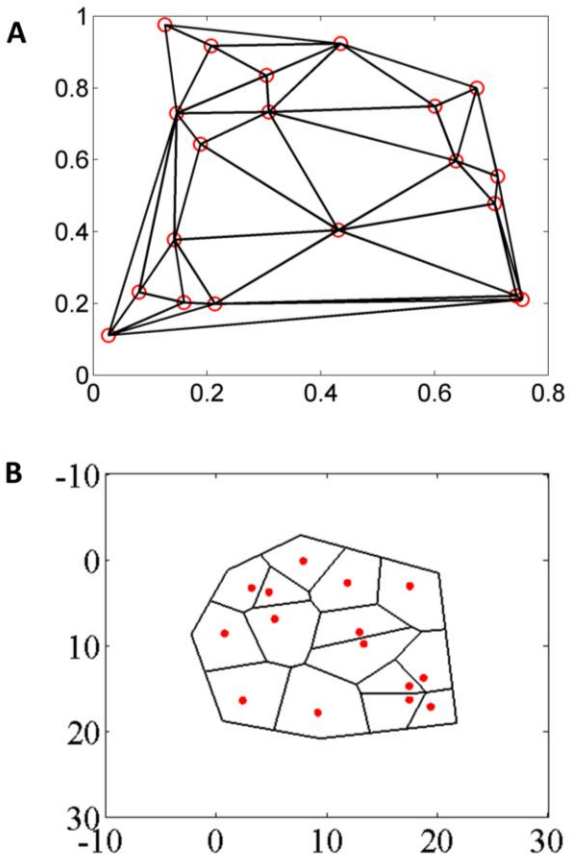


Figure 5.3. (A) Delaunay triangulation of the points shown in red circles (B) Voronoi tessellation (bounded) of the points shown in red dots

5.3.3 Alpha Shaping

Alpha shaping(Edelsbrunner et al., 1983; Edelsbrunner and Mucke, 1994) is a generalization of finding 'hulls' given a set of 2-D and 3-D point cloud, although not every alpha shape is a convex hull. It is a very useful method in computational geometry. Alpha shaping is based on the Delaunay triangulation and can find both convex and concave hull at the same time, so it can determine both the inner and outer edges of the structures.

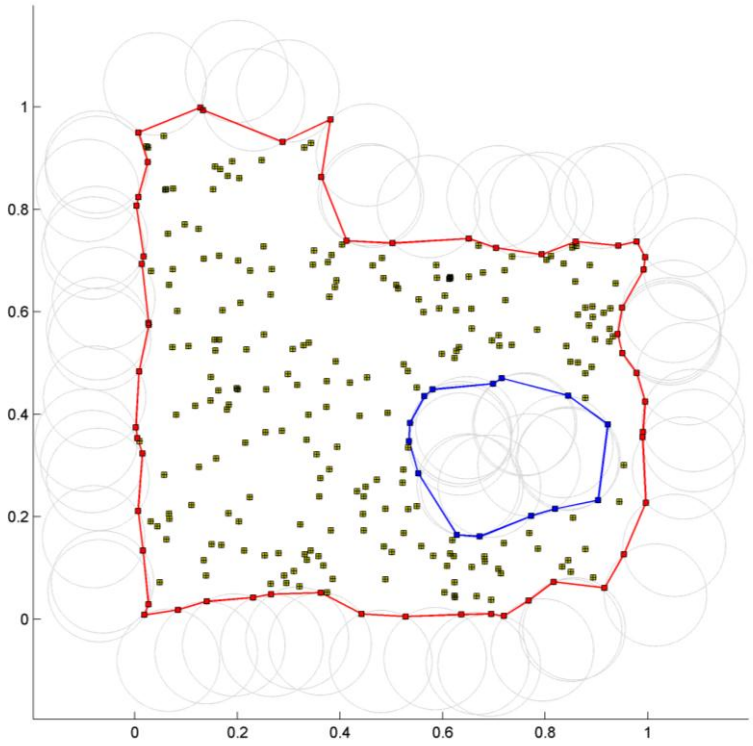


Figure 5.4. Alpha shaping example. Yellow dots are the data points. Red curve is the outer edge and blue curve is the inner edge. Gray circles searches for the boundary for the given set of points.

The alpha shapes are determined based on a parameter value α, which essentially determines how the gray circles in Figure 5.5 searches for the boundary for the given set of points. It controls the desired level of detail. For small $\alpha \to 0$ the alpha shape would degenerate to the original set of points and for large values, the alpha shape would just find the outer boundary since the gray circles cannot find space in between two given points. Hence for $\alpha \to \infty$ the alpha shape is equivalent to finding the convex hull of the set of points.

Below are some examples of the application of alpha shapes on a real single molecule data. We have used the MATLAB toolbox, publicly available at the Mathworks website for alpha shapes (ashape, 2005).

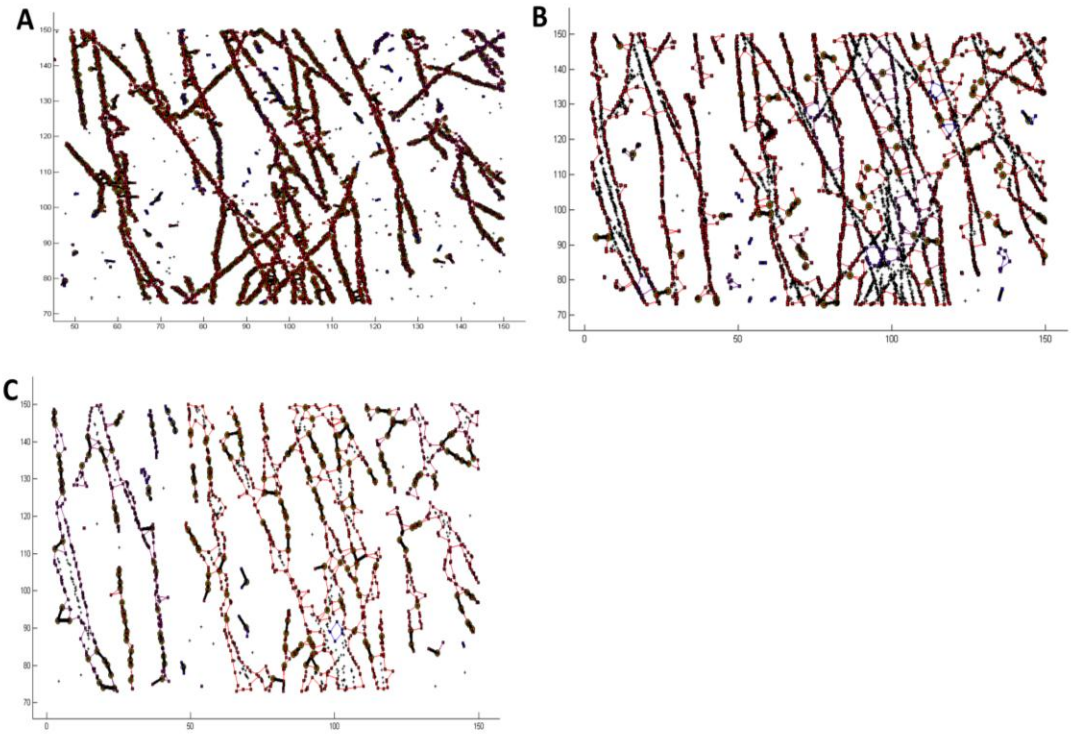


Figure 5.5. (A) Alpha shape at 5% data. r =1 . (B) Alpha shape at 2% data. r = 1. (C) Alpha shape at 1% data, r = 2. r is the search radius.

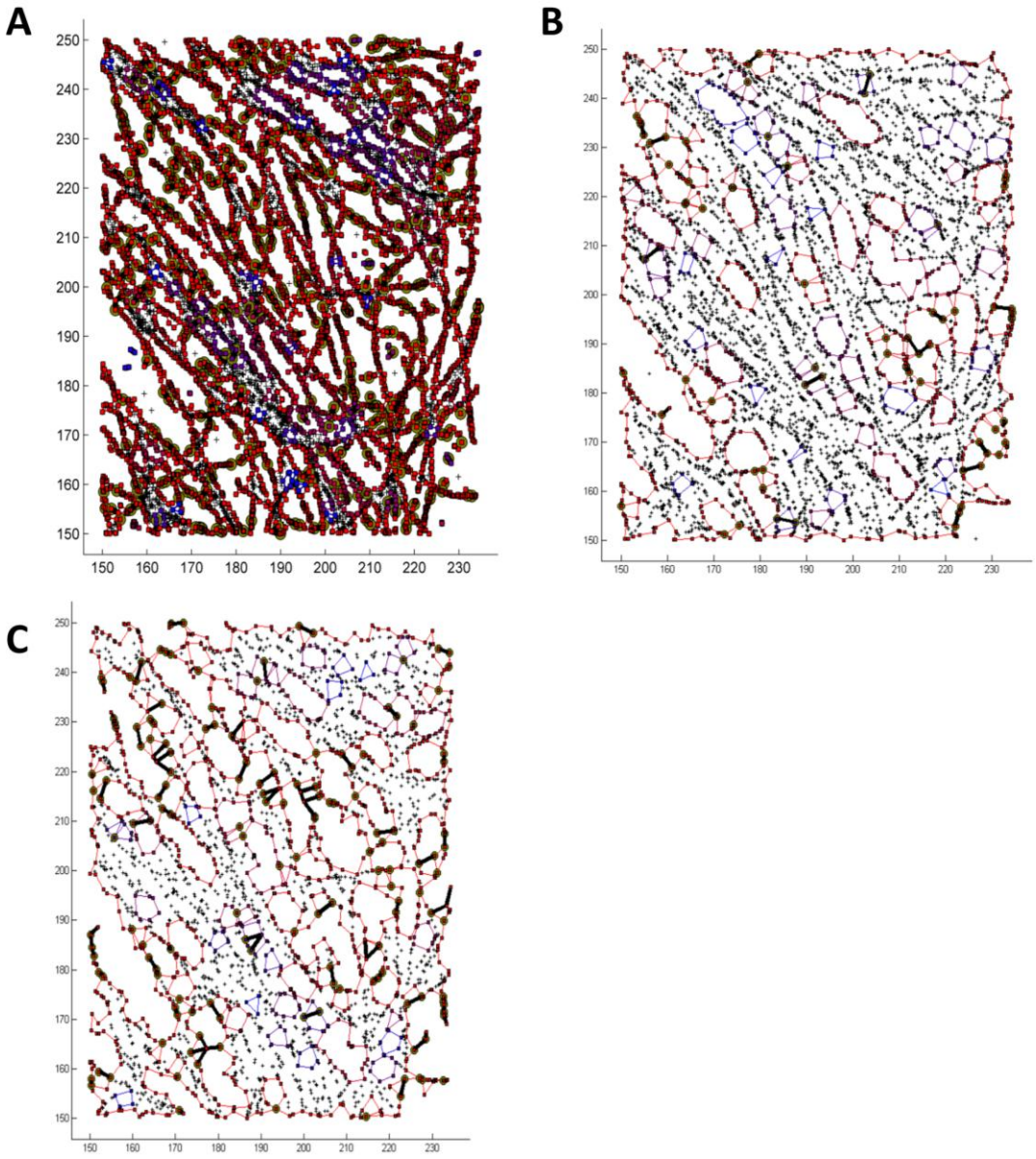


Figure 5.6. (A) Alpha shape at 5% data. r =1. (B) Alpha shape at 2% data. r = 1. (C) Alpha shape at 1% data, r = 2. r is the search radius.

Figure 5.5 and Figure 5.6 shows the alpha shapes for the point datasets in two different region of interests. The first case is region of low density of filaments and the second case is a region of high density of filaments. The alpha shaping works well in the former case even at very low data density, but for the second case due to such close association of points from different filaments the algorithm cannot correctly find the distinct structures and the alpha hull includes most of the structure into one complex. The efficiency of the method will depend on the data density and the search radius r of the circle.

5.4 Manifold Analysis

Manifold Learning (Saul and Roweis, 2004) is an elegant method for learning about the topology of any high dimensional dataset. It has been widely used in machine learning for dimensionality reduction and topology learning. The point dataset can be Non-Riemannian i.e. they need not be smoothly positioned in the manifold. We just have to find the connected clusters and perform the embedding on individual components for preserving the local topology, which will help us to find the local connectivity of the points in a cluster. It can be a very useful property by itself when we are looking at a structure, which comprises of points such as the localization microscopy datasets.

5.4.1 Local Topology Preserving

Given a symmetric graph with edge weight matrix$W = (w_{ij})$, the corresponding graph embedding $\{X_n\}$ is said to preserve the local topology if the following condition holds:

$$(X_l - X_j)^2 \leq (X_p - X_q)^2 \quad \text{if } w_{ij} \geq w_{pq} \forall\ i, j, p, q \tag{5.1}$$

Therefore, the more similar or closer two points in the original metric space are, the more the edge weights will be between them and the closer the points will be in the embedding lower dimensional space.

5.4.2 Graph Embedding

Manifold learning is usually used in machine learning for learning a higher dimensional data pattern by projecting onto a lower dimensional space. Therefore, it is a mapping from $f : \mathbb{R}^n \rightarrow \mathbb{R}^d$ where $d < n$. In our case the data point are positions in 2-dimensional Cartesian coordinates (n= 2) and we want to map it onto 1-dimension (d = 1) and sort the points in order of their magnitudes in the 1-d. This will then provide a sequential order to the points in the 2d space.

Normally manifold learning on graphs are done only on the largest connected component otherwise it is meaningless for dimension reduction, however our objective is different and therefore we want to embed the whole data. Single structure, which can be clustered as one connected component, can be embedded directly otherwise for several connected components, the embedding should be

done separately. The objective of our graph embedding on the structures is to ensure the local connectivity of the point clouds and overall shape of the structures are preserved when we embed onto a lower dimensional space so that if we sort the points on the lower dimension we can traverse the structure in the original space in a sequential manner. We have tried to test several embedding methods such as Locally Linear Embedding (LLE) (Roweis and Saul, 2000) ,Neighborhood preserving embedding (NPE) (Xiaofei et al., 2005) , Hessian Locally Linear Embedding (HLLE) (Donoho and Grimes, 2003) , Laplacian Eigenmaps , etc from the toolbox (van der Maaten, 2012; Wittman, 2005) , some were better than others for different cases. A recent publication (Luo et al., 2011) has proposed a new embedding approach, Cauchy graph embedding which performed significantly better than Laplacian embedding which is generally considered as a local shape preserving embedding. They showed that Laplacian embedding fails to preserve the local shape connectivity in many cases and Cauchy graph embedding should be the appropriate choice. In general, if we choose a proper objective function for mapping then it should be able to produce the local topology or neighborhood preserving feature.

Below are some examples of embedding with different methods showing their working principle using the manifold tutorial (Wittman, 2005) showing how the neighborhood information is preserved in various ways.

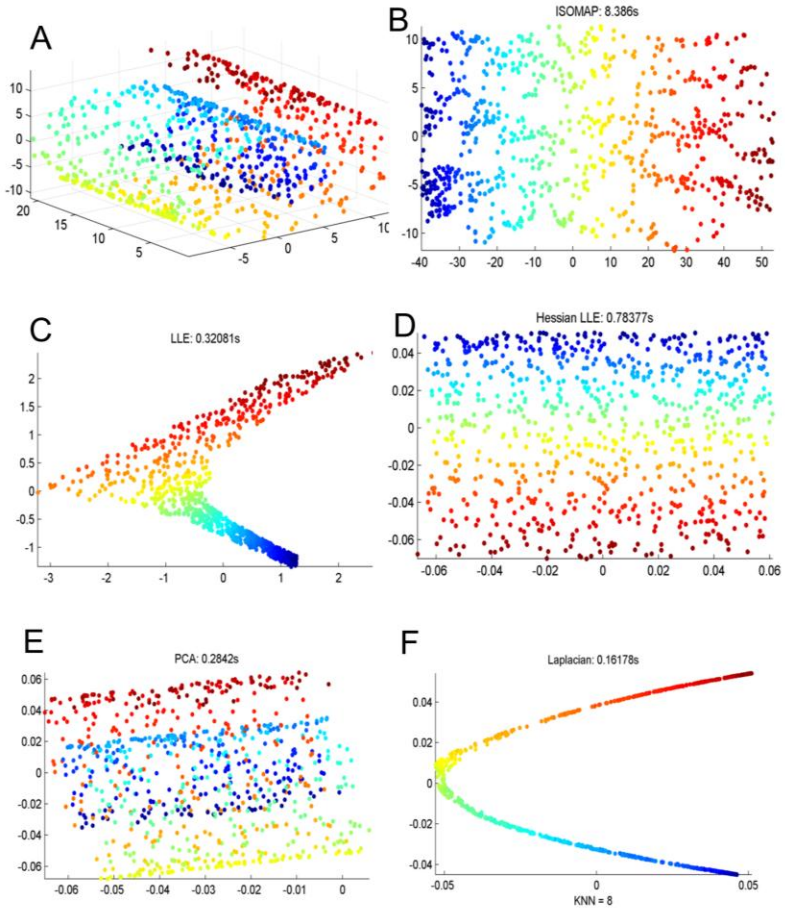


Figure 5.7. The classic example of 3D Swiss roll embedding in 2D using (A) Swiss roll data (B) ISOMAP (C) LLE (D) Hessian LLE (E) PCA (F) Laplacian using the toolbox (Wittman, 2005)

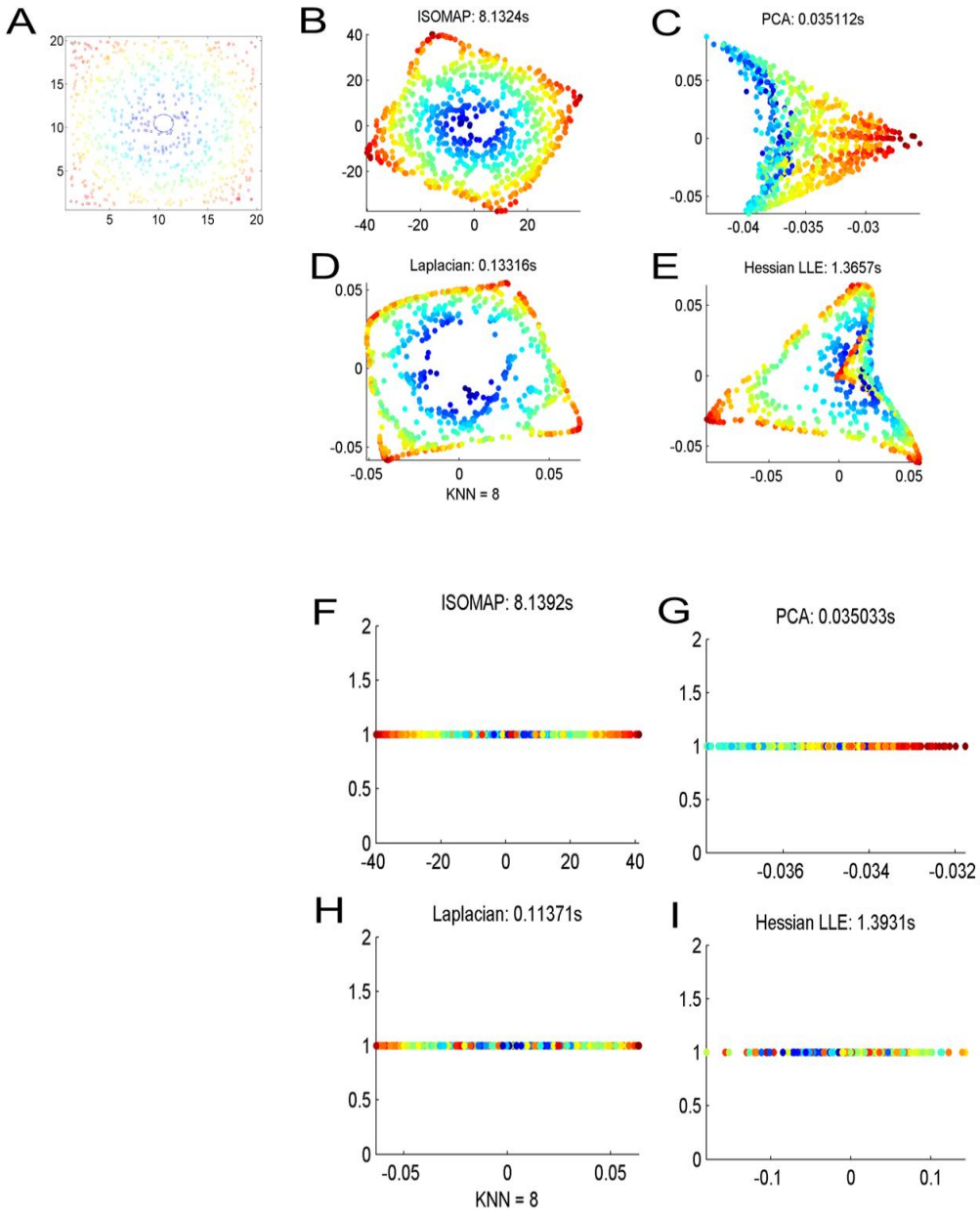


Figure 5.8. An example of 2D manifold embedding of the manifold shown in (A) using (B) ISOMAP (C) PCA (D) Laplacian (E) Hessian LLE for a 2d embedding and same methods in (F)-(I) for a 1d embedding using the toolbox (Wittman, 2005)

The idea of showing these examples is that if are able to get the spatial and structural relationship in lower dimension of the neighboring points of the manifold in higher dimension , then we can learn about the structure more easily.

We have mainly experimented with Laplacian and Cauchy embedding for learning the super-resolution 2d point cloud structure and the details of both methods is provided below.

5.4.2.1 Laplacian Embedding

Laplacian embedding is a widely used method for manifold analysis and has been generally used for topology preserving embedding. Although, Laplacian Eigenmaps may not be the best graph embedding method for neighborhood preservation, we have found reasonable results in many data sets by applying it. The way the Laplacian graph embedding (Belkin and Niyogi, 2003) is performed on the individual components is as follows.

1. Generate a neighborhood graph G using pairwise L2 distance or k-nearest neighbor
2. A weight matrix is computed usually given by

$$W_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{t}\right), t \in \mathbb{R}, if\ i\ and\ j\ are\ connected \\ 0\ , \quad otherwise \end{cases} \tag{5.2}$$

3. Compute eigenvalues and eigenvectors for the generalized eigenvector problem

$$L\,y = \lambda D y \tag{5.3}$$

where D is the diagonal weight matrix with $D_{ii} = \sum_j W_{ij}$ and $L = D - W$ is the Laplacian matrix. L is symmetric and positive semidefinite.

4. Sort the eigenvectors in ascending order of their eigenvalues. If solution of (1) is given by $(y_0, \ldots, y_{k-1})$ then the mapping of a point x_i in the lower dimensional space $\mathbb{R}^d$ is given by $(y_1(i), \ldots, y_d(i))$.

Laplacian embedding is related to the popular spectral clustering. It can preserve the local topology and is robust to noise and outliers in the data, which results in inherent clustering of the data. Therefore, it can be applied to localization datasets consisting of points, resulting in spatial ordering and natural clusters

5.4.2.2 Cauchy Embedding

It has been shown in literature(Luo et al., 2011) that Laplacian Eigenmap are not so efficient in preserving the local topology when the structures are more intricate. The reason is due to the fact that in the case of Laplacian embedding the objective function is minimized while trying to emphasize more on the points which are further away compared to the points which are nearer. This is due to the nature of the objective function, even when a proper weight function such as Gaussian heat kernel is chosen. Therefore, the solution to this drawback is to choose such an objective function, which puts more emphasis on points, which are closer, and tries to minimize the objective function accordingly. The authors

(Luo et al., 2011) suggest that the Cauchy embedding by itself requires a proper initialization and it has been observed that Laplacian embedding results as the initial solution for the gradient algorithm works fairly well. Other possibility is to randomly initialize and choose the best.

Since we were interested in finding methods that can provide spatial ordering of point dataset, in this chapter, we have just introduced some of the methods for clustering and manifold learning with simple examples showing how this approaches can order and provide information about the structures. Manifold learning can be used as a method on its own to learn about biological structures and in general, any object recognition. We will see the application of some these methods on real single molecule super-resolution data and the subsequent application of data association methods in chapter 6 to reconstruct the underlying curves from point data sets.

CHAPTER 6

MONTE CARLO DATA ASSOCIATION FOR SINGLE MOLECULE IMAGES AND PARTICLE TRACKING

Previously we have discussed that generative models can efficiently reconstruct simple biological structures with better spatial and temporal resolution and also provide quantitatively useful biological information using very little input data. Using an implicit Bayesian parametric feature extraction technique called the Hough Transform (HT), simple biological structures such as lines and circles were identified from both simulated and real localization data.

Although, classical HT can be extended to generalized HT and probabilistic HT to reconstruct any parametric or non-parametric objects, I am trying to use a more general unified approach of curve finding. The ultimate goal is to find the generative model that can generate the entire set of molecules given partial single molecule positions and informative constraints (apparently much difficult formulation) using probabilistic generative model such as the probabilistic graphical model (PGM) approach, which we have described in chapter 4. We will however restrict ourselves to improving the molecular positions given localized data.

From the generative model point of view, the data association would provide us a framework for single molecule tracking which has already been shown in (Smal et al., 2008a) There are methods such as (Hill, 2011b), which can possibly determine the structures without any other assumptions. We are trying to experiment if we can use the same framework for single particle tracking can be

applied to interpreting underlying structural features from a set of localized data points.

6.1 Probabilistic Data Association Framework

It has been already shown in literature (Smal et al., 2008b) that Rao Blackwellized Particle Filtering (RBPF) Based Bayesian tracking is more accurate and robust for tracking fluorescent molecules from confocal microscopy images than other probabilistic tracking. Here we apply it to localized super resolution dataset, which are inherently better resolved datasets. Therefore, it should potentially provide significantly better estimates of the trajectories if we are using it for tracking or provide better structural resolution if we are using this concept of data association to trace the underlying structures in single molecule data. Data association problem is generally used for tracking trajectories of multiple targets. It is based on associating trajectory points to the corresponding targets in sequential time instances. Hence, this method is only applicable to super-resolution data of structures, which resembles some kind of arbitrary trajectories. The idea here is to see if the single molecules on a structure can be thought as trajectories of thousands of target molecules. One continuous connected structure will define the trajectory of one target. Points defining the structure will be multiple estimates (points close to the point in consideration) for the same position (uncertainties) and over time (different position) to describe the trajectory. Since the data association works for sequential information, the position information from a super-resolution data, which is generally arranged in a random order, needs to be ordered first as a pre-processing step. If we are using data association to actual single molecule tracking data then this can be

used directly (Sarkka et al., 2007). It would be desirable to utilize a unified approach for trajectory and feature analysis in localization datasets.

6.2 Spatial ordering of points for data association

One of the important steps in this analysis is the ordering of position data based on the inter point distances. Since data-association using particle filters works for dynamic sequential trajectory points and we are interested in using it on position information, which are randomly distributed over the structure, the position needs to be ordered in some manner so that RBPF can perform data association. Although, for linear structures, the data can be simply sorted according to x or y coordinated values, it becomes tricky for arbitrary curvilinear and closed loop structures. We have attempted several different techniques including k-nearest neighbor clustering, spectral graph clustering, k-d tree, Delaunay triangulation, alpha shapes and manifold learning and graph embedding and only found satisfactory results using a combination of alpha shape, Delaunay triangulation to find connected components and graph embedding. The details are described in chapter 5.

We are going to first review the basic theory of applying the RBPF to data association problems. The method is called Rao-Blackwellized Monte Carlo Data Association (RBMCDA) as discussed in (Sarkka et al., 2007).

6.3 Rao-Blackwellized Particle Filter

Data association problem are generally studied in the field of computer vision, machine learning and Robotics such as radar tracking, sensor networks, robot localization and mapping etc. The task is to determine the correct correspondence between sensor measurements and the trajectories. RBMCDA is a relatively new method of solving this problem (Oh et al., 2009; Sarkka et al., 2007). The problem of data association can be formulated as described in (Sarkka et al., 2007) as below.

The state space of the target is given by:

$$X_k = (x_k \;\; y_k \;\; \dot{x}_k \;\; \dot{y}_k)^T \tag{6.1}$$

where (x_k , y_k) is the position and the velocity is $(\dot{x}_k , \dot{y}_k)$. The state variable can be approximated as:

$$X_k \sim p(X_k \,|X_{k-1}) \tag{6.2}$$

which is a stochastic dynamic model and is described below.

6.3.1 Dynamic model for static single molecule structures

If we consider the problem of tracking multiple single particles then RBMCDA or any data association algorithm can be directly applied to the time series position data to determine the particle trajectories. Since, here we have position data on a certain structure; we are trying to transform the structure to some kind of a pseudo time series for unknown number of single particles moving along the

structures. Suppose $(x_{j,k},\ y_{j,k})$ is the position and $(\dot{x}_{j,k},\ \dot{y}_{j,k})$ is the velocity for the jth particle and k[th] time instance , then the dynamic model can be implemented using a discretized Wiener velocity model (Sarkka et al., 2007).

$$\begin{pmatrix} x_{j,k} \\ y_{j,k} \\ \dot{x}_{j,k} \\ \dot{y}_{j,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{j,k-1} \\ y_{j,k-1} \\ \dot{x}_{j,k-1} \\ \dot{y}_{j,k-1} \end{pmatrix} + \boldsymbol{q}_{k-1}$$

(6.3)

where the derivative is w.r.t discretization steps dt along the length of the trajectory (which is actually the structure with some sense of directionality)

$$\dot{x} = \frac{dx}{dt}$$

(6.4)

$\boldsymbol{q}_{k-1}$ is the process noise which can be thought of as the inherent noise in the model. Here the time t is just the step in the discretization or sampling period on a structure. The measurement model tells us how the state space values are measured for a particular actual state space value and can be written as:

$$z_{x,k} = x_k + \sigma_{x,k}$$

$$z_{y,k} = y_k + \sigma_{y,k}$$

(6.5)

Here $\sigma_{x,k} \sim N(0, \sigma^2)$ and the sampling period Δt = 0.1, process noise q = 0.1 and the measurement noise or positional uncertainty $\sigma^2 = 0.0$ has been used as suggested in the RBMCDA toolbox (Sarkka et al., 2007). The maximum number of data associations per target structure is set to 1 to improve the performance of tracking in presence of outlier noise.

The outlier noise can be modeled using any probability density, which is independent of the target states $X_k = (x_k \;\; y_k \;\; \dot{x}_k \;\; \dot{y}_k)^T$.

Then the objective is to recursively compute the current state, i.e. calculate the marginal posterior distribution: $p(X_k|Y_{1:k})$ i.e. find the probability estimate of the state X_k given a measurement value Y_k at time step k. The first step is to initialize with prior $p(X_0)$. Then one can proceed to compute the predictive distribution of the state X_k for the time instance k given all previous measurements $Y_{1:k-1}$ by summing the products of the probability distribution $p(X_k|X_{k-1})$, corresponding to the state transition (k-1)-th to k-th step and the probability distribution of previous state $p(X_{k-1}|Y_{1:k-1})$:

$$p(X_k|Y_{1:k-1}) = \int p(X_k|X_{k-1})\, p(X_{k-1}|Y_{1:k-1})\, dX_{k-1} \tag{6.6}$$

Now, once we get the measurement value Y_k at time step k, the posterior distribution of the state X_k is proportional to the product of the measurement likelihood $p(Y_k|X_k)$ and predicted state (equation 6.6) can be calculated using Bayes rule as follows:

$$p(X_k|Y_k) = \frac{1}{N_k} p(Y_k|X_k)\, p(X_k|Y_{1:k-1}) \tag{6.7}$$

where N_k is a normalization constant given by

$$N_k = p(Y_k|Y_{k-1}) = \int p(Y_k|X_k)\, p(X_k|Y_{1:k-1})\, dX_k \tag{6.8}$$

Since we have the posterior with some uncertainties, we can apply a Bayesian smoothing as follows:

$$p(X_{k+1}|Y_{1:k}) = \int p(X_{k+1}|X_k)\, p(X_k|Y_{1:k})\, dX_k$$

$$p(X_k|Y_{1:T}) = \int \left[\frac{p(X_{k+1}|X_k)p(X_{k+1}|Y_{1:T})}{p(X_{k+1}|Y_{1:k})} \right] dX_{k+1}$$

where T > k. (6.9)

The computation of the integrals described above is a time consuming step and it may have analytical solutions only in special cases. This can be approximated numerically using sequential Monte Carlo or particle filtering method described in the section 6.3.2. For certain situations, if it is possible to solve the integral over the state space, then it should be considered and use sampling for the data association part only. This approach is called as Rao-Blackwellization.

6.3.2 Particle Filter

In literature, sequential Monte Carlo sampling (Doucet et al., 2000) is also known as particle filtering. Given observations y_k and the state space sample x_k, the posterior distribution $p(x_k|y_k)$ is represented using a weighted particle sets $\{ x_k^i,\ w_k^i \}_{i=1}^N$ at time step k as:

$$p(x_k|y_k) = \sum_{i=1}^{N} w_k^i \delta_{x_k^i}$$

(6.10)

where δ is the Dirac delta function, w_k^i is the weight for the ith particle x_k^i at time k and N is the number of particles and $\sum_{i=1}^{N} w_k^i = 1$.

Now using the particles x_k^i that was obtained from the previous time step distribution $p(x_k|y_k)$, the next set of samples and weights can be sampled as follows:

$$\{ x_{k+1}^i, w_{k+1}^i \}_{i=1}^N \sim p(x_{k+1}|y_{k+1}) \tag{6.11}$$

Then the expectations for the samples are given by simple averaging:

$$E(x_k|y_k) = \frac{1}{N}\sum_{i=1}^{N} x_k^i$$

One of the advantages of using the particle filter approach is that it does not need to assume any shape or analytical form of the sample distribution, which makes it applicable to different kind of data.

Below is an example of Monte Carlo sampling (Gibbs sampling) which samples (Figure 6.1A) from a 2d Gaussian distribution shown in Figure 6.1B and C. The Gibbs sampling works by drawing from the full conditionals $p(x^i|x^{-i})$ where x^{-i} is all other variables except i. In general, the samples are drawn from a proposal distribution or importance function.

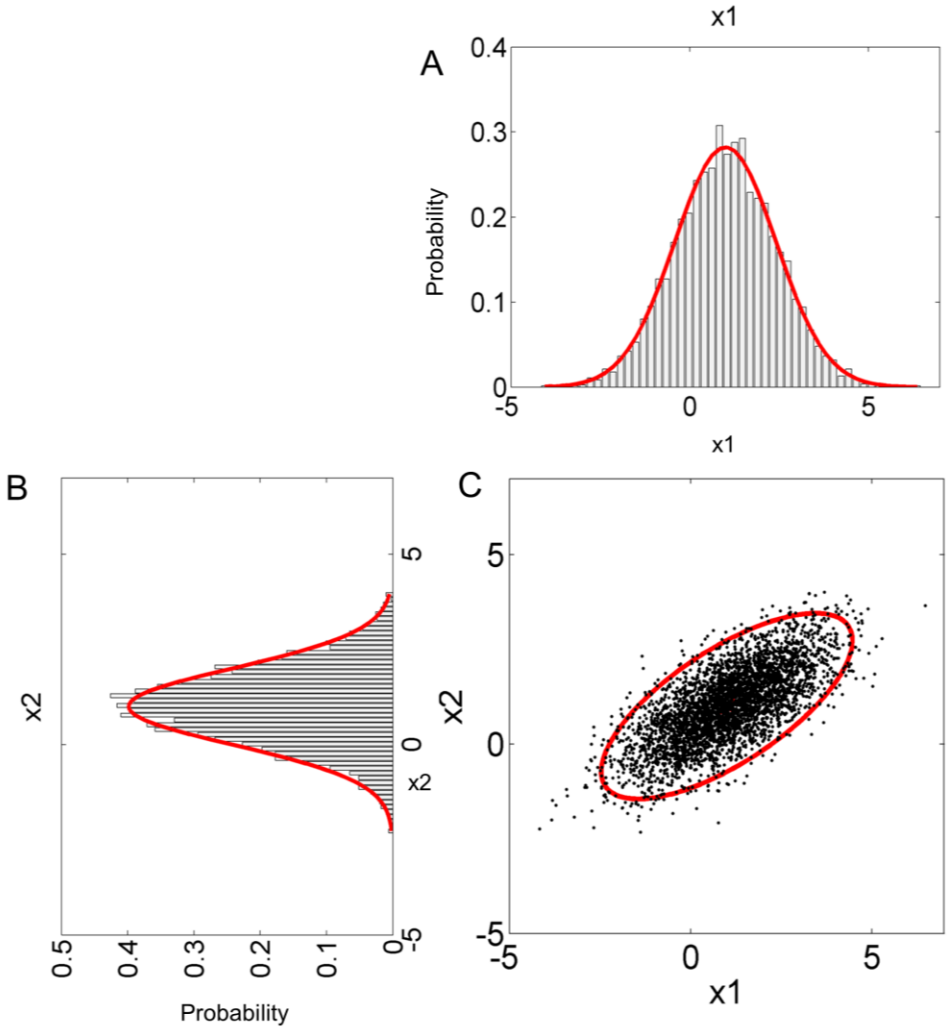


Figure 6.1. 2d Gaussian distribution with Monte Carlo (Gibbs) sampling positions represented by the black dots in (C). (A) and (B) shows the exact (red) and approximate (histogram) 1d marginal distributions in the 2 dimensions.

Although there are several ways to perform particle filtering (Doucet, 2001) , the general algorithm with Sequential Importance Resampling (SIR) is as follows:

1. Samples are drawn

   $x_{k+1}^i \sim p(x_{k+1}|x_k^i)$ (transition prior)

   or the importance function

   $x_{k+1}^i \sim q(x_{k+1}|x_k^i, y_{k+1})$ if high likelihood states are missed by the prior

2. Then they are resampled from

   $x_{k+1}^i \sim Multinomial(\{w_{k+1}^i\}_{i=1}^N)$ (Multinomial resampling)

3. update the weights

   $$w_{k+1}^i = \frac{p(y_{k+1}|x_{k+1}^i)}{\sum_{i=1}^N p(y_{k+1}|x_{k+1}^i)}$$

   $$w_{k+1}^i \propto \frac{p(y_{k+1}|x_{k+1}^i)p(x_{k+1}^i|x_k^i)}{\sum_{i=1}^N p(y_{k+1}|x_{k+1}^i)}$$

   or

   $$w_{k+1}^i \propto \frac{p(y_{k+1}|x_{k+1}^i)p(x_{k+1}^i|x_k^i)}{q(y_{k+1}|x_{k+1}^i)}$$

The sampling method is dependent on the quality of the importance distribution and the one, which minimizes the variances of the importance weights, is

generally preferred. For higher dimension state space, the more particles are needed for sampling.

Since in certain cases the set of particles may get depleted which is known as due to low data points and needs to be replenished.

Figure 6.2 shows a general scheme for particle filtering and resampling strategy as described in (R. van der Merwe, 2000). The particles with higher weights are retained and resampled while particle with very low weights are ignored.

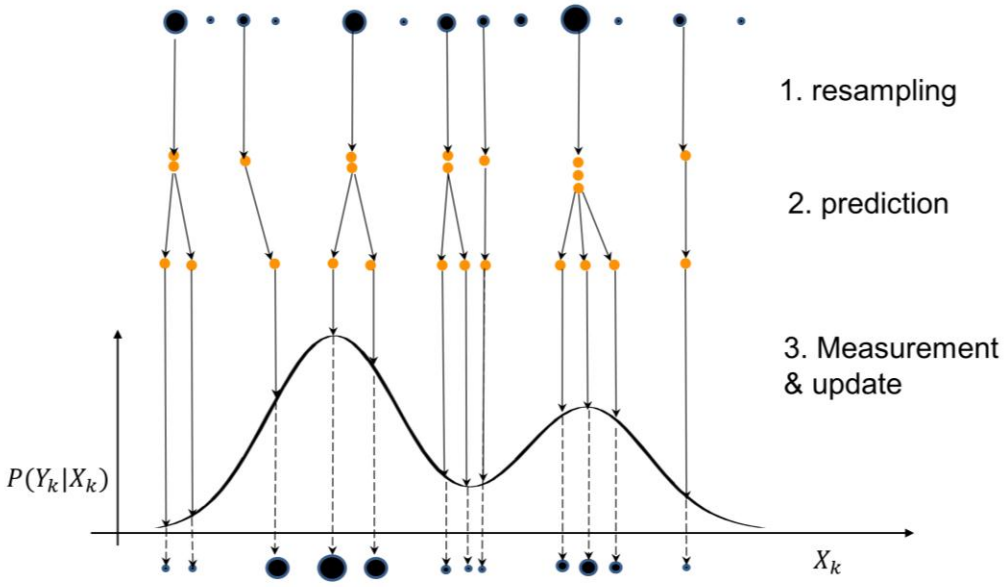


Figure 6.2 Particle Filtering strategy

We are going to revisit particle filtering in the context of non-parametric belief propagation in probabilistic graphical model in chapter 4.

6.4 Results

The data association is applied to the simulated example. First the points are clustered using either Delaunay triangulation or alpha shaping to find the connected components as clearly as possible. It should be noted that the usual graph algorithm for connected component analysis can also be used. We have used other methods for introducing them as other options. Then the points within each cluster or connected component are spatially ordered using another clustering approaches or manifold learning methods including graph embedding.

6.4.1 Simulated data reconstruction

Here we have a simulated structure consisting of points in 2d. The objective is to (1) first cluster the points and then (2) spatially order the points and then (3) apply RBMCDA method to the ordered points to recover the underlying true curves (black) in Figure 6.2A

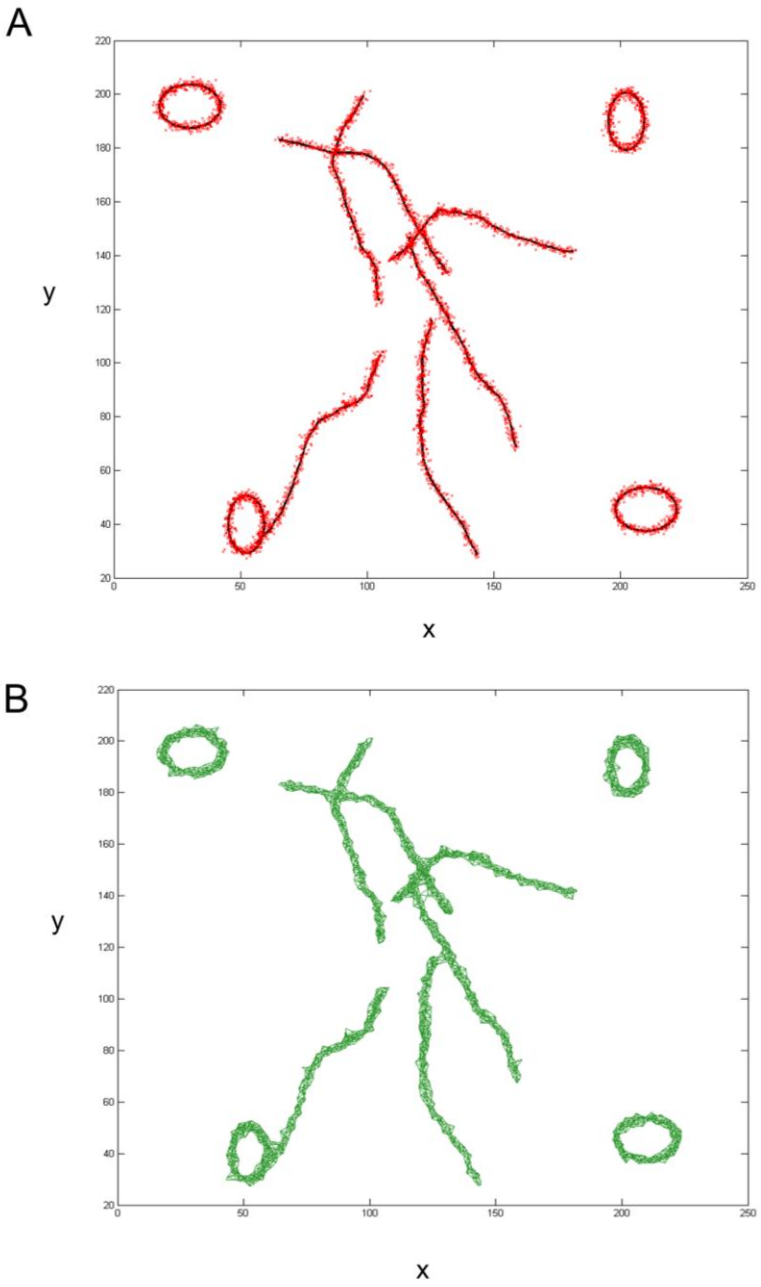


Figure 6.3. (A) Position data is in red. Black curves are the true shapes. (B) Delaunay triangulation of the data shown in 6.3A

Figure 6.3 shows the Delaunay triangulation of the data points shown in Figure 6.2. Once we have the connected components, we proceed to apply the graph embedding on each of the components. The result of embedding is shown for the whole structure in Figure 6.4C. Next, we apply the RBMCDA algorithm on the ordered points and the result is shown in Figure 6.5.

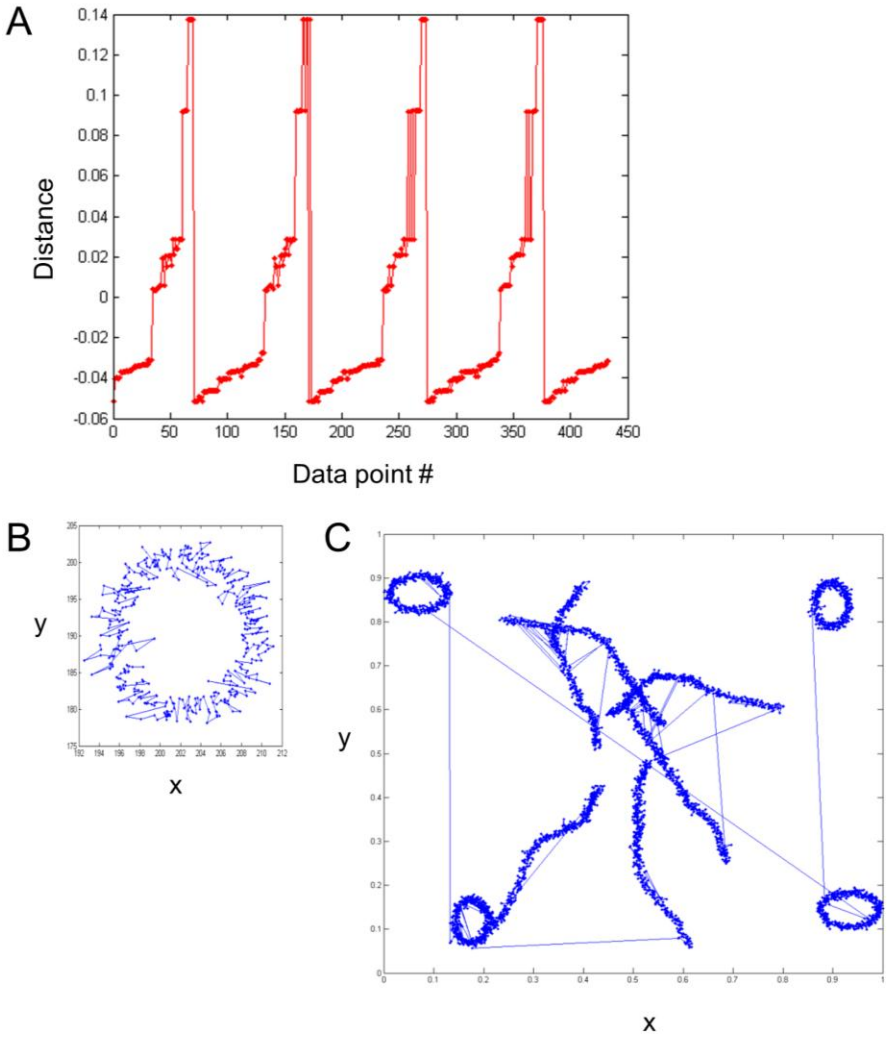


Figure 6.4. (A) Laplacian Eigenmap Embedding in 1d of one of the top right circular structure (Figure 6.3A). (B) sorted points of the circle.(C) entire point cloud in 2D sorted based on the 1-d Embedding.

The particle filtering method of RBMCDA is able to estimate the curves (dark green lines) for the structure quite efficiently, with very small error in data association.

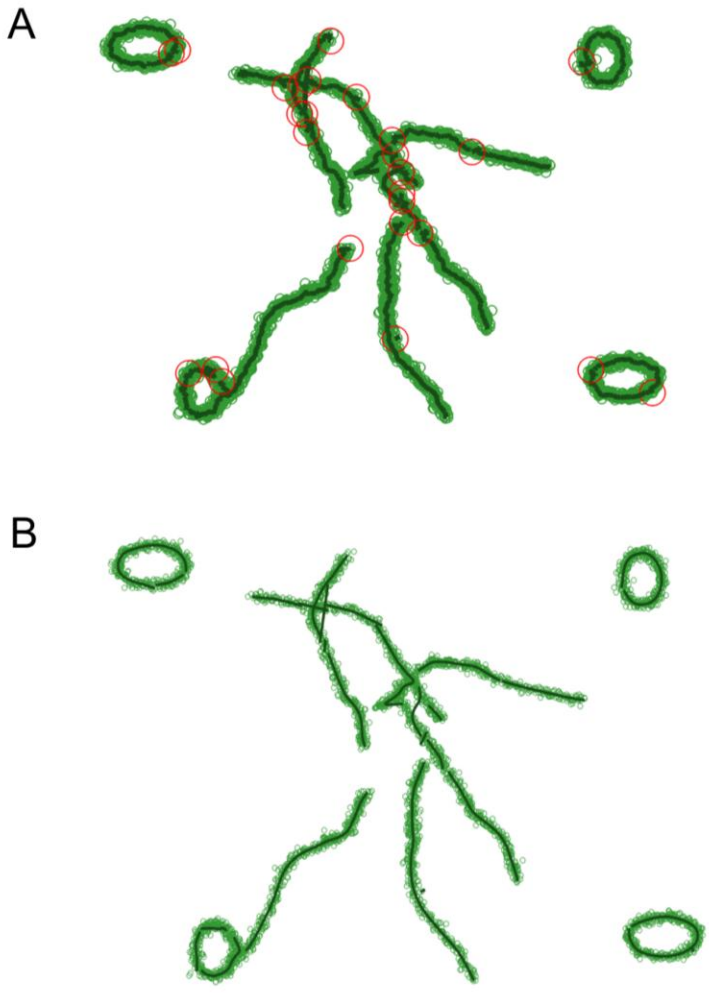


Figure 6.5. (A) Particle filter estimation of the structures. The Estimated smoothed trajectories in bold dark green curves. The red circles indicate the starting point. (B) smoothed using Kalman filter

6.4.2 Application to real data

Reconstruction for real data is performed in the same way. First, find individual components using Delaunay triangulation or alpha shapes. Next, embed the components separately and then sort on the 1-d and then provide an order in 2d space. Next, use particle filter estimate to determine the underlying structure. The major challenge is to provide a proper sequential order when the structure has very densely spaced components.

Figure 6.6. Particle filters estimation of the real data shown with green circles. The estimated smoothed structures are in bold green curves.

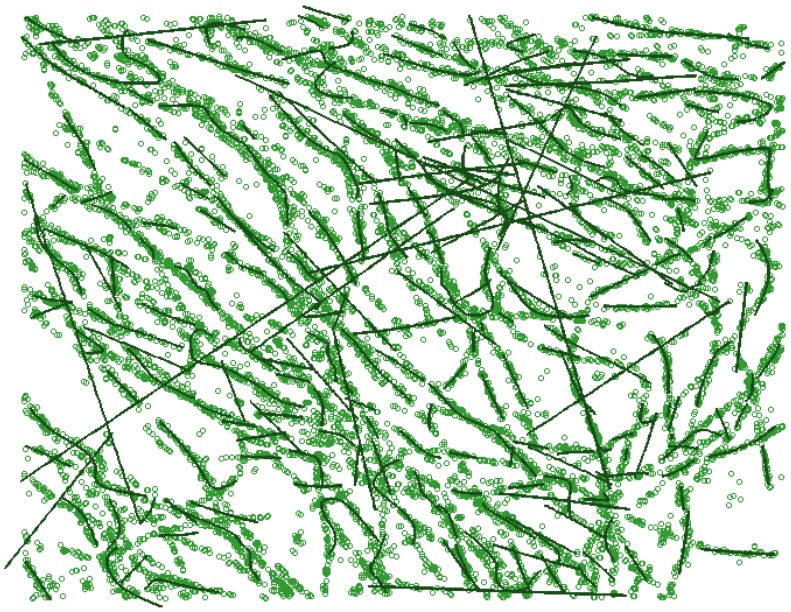


Figure 6.7. Particle filters estimation of the real data shown green circles. The estimated smoothed structures are in dark green curves. The straight lines without having any structural basis are errors in the particle filter estimates for data with high density filaments

For the structure shown in Figure 6.6, RBMCDA was able to find the underlying curves for the most part, whereas for Figure 6.7, the curve finding was not that good. The reason being the sequential ordering of the points was not that good for the latter case due to close structures and ambiguous positions. There are many false data associations due to improper ordering of points. Therefore, the performance of RBMCDA is limited here because of the clustering and embedding method.

6.5 Discussion

The computational complexity of alpha shaping is around O(n log n), whereas for the graph embedding that we are using is ~ $O(n^3)$ which actually limits the data preprocessing. The computational complexity of the RBMCDA method is approximately linear with respect to the number of targets hence for our case it will be proportional to the number of distinct structural segments. When we have more structural segments, the number of particles for particle filtering should be increased to properly sample all the segments, which again will increase the computational complexity proportional to the number of particles.

Here we have used ordering of the points using simple Delaunay triangulation and alpha shaping along with graph embedding for reconstructing both the simulated and real data. With increased structural density, the performance of sequential ordering the data is degraded and hence the performance of data association method suffers too. The ideal method would be the one where we do not require performing the ordering the data separately prior to reconstructing the underlying structures using a generative method. This data association method will work for open filament like structures and tube like structures with or without closed loops (forming 'holes'). However for solid arbitrary shaped structures with no 'holes', it is often informative to just find the outer boundary, in that case we already have several other feature extraction methods available including alpha shapes.

CHAPTER 7

SUMMARY

Recent advancement in super-resolution imaging methods including single molecule analysis has opened up various opportunities to study important problems in biology such as dynamic subcellular processes. Without these methods, it would not have been possible to study the dynamics of complex heterogeneous systems such as living cells at the single molecule level. Even with all these excellent image acquisition capabilities, most of the super-resolution methods cannot reveal faster biological processes or the resolution of the dynamic structures will still be limited.

7.1 Contributions

My dissertation contributes to the field of optical structural biology using super-resolution imaging by developing generative models for reconstruction of biological structures from sparse single molecule super-resolution data. Generating a high-resolution image using localization microscopy requires an extensive sampling of molecular positions, both spatial and temporal, to determine the underlying structure. We hypothesized and demonstrated that generative models could efficiently infer biological structures from single molecule super-resolution datasets, potentially improving dynamic localization imaging of structural components in cells. This initial work was focused on

simple parametric shapes. I have provided a simple demonstration and validated our hypothesis for sparse reconstruction from SR data using a parametric feature extraction method called Hough Transform. To develop a more general approach, I applied biologically relevant informative model constraints to understand the underlying biological structure (tubulin in this case) using a generative probabilistic graphical model. The idea of using such a modeling approach is to be able to incorporate information of different kind into a single framework and with more information. This approach can be related to the integrative approaches for structural biology studies such as integrated modeling platform (IMP) (Yang et al., 2011) and modeling of nuclear pore complex (Alber et al., 2007). The only thing that needs to be modified is the state space in the graphical model and the relevant informative constraints. We have used a non-parametric belief propagation to perform inference on the graphical model. The goal of this approach is to reduce the positional uncertainty and improve upon the structural density at sparse data situation. This can potentially enhance the resolution, which depends on the molecular density as described in chapter 1 equation 1.4. Although right now there is little quantitative information that can be directly extracted out of the PGM method, it can be modeled in a way with additional quantitative steps to obtain other relevant information extracted from the structures such as the features of interest with known shapes or discover newer structural features.

Therefore, the thesis has served two purposes here; one is to provide a computational framework to improve the detailed modeling of biological structures. Secondly, it shows an alternative way to overcome the limitations dynamic imaging of subcellular processes and provide biologically relevant

quantitative information bridging the gap between super-resolution microscopy and structural biology.

In the later chapters, we have also introduced some other methods such as Computational Topology, Computational Geometry such as Delaunay and Voronoi Triangulation, Alpha Shapes, Manifold Learning and some density based clustering approaches, to analyze certain problems in super resolution imaging. We have explored several different methodologies for understanding the underlying biological structures in super-resolution microscopy data and hope that it will help develop better computational methods in the rapidly progressing field of super-resolution imaging of biological structures and processes.

7.1 Future directions

The primary objective of the thesis is to lay a foundation of the application of different computational approaches for reconstruction and improvement of biological structures at sparse single molecule data density and obtain meaningful information in the process. The immediate goal would be to improve upon the methods presented and develop even better and efficient methods. Although, the particular parametric method that we have applied here was chosen as an experimental method, but it can be improved as well, for our purpose:

1. Parameter Optimization

   The choice parameters for generating the parametric curves were chosen manually based on visual accuracy. Therefore, this has room for

improvement using an automatic parameter search using Monte Carlo optimization for the Hough Transform method, or if there is already a parametric curve reconstruction method, which has this ability, then it will be preferable.

2. Computational Complexity

   The time complexity of the parametric methods such as Hough Transform is an issue for 3d parameter space, and also for generalized methods for arbitrary shapes. So, this methods needs to be modified for efficiency or, newer methods needs to be developed.

3. Dynamic Imaging

   We have shown the reconstruction example on static biological structures here, so it would be a necessary next step to perform similar analysis on dynamic super-resolution image data. The reconstruction approach would still be similar.

For the probabilistic graphical model approach, there are certain things we can address:

1. Incorporate the structure reconstruction (the curves) capability within the modeling framework, such as the method shown in (Hill, 2011b)

2. The PGM framework or the data modeling part with PGM needs to be parallelized for handling very large data sets, since currently; we have used only a small set of molecules when we consider the amount of data that is

usually generated from a long time scale super-resolution imaging. We have discussed a way to parallelize, in chapter 4.

3. The modeling framework can be extended to 3-d , with the state space having the 3-d position and 3-d orientation for the molecules, which of course will make the computation more intensive but will be more general and is necessary for keeping up with the challenges in field of structural biology.

The methods that we have discussed in this thesis are mostly experimental and it lays the foundation for future improvements on the algorithms and development of novel computational methods to meet the challenges in the field of optical structural biology.

APPENDIX

Here, we are going to discuss an additional project, which is different from the actual thesis project. We pursued this work on co-translational folding of proteins with a structural bioinformatics approach, since it is a very fundamental and important problem in biology. However, we were able to complete the analysis partially. I am including the work here, hoping that it may be considered for further study in future.

CHAPTER A

SEARCHING FOR SIGNATURES OF ELONGATION RATE ENCODED FOLDING OF PROTEINS

A.1 Introduction

Protein folding mechanism has been the central problem in biology for around half a century. Although it is a well-established fact that protein is synthesized sequentially, the concept of co-translational folding is still controversial (Buchan and Stansfield, 2007; Deane et al., 2007). Recent experimental studies on protein synthesis suggests that mRNA has an additional layer of information along with amino acid sequence, affecting the folding and function of proteins in vivo and could possibly assist in kinetic co-translational protein folding .We are trying to evaluate the hypothesis that "Timing program encoded by the codon usage has a specific impact on the form, folding and function of some newly synthesized proteins."

Elongation of the nascent peptide generally is fastest at codons with highly abundant cognate tRNA and slowest at codons with rare tRNAs. Translational pausing at rare codons might provide the time delay for sequential folding of

nascent proteins. Fast translation may force folding of certain structures (e.g. α helices) within the ribosome exit tunnel.

It has been suggested that the codon usage and charged tRNA abundance affect the ribosomal translation rate, which in turn delays or prompts the emergence of newly synthesized proteins. This translational time variation might be necessary for sequential folding of multidomain nascent proteins, implying that mRNA has an additional layer of information encoded by codon usage that determines the translation dynamics and influences protein function. Structural features that require elongation rate control would show high conservation of the downstream codon selection within an organism. Evolutionarily, proteins that require translational coordination of folding would be expected to show strong correlation of codon selection based on the effective translation rate for a given codon in the respective organism. Clustering of the motifs according to downstream translation rates across proteins and across phyla would reveal conserved translational timing elements. These identified elements would serve as a rich test-bed for investigations of the importance of the translation rate code in directing nascent protein function.

We are developing a single molecule methodology that will allow us to examine the translation of a single mRNA molecule by a fluorescently tagged ribosome, and to correlate these dynamics with the function of newly synthesized, still bound nascent proteins.

A.2 Methods

We want to test if translation rates relate to structural motifs or the correlation is at a domain level and not at a motif level. Relative translational rate (*TR*) is

calculated from the tRNA abundance (A_{tRNA}) or codon usage (U_{codon}) of the protein sequence for each amino acid i as:

$$TR(i) = \left(\frac{1}{w} \sum_{n=i}^{i+w-1} \frac{1}{A(n)} \right)^{-1}$$

$$A = A_{tRNA} \; or \; U_{codon}$$

$$w = sliding \; window \; size$$

(A.1)

tRNA and codon usage data were obtained from tRNA database at (UCSC) .

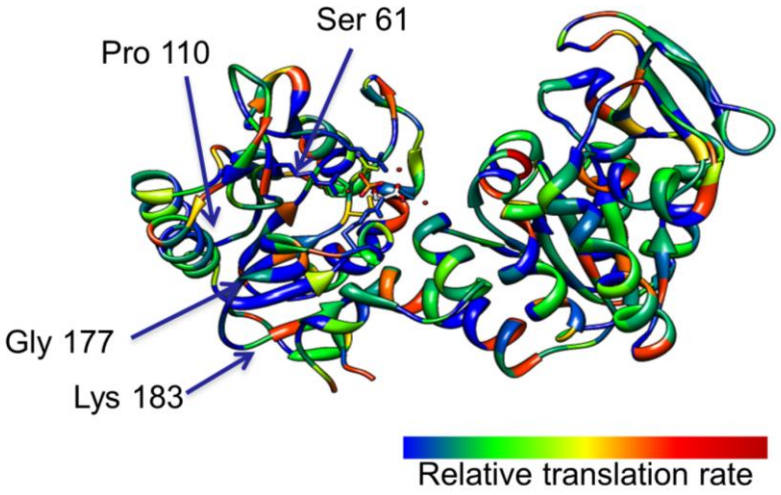
We perform sliding window (w =10) average starting at a motif start and end and continuing till 50 aa downstream to cover the length of the ribosome exit tunnel. We perform the analysis for 32 proteins (unique chains only) across 3 species (mostly human and some mouse and bovine) which shows evidence of co-translational folding (Komar, 2009; Parmley and Huynen, 2009). There are total of 263 α-helices and 316 β-sheets.

A hierarchical clustering analysis over all such profiles from all proteins is performed to detect conserved patterns or signatures of co-translational folding.

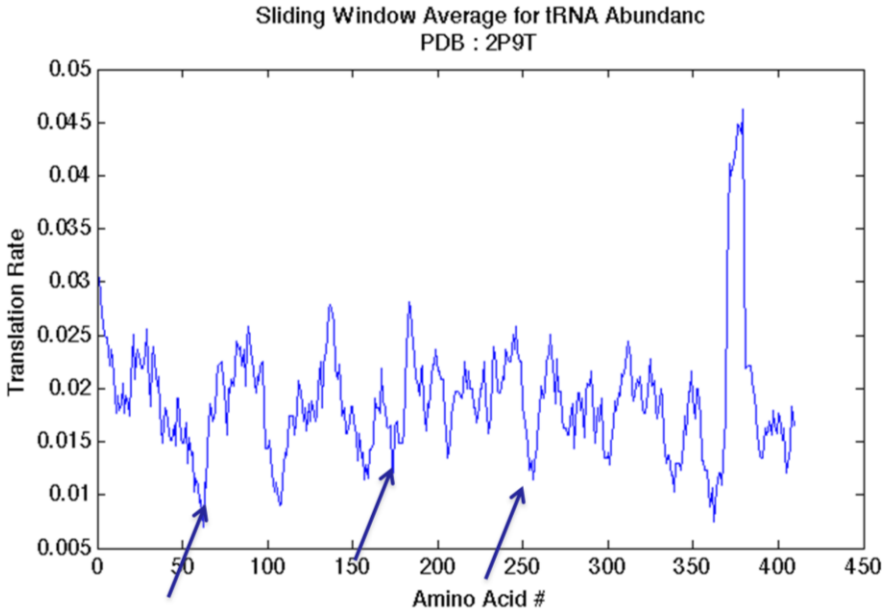
As structural motifs are usually 10-20 amino acids long and ribosomal exit tunnels are around 30 aa, to get the translation rate we have analyzed the sequence up to 50 amino acids downstream. We have assumed here that the number of tRNA genes is a true representation of the tRNA abundance within the cell.

A.3 Results

Previous studies (Komar, 2009) have shown some evidence of correlation of codon usage of proteins to structural motifs. However, tRNA abundance and not codon usage is the more appropriate determinant of translation rate. Here are few example proteins that show evidence of co-translational folding.



A



B

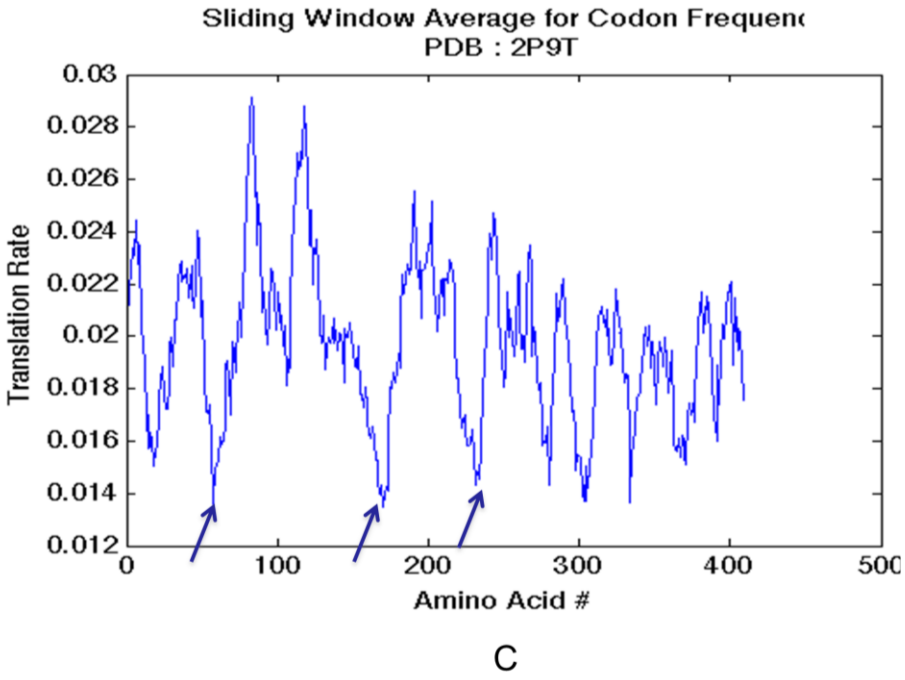
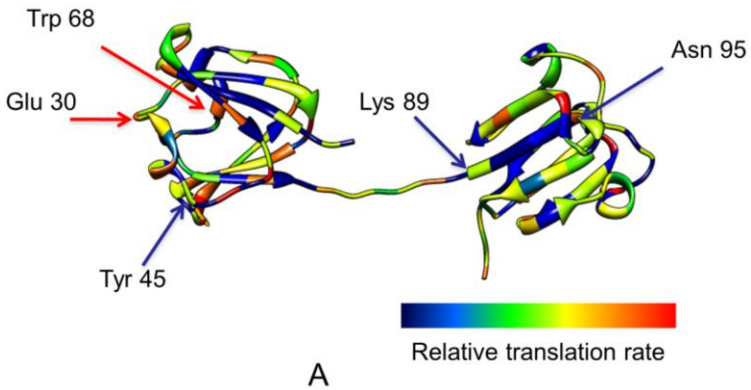
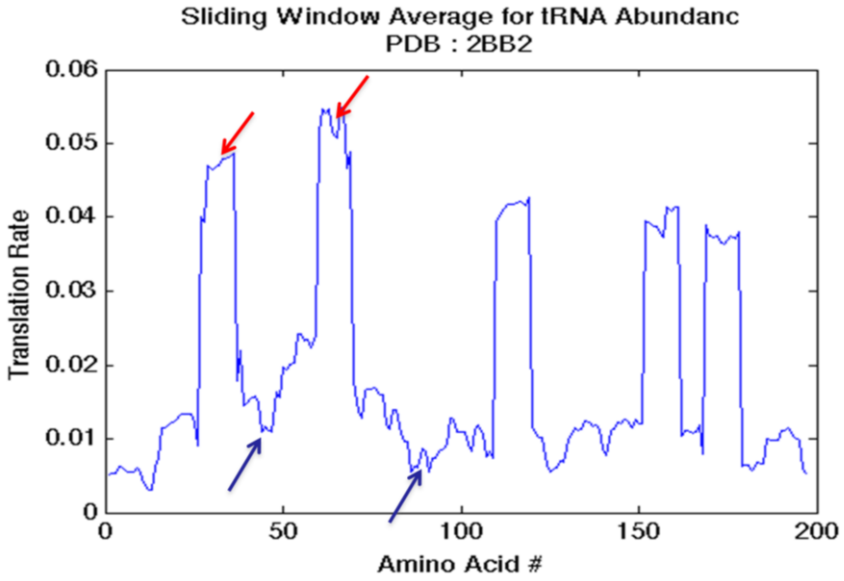


Figure A.1 (a) Mouse phosphoglycerate kinase (PGK) color coded according to the tRNA abundance of the amino acids (b) Sliding window (10) average of the tRNA abundance for the full protein sequence, representing the translation rate. (c) Sliding window average of the codon usage for the full protein sequence


Trp 68
Glu 30
Tyr 45
Lys 89
Asn 95
Relative translation rate

A


Sliding Window Average for tRNA Abundanc
PDB : 2BB2
Translation Rate
0.06
0.05
0.04
0.03
0.02
0.01
0
0
50
100
150
200
Amino Acid #

B

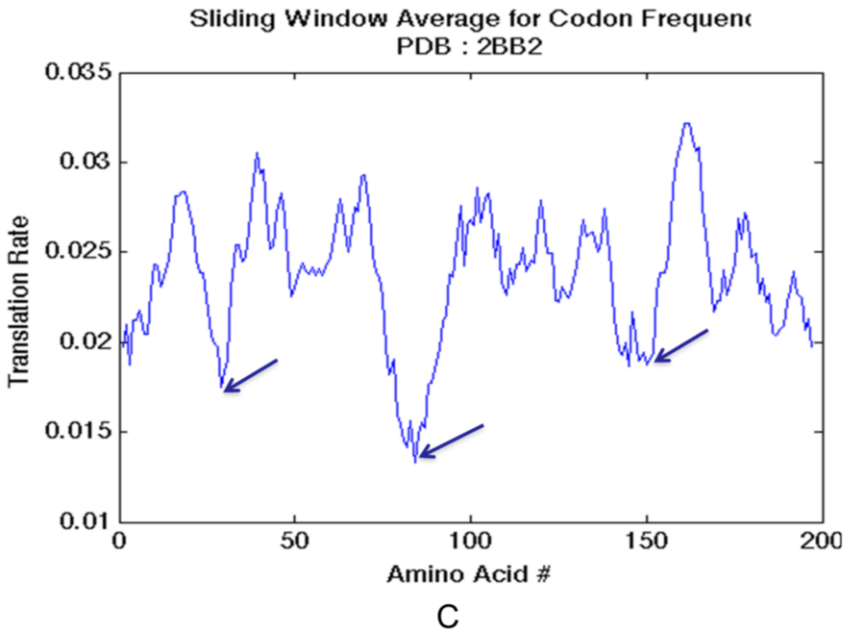


Figure A.2. (a) Bovine β-B2 crystallin (CRYBB2) color coded according to the tRNA abundance of the amino acids (b) Sliding window (10) average of the tRNA abundance for the full protein sequence, representing the translation rate. (c) Sliding window average of the codon usage for the full protein sequence, representing the translation rate.

A.3.1 Clustering Analysis

In order to find any obvious conservation region in the protein sequences, the straightforward approach is to perform clustering of the motif translational rates across several proteins to search for a signature for co-translational folding. Since our starting hypothesis was that tRNA abundance and codon usage for the protein has a direct effect on the ribosomal translation, therefore if we compute the relative tRNA abundance or codon usage for the amino acids for the protein,

we can have a sense for regions of low or high abundance in a particular sequence. This would show up as possible clusters across the protein sequences.

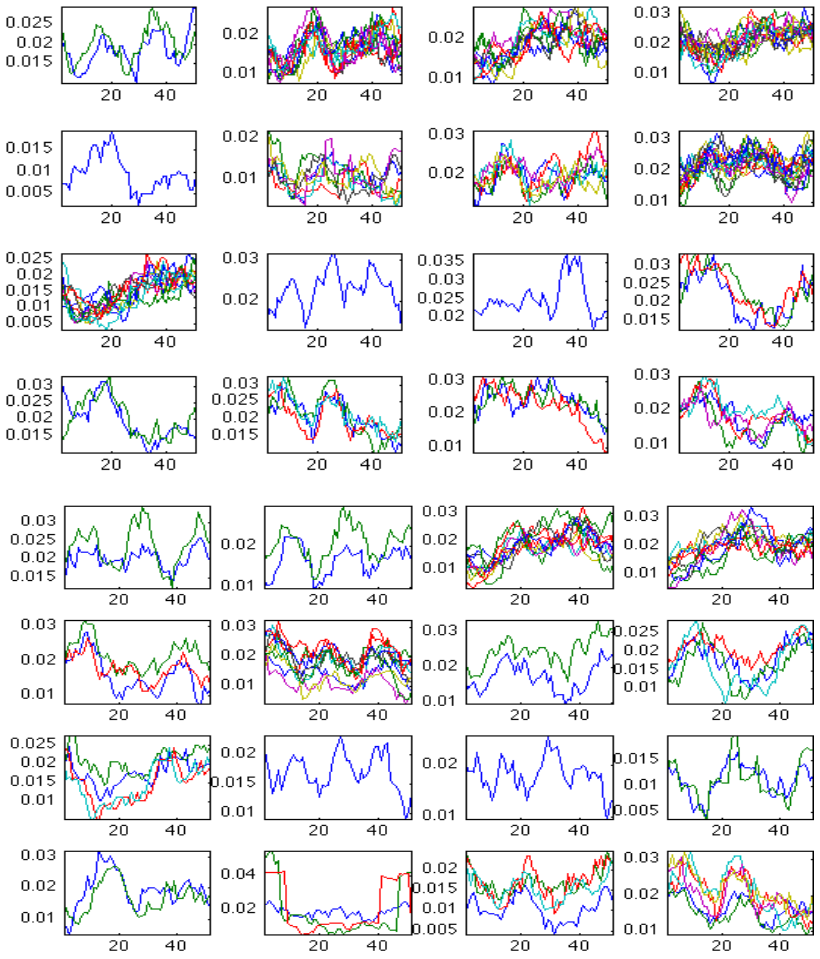


Figure A.3. Clustering analysis (partial results shown) of the Translation rates for all the proteins starting from Alpha Helix and 50 sequence downstream (a) clustering using Euclidean distance as the metric .(b) clustering using correlation as the metric.

The hierarchical clustering using 'Euclidean distance' and 'correlation' clustering return different clusters of the translational profile of the motifs. The analysis of clusters is inconclusive for the intra-species proteins as we did not find any meaningful clusters and the codon usage and tRNA abundance for the amino acids were not very significantly better than random chance at least from the preliminary evidence. We have to perform the same analysis for inter-species proteins to see if we can find an evolutionary conserved region in the proteins.

A.4 Future work

One of the objectives of the translation studies is to verify the functionality of the nascent proteins and test our hypothesis that proteins may acquire its functional form while still attached to the ribosomal subunits asserting the co-translational folding phenomenon. We may propose an experimental technique shown in Figure A6. The underlying idea is to spot the functional nascent proteins within the time scale of folding.

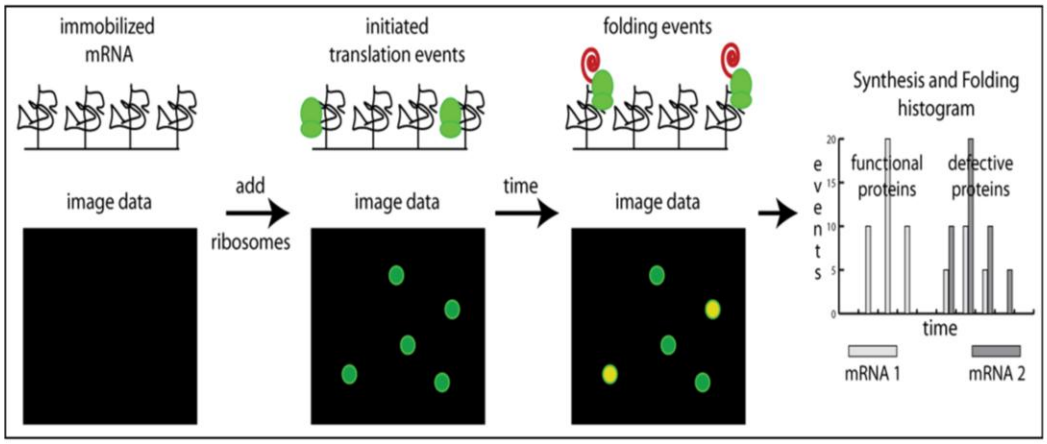


Figure A.4. Example schematics of single molecule imaging study of translation and folding of proteins

The basic steps for the approach are summarized below:

1. The immobilized transcript codes for FAP (fluorogen activating peptide) which binds and activates malachite green dye. The FAP-dye combination can be used as a unique reporter for translation dynamics and nascent chain folding. When the FAP folds on the ribosome, it is able to bind to the dye and activate it signifying completion of translational process and proper folding.

2. The distribution of translation-folding times for the FAP-dye reporter can then be used to measure the effect of codon degeneracy on folding and function of the reporter protein. Synonymous mutations will be made in the transcript to identify changes that alter the folding probability of the reporter.

3. The emergence of GFP signal indicates docking of ribosome on the mRNA whereas malachite green signal indicates proper fold of the nascent chain. Hence, the time difference between appearance of GFP signal and appearance of MG signal establishes the time delay between initiations of translation to maturation of nascent protein. Variation in this time interval will be compared as a function of codon degeneracy, charged tRNA abundance, folding effectors (chaperones) and translation effectors (initiators).

4. The translational rate model allows simulation of expected translation times through Monte Carlo and HMM analysis of the folding trajectories. These predictions can be compared to Single Molecule Detection methods mentioned to determine whether

codon abundance or another factor controls the protein synthesis and folding rate.

A.6 Conclusion

In conclusion, we have developed an approach for investigating the correlation of local translation rate with structural motifs in the search for signatures across library of nearly 40 proteins. We are currently performing the clustering analysis using different methods. The intra-species protein sequence analysis did not reveal any obvious indication of conserved tRNA abundance or codon usage, although we have found evidence in certain individual proteins. To make the analysis complete we have to perform similar analysis on proteins across different species, the idea being that for intra-species proteins the conservation for characteristic features may be more obvious. In addition, if it is possible to develop a practical experimental platform based on single molecule fluorescence to understand translation dynamics and the impact of translation rate or other factors on protein folding.

BIBLIOGRAPHY

Abbe, E. 1873. Beitrage zur Theorie des Mikroskops und der mikroskopischen Wahrmehmung. *Archiv für Mikroskopische Anatomie* 9:412-420.

Aguet, F., D. Van De Ville, and M. Unser. 2005. A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles. *Opt Express*. 13:10503-10522.

Alber, F., S. Dokudovskaya, L.M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B.T. Chait, A. Sali, and M.P. Rout. 2007. The molecular architecture of the nuclear pore complex. *Nature*. 450:695-701.

Albou, L.P., B. Schwarz, O. Poch, J.M. Wurtz, and D. Moras. 2009. Defining and characterizing protein surface using alpha shapes. *Proteins-Structure Function and Bioinformatics*. 76:1-12.

Ankerst, M.B., M.M; Kriegel, H.P; Sander, J; . 1999. OPTICS: Ordering Points To Identify the Clustering Structure. *In* ACM SIGMOD international conference on Management of data. ACM Press. 49-60.

ashape. 2005. ashape: a pedestrian alpha shape extractor.

Babcock, H.S., Y.M ; Zhuang, X. 2012. A high-density 3D localization algorithm for stochastic optical reconstruction microscopy. *Optical Nanoscopy* 1.

Ballard, D.H. 1981. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*. 13:111-122.

Bates, M., B. Huang, G.T. Dempsey, and X.W. Zhuang. 2007. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science*. 317:1749-1753.

Belkin, M., and P. Niyogi. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput*. 15:1373-1396.

Bentley, J.L. 1975. Multidimensional binary search trees used for associative searching. *In* Commun Acm. Vol. 18. 509 - 517

Berlemont, S., R. Tournebize, A. Bensimon, and J. Olivo-Marin. 2008. Detection of full length microtubules in live microscopy images. *5th IEEE International Symposium on Biomedical Imaging From Nano to Macro*:851-854.

Betzig, E., A. Lewis, A. Harootunian, M. Isaacson, and E. Kratschmer. 1986. Near Field Scanning Optical Microscopy (NSOM): Development and Biophysical Applications. *Biophys J*. 49:269-279.

Betzig, E., G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, and H.F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*. 313:1642-1645.

Bicici, E., and D. Yuret. 2007. Locally Scaled Density Based Clustering. *In* Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms, Part I. Springer-Verlag, Warsaw, Poland. 739-748.

Bishop, C.M. 2006. Pattern recognition and machine learning. Springer, New York. xx, 738 p. pp.

Blum, H. 1973. Biological Shape and Visual Science .1. *J Theor Biol*. 38:205-287.

Born, M., and E. Wolf. 1999. Principles of optics : electromagnetic theory of propagation, interference and diffraction of light. Cambridge University Press, Cambridge. 986 pp.

Buchan, J.R., and I. Stansfield. 2007. Halting a cellular production line: responses to ribosomal pausing during translation. *Biology of the cell / under the auspices of the European Cell Biology Organization*. 99:475-487.

Churchman, L.S., Z. Okten, R.S. Rock, J.F. Dawson, and J.A. Spudich. 2005. Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time. *Proc Natl Acad Sci U S A*. 102:1419-1423.

Cootes, T., G. Edwards, and C. Taylor. 1998. Active Appearance Models. *Computer Vision-ECCV*. 2:484-498.

Cootes, T.F., C.J. Taylor, D.H. Cooper, and J. Graham. 1995. Active Shape Models - Their Training and Application. *Comput Vis Image Und*. 61:38-59.

Cox, S., E. Rosten, J. Monypenny, T. Jovanovic-Talisman, D.T. Burnette, J. Lippincott-Schwartz, G.E. Jones, and R. Heintzmann. 2012. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nat Methods*. 9:195-200.

Davies, R.H., C.J. Twining, T.F. Cootes, J.C. Waterton, and C.J. Taylor. 2002. A minimum description length approach to statistical shape modeling. *Ieee T Med Imaging*. 21:525-537.

Deane, C.M., M. Dong, F.P. Huard, B.K. Lance, and G.R. Wood. 2007. Cotranslational protein folding--fact or fiction? *Bioinformatics*. 23:i142-148.

Dertinger, T., R. Colyer, G. Iyer, S. Weiss, and J. Enderlein. 2009. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proc Natl Acad Sci U S A*. 106:22287-22292.

DiMaio, F., D.A. Kondrashov, E. Bitto, A. Soni, C.A. Bingman, G.N. Phillips, Jr., and J.W. Shavlik. 2007. Creating protein models from electron-density maps using particle-filtering methods. *Bioinformatics*. 23:2851-2858.

Donoho, D.L., and C. Grimes. 2003. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *P Natl Acad Sci USA*. 100:5591-5596.

Doucet, A., S. Godsill, and C. Andrieu. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput*. 10:197-208.

Doucet, A.D.F., N; Gordon, N.J. 2001. Sequential Monte Carlo Methods in Practice. Springer.

Duda, R.O., and P.E. Hart. 1972. Use of Hough Transformation to Detect Lines and Curves in Pictures. *Commun Acm*. 15:11-&.

Edelsbrunner, H., D.G. Kirkpatrick, and R. Seidel. 1983. On the Shape of a Set of Points in the Plane. *Ieee T Inform Theory*. 29:551-559.

Edelsbrunner, H., and E.P. Mucke. 1994. 3-Dimensional Alpha-Shapes. *Acm T Graphic*. 13:43-72.

Edelsbrunner, H.F., M; Fu, P; Liang, J;. 1995. Measuring proteins and voids in proteins. *In* In "Proc. 28th Ann. Hawaii Internat Conf. System Sci. 256-264.

Ester, M.K., H ; Sander, J; Xu, X 1996. A density-based algorithm for discovering clusters in large spatial databases with noise *In* Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. 226-231.

Fernandez-Suarez, M., and A.Y. Ting. 2008. Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Bio*. 9:929-943.

Fitzpatrick, J.A., Q. Yan, J.J. Sieber, M. Dyba, U. Schwarz, C. Szent-Gyorgyi, C.A. Woolford, P.B. Berget, A.S. Waggoner, and M.P. Bruchez. 2009. STED nanoscopy in living cells using Fluorogen Activating Proteins. *Bioconjugate chemistry*. 20:1843-1847.

Folling, J., M. Bossi, H. Bock, R. Medda, C.A. Wurm, B. Hein, S. Jakobs, C. Eggeling, and S.W. Hell. 2008. Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nat Methods*. 5:943-945.

Fudenberg, G., and L. Paninski. 2009. Bayesian Image Recovery for Dendritic Structures Under Low Signal-to-Noise Conditions. *Ieee T Image Process*. 18:471-482.

Fung, P., W. Lee, and I. King. 1996. Randomized generalized Hough transform for 2-D gray scale object detection. *Pattern Recognition*. 2:511-515.

Galamhos, C., J. Matas, and J. Kittler. 1999. Progressive probabilistic Hough transform for line detection. *Computer Vision and Pattern Recognition*. 1:554-560.

Geninatti, S., J. Ignacio, B. Benítez, M. Calviño, N. Mata, and J. Luna. 2009. FPGA implementation of the generalized Hough transform. *International Conference on Reconfigurable Computing and FPGAs*:172-177.

Gibson , S.F., and F. Lanni. 1989. Diffraction by a circular aperture as a model for 3D optical microscopy. *J. Opt. Soc. Am. A*. 6:1357-1367.

Gibson , S.F., and F. Lanni. 1991. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *J. Opt. Soc. Am. A*. 8:1601-1613.

Gómez-Luna, J., J. González-Linares, J. Benavides, and N. Guil. 2011. Parallelization of the Generalized Hough Transform on GPU. *Actas XXII Jornadas de Paralelismo*:359-366.

Gordon, M.P., T. Ha, and P.R. Selvin. 2004. Single-molecule high-resolution imaging with photobleaching. *P Natl Acad Sci USA*. 101:6462-6465.

Gustafsson, M.G. 2000. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy*. 198:82-87.

Hell, S.W., and J. Wichmann. 1994. Breaking the Diffraction Resolution Limit by Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt Lett*. 19:780-782.

Hess, S.T., T.P.K. Girirajan, and M.D. Mason. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J*. 91:4258-4272.

Hill, B.J. 2011a. An Orientation Field Approach to Modelling Fibre-Generated Spatial Point Processes. Vol. PhD. University of Warwick.

Hill, B.J.K., W.S ; Thonnes , E. 2011b. Fibre-Generated point process and fields of orientations. *CRISM Report*.

Hinz, S. 2005. Fast and subpixel precise blob detection and attribution. *IEEE International Conference on Image Processing*. 3:III-457-460.

Holden, S.J., S. Uphoff, and A.N. Kapanidis. 2011. DAOSTORM: an algorithm for high-density super-resolution microscopy. *Nat Methods*. 8:279-280.

Holtzer, L., T. Meckel, and T. Schmidt. 2007. Nanometric three-dimensional tracking of individual quantum dots in cells. *Appl Phys Lett*. 90.

Huang, B., H. Babcock, and X. Zhuang. 2010. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell*. 143:1047-1058.

Huang, B., W.Q. Wang, M. Bates, and X.W. Zhuang. 2008. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*. 319:810-813.

Huang, F., S.L. Schwartz, J.M. Byars, and K.A. Lidke. 2011. Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomedical optics express*. 2:1377-1393.

Ihler, A.T., J.W. Fisher, and A.S. Willsky. 2005. Loopy belief propagation: Convergence and effects of message errors. *J Mach Learn Res*. 6:905-936.

Illingworth, J., and J. Kittler. 1987. The Adaptive Hough Transform. *Ieee T Pattern Anal*. 9:690-698.

Isard, M. 2003. PAMPAS: real-valued graphical models for computer vision. *In* Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 1. I-613-I-620 vol.611.

Jones, S.A., S.H. Shim, J. He, and X.W. Zhuang. 2011. Fast, three-dimensional super-resolution imaging of live cells. *Nat Methods*. 8:499-U496.

Jordan, M.I. 2004. Graphical models. *Stat Sci*. 19:140-155.

Juette, M.F., T.J. Gould, M.D. Lessard, M.J. Mlodzianoski, B.S. Nagpure, B.T. Bennett, S.T. Hess, and J. Bewersdorf. 2008. Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. *Nat Methods*. 5:527-529.

Kao, H.P., and A.S. Verkman. 1994. Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position. *Biophys J*. 67:1291-1300.

Kassim, A.A., T. Tan, and K.H. Tan. 1999. A comparative study of efficient generalised Hough transform techniques. *Image Vision Comput*. 17:737-748.

Kaufmann, R., J. Piontek, F. Grull, M. Kirchgessner, J. Rossa, H. Wolburg, I.E. Blasig, and C. Cremer. 2012. Visualization and Quantitative Analysis of Reconstituted Tight Junctions Using Localization Microscopy. *PloS one*. 7.

Khoshelham, K. 2007. Extending Generalized Hough Transform to Detect 3D Objects in Laser Range Data. *Transform*. XXXVI:206-210.

Kimura, A., and T. Watanabe. 2000. Fast Generalized Hough Transform that Improves its Robustness of Shape Detection. *IEICE*. J83-D-II:1256-1265.

Kimura, A., and T. Watanabe. 2004. Generalized Hough transform to be extended as an affine-invariant detector of arbitrary shapes. *Electron Comm Jpn 2*. 87:58-68.

Knappe, R., H. Bulskov, and T. Andreasen. 2003. Similarity graphs. *Foundations of Intelligent Systems*. 2871:668-672.

Komar, A.A. 2009. A pause for thought along the co-translational folding pathway. *Trends in biochemical sciences*. 34:16-24.

Kroug, S.W.H.M. 1995. Ground-state-depletion fluorescence microscopy: a concept for breaking the diffraction resolution limit. *Applied Physics B: Lasers and Optics*. 60:495-497.

Lau, L., Y.L. Lee, S.J. Sahl, T. Stearns, and W.E. Moerner. 2012. STED Microscopy with Optimized Labeling Density Reveals 9-Fold Arrangement of a Centriole Protein. *Biophys J*. 102:2926-2935.

Li, H., T. Shen, D. Vavylonis, and X. Huang. 2009a. Actin filament tracking based on particle filters and stretching open active contour models. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*. 12:673-681.

Li, L.Q., W.X. Me, J.X. Huang, and J.J. Huang. 2009b. Multiple Model Rao-Blackwellized Particle Filter for Manoeuvring Target Tracking. *Defence Sci J*. 59:197-204.

Liang, J., H. Edelsbrunner, P. Fu, P.V. Sudhakar, and S. Subramaniam. 1998a. Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins*. 33:1-17.

Liang, J., H. Edelsbrunner, P. Fu, P.V. Sudhakar, and S. Subramaniam. 1998b. Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. *Proteins*. 33:18-29.

Liang, J., H. Edelsbrunner, S. Pamidghantam, and S. Subramaniam. 1996. Analytical method for molecular shapes: Area, volume, cavities, interface and pockets. *Biophys J*. 70:Waml6-Waml6.

Lidke, K.A. 2012. Super resolution for common probes and common microscopes. *Nat Methods*. 9:139-+.

Luo, D., C.H.Q. Ding, F. Nie, and H. Huang. 2011. Cauchy Graph Embedding. *In* ICML. L. Getoor and T. Scheffer, editors. Omnipress. 553-560.

Maji, S., and M.P. Bruchez. 2012. Inferring biological structures from super-resolution single molecule images using generative models. *PloS one*. 7:e36973.

Maly, I.V., and G.G. Borisy. 2001. Self-organization of a propulsive actin network as an evolutionary process. *P Natl Acad Sci USA*. 98:11324-11329.

Maskell, S.R., R.G. Everitt, R. Wright, and M. Briers. 2006. Multi-target out-of-sequence data association: Tracking using graphical models. *Inform Fusion*. 7:434-447.

Mokhtarian, F., and A.K. Mackworth. 1992. A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves. *Ieee T Pattern Anal*. 14:789-805.

Murphy, K. 2010. Review of "Probabilistic graphical models" by Koller and Friedman. *Artif Intell*. 174:145-146.

Murphy, k.P., Y. Weiss , and M.I. Jordan. 1999. Loopy belief propagation for approximate inference: an empirical study. *Uncertainty in AI*.

Nogales, E., S.G. Wolf, and K.H. Downing. 1998. Structure of the alpha beta tubulin dimer by electron crystallography. *Nature*. 391:199-203.

Oh, S., S. Russell, and S. Sastry. 2009. Markov Chain Monte Carlo Data Association for Multi-Target Tracking. *Ieee T Automat Contr*. 54:481-497.

Olson, C.F. 1998. Improving the generalized Hough transform through imperfect grouping. *Image Vision Comput*. 16:627-634.

Parmley, J.L., and M.A. Huynen. 2009. Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. *PLoS genetics*. 5:e1000548.

Pavani, S.R.P., M.A. Thompson, J.S. Biteen, S.J. Lord, N. Liu, R.J. Twieg, R. Piestun, and W.E. Moerner. 2009. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *P Natl Acad Sci USA*. 106:2995-2999.

R. van der Merwe, A.D., J. F. G. de Freitas, and E. Wan. 2000. The Unscented Particle Filter. *Adv. Neural Inform. Process. Syst*.

Ram, S., P. Prabhat, J. Chao, E.S. Ward, and R.J. Ober. 2008. High Accuracy 3D Quantum Dot Tracking with Multifocal Plane Microscopy for the Study of Fast Intracellular Dynamics in Live Cells. *Biophys J*. 95:6025-6043.

Ram, S., P. Prabhat, E.S. Ward, and R.J. Ober. 2009a. Improved single particle localization accuracy with dual objective multifocal plane microscopy. *Opt Express*. 17:6881-6898.

Ram, S., P. Prabhat, E.S. Ward, and R.J. Ober. 2009b. Improved single particle localization accuracy with dual objective multifocal plane microscopy. *Opt Express*. 17:6881-6898.

Resch, G.P., K.N. Goldie, A. Krebs, A. Hoenger, and J.V. Small. 2002. Visualisation of the actin cytoskeleton by cryo-electron microscopy. *Journal of cell science*. 115:1877-1882.

Reymann, J., D. Baddeley, M. Gunkel, P. Lemmer, W. Stadter, T. Jegou, K. Rippe, C. Cremer, and U. Birk. 2008. High-precision structural analysis of subnuclear complexes in fixed and live cells via spatially modulated illumination (SMI) microscopy. *Chromosome Res*. 16:367-382.

Rezaii, T.Y., and M.A. Tinati. 2011. Distributed multi-target tracking using joint probabilistic data association and average consensus filter. *Ann Telecommun*. 66:553-566.

Roweis, S.T., and L.K. Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 290:2323-+.

Rust, M.J., M. Bates, and X.W. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods*. 3:793-795.

Sampat, M.P., Z. Wang, S. Gupta, A.C. Bovik, and M.K. Markey. 2009. Complex Wavelet Structural Similarity: A New Image Similarity Index. *Ieee T Image Process*. 18:2385-2401.

Sander, J., M. Ester, H.P. Kriegel, and X.W. Xu. 1998. Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Min Knowl Disc*. 2:169-194.

Sarkka, S., A. Vehtari, and J. Lampinen. 2007. Rao-Blackwellized particle filter for multiple target tracking. *Inform Fusion*. 8:2-15.

Saul, L.K., and S.T. Roweis. 2004. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J Mach Learn Res*. 4:119-155.

Schaub, S., J.J. Meister, and A.B. Verkhovsky. 2007. Analysis of actin filament network organization in lamellipodia by comparing experimental and simulated images. *Journal of cell science*. 120:1491-1500.

Sharonov, A., and R.M. Hochstrasser. 2006. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc Natl Acad Sci U S A*. 103:18911-18916.

Shi, J.B., and J. Malik. 2000. Normalized cuts and image segmentation. *Ieee T Pattern Anal*. 22:888-905.

Sigal, L. 2005. http://www.cs.brown.edu/~ls/download.html.

Sigal, L., S. Bhatia, S. Roth, M.J. Black, and M. Isard. 2004. Tracking loose-limbed people. *In* Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 1. I-421-I-428 Vol.421.

Simoncelli, E.P., W.T. Freeman, E.H. Adelson, and D.J. Heeger. 1992. Shiftable Multiscale Transforms. *Ieee T Inform Theory*. 38:587-607.

Smal, I., K. Draegestein, N. Galjart, W. Niessen, and E. Meijering. 2008a. Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: Application to microtubule growth analysis. *Ieee T Med Imaging*. 27:789-804.

Smal, I., E. Meijering, K. Draegestein, N. Galjart, I. Grigoriev, A. Akhmanova, M.E. van Royen, A.B. Houtsmuller, and W. Niessen. 2008b. Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering. *Med Image Anal*. 12:764-777.

Smith, C.S., N. Joseph, B. Rieger, and K.A. Lidke. 2010. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat Methods*. 7:373-U352.

Speidel, M., A. Jonas, and E.L. Florin. 2003. Three-dimensional tracking of fluorescent nanoparticles with subnanometer precision by use of off-focus imaging. *Opt Lett*. 28:69-71.

Staib, L.H., and J.S. Duncan. 1992. Boundary Finding with Parametrically Deformable Models. *Ieee T Pattern Anal*. 14:1061-1075.

Stetson, P.B. 1987. Daophot - a Computer-Program for Crowded-Field Stellar Photometry. *Publ Astron Soc Pac*. 99:191-222.

Stoitsis, J., S. Golemati, S. Kendros, and K.S. Nikita. 2008. Automated detection of the carotid artery wall in B-mode ultrasound images using active contours initialized by the Hough Transform. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. 2008:3146-3149.

Su, J.H., B; Kendall, E ; Thonnes, E. 2008. Inference for point processes with unobserved, one-dimensional reference structure. *CRISM Research Report*

Sudderth, E.B., A.T. Ihler, M. Isard, W.T. Freeman, and A.S. Willsky. 2010. Nonparametric Belief Propagation. *Commun Acm*. 53:95-103.

Suetake, N., E. Uchino, and K. Hirata. 2006. Generalized fuzzy Hough transform for detecting arbitrary shapes in a vague and noisy image. *Soft Comput*. 10:1161-1168.

Svoboda, D., M. Kozubek, and S. Stejskal. 2009. Generation of Digital Phantoms of Cell Nuclei and Simulation of Image Formation in 3D Image Cytometry. *Cytom Part A*. 75A:494-509.

Taylor, M.J., D. Perrais, and C.J. Merrifield. 2011. A high precision survey of the molecular dynamics of mammalian clathrin-mediated endocytosis. *PLoS biology*. 9:e1000604.

Thomann, D., J. Dorn, P.K. Sorger, and G. Danuser. 2003. Automatic fluorescent tag localization II: Improvement in super-resolution by relative tracking. *Journal of microscopy*. 211:230-248.

Thompson, M.A., J.M. Casolari, P.O. Brown, and W.E. Moerner. 2011. 3D tracking of single mRNA particles in S. cerevisiae using a double-helix point spread function. *Abstr Pap Am Chem S*. 241.

Thompson, R.E., D.R. Larson, and W.W. Webb. 2002. Precise nanometer localization analysis for individual fluorescent probes. *Biophys J*. 82:2775-2783.

Toprak, E., H. Balci, B.H. Blehm, and P.R. Selvin. 2007. Three-dimensional particle tracking via bifocal imaging. *Nano Lett*. 7:2043-2045.

Toronto, N., B. Morse, D. Ventura, and K. Seppi. 2007. The Hough Transform's Implicit Bayesian Foundation. *IEEE International Conference on Image Processing*:377-380.

UCSC. gtrnadb.

van der Maaten, L.J.P. 2012. MATLAB Toolbox for Dimensionlity Reduction.

Verkhovsky, A.B., O.Y. Chaga, S. Schaub, T.M. Svitkina, J.J. Meister, and G.G. Borisy. 2003. Orientational order of the lamellipodial actin network as demonstrated in living motile cells. *Mol Biol Cell*. 14:4667-4675.

Watanabe, T.M., S. Fukui, T. Jin, F. Fujii, and T. Yanagida. 2010. Real-time nanoscopy by using blinking enhanced quantum dots. *Biophys J*. 99:L50-52.

Watanabe, T.M., T. Sato, K. Gonda, and H. Higuchi. 2007. Three-dimensional nanometry of vesicle transport in living cells using dual-focus imaging optics. *Biochem Bioph Res Co*. 359:1-7.

Wei-Kai, L., and I. Cohen. 2006. Belief Propagation Driven Method for Facial Gestures Recognition in Presence of Occlusions. *In* Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on. 158-158.

Weiss, Y., and J. Pearl. 2010. Belief Propagation. *Commun Acm*. 53:94-94.

Wittman, T. 2005. MANIfold Learning Matlab Demo.

Xiaofei, H., C. Deng, Y. Shuicheng, and Z. Hong-Jiang. 2005. Neighborhood preserving embedding. *In* Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 2. 1208-1213 Vol. 1202.

Xu, L., E. Oja, and P. Kultanen. 1990. A New Curve Detection Method - Randomized Hough Transform (Rht). *Pattern Recogn Lett*. 11:331-338.

Yang, Z., K. Lasker, D. Schneidman-Duhovny, B. Webb, C.C. Huang, E.F. Pettersen, T.D. Goddard, E.C. Meng, A. Sali, and T.E. Ferrin. 2011. UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *Journal of structural biology*.

Yanover, C., and Y. Weiss. 2003. Finding the M Most Probable Configurations using Loopy Belief Propagation. *NIPS*.

Yedidia, J.S., W.T. Freeman, and Y. Weiss. 2003. Understanding belief propagation and its generalizations. *In* Exploring artificial intelligence in the new millennium. Morgan Kaufmann Publishers Inc. 239-269.

Yildiz, A., J.N. Forkey, S.A. McKinney, T. Ha, Y.E. Goldman, and P.R. Selvin. 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science*. 300:2061-2065.

Yildiz, A., and P.R. Selvin. 2005. Fluorescence imaging with one nanometer accuracy: application to molecular motors. *Accounts of chemical research*. 38:574-582.

Yu, Q., and G. Medioni. 2009. Multiple-Target Tracking by Spatiotemporal Monte Carlo Markov Chain Data Association. *Ieee T Pattern Anal*. 31:2196-2210.

Zelniker, E.E., and I.V.L. Clarkson. 2006. Maximum-likelihood estimation of circle parameters via convolution. *Ieee T Image Process*. 15:865-876.

Zhang, Q.P., and I. Couloigner. 2007. Accurate centerline detection and line width estimation of thick lines using the radon transform. *Ieee T Image Process*. 16:310-316.

Zhang, Z.P., and C.H. Menq. 2008. Three-dimensional particle tracking with subnanometer resolution using off-focus images. *Appl Optics*. 47:2361-2370.

Zhao, T., and R.F. Murphy. 2007. Automated learning of generative models for subcellular location: Building blocks for systems biology. *Cytom Part A*. 71A:978-990.

Zhou, Y.J., and Y.P. Zheng. 2008. Estimation of muscle fiber orientation in ultrasound images using revoting Hough transform (RVHT). *Ultrasound Med Biol*. 34:1474-1481.

Zhu, L., W. Zhang, D. Elnatan, and B. Huang. 2012. Faster STORM using compressed sensing. *Nat Methods*. 9:721-723.