3D Reconstruction from 2D Images and Applications to Cell Cytoskeleton

by

Yuan Cheng

M.S., Mechanical Engineering (1998)

Massachusetts Institute of Technology

Submitted to the Department of Mechanical Engineering

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Mechanical Engineering

at the

Massachusetts Institute of Technology

December, 2000

[February 2001]

Signature of Author _____

Department of Mechanical Engineering

December, 2000

Certified by _____

C.Forbes Dewey, Jr.

Professor of Mechanical Engineering

Thesis Supervisor

Accepted by _____

Ain A. Sonin

Chairman, Department Committee on Graduate Student

3D Reconstruction from 2D Images and Applications to Cell Cytoskeleton

by

Yuan Cheng

Submitted to the Department of Mechanical Engineering
on December 31st, 2000, in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Approaches to achieve three dimensional (3D) reconstruction from 2D images can be grouped into two categories: computer-vision-based reconstruction and tomographic reconstruction. By exploring both the differences and connections between these two types of reconstruction, the thesis attempts to develop a new technique that can be applied to 3D reconstruction of biological structures. Specific attention is given to the reconstruction of the cell cytoskeleton from electron microscope images.

The thesis is composed of two parts. The first part studies computer-vision-based reconstruction methods that extract 3D information from geometric relationship among images. First, a multiple-feature-based stereo reconstruction algorithm that recovers the 3D structure of an object from two images is presented. A volumetric reconstruction method is then developed by extending the algorithm to multiple images. The method integrates a sequence of 3D reconstruction from different stereo pairs. It achieves a globally optimized reconstruction by evaluating certainty values of each stereo reconstruction. This method is tuned and applied to 3D reconstruction of the cell cytoskeleton. Feasibility, reliability and flexibility of the method are explored.

The second part of the thesis focuses on a special tomographic reconstruction, discrete tomography, where the object to be reconstructed is composed of a discrete set of materials each with uniform values. A Bayesian labeling process is proposed as a framework for discrete tomography. The process uses an expectation-maximization (EM) algorithm with which the reconstruction is obtained efficiently. Results demonstrate that the proposed algorithm achieves high reconstruction quality even with a small number of projections. An interesting relationship between discrete tomography and conventional tomography is also derived, showing that discrete tomography is a more generalized form of tomography and conventional tomography is only a special case of such generalization.

Thesis Committee:

C. Forbes Dewey, Jr., Professor of Mechanical Engineering and Bioengineering

Eric Grimson, Professor of Computer Science and Engineering

David Gossard, Professor of Mechanical Engineering

John Hartwig, Associate Professor of Brigham and Women's Hospital

Acknowledgements

I am very proud of having the most important part of my life associated with the world's most prestigious engineering school, MIT. I am more grateful for the helps and supports that I received during this period of time from so many wonderful people around me.

First and foremost, I would like to thank my advisor, Professor C. Forbes Dewey for bringing me into his group and showing confidence in me and my work. His continuous encouragement, patience and guidance have shaped my research and helped me to overcome hurdles one after the other. I also thank him for a number of inspiring and productive discussions on the way to and back from Royal East.

I am thankful to my thesis committee members: Professor Eric Grimson, Professor David Gossard and Professor John Hartwig. It has been my special honor to have these world-known scientists in my thesis committee. I thank them for their presence, their support and numerous suggestions that they made towards my work. Special thanks to Professor Hartwig for tolerating my near ignorance of cell biology and helping me in preparing the specimen.

I am also grateful to my present and past lab-mates for their helps. My improvement in English, if any, largely attributes to them. I would also like to thank Ms. Donna Wilker for her helps in past years and for her successful attempts to balance my research life by lending me her movie collections.

Finally, I dedicate this thesis to my parents who supports me unconditionally with their hearts and also to my loving wife Yan Feng who has always stood by me and sacrificed so much so that I could fulfill my dream. I am fortunate to have them as part of my life.

Table of Contents

List of Figures

Chapter 1

Background and Introduction

This thesis is intended to summarize the research in which I have participated over the past four and half years. The goal of my project mainly aims to explore and develop a new technique that is capable of reconstructing structural information from 2D electron microscope images and furthermore to provide quantitative measurements about some biological structures, e.g., the cell cytoskeleton. This Chapter gives an overview of the thesis and some background that this thesis will rely on.

1.1 Origin of Research

The motivation of this project originated from the need to obtain 3D cellular structural properties in our studies of cell cytoskeleton. The cytoskeleton of eucaryotic cells is primarily composed of three types of polymers: actin filaments, microtubules and intermediate filaments. Actin filaments are the most abundant components and they are arranged into a 3D structural network that gives the cytoplasm its shape, form, and mechanical properties. Considerable effort has gone into defining the structural and biochemical properties of the 3D polymer systems that comprise the cytoskeleton. One technique that has been widely applied to understanding this architecture is electron microscopy (EM). Figure 1 exhibits an example of cell cytoskeleton image. The images (or micrographs) taken from EM provide information on the length, geometry, and interaction and location of various cytoskeletal components. However, the images are only a 2D representation of 3D objects. Structural studies are often hampered by the inability to faithfully obtain the complicated 3D geometric relationships made by actin filaments as they course throughout the cytoplasmic space. Over the past years, 3D information is obtained largely by manual measurements. Not only is such manual work

tedious, but also the measurements are very subjective and inaccurate. Therefore, it is very desirable to develop a compute-automated system that reconstructs the 3D structures from 2D images by computer and ultimately makes the measurements on the reconstructed structures. The thesis will mainly focus on the first part of the problem, i.e., the reconstruction of 3D structures from given 2D images.

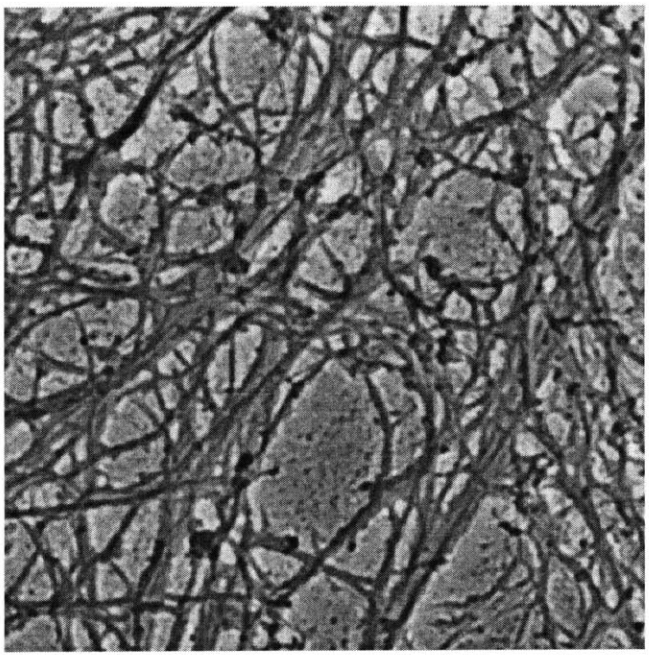


Figure 1 An example of cell cytoskeleton image from electron microscope

1.2 The World of 3D Reconstruction

3D reconstruction from 2D images in general has been of interest in a number of fields, including computer vision, robotics, medical imaging, structural biology, etc. The methods used for 3D reconstruction can roughly be categorized into two groups: one is a computer vision approach and the other is a tomographic approach. This separation is mainly due to their different treatments of the imaging function. The imaging function is defined as a function that maps the relationship between image and object. It describes how the brightness or intensity value in the image is related to the object value or some property value of the object in 3D space. The tomographic approach focuses on the

reconstruction in which the image records the transmission or emission property of the object. The imaging function is relatively simple. For instance, in linear tomography, the value in the image is represented as the integral of the object's property values along the imaging direction. On the other hand, the computer vision approach often deals with the reconstruction problem in which the image primarily describes the reflectance of the object, e.g. an image taken by a camera. The related imaging function is usually very complicated, which may involve the object's shape, its reflectance properties, position, and illumination.

1.2.1 Computer Vision

Computer vision has emerged over the years as a discipline that attempts to enable the machine or computer to sense and interact with the environment. Major efforts have been directed towards the reconstruction of 3D structure of objects using machine analysis of images (Dhond and Aggarwal 1989). However, unlike the tomographic reconstruction, there is no mathematically sound inverse reconstruction method. This is mainly because computer vision has much more complicated relationship between image and object than tomography does. This complexity involves the object's shape, its reflectance properties, position, light sources, etc. The imaging function is sometimes very difficult to obtain.

In computer vision, the understanding of 3D structure is primarily extracted from the geometric relationships between images (Grimson 1980; Grimson 1985; Dhond and Aggarwal 1989; Faugeras 1993; Okutomi and Kanada 1993; Kanada and Okutomi 1994). The images of a 3D object are typically taken at the different locations. The positions of the object in the images are therefore different. This difference, termed disparity, is directly related to the position of the object in 3D space. In other words, the 3D information about the object is embedded in the images in the form of disparity. For instance, Figure 2 illustrates a famous stereo vision model of parallel axis geometry. The camera is simply represented by a pin-hole model. Let us denote (X,Y,Z) as a point on an object in 3D world coordinates with origin at O. Its positions in two images are denoted by (x_l, y_l) and (x_r, y_r) with respect to their own image coordinate system. Let f be the focal

length of both cameras and d be the baseline distance between two camera. By simple triangulation, we have

$$\frac{x_l}{f} = \frac{X + d/2}{Z} \tag{1.1}$$

$$\frac{x_r}{f} = \frac{X - d/2}{Z} \tag{1.2}$$

$$\frac{y_l}{f} = \frac{y_r}{f} = \frac{Y}{Z} \quad . \tag{1.3}$$

Solving for (X,Y,Z) gives:

$$X = \frac{d(x_l + x_r)}{2(x_l - x_r)} \tag{1.4}$$

$$Y = \frac{d(y_l + y_r)}{2(x_l - x_r)} \tag{1.5}$$

$$Z = \frac{df}{(x_l - x_r)} \quad , \tag{1.6}$$

where $(x_l\text{-}x_r)$ in the equations are often referred as the disparity. The equations indicate that if we know the object's projection positions in two images, we can fully determine the object position in 3D space. The implication is very powerful. The reconstruction problem is therefore simplified to find the object's projection positions in images. However, finding the corresponding projection position in images, called matching, is usually not trivial. Over the years, a number of matching techniques have been developed. We will discuss them in details in Chapter 3. Also in Chapter 3, we will study a different type of imaging model that results in a completely different set of equations but the concept remains the same.

One of the advantages of this reconstruction method is that we don't need to worry about the complicated imaging function between the image and the object. The reconstruction can be applicable to any type of imaging systems. However, since no imaging function is used, the exact values of the object cannot be obtained. The obtained reconstruction only provides the locations or the shapes of the 3D object. In other words, we may only tell whether there exists an object or not. We cannot know what value the object may take.

The tomographic reconstruction that will be discussed later is completely different in this point. It can give both location of the object and the value of the object because the reconstruction is derived from the given imaging function.

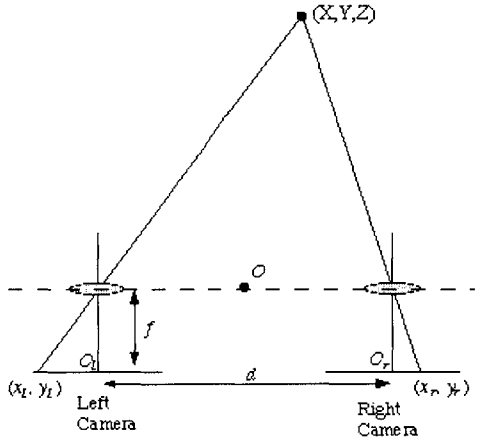


Figure 2 Parallel axis stereo geometry

Besides the reconstruction from disparity, there are several other 3D reconstruction techniques in computer vision. For example, shape from shading (Horn 1986; Horn and Brooks 1989), which requires a manageable imaging function, reconstructs 3D surface from 2D images with different shades. The technique is based on the observation that under different illumination conditions the image brightness responds differently based on the surface orientations. There have been some successful stories of this technique but its use is still very limited because controlling illumination condition and surface properties are practically very difficult. Interested readers are referred to related literature (Horn 1986; Horn and Brooks 1989). Another alternative reconstruction method in computer vision is optical flow (Horn 1986). Optical flow provides a reliable approximation to the image motion. Barron et al (Barron, Fleet et al. 1994) did a very good review on various optical flow algorithms and their performances. Optical flow can be used to estimate the disparity map between images. With disparity map, 3D surface or object may be recovered via geometric relationships such as Equations (1.4) ~ (1.6). This two-step procedure is often referred as indirect method. In contrast, Horn and Weldon (Horn and Weldon 1988) developed a direct method for scene reconstruction using

optical flow. Shashua and Stein (Shashua 1995; Stein and Shashua 1996; Stein 1998; Stein and Shashua 2000) further expanded the approach to three views. In both the direct and the indirect method, the optical flow approach is based on a fundamental assumption: the constant brightness assumption. It assumes that the brightness pattern of the object remains the same between images even though their positions may change. In order for this assumption to be applicable, there are some strict requirements, such as Lambertian surface, uniform illumination, etc. These conditions are normally hard to satisfy in practice.

As a summary, due to the complexity of the imaging function in computer vision, there is no direct inverse reconstruction from the image to 3D object. The reconstruction is primarily obtained by establishing the geometric relationship between image and object. Reconstruction from disparity is one of the most used methods in computer vision because it shields us from the complexity of the imaging function. Other methods, such as shape from shading and structure from motion (optical flow), are alternative methods of reconstruction in computer vision, applicable to some controlled conditions.

Note that in this thesis we will mainly focus on the method of the reconstruction from disparity.

1.2.2 Tomography

Tomography solves a special reconstruction problem in which the imaging function is known and has a strict form. The reconstruction can be derived mathematically via an inverse mapping function. 3D reconstruction in tomography can be considered a simple extension of the 2D tomographic reconstruction. In most situations, 3D reconstruction is obtained by stacking a series of 2D reconstructed slices. Let us begin our discussion on 2D reconstruction first.

The fundamental problem that needs to solve is the reconstruction from projections. The solution to this problem is tomography. Over the past 40 years, it has seen its rapid

developments and applications expanding into a large number of scientific, medical, and technical fields. Of all the applications, probably the greatest impact has been in the areas of diagnostic medicine (Herman 1980). It earned itself two Nobel prizes: one in 1979 for computerized tomography (CT) and the other in 1982 for electron tomography (which will be discussed separately in next subsection). In recent years, some newly emerged technologies in medicine, such as MRI (Magnetic Resonance Imaging), SPECT (Single Photon Emission Computed Tomography) and PET (Positron Emission Tomography) are all based on tomography.

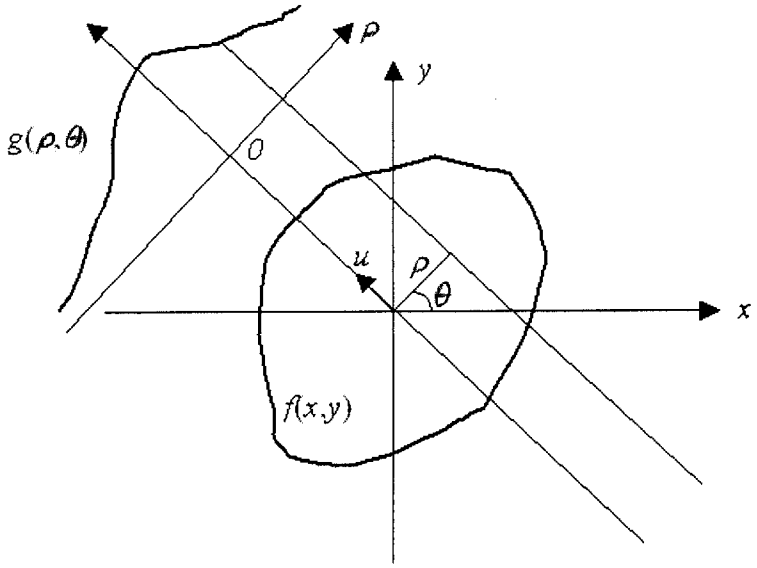


Figure 3 Projection imaging geometry of tomography

Interestingly, as early as 1917, Radon derived mathematical theories of Radon transform and inverse Radon transform, which turned out to be the underlying fundamentals of tomography. In linear tomography, the imaging function is defined as follows: the value in each projection is the integral of the object values along projection direction. Figure 3 illustrates a 2D case of the projection geometry. The relationship is expressed as:

$$g(\rho,\theta) = \int_{u} f(x,y)du \tag{1.7}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy, \tag{1.8}$$

where u is the projection direction and $\delta(\cdot)$ is a Dirac delta function (impulse function). The projections $g(\rho,\theta)$ can be re-organized into an image with ρ and θ as two coordinates, and often called a sinogram. Therefore, the goal in tomography is to reconstruct the object function $f(x,y)$ from its sinogram $g(\rho,\theta)$. Equation (1.8) is actually the exact form of Radon transform. The object $f(x,y)$ can therefore be obtained by inverse Radon transform. Inversion of the Radon transform can be done in several ways. One standard algorithm is based on the projection-slice theorem (Mersereau 1973; Jain 1989). The theorem indicates that the one-dimensional Fourier transform of the projection $g(\rho,\theta)$ with respect to ρ is equal to the central slice, at angle θ, of the two-dimensional Fourier transform of the object $f(x,y)$, i.e.,

$$G(\xi,\theta) = F_p(\xi,\theta) \quad , \tag{1.9}$$

where capital names, such as G or F, represent the function in Fourier domain and ξ is the coordinate in Fourier space. Subscript p means the variables are in the polar coordinate system, such as F_p. If we fill the whole Fourier domain at any angle with corresponding projection, the object is then obtained by inverse Fourier transform. Mathematically, it can be derived as follows (Jain 1989; Toft 1996). The inverse Fourier transform is given by:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\xi_x,\xi_y)\exp\left[j2\pi(\xi_x x + \xi_y y)\right]d\xi_x d\xi_y \tag{1.10}$$

When written in polar coordinates in Fourier domain, Equation (1.10) gives:

$$\begin{aligned} f(x,y) &= \int_{0}^{\pi}\int_{-\infty}^{\infty} |\xi| F_p(\xi,\theta)\exp\left[j2\pi\xi(x\cos\theta + y\sin\theta)\right]d\xi\, d\theta \\ &= \int_{0}^{\pi}\int_{-\infty}^{\infty} |\xi| G(\xi,\theta)\exp\left[j2\pi\xi(x\cos\theta + y\sin\theta)\right]d\xi\, d\theta \end{aligned} \tag{1.11}$$

$$= \int_{0}^{\pi} \hat{g}(x\cos\theta + y\sin\theta,\theta)d\theta \quad , \tag{1.12}$$

where

$$\hat{g}(\rho,\theta) = \int_{-\infty}^{\infty} |\xi| G(\xi,\theta)\exp\left[j2\pi\xi\rho\right]d\xi \quad . \tag{1.13}$$

Equation (1.13) is an inverse Fourier transform of function $|\xi|G(\xi, \theta)$, which can be considered a filtered version of $G(\xi, \theta)$ with the high-pass filter $|\xi|$ in Fourier domain. Equation (1.12) is called back-projection. In fact, these two equations, Equations (1.12) and (1.13) form the famous filtered-backprojection (FBP) method, which involves two steps. First, each projection $g(\rho,\theta)$ is filtered by the one-dimensional filter whose Fourier transform is $|\xi|$ (which is Equation (1.13)). The result, $\hat{g}(\rho,\theta)$, is then back-projected to yield $f(x,y)$ (which is Equation (1.12)).

It is also possible to make the backprojection before the filtering (Herman 1980; Jain 1989). The method is called filtering after backprojection (FABP), as well as rho-filtered layergram in some early literature. In this method, a different filter must be used. The procedures are given as follows:

$$\hat{f}(x,y) = \int_0^{\pi} g(x\cos\theta + y\sin\theta, \theta)\, d\theta \tag{1.14}$$

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{\xi_x^2 + \xi_y^2}\, \hat{F}(\xi_x,\xi_y) \exp\left[j2\pi(\xi_x x + \xi_y y)\right] d\xi_x d\xi_y \tag{1.15}$$

where $\hat{F}(\xi_x,\xi_y)$ is the Fourier transform of $\hat{f}(x,y)$. The filter is a two-dimensional filter with the form of $\sqrt{\xi_x^2 + \xi_y^2}$ in Fourier domain. Although FABP method is not used as popularly as FBP method in practice, it achieves comparable performance to FBP (Suzuki and Yamaguchi 1988). In Chapter 5 of the thesis, we will come back to FABP method which helped us to derive an efficient algorithm for discrete tomography. One common property of both FBP and FABP methods, as indicated in Equation (1.12) or (1.14), is that they require a full range of projections (θ from 0 to π). The incomplete projections may lead to incorrect reconstruction.

In practice, perfect reconstruction described by the above methods can not be achieved because we cannot obtain an infinite number of projections for the integral operation in Equations (1.12) and (1.14). Furthermore, the methods have to be implemented in the computer and all operations are approximated by their discrete versions. Therefore, each

method has its reconstruction limit. The resolution is typically dependant on the angle between adjacent projections ($\Delta\theta$) and the point resolution in each projection ($\Delta\rho$).

Besides the direct inverse formulas above, inversion of the Radon transform can also be obtained via some iterative methods, such as algebraic reconstruction techniques (ART) (Gordon, Bender et al. 1970) or expectation maximization (EM) methods (Shepp and Vardi 1982; Green 1990). In comparison to the direct reconstruction method, iterative methods are usually very expensive in computation. However, they may achieve better reconstruction quality. More importantly, iterative methods offer a great deal of flexibility and robustness. They can deal with situations like incomplete projection data, uneven-sampled projections, noise-deteriorated projections, reconstruction with constraints, etc. None of the direct inverse methods can easily handle these common but difficult situations. It is because of these advantages that iterative methods have gained a lot of attention and popularity in many medical applications, such as PET and SPECT. We will review some of the well-known iterative reconstruction techniques later in the thesis. Those iterative methods form the basis to our algorithm for discrete tomographic reconstruction.

One important observation worthy of mentioning is the duality between linear algebraic operations and the Radon / inverse Radon transforms. In discrete implementation, the linear projection of Equation (1.7) can be represented by the sum of the object points along the projection path. If we represent the all object points in a vector form denoted by $\mathbf{x}$ and all projection points in a vector form denoted by $\mathbf{b}$, the projections can then be written in a matrix vector formulation:

$$\mathbf{b} = \mathbf{A}\mathbf{x} \tag{1.16}$$

where $\mathbf{A}$ is coefficient matrix containing the weighting factors between each of object points and projection direction. With this linear algebra formula, the reconstruction problem becomes to solve the linear equation of Equation (1.16). Since matrix $\mathbf{A}$ is non-square, one approach to solve Equation (1.16) is to form a normal equation:

$$\mathbf{A}^T\mathbf{b} = \mathbf{A}^T\mathbf{A}\mathbf{x} \tag{1.17}$$

where $\mathbf{A}^T$ is the transpose of $\mathbf{A}$, and $\mathbf{A}^T\mathbf{A}$ is a square matrix. If we assume $\mathbf{A}^T\mathbf{A}$ is invertible, the solution of the linear equation is given by:

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \tag{1.18}$$

Let us draw some analogy between this linear algebraic form and the direct inverse Radon transform (Kawata and Nalcioglu 1985; Toft 1996). Equation (1.16) corresponds to the Radon transform, which does the forward projection. Thus, Equation (1.18) essentially represents the inverse Radon transform. More interestingly, if matrix $\mathbf{A}$ is considered a projection operator, its transpose $\mathbf{A}^T$ is actually a backprojection operator. Therefore, Equation (1.18) can be interpreted as filtering after backprojection (FABP) with $\mathbf{A}^T$ representing backprojection and $(\mathbf{A}^T\mathbf{A})^{-1}$ representing the filter. This filter corresponds to the one given in Equation (1.15). Similarly, filtered backprojection (FBP) can be matched to a different form of linear equation solution:

$$\begin{aligned} \mathbf{x} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \\ &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} \end{aligned} \tag{1.19}$$

$$= \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} \quad . \tag{1.20}$$

In Equation (1.20), $(\mathbf{A}\mathbf{A}^T)^{-1}$ corresponds to the high-pass filter in Equation (1.13). The backprojection $\mathbf{A}^T$ is performed after filtering. In summary, Figure 4 lists these duality mappings.

$\mathbf{Ax} = \mathbf{b}$	⟺	**b** is projection
$\mathbf{A}$	⟺	projection operator
$\mathbf{A}^T$	⟺	backprojection operator
$\mathbf{x} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$	⟺	FBP method
$(\mathbf{A}\mathbf{A}^T)^{-1}$	⟺	filter in FBP method
$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$	⟺	FABP method
$(\mathbf{A}^T\mathbf{A})^{-1}$	⟺	filter in FABP method

Figure 4 Duality relationships between algebraic method and direct method

So far, the reconstruction from projections has been focused on 2D tomography. 3D reconstruction can simply be constructed from 2D case slice by slice. Alternatively, 3D

tomography can also be obtained by 3D Radon transform and inverse 3D Radon transform (Deans 1993). Both direct inverse and iterative methods are applicable in 3D but may have a different form. Some types of 3D projection and reconstruction are described in (Toft 1996).

1.2.3 Electron Tomography

Electron tomography specifically refers to the technique that employs an electron microscope to collect projections of an object and uses these projections to reconstruct the object (Frank 1996). It inherits all the theories developed in tomography. In structural biology, it is currently the primary approach of 3D reconstruction for macromolecular or sub-cellular level structures, such as virus or ribosome (Frank 1996). In some literature, one would prefer to exclude crystallographic methods from electron tomography, more or less for historical reasons. Crystallography specifically refers to the technique that makes use of inherent order or symmetry properties of the object in reconstruction (DeRosier and Klug 1968). Electron tomography is considered a more general theory of reconstruction that makes no assumption of the presence of the order or symmetry. Fundamentally, both crystallography and electron tomography are based on the same underlying theorem: the projection-slice theorem, developed in tomography. The goal of the reconstruction is to use projections to fill the entire Fourier space as the theorem suggests.

Unlike the tomography in medical applications, electron tomography faces a number of its own obstacles. The first is the restricted tilt angles of the electron microscope. The angular range in most electron microscopes does not exceed ±60°. A significant portion of the Fourier transform is simply missing, which may result in a great deal of distortion in the reconstruction. However, crystallography is one exception. Because of the symmetries, crystallography has been very successful in dealing with this issue when reconstructing crystalline objects. Due to the same symmetries in Fourier space, the missing part of the Fourier transform can be inferred and derived from other parts of the Fourier space and the entire Fourier transform can therefore be obtained. Another special

case is some macromolecules that exist structurally identical copies in the volume considered. These identical copies typically have random orientations in each projection. The projection copies at these orientations can be used to make up the missing projection at unreachable tilt angles. Therefore, the missing part of Fourier transform can be filled as well. The current applications of Electron tomography have primarily focused on these macromolecular structures (Frank 1992; Frank 1995). However, in nature, the biological structures or entities that fit these special criteria are very limited. The development of electron tomography has been hampered greatly because of this.

The second obstacle is the strict imaging function requirement. Since both tomography and electron tomography are based on projection-slice theorem. The applicability of the theorem relies on the imaging relationship between projection and object being the form of Equation (1.7). In the case of electron-microscope imaging system, such relationship usually doesn't hold well. Some approximation has to be made. The condition for approximation is dependant on the specimen preparation, imaging conditions, etc. (Frank 1992)

Another issue is the difficulty of preserving the native structure during specimen preparation and image collection. The associated problems are two folds. One is the ability of the electron beam to penetrate the specimen. The second is the structural damages induced by multiple exposure to electron beams during imaging. The former one tends to suggest the use of intermediate to high voltage electron microscope while the latter one suggests the opposite. In addition, in order to improve the contrast, the specimen is often carefully stained or gold-labeled (Hartwig 1992; Frank 1996). However, the presence of heavy atoms in stained specimen makes the approximation of the imaging function untenable.

Notwithstanding these obstacles, electron tomographic techniques have had tremendous success and impact in structural biology. In recent years, many new developments have been witnessed. For instance, the technique of double-tilt is used to reduce the size of the angular gap (Penczek, Marko et al. 1995). However, obstacles still exist. Electron

tomography is still not generally applicable to a large number of biological structures or entities. The thesis, in some ways, attempts to explore an alternative reconstruction technique that would tackle some of these obstacles.

1.3 Bayesian Estimation

Let us look at the reconstruction problem from a different perspective. The central problem of reconstruction is to recover the object from its projections. It perfectly fits into the estimation methodology where we want to estimate the unknown input (the object) from the given output (the projections). Estimation of input from output has been a very classical problem in many areas. In statistic estimation, the problem can be approached by attempting to answer this question: what is the statistically best input that generates the given output?

In this section, we will first briefly introduce some mathematics and terminology of statistical estimation. Focus will be on discussing estimation methods under a Bayesian framework. We assume that readers are familiar with basic probability and statistics, such as random variables, random vectors, probability function, mean, variance, etc. In the thesis, a prevailing convention is adopted: probability distribution function $P(\cdot)$ (capital letter P) is used when the considered random variable is discrete, and probability density function $p(\cdot)$ (low-case letter p) is used when the random variable is continuous. Bold case is used to refer a random vector, such as $\boldsymbol{x}$, while non-bold represents a scalar random variable, such as x.

1.3.1 Conditional Probability and Bayes' Rule

We will frequently deal with several random variables. These random variables may or may not be related to each other. If they are, we hope that they may be predictable from each other. However, the predictability is not guaranteed. For instance, when two random variables are completely independent, knowing one random variable does not help to predict the other.

The relationship between random variables can be characterized by the joint distribution and conditional distribution. The joint distribution of two random variables, $P(X,Y)$, describes the co-occurrence of both events X and Y. The marginal distribution, which defines the occurrence of either X or Y, can be obtained from their joint distribution:

$$P(X) = \sum_{y \in \Omega_Y} P(X, Y = y) \tag{1.21}$$

$$P(Y) = \sum_{x \in \Omega_X} P(X = x, Y) \tag{1.22}$$

where Ω_X and Ω_Y are sample spaces of random variable X and Y, respectively.

The conditional distribution, $P(X|Y)$, describes the probability of random variable X given Y. It is defined as:

$$P(X \mid Y) = \frac{P(X,Y)}{P(Y)} \quad . \tag{1.23}$$

Likewise, the conditional distribution, $P(Y|X)$, is given by:

$$P(Y \mid X) = \frac{P(X,Y)}{P(X)} \quad . \tag{1.24}$$

A straightforward extension of Equations (1.23) and (1.24) is the well-known Bayes' Rule, which inverts the conditional distributions:

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)} \quad . \tag{1.25}$$

When random variables are continuous variables, similar definitions are given to the probability density functions. In brief, the marginal density functions are obtained from:

$$p_x(x) = \int_{-\infty}^{+\infty} p_{x,y}(x, y)dy \tag{1.26}$$

$$p_y(y) = \int_{-\infty}^{+\infty} p_{x,y}(x, y)dx \, . \tag{1.27}$$

The relationship between joint density function and conditional density functions are given by:

$$p_{x,y}(x, y) = p_{x|y}(x \mid y)p_y(y) = p_{y|x}(y \mid x)p_x(x) \tag{1.28}$$

Therefore, Bayes' Rule is easily derived from Equation (1.28):

$$p_{x|y}(x \mid y) = \frac{p_{y|x}(y \mid x)p_x(x)}{p_y(y)} \quad . \tag{1.29}$$

Bayes' Rule has been very useful in decision theory, estimation, learning, pattern classification, etc. For instance, in estimation problem, Equation (1.29) can be used to find the best estimation of unknown variable x from given observations of known variable y.

Finally, we would like to stress that the above definitions and derivations hold not only for random variables but also for random vectors.

1.3.2 Bayesian Estimation Methods

In the Bayesian framework, we refer to the probability density $p_{\boldsymbol{x}}(\boldsymbol{x})$ for the vector $\boldsymbol{x}$ to be estimated as the *a priori* density. This is because this density fully specifies our knowledge about $\boldsymbol{x}$ prior to any observation of the measurement $\boldsymbol{y}$. The conditional density $p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{x})$, which fully specifies how $\boldsymbol{y}$ contains the information about $\boldsymbol{x}$, is typically called the likelihood function. Generally, this likelihood function, $p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{x})$, is not directly given but can be inferred from the measurement model. The conditional density $p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ is called the *a posteriori* density because it specifies the behavior of $\boldsymbol{x}$ in accordance to the measurement $\boldsymbol{y}$.

We refer to the estimation of $\boldsymbol{x}$ as the estimator denoted by $\hat{\boldsymbol{x}}$. Let $\hat{\boldsymbol{x}}(\boldsymbol{y})$ denote the estimation of $\boldsymbol{x}$ based on measurements $\boldsymbol{y}$. In a Bayesian framework, we choose the estimator so that it optimizes some performance criterion. For instance, a cost function $C(\boldsymbol{x}, \hat{\boldsymbol{x}})$ is often used to evaluate the cost of estimating an arbitrary vector $\boldsymbol{x}$ as $\hat{\boldsymbol{x}}$. The best estimator is defined as the one that minimizes the average cost given $\boldsymbol{y}$:

$$\begin{aligned} \hat{\boldsymbol{x}}(\boldsymbol{y}) &= \arg\min_{\boldsymbol{a}} E[C(\boldsymbol{x}, \boldsymbol{a}) \mid \boldsymbol{y}] \\ &= \arg\min_{\boldsymbol{a}} \int_{-\infty}^{+\infty} C(\boldsymbol{x}, \boldsymbol{a}) p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x} \mid \boldsymbol{y}) d\boldsymbol{x} \end{aligned} \tag{1.30}$$

As Equation (1.30) indicates, the *a posteriori* density function plays an important role in Bayesian estimation.

Consider the cost function given by:

$$C(a,\hat{a}) = \begin{cases} 1 & |a-\hat{a}| > \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{1.31}$$

which uniformly penalizes all estimation errors with magnitude bigger than ε. Substitution of Equation (1.31) into Equation (1.30) gives:

$$\begin{aligned} \hat{\boldsymbol{x}}(\boldsymbol{y}) &= \arg\min_{\boldsymbol{a}}[1 - \int_{a-\infty}^{a+\infty} p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x} \mid \boldsymbol{y})d\boldsymbol{x}] \\ &= \arg\max_{\boldsymbol{a}} \int_{a-\infty}^{a+\infty} p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x} \mid \boldsymbol{y})d\boldsymbol{x} \quad . \end{aligned} \tag{1.32}$$

This estimator is called minimum uniform cost (MUC) estimation. It finds the interval of length 2ε where the *a posteriori* density is the most concentrated. If we let ε get sufficient small, MUC estimation essentially approaches the point corresponding to the peak of the *a posteriori* density function. This is the so called maximum *a posteriori* (MAP) estimation:

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}}(\boldsymbol{y}) = \arg\max_{\boldsymbol{a}} p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{a} \mid \boldsymbol{y}) . \tag{1.33}$$

The MAP method has been one of the most widely used estimation techniques because of its simplicity and computability. The maximum of $p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ may be obtained by differentiating the function:

$$\frac{\partial}{\partial \boldsymbol{x}} p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x} \mid \boldsymbol{y}) = 0 . \tag{1.34}$$

Furthermore, according to Bayes' Rule, substituting Equation (1.29) into Equation (1.34) leads to:

$$\frac{\partial}{\partial \boldsymbol{x}} \left(\frac{p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y} \mid \boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{x})}{p_{\boldsymbol{y}}(\boldsymbol{y})} \right) = 0 . \tag{1.35}$$

Since $p_{\boldsymbol{y}}(\boldsymbol{y})$ is not a function of x, Equation is equivalent to

$$\frac{\partial}{\partial \boldsymbol{x}} \left(p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y} \mid \boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{x}) \right) = 0 . \tag{1.36}$$

It is sometimes more convenient to maximize some monotonic function of *a posteriori* density than itself. For example, a normal log function is often used. Maximizing $\ln p_{x|y}(\boldsymbol{x}|\boldsymbol{y})$ is often easier than maximizing $p_{x|y}(\boldsymbol{x}|\boldsymbol{y})$ directly. In this case, the MAP equation of (1.36) is rewritten as:

$$\frac{\partial}{\partial \boldsymbol{x}} \ln p_{y|x}(\boldsymbol{y} \mid \boldsymbol{x}) + \frac{\partial}{\partial \boldsymbol{x}} \ln p_x(\boldsymbol{x}) = 0. \tag{1.37}$$

Equation (1.37) indicates that MAP estimation is only dependent on *a priori* density $p_x(\boldsymbol{x})$ of $\boldsymbol{x}$ and likelihood function $p_{y|x}(\boldsymbol{y}|\boldsymbol{x})$.

Let us consider a rather special case. If our knowledge indicates that $\boldsymbol{x}$ is only an unknown constant (such as some parameter), instead of a random vector, the MAP method reduces to be the maximum likelihood (ML) method. Because in this situation the *a priori* function $p_x(\boldsymbol{x})$ becomes a constant, ML estimation can then be obtained by:

$$\frac{\partial}{\partial \boldsymbol{x}} \ln p_{y|x}(\boldsymbol{y} \mid \boldsymbol{x}) = 0. \tag{1.38}$$

This illustrates the close relationship between ML and MAP methods. A formal derivation of ML estimation can be found in any good book on statistics (Papoulis 1991).

As a final remark, with a different cost function, we may achieve many other estimation methods, such as minimum absolute-error (MAE) estimation, minimum mean square error (MMSE) estimation, etc. Choosing a suitable method for a particular problem depends on a variety of factors, including the knowledge about the unknown $\boldsymbol{x}$ and observation $\boldsymbol{y}$, the importance of errors, the computability of optimization, etc.

1.3.3 Markov Random Field (MRF)

A *priori* information about $\boldsymbol{x}$ is very important and useful in estimating $\boldsymbol{x}$. Intuitively, the more *a priori* information we know about $\boldsymbol{x}$, the better estimation we can expect to obtain. In addition, as Equation (1.37) indicates, knowing the exact form of the *a priori* function is necessary for MAP method to be applied. In many problems, the *a priori* function is not always given, or is very difficult to obtain. Markov Random Field (MRF) can be used in many problems to model the *a priori* function based on some assumptions.

MRF defines a field of random variables whose relationships are represented by a neighborhood structure. In discrete case, the field is often represented by a lattice, composed of a set of sites or nodes in a graph. For example, an m×n image can be represented as an m×n lattice where each site corresponds to a pixel in the image. One of the most important properties of MRF is local characteristics, which asserts conditional independence: given its immediate neighbors, a random variable of MRF is independent of all other variables. Mathematically, this property is expressed by:

$$P(\boldsymbol{x}_s \mid \boldsymbol{x}_r, \forall r \neq s) = P(\boldsymbol{x}_s \mid \boldsymbol{x}_r, \forall r \in \boldsymbol{N}_s) \tag{1.39}$$

where s (a site) represent a variable under consideration and $\boldsymbol{N}_s$ is its neighborhood. The left-hand side of Equation (1.39) is the conditional probability of the variable given the rest of other variables in the lattice. The right-hand side of Equation (1.39) is the conditional probability of that variable given the states of all variables in its neighborhood. The neighborhood $\boldsymbol{N}_s$ is defined as a collection of "neighbors" of the site s. Neighbors are dependent on the order of the neighborhood to be considered. For instance, a first-order neighborhood of a pixel (i, j) in an image is composed of pixels immediately above, below, to the left and to the right of the given pixel (i, j), i.e., $\boldsymbol{N}_{ij}=\{(i\text{-}1, j), (i, j\text{-}1), (i, j+1), (i+1, j)\}$. A second-order neighborhood includes the first order neighbors as well as pixels that are diagonally across from the given pixel, i.e., $\boldsymbol{N}_{ij}=\{(i\text{-}1, j), (i, j\text{-}1), (i, j+1), (i+1, j), (i\text{-}1, j\text{-}1), (i\text{-}1, j+1), (i+1, j\text{-}1), (i+1, j+1)\}$. Higher order neighborhood can be similarly defined, but they are rarely used in practical applications. Let us introduce another term, clique. A *clique* c is defined as a subset of sites in $\boldsymbol{S}$ (the space of s) and neighborhood $\boldsymbol{N}$ (Li 1995). It consists either of a single site $c=\{s\}$, or of a pair of neighboring sites $c=\{s, s'\}$, or of a triple of neighboring sites $c=\{s, s', s''\}$, and so on. The collections of single-site, pair-site and triple-site cliques are denoted by C_1, C_2, C_3, respectively:

$$C_1 = \{s \mid s \in \boldsymbol{S}\} \tag{1.40}$$

$$C_2 = \{\{s, s'\} \mid s \in \boldsymbol{S}, s' \in \boldsymbol{N}_s\} \tag{1.41}$$

$$C_2 = \{\{s, s', s''\} \mid s, s', s'' \in \boldsymbol{S} \text{ are neighbors to each other}\} \ . \tag{1.42}$$

The collection of all cliques is denoted by $\boldsymbol{C}$:

$$C = C_1 \cup C_2 \cup C_3 \cdots \tag{1.43}$$

The Hammersley-Clifford theorem provides us a systematic way for modeling MRF (Besag 1974; Geman and Geman 1984; Besag 1986; Li 1995). The theorem establishes equivalence between MRF and Gibbs random fields (GRF) and allows us to construct MRF using a GRF model. With this equivalence, the conditional distribution model of MRF is converted to a joint distribution model of GRF. The Gibbs distribution is given as:

$$P(\boldsymbol{x}) = \frac{1}{Z}\exp(-\frac{E(\boldsymbol{x})}{T}) \ , \tag{1.44}$$

where T is a constant called the temperature. Z is a normalizing constant,

$$Z = \sum_{x}\exp(-\frac{E(\boldsymbol{x})}{T}) \tag{1.45}$$

and also known as partition function. It ensures that the sum of all probabilities $P(\boldsymbol{x})$ is equal to 1. $E(\boldsymbol{x})$ is called the energy function and expressed as a sum of clique potentials over all possible cliques $\boldsymbol{C}$:

$$E(\boldsymbol{x}) = \sum_{c\in C} V_c(\boldsymbol{x}) \ . \tag{1.46}$$

The value of $V_c(\boldsymbol{x})$ depends on local configuration on the clique c.

In recent years, MRF has been of great interests in many areas, such as computer vision and image processing (Geman and Geman 1984; Geiger and Girosi 1991; Kapur, Grimson et al. 1998). It works particularly well when a problem involves uncertainties or constraints that are difficult to solve by other methods. MRF offers a great deal of flexibility for modeling different types of *a priori* information, such as local smoothness (Geman and Geman 1984; Geiger and Girosi 1991), geometric relationship (Kapur, Grimson et al. 1998), etc.

1.4 Overview and Contributions

Due to limitations of electron microscopy, a lot of biological structures can not be reconstructed directly by electron tomography. Furthermore, when a specimen is stained with heavy metals, the imaging function no longer satisfies the condition of the projection-slice theorem. Tomographic reconstruction methods won't work well in these situations. This thesis attempts to explore and develop a new reconstruction technique that can be applied to a large variety of structures, especially to ones that are intractable using electron tomography. It may serve as an alternative method to electron tomography in structural biology.

This thesis addresses reconstruction problems in a more general context, covering both computer vision based reconstruction and tomographic reconstruction. However, specific attention is given to 3D reconstruction of biological cellular structures, such as cell cytoskeleton, from electron microscope images taken at different tilt angles. The thesis is composed of two parts. At the first part of the thesis, we study reconstruction from a computer vision perspective. As mentioned earlier, the reconstruction by computer vision approaches doesn't require exact knowledge of an imaging function, and 3D information is extracted mainly from the geometric relationship between images. This shields us from complexity of the imaging function. Interestingly, computer vision based reconstruction has rarely been used in biological applications. As one of our goals, we hope that this thesis will bring computer vision based techniques to the attention of structural biology communities.

We first derive the geometric relationship that describes the imaging system of electron microscope. Then, the thesis focuses on finding structural relationships between 3D objects and 2D images. Based on a stereo vision idea, a stereo reconstruction method is presented, which reconstructs 3D structure from only two images taken at two different tilt angles. A multiple-feature matching algorithm is employed. The algorithm introduces a complex-value representation that takes into account both feature value and the certainty of this value. A volumetric reconstruction method is then developed by

extending stereo reconstruction to reconstruction from multiple images. The method integrates a sequence of 3D reconstruction from different image pair. The integration is an optimization process that attempts to find the best estimation for each location in a 3D object space. As an example of applications to biological structure, the proposed approach is tuned and applied to 3D reconstruction of cell cytoskeleton structure from multiple images taken at different tilt angles on an electron microscope. The reconstruction demonstrates the feasibility, reliability and flexibility of the proposed approach.

The second part of this thesis focuses on a special tomographic reconstruction, namely discrete tomography. Discrete tomography deals with a reconstruction problem in which the object to be reconstructed is composed of a discrete set of materials each with uniform values. Such condition or constraint is termed discreteness in this thesis. The discreteness can be confirmed from observations that each type of biological structure tends to have one uniform value in its electron microscopy images.

Discrete tomography has recently been of interest to many fields, such as medical imaging (Frese, Bouman et al. 1998; Vardi and Lee 1998; Chan, Herman et al. 1999; Herman and Kuba 1999). This thesis studies discrete tomography in a general format. The proposed approach starts with an explicit model of the discreteness constraint. within a Bayesian framework. The problem of discrete tomography is converted into a Bayesian labeling process, assigning a unique label to each object point. The formula also establishes a mathematical relationship between discrete tomography and conventional tomography, suggesting that discrete tomography may be a more generalized form of tomography and conventional tomography is only a special case of such generalization. An expectation-maximization (EM) algorithm is developed to solve discrete tomography. An efficient solution for discrete tomography is proposed in the light of the duality of tomographic reconstruction methods discussed in earlier section.

As a summary, this thesis studies 3D reconstruction from 2D images using imaging function as guidance. The thesis is primarily composed of two parts that corresponds to

two different treatments to the imaging function. In the first part of thesis, a computer vision based reconstruction is developed, which doesn't require any knowledge of the imaging function. The second part of thesis deals with discrete tomography in which an approximation of the imaging function is used and some constraints are considered.

The key contributions of this thesis are closely related to these two parts as well:

- The thesis develops computer vision based reconstruction methods for both stereo images and multiple images. The proposed volumetric reconstruction method can be applied to a large number of biological structures, demonstrated by the reconstruction of the cell cytoskeleton from multiple images. The approach offers an alternative reconstruction technique to electron tomography in structural biology.
- The thesis proposes an integrated model for discrete tomography and develops a Bayesian labeling method and an efficient algorithm. A mathematical relationship between discrete tomography and conventional tomography is established and suggests a close tie between them.

Chapter 2

Alignment

2.1 Overview

As described in the previous Chapter, 3D information of objects is embedded within 2D images in the form of disparities. For instance, 3D positions can be recovered by triangulation from the disparities among 2D images (Marr and Poggio 1979; Grimson 1980; Grimson 1985). However, due to inherent deviations of imaging device or movements introduced during image acquisition process, images obtained are rarely in perfect alignment with respect to each other. Disparities among images are often contaminated and contain in-plane translations, rotations, and/or even scaling change of each image with respect to the other. The alignment is a process that compensates these movements so that the images will share a common coordinate system and the disparities will truly reflect 3D information of the objects. The alignment is a necessary pre-process before reconstruction can be conducted.

The fiducial-marker alignment method, developed by Lawrence (Luther, Lawrence et al. 1988; Lawrence 1992), is used widely in electron tomography field. It requires users to manually pick several markers from each image. In biological applications, markers are often made by embedding gold particles into specimen before taking images. Image movements, including in-plane translation, rotation and scaling, are then calculated by minimizing a least-square error between measured marker coordinates in images and their true coordinates projected from 3D space. However, this least-mean-square-error (LMSE) method results in a set of non-linear equations. Solving these non-linear equations is often computationally difficult and time-consuming.

In this Chapter, we propose a new and iterative algorithm with which the solutions for alignment can be obtained from a set of linear equations, as opposed to non-linear ones. The proposed algorithm proves to be fast and robust.

2.2 Iterative Algorithm

The proposed algorithm also adopts the idea of using fiducial-markers. Markers need to be picked manually in each image. The number of markers needed can be as few as two in each image. However, more markers improve accuracy and quality of the alignment. We have developed a graphic user interface (GUI) program to help users with this process. The program is described in Appendix A. In addition, since electron microscope imaging system is modeled by parallel projection, the scaling factor is negligible. Thus, only in-plane translations and rotations need to be computed in the alignment.

Let us assume markers have been picked from each image and denoted by $(x_{i,j}, y_{i,j})$, where (x, y) is the coordinate in each image plane. Let i=1,2,..,m denote the ith image plane with corresponding tilt angle θ_i . Let j=1,2,...,n denote the jth marker. Denote XYZ as the common 3D coordinate reference frame where 3D objects are imaged. With a parallel projection model, the geometric relationship of a point's position in 3D space and its coordinates in the image plane is given as follows:

$$\underline{\mathbf{x}}_{i,j} = \mathbf{R}_i \ \mathbf{P} \ \mathbf{T}_i \ \underline{\mathbf{X}}_j + \underline{\mathbf{d}}_i \ , \qquad (i = 1,2,...,m \ ; \ j = 1,2,...,n.) \tag{2.1}$$

where $\underline{\mathbf{x}}_{i,j}$ is the image coordinate vector $[x_{i,j}, y_{i,j}]^T$ (where ^{T} represents transpose) of the projection of the jth marker in the ith image plane, where ^{T} represents transpose. $\underline{\mathbf{X}}_j$ is the position vector $[X_j, Y_j, Z_j]^T$ of the jth marker in the 3D common coordinate system. T_i is a 3×3 rotation matrix representing the tilting operation with tilt angle θ_i. $\mathbf{P}$ is a 2×3 projection matrix. $\mathbf{R}_i$ is a 2×2 rotation matrix describing the in-plane rotation of the ith image plane with angle α_i. $\underline{\mathbf{d}}_i$ is a two-dimensional vector representing the in-plane translation of the ith image plane. Let us assume the tilting axis is always the Y-axis and parallel to image planes. The transformation matrices in Equation (2.1) have this form:

$$\mathbf{T}_i = \begin{bmatrix} \cos\theta_i & 0 & -\sin\theta_i \\ 0 & 1 & 0 \\ \sin\theta_i & 0 & \cos\theta_i \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{R}_i = \begin{bmatrix} \cos\alpha_i & -\sin\alpha_i \\ \sin\alpha_i & \cos\alpha_i \end{bmatrix}. \qquad (2.2)$$

There is one very important observation from Equation (2.1): the common coordinate system is translation-invariant. It means that if the common coordinate system is translated, we can always find in-plane translations $\underline{\mathbf{d}}_i$ such that the image coordinates $\underline{\mathbf{x}}_{i,j}$ remains the same. Therefore, the in-plane translations $\underline{\mathbf{d}}_i$ are merely relative values with respect to the common coordinate system we choose. This property holds mainly because of the parallel projection model of our imaging system.

Bearing this property in mind, let us proceed to solve the alignment. The alignment can be considered to solve an inverse estimation problem. It seeks to retrieve unknown inputs $\underline{\mathbf{X}}_j$ and the best set of parameters, $\mathbf{R}_i$'s and $\underline{\mathbf{d}}_i$'s, from known output markers' position $\underline{\mathbf{x}}_{i,j}$ in each image plane. The LMS method is adopted to minimize the mean square error (MSE) between the measured marker coordinates in the images and their true coordinates projected from the 3D space:

$$\text{MSE} = \sum_i \sum_j \left\| (\underline{\mathbf{x}}_{i,j} - \mathbf{R}_i \, \mathbf{P} \, \mathbf{T}_i \, \underline{\mathbf{X}}_j - \underline{\mathbf{d}}_i) \right\| . \qquad (2.3)$$

To make the problem simpler, let us assume the tilt angle θ_i 's are known and accurate. Furthermore, due to the translation-invariant property, the in-plane translations $\underline{\mathbf{d}}_i$ in Equation (2.3) can be eliminated by choosing a proper reference point. It is easy to derive that such reference point is the centroid of the picked markers. Its projection in each image plane is also the centroid of the markers' projections in each image plane. They can be obtained by:

$$\underline{\mathbf{X}}_r = \frac{\sum_{j=1}^{n} \underline{\mathbf{X}}_j}{n} \quad ; \qquad \underline{\mathbf{x}}_{i,r} = \frac{\sum_{j=1}^{n} \underline{\mathbf{x}}_{i,j}}{n} , \quad \text{for } i = 1, 2, \ldots, m. \qquad (2.4)$$

These reference points are then treated as the origins of the common coordinate system and each image planes. New coordinates are obtained by translations:

$$\underline{\mathbf{X}}'_j = \underline{\mathbf{X}}_j - \underline{\mathbf{X}}_r \quad ; \qquad \underline{\mathbf{x}}'_{i,j} = \underline{\mathbf{x}}_{i,j} - \underline{\mathbf{x}}_{i,r} \qquad \text{for } i = 1,2,...,m;\ \text{j} = 1,2,...,\text{n}. \tag{2.5}$$

Under new coordinate systems, Equation (2.1) then becomes

$$\underline{\mathbf{x}}'_{i,j} = \mathbf{R}_i\ \mathbf{P}\ \mathbf{T}_i\ \underline{\mathbf{X}}'_j \ . \qquad (i = 1,2,...,m\ ;\ j = 1,2,...,n.) \tag{2.6}$$

Unknown variables in Equation (2.6) are in-plane rotation R_i and the 3D coordinate $\underline{\mathbf{X}}'_j$ in the new common coordinate system. The corresponding MSE equation is given by:

$$\text{MSE} = \sum_i \sum_j \left\| (\underline{\mathbf{x}}'_{i,j} - \mathbf{R}_i\ \mathbf{P}\ \mathbf{T}_i\ \underline{\mathbf{X}}'_j) \right\| \ . \tag{2.7}$$

The direct differentiation on Equation (2.7) will lead to a set of non-linear equations, which is not good for computation. However, by some manipulation of Equation (2.6), we find a linearization scheme to get around this non-linear issue. By moving the rotation matrix R_i to the left hand side of Equation (2.6), a new MSE function is formed and given by:

$$\text{MSE} = \sum_i \sum_j \left\| (\mathbf{R}_j^{-1}\ \underline{\mathbf{x}}'_{i,j} - \ \mathbf{P}\ \mathbf{T}_i\ \underline{\mathbf{X}}'_j) \right\| \ , \tag{2.8}$$

where $\mathbf{R}_j^{-1}$ is the inverse matrix of $\mathbf{R}_j$. Furthermore, let us assume the rotation angle α_i is very small for now. $\mathbf{R}_j^{-1}$ can be approximated by:

$$\mathbf{R}_j^{-1} = \begin{bmatrix} \cos\alpha_i & \sin\alpha_i \\ -\sin\alpha_i & \cos\alpha_i \end{bmatrix} \approx \begin{bmatrix} 1 & \alpha_i \\ -\alpha_i & 1 \end{bmatrix} . \tag{2.9}$$

The LMS method on this MSE function of Equation (2.8) will then result in a set of linear equations, obtained by setting partial differentiation of α_i, X'_j, Y'_j, Z'_j to be zeros:

$$\begin{cases} -\sum_j^n \left(x_{i,j}^2 + y_{i,j}^2\right)\alpha_i + \sum_j^n \left(y_{i,j}\cos\theta_i \cdot X'_j\right) - \sum_j^n \left(x_{i,j} \cdot Y'_j\right) + \sum_j^n \left(y_{i,j}\sin\theta_i \cdot Z'_j\right) = 0 \quad , i = 1,2,...,m. \\ -\sum_i^m \left(y_{i,j}\cos\theta_i \cdot \alpha_i\right) + \left(\sum_i^m \cos^2\theta_i\right) \cdot X'_j + \left(\sum_i^m \sin\theta_i \cos\theta_i\right) \cdot Z'_j = \sum_i^m x_{i,j}\cos\theta_i \quad , j = 1,2,...,n. \\ -\sum_i^m \left(x_{i,j} \cdot \alpha_i\right) - nY'_j = -\sum_i^m y_{i,j} \quad , j = 1,2,...,n. \\ -\sum_i^m \left(y_{i,j}\sin\theta_i \cdot \alpha_i\right) + \left(\sum_i^m \sin\theta_i \sin\theta_i\right) \cdot X'_j + \left(\sum_i^m \sin^2\theta_i\right) \cdot Z'_j = \sum_i^m x_{i,j}\sin\theta_i \quad , j = 1,2,...,n. \end{cases} \tag{2.10}$$

This (m+3n)×(m+3n) linear equation simply has the form of:

$$\mathbf{Av} = \mathbf{b}, \tag{2.11}$$

where the unknown vector v = $[\alpha_1, \alpha_2, \ldots, \alpha_m, X'_1, X'_2, \ldots, X'_n, Y'_1, Y'_2, \ldots, Y'_n, Z'_1, Z'_2, \ldots, Z'_n]^T$, including the rotation angles α_i and 3D coordinates of the markers. **A** is coefficient matrix and **b** is some constant vector. By solving the linear equation, both rotation angles α_i and 3-D coordinates of the markers are obtained. Even though 3-D coordinates of the markers are of no use in the image alignment, they can yield some useful information about the range of the object space.

One important issue here is the approximation introduced in Equation (2.9). In practice, since in-plane rotations can be arbitrary, the approximation may not be correct for large angle of α_i. In order to get around this issue, the algorithm introduces an iterative procedure that can be used to obtain solutions for any rotation angles. The iterative procedure is described as follows:

Step 1: Assume small angle of α_i and then solve for the α_i's from linear equation of Equation (2.11).

Step 2: Rotate the marker's coordinates $\underline{\mathbf{x}}'_{i,j}$ based on the obtained α_i's. New set of coordinates is obtained by rotating each image plane with corresponding α_i. These coordinates are considered new positions of the markers.

Step 3: Perform the LMS method based on Equation (2.8) and Equation (2.11) to obtain the new set of rotation angle α_i. Total rotation angle is the sum of the new α_i and the previous ones. In other words, the rotation angle α_i is accumulated during each iteration.

Step 4: Iterate the Step 2 and Step 3. The iteration stops when MSE is less than some given threshold.

With this iterative procedure, our alignment method has the ability to recover the incorrect alignment of very large rotation angle α_i, even up to ±90°. The method proves to converge very fast. For instance, given the synthetic data with α_i ranging between -90° to +90°, the proposed iterative algorithm reaches ±0.01° precision within only 4 to 5 iterations.

2.3 Performance Studies

Experiments are designed to study the performance of the proposed algorithm. In all experiments, synthetic data are used so that their true positions or values are known. We use root mean square errors (RMSE) to measure the accuracy of obtained results with respect to their true values. RMSE is given by:

$$\text{rmse} = \sqrt{\frac{1}{n}\sum_{i}\left(u_{im} - u_{it}\right)^2} \tag{2.12}$$

where u_m is the measured value and u_t is the true value. Due to the property of translation-invariance, obtained in-plane translations are relative values. Therefore, we primarily measure the RMSE of obtained in-plane rotations in following experiments.

2.3.1 Wide range of rotation angles α

One important assumption in our algorithm is that the rotation angles α_i are small so that non-linear equations can be approximated by linear ones. The algorithm later introduces an iterative procedure to deal with large rotation angles.

This experiment uses synthetic data of 10 marker points, 6 projected images with tilt angles between [-20°, 20°]. Figure 5 exhibits the performance measure of our algorithm on different range of rotation angles. The result clearly demonstrates the proposed iterative algorithm is capable of dealing with a wide range of rotation angles. The data with a small range of rotation angles need less iterations than ones with a large range of rotation angles. The convergence is very fast even in the case of large angles. The accurate solutions are obtained within 4 or 5 iterations.

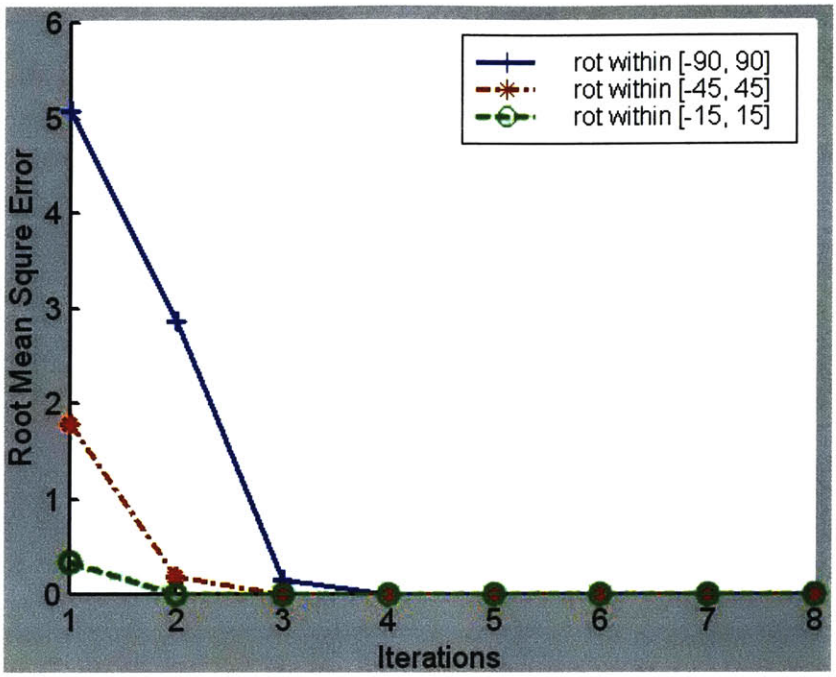


Figure 5 Performance on different range of rotation angles

2.3.2 Imperfect tilt angles

As described in previous section, the algorithm also assumes that the tilt angles are known and accurate. In practice, the tilt angles is read out from the goniometer, and often not very accurate. One would wonder how well the proposed alignment algorithm performs on imperfect tilt angles. This experiment is intended to address this concern.

This experiment uses a set of synthetic data similar to the previous experiment. The only difference is that some random noise is added into each tilt angle. The level of noise is set to be 20% of interval tilt angles between neighboring images. Figure 6 shows the performance measure of the algorithm on these data. Interestingly, the result or the performance of the algorithm doesn't seem to be affected much. We barely see any difference in comparison to the previous experiment in which the tilt angles are accurate. This interesting observation seems to imply that the estimation of in-plane translations and rotations may be irrelevant to tilt angles.

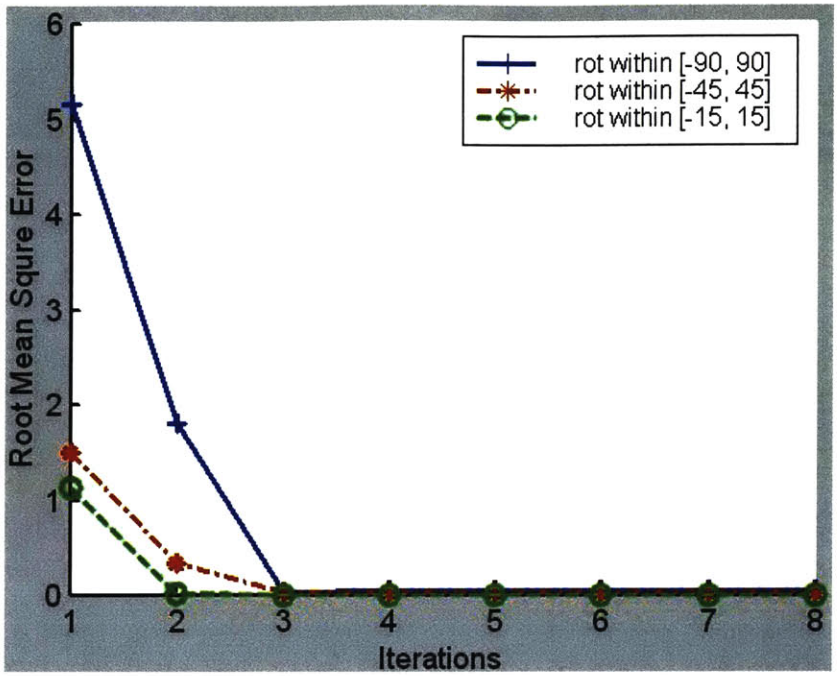


Figure 6 Performance on imperfect tilt angles

In fact, the LMS method can be used to estimate true tilt angles (*i.e.* θ_i) in Equation (2.7), along with in-plane translations and rotations, if tilt angles are unknown. In other words, the picked marker points already contain all information about tilt angles. They alone are sufficient to estimate either the tilt angles, or alignment-associated in-plane translations or rotations. Therefore, imperfect tilt angles won't affect the estimation of in-plane translations or rotations, which is in accordance with our experiment. However, it is not meaning that we don't need tilt angles. Knowing, even imperfect, tilt angles helps greatly in estimating in-plane translations or rotations. It improves the algorithm's convergence and helps to find the best solutions quickly. If tilt angles, along with in-plane rotations and translations, are completely unknown, the LMS method will lead to a set of complicated highly non-linear equations, which would be very difficult to solve.

2.3.3 Number of markers

Another interesting question in the marker-based alignment method is that how many markers one need to pick in each image. This experiment is designed to help to answer this question.

This experiment also uses similar synthetic data as previous experiments, except that we vary the number of markers this time and randomly choose the rotation angle within the range of [-90°, 90°]. The results are exhibited in Figure 7. The observation seems to suggest that more marker points improve the performance and convergence of the algorithm. This observation matches our expectation. In addition, Figure 7 also depicts that we obtain almost the same results in each case after a certain number of iterations. It implies that more marker points may not necessarily help a lot. These two seemly contrary observations are both true in practice. Since the markers need to be picked manually, this process is very error-prone. The markers picked in one image may not perfectly match to the markers picked from the other. These imperfect matches will lead to incorrect alignment. More markers may help to build up statistics and reduce the incorrectness, but in the mean time, more markers also mean more chance to expose to that error-prone process. As a tentative conclusion, we recommend the number of marker points be around 6 for each image.

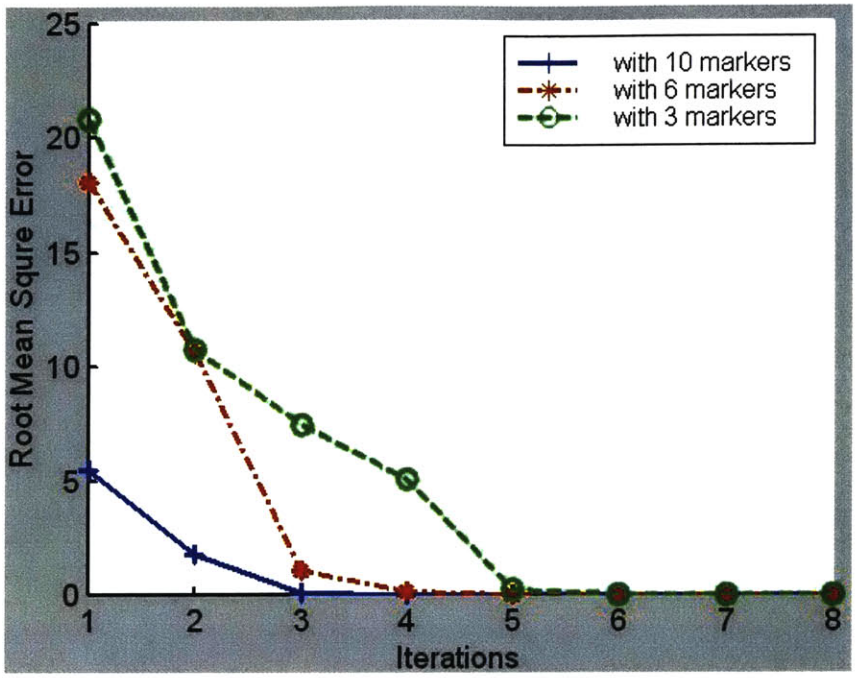


Figure 7 Performance on different number of marker points

2.4 Summary

Alignment is a necessary process before 3D reconstruction. This Chapter proposed an iterative algorithm to solve the alignment problem, including in-plane translations and rotations. By making some approximations and reorganizing the error function, the algorithm converts an non-linear problem to a linear one. An iterative scheme is employed to compensate or even eliminate errors introduced by approximation. Verified by a number of experiments, the proposed algirthm proves to be fast, robust and accurate.

Chapter 3

Reconstruction from Two Images

The alignment in Chapter 2 establishes a necessary starting point for reconstruction. Hereafter, we will assume all images have been aligned correctly.

In this Chapter, the most basic form of computer vision based reconstruction, stereo reconstruction (from two images), is reviewed and explored. A new algorithm is proposed. The algorithm is designed to take advantage of various structural characteristics of the objects. The work presented in this Chapter has also been published in the paper of (Cheng, Hartmink et al. 2000).

3.1 Introduction

Based on triangulation, 3D information, such as depth, of the objects can be reconstructed from images taken from different perspectives (Barnard and Fischler 1982; Dhond and Aggarwal 1989; Faugeras 1993). In its simplest form, two images may be sufficient to recover the depth information of the objects. The depth is actually embedded within 2D images in the form of coordinate differences, or formally termed the disparity (Marr and Poggio 1979; Grimson 1980; Grimson 1985). Reconstruction from two images, called stereo reconstruction or stereo vision, has been studied extensively in computer vision field. Many approaches have been proposed to find disparities between paired images. They can roughly be grouped into two categories (Barnard and Fischler 1982; Vosselman 1992; Fua 1993): feature-based and area-based methods. The feature-based method first extracts features, such as edges/lines (Ohta and Kanada 1985; McIntosh and Mutch 1988; Horaud and Skordas 1989), zero-crossings (Grimson 1980; Grimson 1985), or contours

(Marapane and Trivedi 1994) from each image and then matches these features between two images. In general, the feature-based method is more accurate and less sensitive to noise. However, these features, called high-level descriptors, provide only an abstract description about the objects in the image. The features often only exist at certain pixels in the image. The disparities obtained by the feature-based method are sparsely distributed. In order to obtain a disparity map of the whole image, interpolation techniques are used. On the other hand, the area-based method uses only the intensity value as the matching element. To reduce the noise sensitivity, correlation methods, such as cross-correlation or sum of squared difference (SSD), are often employed (Hannah 1989; Fua 1993; Okutomi and Kanada 1993). Since the intensity value is a low-level descriptor, it is not necessarily related to the presence of the objects in the image. Thus, the structural information is not explicitly considered in the matching. The advantage, however, is that a dense disparity map can be obtained without interpolation because the intensity value is available for each pixel.

The stereo reconstruction approach proposed in this chapter attempts to combine advantages of both feature-based and area-based methods. It finds multiple features using both high-level and low-level descriptors of each image. Stereo matching is then conducted on a point-to-point basis. The approach allows us to obtain a dense disparity map without interpolation while the structural information is considered explicitly so as to reduce noise. Furthermore, the approach introduces a complex representation so that the certainty information is always naturally incorporated and considered in the process. Finally, the approach is specifically tuned to the reconstruction of the actin-based cytoskeleton structures. Their structural information is explicitly incorporated into the analysis and reconstruction process.

3.2 Geometry of the Imaging System

Cell cytoskeleton micrographs are obtained in a Transmission Electron Microscope (TEM) at different tilt angles. Each image is a projection of the 3D structure onto a 2D

plane. Figure 8 shows a schematic diagram of an electron microscope imaging system. Although there are two models for imaging systems, parallel projection and perspective projection (Horn 1986; Faugeras 1993), parallel projection is adopted because of the small field of view in our imaging system.

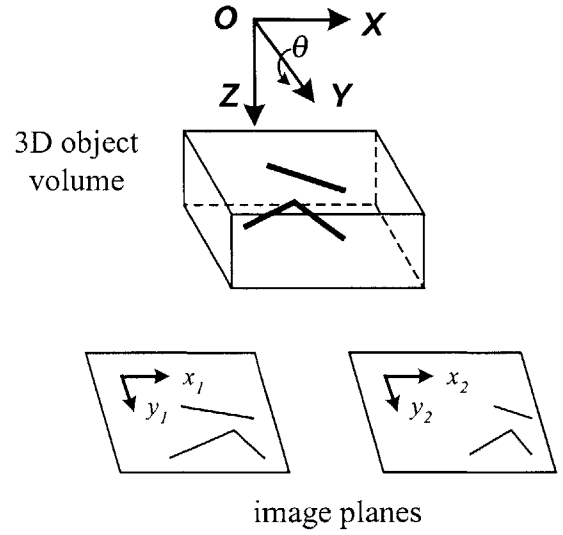


Figure 8 Geometry of the imaging system

As illustrated in Figure 8, the specimen coordinate reference frame is denoted XYZ and the image planes are denoted x_iy_i (subscript i is the index of the images). Let us assume that the tilting axis is defined to be the Y-axis of the specimen coordinates and the images have been aligned correctly via algorithm presented in the previous Chapter. With the parallel projection model, the geometric relationship of the point and its projection in the image plane is simply obtained, which can be expressed as follows:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} X\cos\theta_1 + Z\sin\theta_1 \\ Y \end{bmatrix} \tag{3.1}$$

where θ_1 is the tilt angle. If the specimen is tilted to another angle θ_2, the projection of the same point $[X, Y, Z]^T$ is given as:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} X\cos\theta_2 + Z\sin\theta_2 \\ Y \end{bmatrix}. \tag{3.2}$$

By solving X, Y, Z from Equations (3.1) and (3.2), we obtain

$$X = \frac{x_2 \sin\theta_1 - x_1 \sin\theta_2}{\sin(\theta_1 - \theta_2)} \tag{3.3}$$

$$Y = y_1 = y_2 \tag{3.4}$$

$$Z = \frac{x_1 \cos\theta_2 - x_2 \cos\theta_1}{\sin(\theta_1 - \theta_2)}. \tag{3.5}$$

Without the loss of generality, let us consider the situation where two projections (x_1, y_1) and (x_2, y_2) are taken from two symmetric tilt angles, i.e., $\theta = \theta_1 = -\theta_2$. Equations (3.3) and (3.5) are simplified as

$$X = \frac{x_2 + x_1}{2\cos\theta} \tag{3.6}$$

$$Z = \frac{x_1 - x_2}{2\sin\theta} = \frac{d}{2\sin\theta}, \tag{3.7}$$

where $d=(x_1 - x_2)$ is the disparity. The Equation (3.7) indicates that the depth Z is proportional to the disparity d. With Equation (3.7), the problem of the reconstruction of the 3D objects now becomes one of computing the disparity d.

3.3 Description of Method

3.3.1 Feature representation

Three features are extracted from the images: local direction, local phase, and intensity. Local direction represents the orientation of a structure where the point lies, and is denoted by the angle α ranged between 0 to π. Local phase indicates the degree to which the point is part of an edge or a line in the object (more details are discussed in the following subsections A and B), and is denoted by the angle β ranged between 0 and 2π. Both local direction and local phase are related to the line/edge information of the object or structure in the image. They are therefore high-level descriptors implying that local direction and local phase may not be available at every pixel in the image. To overcome this problem, we introduce a complex value for each of the features. The phase of the complex value represents each feature value, such as the angle α or β. The magnitude, normalized between 0 to 1, of the complex value indicates the certainty of the feature

value. Small magnitude values mean that the feature value has a low degree of certainty. Therefore, a complex value with magnitude equal to zero can be used to represent the points lacking local direction and local phase. The advantage of using complex representation is that the certainty information can be carried into the subsequent calculations.

3.3.1.1 Local direction

Local direction is estimated by vector representation as proposed by Granlund[3.1] (Granlund 1978; Granlund and Knutsson 1995). "Local" means that the direction of a point is calculated based on its neighboring area, hereafter referred to as its "neighborhood." The estimation operations are designed to detect the directional information in a neighborhood with a dominant orientation, constituted by a line and/or an edge. The direction is obtained by combining a number of outputs of directional quadrature filters. The quadrature filter is defined in the frequency domain and only lies on the half plane of the frequency domain. In the spatial domain, the quadrature filter is a complex function. Its real component is an even function like a line filter and its imaginary component is an odd function like an edge filter. The output of each quadrature filter collects the energy of a structure along the direction that the filter is defined. Different directional quadrature filters collect the energy of the structure from multiple directions. A proper combination of their outputs ensures a continuous representation of angles, i.e., the operations are valid for any angle value that the structure may have. The representation of the local direction at a point is a complex value: its phase is the double of the local direction value (in radians) and its magnitude is the certainty. It can be expressed as

$$D_c = C_d \, e^{i2\alpha} \tag{3.8}$$

where D_c is the complex-valued directional representation, α is the local direction value ranged between $[0, \pi]$, and C_d is the certainty value normalized within $[0,1]$. Interested

---

[3.1] We would like to thank Professor Gösta Granlund and his group at Linköping University for their filter kernels and many insightful comments.

readers are referred to more comprehensive descriptions about directional estimation in the literature (Granlund 1978); (Granlund and Knutsson 1995).

3.3.1.2 Local phase

Phase information alone has been often used to estimate the disparity in stereo vision (Sanger 1988; Fleet, Jepson et al. 1991; Westelius 1995). However, the proposed reconstruction approach in this Chapter uses local phase as only one type of feature. The local phase is the phase of the output of a complex filter, such as a Gabor filter (Sanger 1988) or quadrature filter (Westelius 1995). Westelius (Westelius 1995) studied and compared a number of complex filters for the extraction of local phase information. By using quadrature filters, the phase information can be interpreted as a measure of the degree of a point being in a line or an edge. Figure 9 illustrates such a representation of local phase. Local phase is also represented by a complex value. The angle β is the value of local phase and the magnitude of the complex value depicts the certainty. The expression is given as:

$$P_c = C_p \, e^{i\beta}, \tag{3.9}$$

where P_c is a complex-valued representation for the local phase, β is the local phase value ranged between $[0, 2\pi]$, and C_p is the certainty value normalized within $[0,1]$.

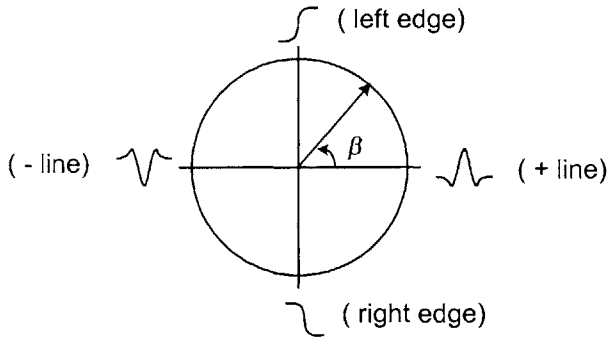


Figure 9 A representation of the local phase β as a complex vector (Reproduced from Westelius (1995))

3.3.1.3 Intensity

As discussed in preceding subsections, the complex representation provides us a certainty measure of the local direction and local phase. To maintain consistency with this certainty measure, a number is created for each intensity value and treated as its certainty value. Therefore, the intensity is represented as $I_c\{C_I, I\}$ [3.2], where I_c is the complex representation containing two components C_I and I. C_I is the certainty value of the intensity value I. In the implementation, since the actin filaments have higher intensity (brightness) than the rest part of the images, we assign large certainty values to the points with high intensity. The mapping function between the intensity value and its certainty is given by

$$C_I = (1 - e^{-\frac{I^2}{2\sigma_I^2}})^2 \tag{3.10}$$

where I is the intensity value of a point. σ_I is the parameter used to adjust the mapping relationship. The mapping function with σ_I equal to 40 is used in our calculation.

3.3.2 Stereo matching

Stereo matching uses the features described above to find the corresponding points between stereo images. The matching process is carried out on a point-to-point basis so that a dense disparity map can be obtained. However, due to ambiguities during matching, one point in one image may have several possible corresponding points in the other images. In order to distinguish among these candidates, a metric function is necessary to evaluate and compare them.

From feature representation, each point in a stereo image is described by three features: local direction (D_c), local phase (P_c) and intensity (I_c). If two points from each of the stereo images are matching points, they should have the same, or the closest, feature

---

[3.2] The complex representation of the form $Ae^{i\alpha}$ implies that the phase has a period of 2π. However, intensity doesn't have this property so that its complex representation cannot be expressed in the same form as local direction and local phase. We represent it as $I_c\{C_I, I\}$ to indicate that it also has two components similar to the form $Ae^{i\alpha}$.

values for all features. The reconstruction approach utilized here defines a similarity function to measure this proximity. The similarity function integrates all contributions from three feature values defined above and their certainty values. It yields a large similarity value when the feature values of two points are close and their corresponding certainty values are high. Mathematically, the similarity function is given as:

$$S_{LR} = S_{intensity}\ (I_L, I_R) + S_{direction}\ (D_L, D_R) + S_{phase}\ (P_L, P_R)$$
$$= \left| \mathbf{C}_I^{\gamma}\, e^{-\frac{(\Delta I)^2}{2\sigma_1^2}} \right| + \left| \mathbf{C}_D^{\gamma}\, e^{-\frac{(\Delta \alpha)^2}{2\sigma_2^2}} \right| + \left| \mathbf{C}_P^{\gamma}\, e^{-\frac{(\Delta \beta)^2}{2\sigma_3^2}} \right| \tag{3.11}$$

where $\Delta I = I_L - I_R$, $\Delta\alpha = \alpha_L - \alpha_R$ and $\Delta\beta = \beta_L - \beta_R$ are the differences of each feature value between stereo pairs denoted as the left (L) and right (R) images. C_I, C_D and C_P are three coefficients depicting the contributions of the certainty values on the similarity measure. They are calculated from the certainty values of the features between two images. σ_1, σ_2 and σ_3 are three predefined parameters controlling the normalization of the three features. The exponent γ adjusts the weight of the contribution from certainty values. It is normally set at 1. Since the features are represented as complex values, their differences can be calculated as follows. In a general form, the difference between two complex values $\mathbf{Z}_L$ and $\mathbf{Z}_R$ is given by (Westelius 1995):

$$\mathbf{d} = \mathbf{Z}_L \mathbf{Z}_R^* \tag{3.12}$$

$$\|\mathbf{d}\| = \|\mathbf{Z}_L\|\ \|\mathbf{Z}_R\| \tag{3.13}$$

$$\arg(\mathbf{d}) = \arg(\mathbf{Z}_L) - \arg(\mathbf{Z}_R)\,, \tag{3.14}$$

where $\mathbf{Z}^*$ denotes the complex conjugate of $\mathbf{Z}$. Equation (3.14) generates the feature value difference Δ used in Equation (3.11). Equation (3.13) calculates the new certainty value, which is the coefficient C in Equation (3.11). The coefficient C is large only if both certainty values are large.

During the stereo matching, corresponding points are determined by comparing all candidate points according to Equation (3.11). Corresponding points are chosen that exhibit the maximum similarity value. The coordinate differences of the corresponding points give the disparity value. On the other hand, the corresponding similarity value

from Equation (3.11) provides the certainty measure of this disparity estimation. It is utilized in the disparity-refining process.

3.3.3 Disparity refinement

Because the depth of an object or structure usually does not change significantly from one pixel to another, the disparity obtained at one point is expected to be very similar to the disparities of the points in the neighboring area. However, because of the ambiguities in the matching procedure and signal noise present in the imaging process, there are often variations in the disparity map obtained. Therefore, a disparity-refining process is carried out as a final step to improve the disparity map obtained from the stereo matching.

If the structure composes a continuous smooth surface, refinement could be accomplished by simply averaging the disparity map over a neighborhood to reduce the variations. However, the filament structure that we are concerned with has continuity primarily in one direction (the filament's contour direction). Simple averaging might miscalculate the disparities when two filaments with different depths and different directions are in the same neighborhood. In the proposed reconstruction approach, a more precise refining process is utilized. The process first distinguishes the different structures in a neighborhood. The disparity is then smoothed on each individual structure with little influence from other structures present in the same neighborhood. Therefore, the process still preserves the structural difference after refinement. In the actin-filament structures, the directional feature is the dominant structural information. The difference of the local directions is therefore considered the indicator of the structural difference. To measure this structural difference, a metric function is defined and given as follows:

$$w(i,j) = e^{-\frac{(\alpha_j - \alpha_i)^2}{2\sigma_w^2}} \quad , \; w \in (0,1] \, , \tag{3.15}$$

where α_i and α_j are two local directions of the points i, j, respectively. The metric function of Equation (3.15) exhibits the shape of a Gaussian function controlled by parameter σ_w. The function gives the maximum value 1 when two local directions are the same. The value decreases as the structural difference increases.

As discussed earlier, the similarity value is considered the certainty of the disparity estimation. It is also included in the disparity-refinement. By extending the normalization concept (Knutsson and Westin 1993), the disparity-refining process used in the proposed reconstruction approach takes into account the disparity value, the similarity value and the structural difference measured by the metric function of Equation (3.15). If we denote d a disparity value and S its corresponding similarity value (the certainty value of d), the disparity-refining process is conducted as follows:

$$d_{new} = \frac{\sum_{m,n}(w \cdot S \cdot d)}{\sum_{m,n}(w \cdot S)}, \tag{3.16}$$

$$S_{new} = \frac{\sum_{m,n}(w \cdot S)}{\sum_{m,n} w}, \tag{3.17}$$

where m, n defines the size of the neighborhood of a point under consideration. d_{new} and S_{new} are the new disparity value and its new similarity value, respectively. The disparity-refining process is performed at each point of the disparity map. The advantage of this refining process is that it smoothes the disparity map while it preserves the structural difference.

3.3.4 Hierarchical coarse-to-fine strategy

The proposed reconstruction approach is implemented in a multi-resolution manner. A hierarchical coarse-to-fine strategy allows for the successive improvement of the disparity estimation and reduces the computation of the stereo matching (Grimson 1985; Westelius 1995). Figure 10 illustrates the computational framework of the hierarchical implementation. The pyramid in Figure 10 is formed by successively downsampling the image, according to the Laplacian pyramid method proposed by Burt (Burt and Adelson 1983). With downsampling, the resolution of the image is reduced by a factor of 2 between two consecutive levels. Accordingly, the resolution of the disparity map between

two consecutive levels is also downsampled by a factor of 2. Therefore, the highest level in the pyramid has the coarsest resolution in both image size and disparity map.

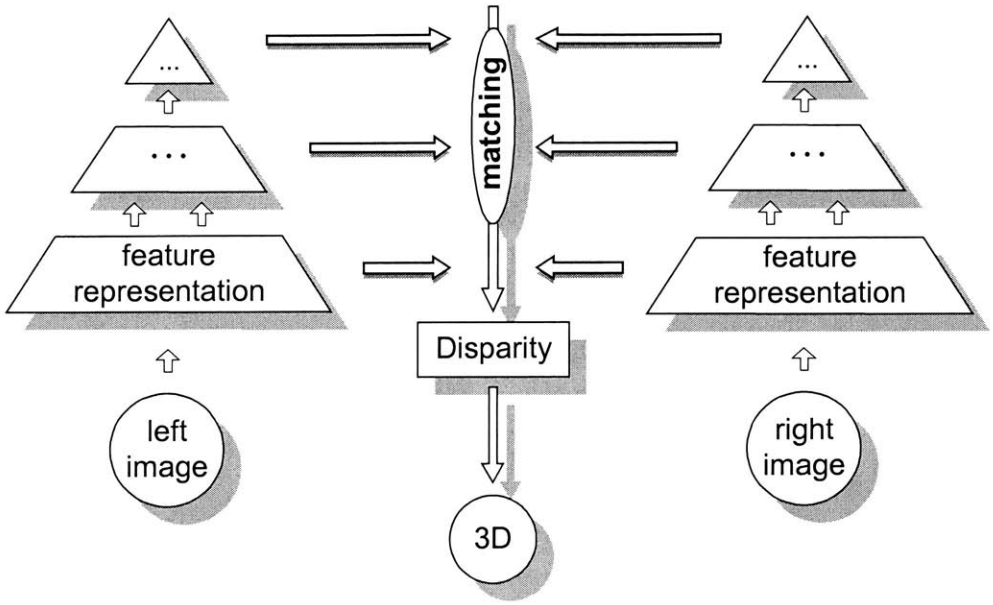


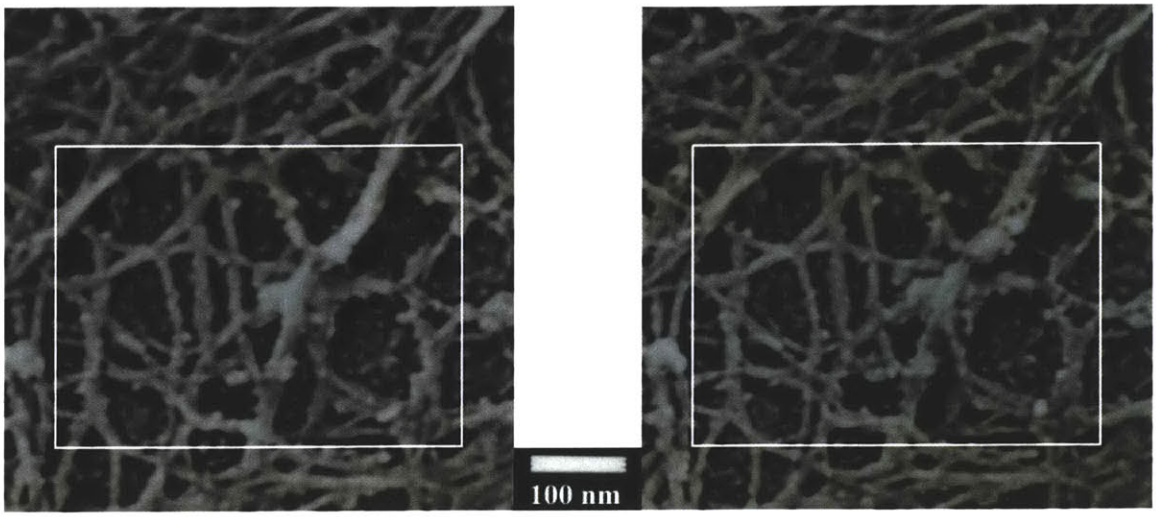
Figure 10 Hierarchical architecture of the computation

With this coarse-to-fine strategy, the algorithm starts at the highest level (lowest resolution). Feature representation is first applied in the downsampled images at this level, followed by the stereo matching and the disparity-refinement. The disparity map obtained at this level is then inputted into a disparity accumulator that updates the disparities from level to level. The disparity is then upsampled and fed into the next lower level (higher resolution) of the pyramid. Thus, the algorithm goes down one level in the pyramid. The images at the corresponding level are warped according to the disparity map fed from the higher level. Warping moves one image relative to the other according to the current disparity map. After the warping, a new disparity map at this level is obtained by performing the feature representation, stereo matching and disparity-refining process and inputted into disparity accumulator. The updated disparity map in the disparity accumulator is then upsampled and fed into the next lower level of the pyramid as before. This procedure is reiterated until the algorithm comes to the lowest level (finest resolution) of the pyramid. Finally, the disparity map in the disparity accumulator at the

lowest level is the final disparity map that now is used to compute the depth information based on Equation (3.7).

3.4 Experiment and Results

To demonstrate the algorithm described in previous sections, 3D reconstruction of a cortical portion of a cell cytoskeleton is obtained and shown in Figure 11 and Figure 12. Figure 11 depicts the paired images of the cell cytoskeleton taken at ±10° tilt angles. The scale bar indicates that one pixel in the image roughly corresponds to 1.33 nanometers. (The specimen used in this experiment is prepared by Dr. Hartwig, Brigham and Women's Hospital, Harvard Medical School.) Since the validity of the reconstruction requires faithful preservation of the specimen in the electron microscope, all efforts have been made to preserve the 3D structure of the cytoskeleton (Heuser and Evans 1980; Niederman, Amrein et al. 1983; Hartwig and Shevlin 1986; Hartwig 1992). Generally, the specimen is prepared such that the cytoskeleton is completely separated from the membrane and other soluble proteins under conditions where actin filaments are stabilized (Hartwig 1992). The cortical cytoskeleton observed consists mostly of actin filaments (F-actin), a small number of actin-binding-proteins, and a negligible amount of monomeric actin. In order to enhance the contrast, the specimen is further coated using metal. With coating, the filament is thickened by 2~4 nm. The background regions in the micrograph are cytoplasmic space where the electron beam has little attenuation because they lack actin filaments. The regions (see Figure 11) exhibit low intensity and lack high-level features, such as local direction or local phase. The certainty values of the features within these regions are therefore very low. Using these properties, background regions can be readily detected and separated from the actin filaments.



(a) (b) (c)

Figure 11 Example of electron microscope stereo images (400×400 in pixels) of cell cytoskeleton which were taken at ±10° of tilt angles: (a) left image; (b) scale bar; (c) right image. The reconstruction is performed on the region indicated by the white box (320×240 in pixels).

The proposed reconstruction approach described in previous section was applied to these images. Only the region indicated by the white box in Figure 11 is considered in the reconstruction. In this example, the coarse-to-fine "pyramid" implementation consisted of three levels. Figure 12 depicts the 3D reconstructed volume visualized by the isosurface method. Background regions are removed from the structure.

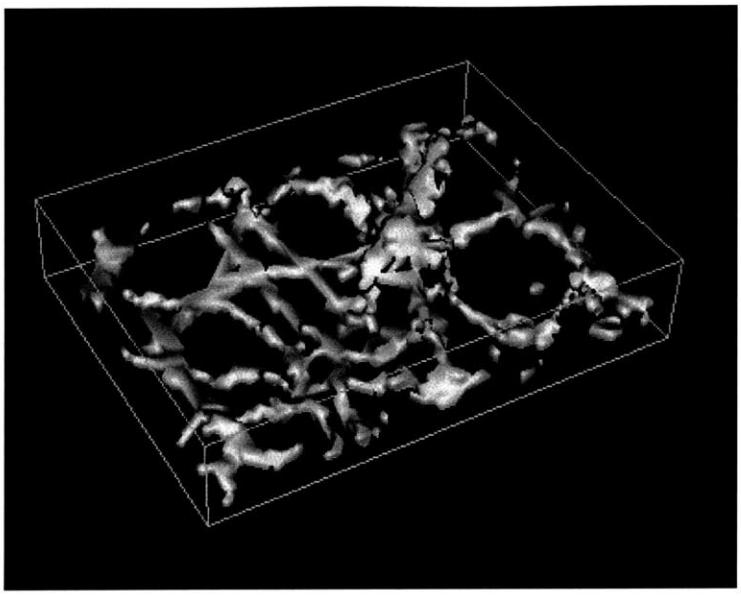


Figure 12 Reconstructed 3D cytoskeleton structure visualized by isosurface method.

After 3D reconstruction, new structural information becomes available. 3D structures provide realistic and accurate descriptions about the intersection points of actin filaments, the geometry of filaments at intersections, the pore size formed by intersecting filaments, the cytoskeleton thickness, and the filament concentration in a cytoskeleton volume. These properties can be measured quantitatively with additional modeling.

For example, the estimations of the cytoskeleton thickness and the actin filament concentration in the volume can be estimated easily from the reconstruction. The thickness of the reconstructed volume is considered the cytoskeleton thickness. The filament concentration in the volume is defined as the volume fraction of the filament in the reconstructed volume. In Figure 12, the cytoskeleton thickness of the volume is determined to be ≅ 90 nm. Furthermore, we model each filament as a solid cylinder-type structure with the diameter about 14 nm as measured from Figure 11. The filament concentration is then calculated as ≅ 5.2%, i.e., filaments occupy ≅ 5.2% of the volume. This number is actually the combined volume fraction of the actin filaments and their metal coat. The true volume fraction of the filaments should be corrected according to the amount of the metal coat. If we assume that the metal coat uniformly increases the diameter of each filament by 2 nm, the filament concentration decreases to 3.8%. The

value is in reasonable agreement with what is currently known about the cytoskeleton. For example, Hartwig and Shevlin (1986) manually calculated a 10 mg/ml concentration of actin filaments in the cortex of a white cell, which corresponds to a filament concentration by volume of $\cong 1.7\%$.[3.3]

3.5 Discussion and Conclusion

In this Chapter, we have presented a stereo reconstruction approach that reconstructs 3D structure from two images. The approach introduced a complex-valued representation of features, incorporating both feature values and corresponding certainty values into the reconstruction. In the experiment, the algorithm is applied to reconstruct the cell cytoskeleton structure from stereo images. The results demonstrate that the reconstruction offers us more opportunities to make valuable quantitative measurements in a 3D context. The analysis of these measurements will furthermore provide useful insights to the understanding and modeling of cell mechanical properties and cytoskeletal dynamics. More experiments and measurements would be conducted in the future for the algorithm to be used routinely.

Even though the algorithm is tuned for the cell cytoskeleton structure, the proposed approach has its generality. It is applicable to other applications, especially in the situations involving filament-like structures such as blood vessel networks or neural networks (Kim, Aggarwal et al. 1990; Ramirez, Mitra et al. 1999).

[3.3] The concentration of actin filaments in the cell cortex is approximately 10 mg/ml. The following estimation is based on the size of the monomers even though the actin filaments we consider are polymers. If we consider 1 ml, the volume contains 10 mg of actin. Because one mole of actin monomers weighs 42000 g, 1 ml contains 2.5×10^{-7} moles of monomers. Since there are 6×10^{23} monomers in one mole, 1 ml contains 1.5×10^{17} monomers of actin. When a monomer of actin is modeled as a sphere of diameter 6 nm, its volume is approximately 113 nm^3, or 1.3×10^{-19} ml. Therefore, the total volume of monomers is approximately 0.017 ml, or 1.7%.

The limitations of the reconstruction from only two images are also visible, especially in situations of occlusion or noise contaminated images. The ambiguities in the matching process are inevitable. One remedy is the extension to 3D reconstruction from multiple images. Reconstruction from multiple images will alleviate the limitations and reduce the ambiguities.

Chapter 4

Reconstruction from Multiple Images

A natural extension of stereo reconstruction is the reconstruction from multiple images. In this Chapter, we extend the reconstruction from two images developed in the previous Chapter to reconstruction from multiple images. A volumetric reconstruction method is proposed. It consists of two main processes: moving stereo reconstruction and integration.

4.1 Introduction

In theory, 3D information can be retrieved from as few as two images by stereo reconstruction. However, there usually exist a great deal of ambiguities in stereo reconstruction. For example, because of object occlusion, some part of the object may only appear in one image, not the other. Stereo reconstruction cannot recover accurate information about the structures in these parts. One remedy is to perform reconstruction from multiple images. With multiple images, occluded object missing in some images can be reconstructed from other images that contain the view of the missing parts.

In addition, reconstruction from multiple images improves both reconstruction precision and reconstruction accuracy. As discussed in previous Chapter, the 3D depth Z is related to the disparity d by:

$$Z = \frac{x_1 - x_2}{2\sin\theta} = \frac{d}{2\sin\theta}, \tag{4.1}$$

where x_1 and x_2 are positions in two images with tilt angles θ and $-\theta$, respectively. Besides the proportional relationship between depth Z and disparity d, this equation also

indicates that Z is inverse proportional to tilt angle θ. The term $1/(2\sin\theta)$ acts as a magnification factor in Equation (4.1). With unit change in disparity, depth estimation is more precise when tilt angle θ is larger, i.e., larger tilt angle produces better precision. However, images with larger tilt angle often have fewer similarities between them. This poses a big problem in stereo matching. Since matching algorithm in stereo reconstruction relies largely on the similarity between two images, there is greater possibility of false match when tilt angle is larger. In other words, reconstruction is less accurate when using larger tilt angle. A tradeoff exists between accuracy and precision when only two images are used. Reconstruction from multiple images can help to improve such tradeoff and achieve both accuracy and precision at the same time because multiple images can span a wide range of angles with small increment between sequential images.

4.1.1 Related Work

Recently, there have been many studies on reconstruction from multiple images in computer vision. One special case is reconstruction from triple images, often called trinocular stereo. The three views are taken from two motions, usually perpendicular to each other. (Motion means the change of camera position, or equivalently using multiple cameras at multiple positions.) Shashua (Shashua 1995) introduced a trilinear tensor that represents geometric constraints among three views. Stein (Stein and Shashua 1996; Stein 1998; Stein and Shashua 2000) later developed a direct reconstruction method from three views by combining trilinear tensor with optical flow constraints. The three-view scheme reduces ambiguity of the reconstruction. Disadvantage of this method is that it requires two motions in two perpendicular directions and involves 27 undetermined parameters.

A more popular case is that multiple images are taken from a monocular motion along one direction, forming a sequence of images. Computational procedures used to reconstruct the object or surface from these images can be roughly grouped into two categories (Heel 1991; Heel 1991): incremental and batch algorithms. Incremental algorithms attempt to estimate the reconstruction incrementally from multiple images.

When a new image becomes available, reconstruction is updated from previous reconstruction result. For example, Matthies et al (Matthies, Szeliski et al. 1989) and Heel (Heel 1991; Heel 1991) developed Kalman-filter based algorithms for scene reconstruction from a sequence of images. The Kalman-filter technique is a recursive algorithm consisting of two stages: prediction and update. The prediction stage predicts upcoming new image based on current estimation and system dynamics. The update stage updates the estimation by correcting the difference between expected new image and the actual new image. Such recursive algorithms are very efficient in computation because only the latest frame of image information is required to be present in memory at any time. The disadvantage is that the reconstruction is not guaranteed to be the best one because the estimation is not globally optimized.

Batch algorithms attempt to achieve optimal reconstruction by processing all images together. For instance, Iu and Wohn (Iu and Wohn March, 1989) modeled each 3D surface point with a truncated Taylor series. Coefficients of the Taylor series are estimated from feature locations in all images. The estimation is conducted in a least-square fashion. Okutomi et al (Okutomi and Kanada 1993) proposed a multiple-baseline stereo approach in which a cost function, called SSSD-in-inverse-distance function in the paper, is defined to evaluate the reconstruction. The best reconstruction is chosen to be the one that minimizes this function. More recently, Fua (Fua 1997) presented a framework of 3D surface reconstruction from an arbitrary number of stereo views. This approach introduces a particle-based representation from which a surface patch is generated locally by clustering and function-fitting. In general, most batch algorithms require a lot of computation and are not very favorable in real-time applications. The advantage is that these algorithms usually have a global cost function and the reconstruction obtained is globally optimized with respect to this function. In other words, given the available images and defined cost function, the batch algorithms achieve the best reconstruction.

Even though there is clear distinction between incremental and batch algorithms, there is no clear-cut criterion on choosing one particular type of algorithm over the other. The

decision typically depends on several factors, such as computation, type of images, type of applications, etc. For example, in a navigation system, incremental algorithms are often preferred because images are obtained sequentially and computation requires real-time process. On the other hand, a recognition task may need to process all images together to make a decision. In this case, batch algorithms may be a better choice.

4.1.2 Overview

In this Chapter, we present a new batch algorithm for reconstruction from multiple images. Similar to other batch algorithms, optimal reconstruction is obtained by evaluating available images altogether. There is no preference to the order of images. The proposed reconstruction method consists of two main processes: moving stereo reconstruction and integration. Moving stereo reconstruction performs reconstruction sequentially on each image pair. We extend stereo reconstruction developed in previous Chapter into a volumetric representation. In each reconstruction, a probability volume is generated by evaluating a multi-feature based correlation between two images. The probability at each voxel represents certainty of the voxel being an object point. The integration process determines optimal reconstruction by evaluating all probability volumes from individual stereo reconstruction. The process is measured by a global cost function. Optimal reconstruction is the one that minimizes the defined cost function.

Differences between our method and the algorithms discussed in the previous section are listed below. First and foremost, the imaging system is different. Our method deals with electron microscope imaging system that adopts a parallel projection model. The geometric relationship between the object and images are different from the camera system used in other algorithms. Second, the image sequence is different. In our algorithm, the image sequence is obtained via a rotational motion while most algorithms in literature deal with a translational sequence of images. Third, the focus is different. Most algorithms in computer vision focus on the surface reconstruction of a scene or object. Our proposed method attempts to develop a volumetric representation of the 3D object by employing the concepts of projection and backprojection. Finally, the

application is different. Our algorithm is specially designed for 3D reconstruction of cell cytoskeleton structure. The specific features of cell cytoskeleton are incorporated into the algorithm.

4.2 Description of Method

The diagram in Figure 13 illustrates procedures of the proposed method. Besides alignment, the method includes two other processes: moving stereo reconstruction and integration. By using the alignment algorithm presented in Chapter 2, we assume all images have been aligned correctly. This Chapter will focus on the other two processes.

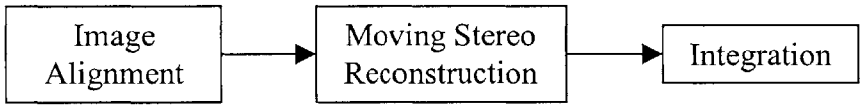


Figure 13 Procedures of the reconstruction from multiple images

4.2.1 Moving Stereo Reconstruction

Stereo reconstruction is the building block of our method. "Moving" means that stereo reconstruction is conducted sequentially at each image pair. The stereo reconstruction method proposed here is slightly different from the one discussed in Chapter 3. Instead of reconstructing object or surface directly, moving stereo reconstruction generates a volumetric representation of the reconstruction. Moving stereo reconstruction starts with a pre-defined volume and attempts to determine at each voxel whether the voxel belongs to the object or not. The evaluation is measured by correlation between two projections including their associated neighbors. In the correlation, multiple features of each point are extracted and used. The features include local direction, local phase and intensity. The algorithm for extracting these features has been presented in Chapter 3.

4.2.1.1 System Configuration

As discussed in Chapter 3, our imaging system is modeled as parallel projection. The image sequence that our method considers is obtained from pure rotational motions. Without loss of generality, let us assume that a 3D coordinate system *O-XYZ* is placed in the center of the volume to be reconstructed, and the images are taken at a sequence of angles (θ_1, θ_2, ..., θ_n) by rotating the object along *Y*-axis. Figure 14 demonstrates such configuration. Since rotating the object is equivalent to rotating the position of the image plane, we may use both concepts interchangeably. Furthermore, we assume the coordinate system *o-xy* in each image is the projection of 3D coordinate system *O-XYZ* to the image plane. As derived in Chapter 3, the geometric relationship between an object point and its projection in the image can be given by:

$$X = \frac{x_j \sin\theta_i - x_i \sin\theta_j}{\sin(\theta_i - \theta_j)} \tag{4.2}$$

$$Y = y_i = y_j \tag{4.3}$$

$$Z = \frac{x_i \cos\theta_j - x_j \cos\theta_i}{\sin(\theta_i - \theta_j)}, \tag{4.4}$$

where i and j correspond to the images with tilt angles of θ_i, θ_j, respectively.

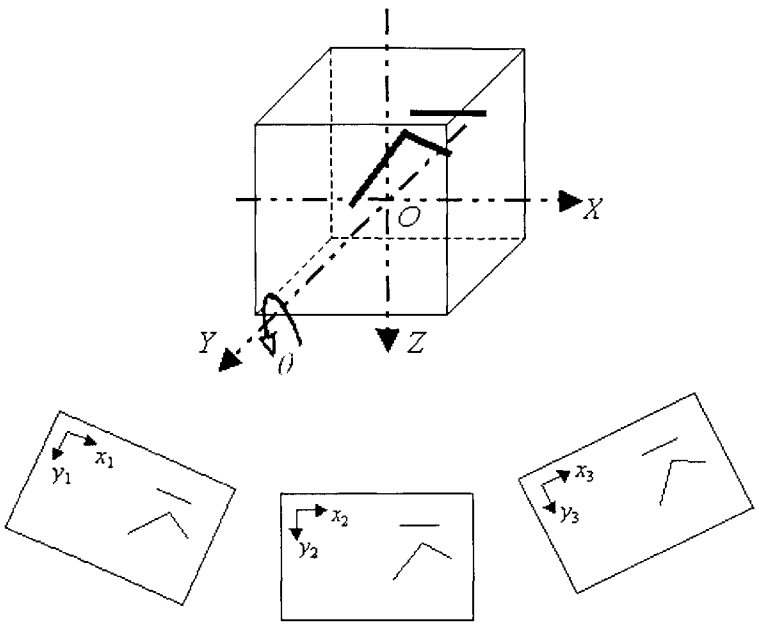


Figure 14 The image sequence is obtained by a rotational motion.

The Equations indicate that the y-coordinate of a point remains the same in every image. Thus, the geometric relationship can be considered two-dimensional and each XZ slice is independent to each other. In fact, the same geometric relationship exists in tomography even though the imaging relationship between intensity value and object value is different from that of tomography. Let us examine this geometric relationship from a projection/backprojection viewpoint. Figure 15 plots an object point P and its projection points p_i in each image. R_i is a ray corresponding to the projection direction θ_i. If we know the object point and the projection direction θ_i, we can easily calculate the projection point p_i in corresponding image plane. On the other hand, if we have a projection point, we know that the corresponding object point must lie somewhere on its projection direction. This is called backprojection. Furthermore, if we have another projection point in another image and know that this projection point is from the same object point, we can find out the object point by backprojection of both projection points. The key here is that two projection points must be from the same object point. They are often called "matched projection points". For example in Figure 15, the location of the object point is the intersection of two rays R_1 and R_2 of matched projection points p_1 and p_2.

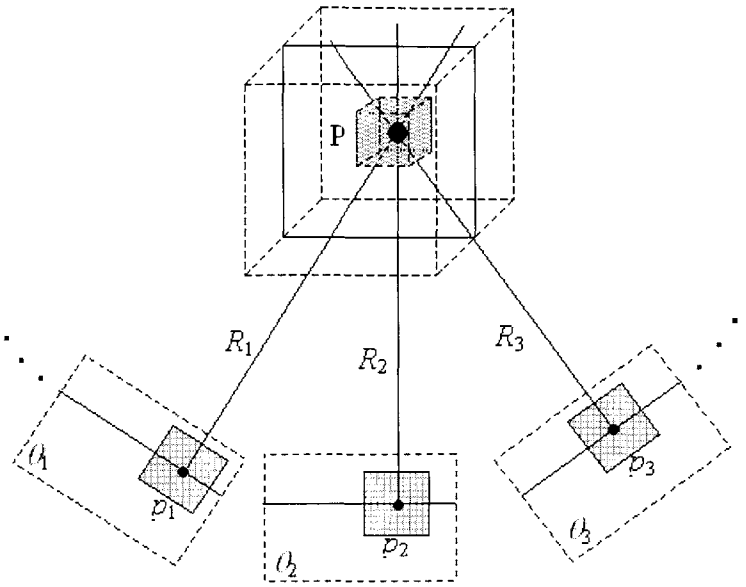


Figure 15 Projection and backprojection of an object point

There are several important observations in this system that we would like to point out:

Observation I: ambiguity property

If only two projection points are given and we don't know whether they are matched projection points or not, there could be unlimited solutions. For example in Figure 15, any two object points on ray R_1 and R_2 could have the same projection points p_1 and p_2. Thus, one of the biggest tasks in reconstruction is to determine whether two projection points are matched projection points or not.

Observation II: uniqueness property

There is only one, if exist, projection point in each image for a given object point. In the occlusion situation, some object points may block others so that there are no corresponding projection points in the image. In the mean time, there is only one, if exist, object point corresponding to any two matched projection points from any image pair.

Observation III: neighborhood property

A neighborhood remains as the neighborhood in the projection. In other words, if some object points are neighbors to each other in 3D space, their projection points in each image are also neighbors to each other. However, the reverse is not always

true. A neighborhood in the image may not necessarily be the projections from a neighborhood in the 3D object. The neighborhood property is demonstrated in Figure 15 where shade area means the neighbors.

4.2.1.2 Volumetric Representation

With the above analysis of projection and backprojection, the proposed method extends the stereo reconstruction approach developed in previous Chapter and proposes a volumetric method of 3D reconstruction. The method starts with a pre-defined empty volume. Reconstruction of the 3D object is equivalent to determining whether each voxel in the volume belongs to the object or not. A probability volume is introduced for the purpose of evaluation at each voxel. Probability at each voxel indicates the certainty level of the voxel being an object point. It is computed based on the correlation between two corresponding projections and their associated neighbors.

For each voxel, we first identify the location of its projections in two images by forward projection. If a voxel is an object point, projections of this voxel should have similar characteristics in two images, i.e., they should have high correlation. Furthermore, according to Observation III, neighbors of these projections should also have high correlation between these two images, except for occlusion situations. Based on this observation, we employ a multiple-feature based correlation algorithm and compute the correlation on a neighborhood system. For example in Figure 15, the shaded areas treated as neighborhood of the projections are used for computing correlation between p_1 and p_2 in images of θ_1 and θ_2, respectively.

Multiple features used include local direction, local phase and intensity. The extraction of these features can be obtained by the method presented in Chapter 3. With multiple features, each point is represented by a feature vector:

$$f = \begin{bmatrix} D_c \\ P_c \\ I \end{bmatrix}, \tag{4.5}$$

where D_c is the directional representation. P_c is the representation of local phase. (Please refer to Chapter 3 for more details.) I is the intensity value in an image. Since both D_c and P_c are complex-valued representations, a more computable feature vector can be adopted:

$$\boldsymbol{f} = \begin{bmatrix} \mathrm{Re}(D_c) \\ \mathrm{Im}(D_c) \\ \mathrm{Re}(P_c) \\ \mathrm{Im}(P_c) \\ I \end{bmatrix}, \tag{4.6}$$

where Re(·) means the real part and Im(·) is the imaginary part of a complex value. This vector variable will be used for correlation in the practical implementation.

Correlation between two projections is computed on a neighborhood system where neighbors are considered the sample source of the vector variable. Let us denote the neighborhood window as N, or $n \times n$, centered at the projection point. Denote $\boldsymbol{f}_{\theta 1}$ and $\boldsymbol{f}_{\theta 2}$ two vector variables corresponding to the feature representations at the projection points and their neighbors in two images of tilt angles θ_1 and θ_2, respectively. Mathematically, correlation between two vector variables can be obtained by canonical correlation. Canonical correlation finds the linear combinations $f_{\theta 1}=\boldsymbol{f}_{\theta 1}{}^{\mathrm{T}} \cdot \boldsymbol{w}_{\theta 1}$ and $f_{\theta 2}=\boldsymbol{f}_{\theta 2}{}^{\mathrm{T}} \cdot \boldsymbol{w}_{\theta 2}$ such that the scalar variables $f_{\theta 1}$ and $f_{\theta 2}$ have the maximum correlation[4.1]:

$$\rho = \max_{\boldsymbol{w}_{\theta 1}, \boldsymbol{w}_{\theta 2}} \frac{E\left[\left(f_{\theta 1} - \bar{f}_{\theta 1}\right)\left(f_{\theta 2} - \bar{f}_{\theta 2}\right)\right]}{\sqrt{E\left[\left(f_{\theta 1} - \bar{f}_{\theta 1}\right)^2\right] E\left[\left(f_{\theta 2} - \bar{f}_{\theta 2}\right)^2\right]}} \tag{4.7}$$

$$\text{or} \quad \rho = \max_{\boldsymbol{w}_{\theta 1}, \boldsymbol{w}_{\theta 2}} \frac{\boldsymbol{w}_{\theta 1}^T E\left[\left(\boldsymbol{f}_{\theta 1} - \bar{\boldsymbol{f}}_{\theta 1}\right)\left(\boldsymbol{f}_{\theta 2} - \bar{\boldsymbol{f}}_{\theta 2}\right)^T\right] \boldsymbol{w}_{\theta 2}}{\sqrt{\boldsymbol{w}_{\theta 1}^T E\left[\left(\boldsymbol{f}_{\theta 1} - \bar{\boldsymbol{f}}_{\theta 1}\right)\left(\boldsymbol{f}_{\theta 1} - \bar{\boldsymbol{f}}_{\theta 1}\right)^T\right] \boldsymbol{w}_{\theta 1} \boldsymbol{w}_{\theta 2}^T E\left[\left(\boldsymbol{f}_{\theta 2} - \bar{\boldsymbol{f}}_{\theta 2}\right)\left(\boldsymbol{f}_{\theta 2} - \bar{\boldsymbol{f}}_{\theta 2}\right)^T\right] \boldsymbol{w}_{\theta 2}}}, \tag{4.8}$$

where $E[\cdot]$ is the expectation operator and $^{\mathrm{T}}$ is the transpose operator. $\bar{\boldsymbol{f}}$ is the mean of the variable $\boldsymbol{f}$. $\boldsymbol{w}_{\theta 1}$ and $\boldsymbol{w}_{\theta 2}$ are coefficient vectors that maximizes Equation (4.8). The correlation coefficient ρ ranges between −1 to 1.

[4.1] Note that non-bold font means a scalar variable and **bold** case means a vector variable.

Solving Equation (4.8) is not trivial at all. To make the computation simple, we introduce a compromised approach to calculate the correlation. The basic idea is to convert the vector variable in Equation (4.6) to a scalar variable. First, each element in the feature vector is normalized with respect to its own mean and variance. Normalization of a random variable x is defined as:

$$x_{norm} = \frac{x - \bar{x}}{\sigma_x}, \tag{4.9}$$

where $\bar{x}$ is the mean and σ_x^2 is the variance of x. After normalization, new variable x_{norm} has the mean of 0 and the variance of 1. In other words, normalization brings variables to the same scale so that they are comparable. After normalization, we have a new feature vector $\boldsymbol{f}_{norm}$ in which each element is comparable to each other. We then reorganize all feature vectors in the neighborhood into a large one-dimensional vector as illustrated in Figure 16. Each element of this large vector is considered a sample of some random variable. Correlation coefficient is then calculated by using normalized cross-correlation between two generated large vectors from two images. For example, if a neighborhood is $n \times n$ in size and the feature vector has 5 elements as in Equation (4.6), the generated large vector has the size of $1 \times (5n^2)$. If variables in the large vectors are denoted by u and v for two images of tilt angles θ_1 and θ_2, respectively, the correlation coefficient is computed by:

$$\rho = \frac{\sum_{i=0}^{5n^2} (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=0}^{5n^2} (u_i - \bar{u})^2} \sqrt{\sum_{i=0}^{5n^2} (v_i - \bar{v})^2}}, \tag{4.10}$$

where $\bar{u}$ and $\bar{v}$ are the means of the corresponding variables. The coefficient ρ, ranging between –1 and 1, reflects the similarity of two projection points. If they have very high correlation, it gives a very good indication that two projection points may be matched projection points, and therefore the voxel under consideration may be an object point.

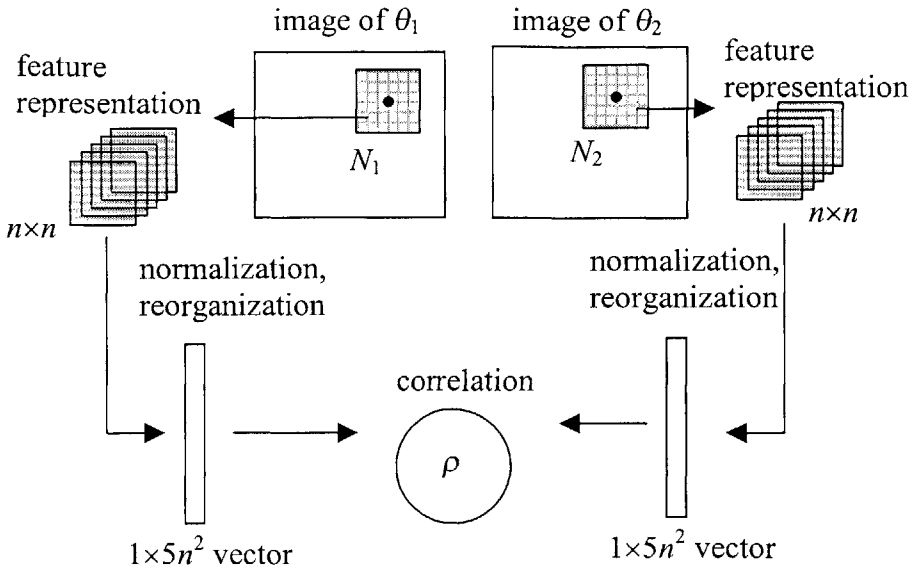


Figure 16 Illustration of the generation of the correlation between two images

We repeat above computation for each voxel in the volume and obtain a correlation coefficient ρ at each voxel. We thus obtain a ρ-volume which provides a reliable indicator of whether each voxel in the volume is an object point or not. Furthermore, the value of the coefficient ρ in some sense reflects the probability of the voxel being an object point. The larger the coefficient, the higher the probability. Since the value of the coefficient ranges from 1 to –1, we map it to a probability value by simply using some threshold or linear transformation. Therefore, a probability volume (p-volume) is obtained from a ρ-volume. Repeating this computation for each pair of images, we obtain a sequence of p-volumes which will then be used in the integration process.

In implementation, the projection of a voxel is not necessarily on the grid of pixel. We employ a bilinear interpolation method to accurately locate the projections in the images. This interpolation is similarly applied to all neighbors as well. The correlation is then calculated on the interpolated neighborhood windows. Such interpolation process improves the resolution of reconstruction.

4.2.1.3 "Moving" volumetric representation

To accomplish the "moving" concept defined earlier, a volumetric reconstruction is obtained sequentially for each pair of images. The image pair can be chosen arbitrarily. For example in Figure 15, we could choose image pair from sequential images, i.e., $\{(\theta_1, \theta_2), (\theta_2, \theta_3), \ldots\}$, or non-sequential images, e.g., $\{(\theta_1, \theta_3), (\theta_2, \theta_4), \ldots\}$. There is no clear criterion on how to choose the type of image pair. The choice may depend on the total number of images, the tilt angle difference between two images, the reconstruction resolution, etc.

4.2.1.4 Some Computation Considerations

The computation in the proposed method above is very expensive, especially in the steps of feature representation and correlation. In order to reduce the computation, some approximation schemes or fast algorithms can be employed. For example in feature representation, local direction can be roughly estimated from gradient. If the intensity value of an image is represented by function $I(x, y)$, the gradient at point (x, y) is given by:

$$\boldsymbol{g} = \begin{bmatrix} \partial I/\partial x \\ \partial I/\partial y \end{bmatrix}. \tag{4.11}$$

Since the local direction α is perpendicular to the direction of gradient $\boldsymbol{g}$, it can be simply obtained by:

$$\alpha = \arctan\left(-\frac{\partial I/\partial y}{\partial I/\partial x}\right). \tag{4.12}$$

In comparison to the method discussed in Chapter 3, using gradient to estimate local direction is computationally more efficient.

The computation of correlation can also be reduced by using a fast algorithm. We simply adopt the fast algorithm developed by Sun (Sun 1997). In brief, the algorithm uses a box-filtering procedure for mean calculation (McDonnell 1981), and reorganizes the equation for fast variance calculation.

4.2.2 Integration Process

In moving stereo reconstruction, we obtain a p-volume from each image pair. Each p-volume describes the same object volume. However, the description from different p-volumes may be different. This is because different image pairs have different angles of view and contain different information. In addition, ambiguities in stereo reconstruction from different image pairs may be different too. The goal of the integration process is to find the best description of the object volume from a sequence of p-volumes.

4.2.2.1 Integration by Average

One straightforward way of integration is to simply take the average of all p-volumes. The theory behind this integration is as follows. First we assume that the probability value at each voxel is an independent random variable, i.e., each voxel can be estimated independently. Then, the value in each p-volume is considered a sample of the random variable. Thus, the problem of finding the best probability value at each voxel becomes the problem of finding the best estimation of a variable given some samples. In other words, we want to estimate the variable p, from samples $\{p^{(1)}, p^{(2)}, \ldots, p^{(n)}\}$, where superscript (n) denotes the n-th p-volume obtained from n-th stereo reconstruction. Based on the least square error method, the best estimation of p is the mean of the samples, i.e.,

$$p = \arg\min_{p} \sum_{s} \left(p - p^{(s)}\right)^2 \tag{4.13}$$

or,

$$p = \frac{1}{n} \sum_{s} p^{(s)} \quad . \tag{4.14}$$

The integrated value p represents the best estimation as we consider all images together, given the error function in (4.13). This probability value indicates the overall certainty level of a voxel being an object point based on the information from all images. With some proper threshold on this probability volume, all object points in the volume can be identified and therefore the object can be reconstructed.

4.2.2.2 Integration with Local Interaction

The assumption that each voxel is independent to each other is usually not true. In practice, there are always interactions among voxels. In this section, we introduce a local interaction model that encourages the continuity or smoothness in a neighborhood.

Let us consider a first order neighborhood system. In other words, for a voxel point (i, j, k) in the volume, six neighbors are considered, including $N_{i,j,k} = \{(i\text{-}1, j, k), (i\text{+}1, j, k), (i, j\text{-}1, k), (i, j\text{+}1, k), (i, j, k\text{-}1), (i, j, k\text{+}1)\}$. We introduce a penality function that measures the departure from smoothness in the neighborhood:

$$e_l = \sum_i \sum_j \sum_k \sum_{N_{i,j,k}} \left(p_{i,j,k} - p_{N_{i,j,k}} \right)^2 . \tag{4.15}$$

This penality function encourages the local smoothness and is minimized when the neighborhood is uniform. This function is incorporated as an additional constraint into the overall error function, i.e.,

$$e = \sum_i \sum_j \sum_k \left(\sum_s \left(p_{i,j,k} - p_{i,j,k}^{(s)} \right)^2 + \lambda \sum_{N_{i,j,k}} \left(p_{i,j,k} - p_{N_{i,j,k}} \right)^2 \right), \tag{4.16}$$

where λ is a parameter that weights the smoothness penality in the overall error function. This parameter will be large if we want to enforce the local smoothness. The best estimation of $p_{i,j,k}$ is the one that minimizes this error function, i.e.,

$$p_{i,j,k} = \arg\min_{p_{i,j,k}} \sum_i \sum_j \sum_k \left(\sum_s \left(p_{i,j,k} - p_{i,j,k}^{(s)} \right)^2 + \lambda \sum_{N_{i,j,k}} \left(p_{i,j,k} - p_{N_{i,j,k}} \right)^2 \right). \tag{4.17}$$

It can be obtained by solving

$$\frac{\partial e}{\partial p_{i,j,k}} = 0, \quad \forall i, j, k . \tag{4.18}$$

This results in a set of linear equations:

$$\frac{(n+6\lambda)}{\lambda} p_{i,j,k} - p_{i-1,j,k} - p_{i+1,j,k} - p_{i,j-1,k} - p_{i,j+1,k} - p_{i,j,k-1} - p_{i,j,k+1} = \frac{1}{\lambda} \sum_s p_{i,j,k}^{(s)} , \forall i, j, k . \tag{4.19}$$

Equation (4.19) forms large sparse (not band) linear equations. A number of commercial and non-commercial software packages are available that can solve such linear system

accurately and efficiently. In our implementation, we use the LAPACK (Linear Algebra Package) developed at the University of Tennessee (LAPACK ; Anderson, Bai et al. 1999).

4.2.2.3 Integration with Consideration of Object Values

In the dicussions so far, we only evaluate whether a voxel is an object point or not. We don't know what value the object point may take. Thus, the obtained reconstruction only tells the presence and shape of the object in the volume. In this sense, the reconstruction is equivalent to assigning either 0 or 1 to each voxel in the volume where 1 means an object point and 0 means not. Certainly, this is not good to distinguish different objects from each other when there are several objects in the volume and each object has a different value. In this section, we present a model that makes use of object values in the reconstruction.

Even though we don't know the imaging relationship between the object and its projections in the images, some assumptions ususally hold true. One of the assumptions is particularly useful: if different objects take different values, their projections in an image also exhibit different values. Figure 17 illustrates this assumption. In the reconstruction, the opposite is more important, i.e., if projections in an image have different values, it implies that they are from different objects. These assumptions establish some mapping relationship between values of different objects in the volume and values of their projections in the images. This relationship is very helpful in distinguishing different objects in the reconstruction.

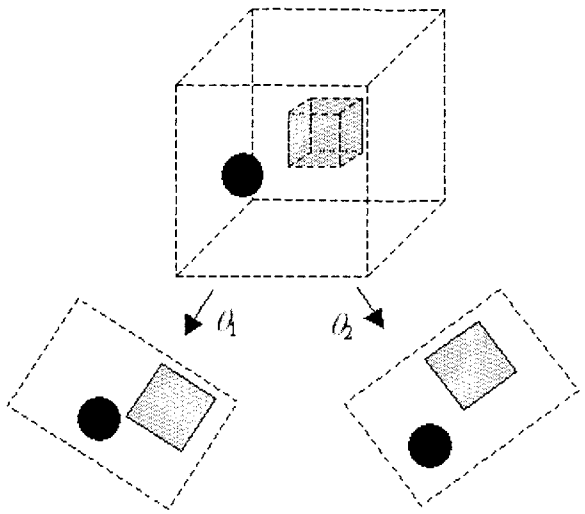


Figure 17 Different-valued objects have different-valued projections in the image

As discussed earlier in moving stereo reconstruction, we generate a probability value at each voxel based on the canonical correlation of its projections in two images. The probability value describes the certainty level of a voxel being an object point. Furthermore, based on our assumption, the intensity value of the projection points map to some object value. Therefore, we may infer that the probability value actually describes the certainty of the voxel taking a specific object value. Thus, we have two sets of information from each stereo reconstruction: a probability value and the object value that the voxel may take. Both sets of information are useful in determining the object volume. The goal of the integration process is then to estimate the best object value that each voxel takes in the volume. Since we have some knowledge of mapping between the object values and the intensity values of the projection, we don't need to know the exact relationship function in order to distinguish the different objects in the volume. For example, if we know large object value maps to high intensity value of the projection point in the image, we can simply assume that the object value takes the intensity value of its projection in the image.

Let us denote v as the object value, and $v^{(s)}$ as the object value obtained from s-th stereo reconstruction. Also denote $p^{(s)}$ as the probability value that indicates the certainty of corresponding object value $v^{(s)}$. The integration process estimates the best v from pairs

$(p^{(s)}, v^{(s)})$ obtained by moving stereo reconstruction. If voxels in the volume are independent to each other, each voxel can be estimated individually. We formulate an error function given by:

$$e = \sum_s p^{(s)} \left(v - v^{(s)} \right)^2 , \tag{4.20}$$

where $p^{(s)}$ is treated as the weighting coefficient. By minimizing this error function, the best estimation of v is obtained and can be written as:

$$v = \frac{1}{\sum_s p^{(s)}} \sum_s p^{(s)} v^{(s)} . \tag{4.21}$$

Essentially, Equation (4.21) computes the weighted average of $v^{(s)}$.

Similar to the discussions presented earlier, we may bring local interaction into the integration. If only first-order neighborhood system is considered, a similar smoothness error function can be employed:

$$e_l = \sum_i \sum_j \sum_k \sum_{N_{i,j,k}} \left(v_{i,j,k} - v_{N_{i,j,k}} \right)^2 . \tag{4.22}$$

The overall error function is given by:

$$e = \sum_i \sum_j \sum_k \left(\sum_s p_{i,j,k}^{(s)} \left(v_{i,j,k} - v_{i,j,k}^{(s)} \right)^2 + \lambda \sum_{N_{i,j,k}} \left(v_{i,j,k} - v_{N_{i,j,k}} \right)^2 \right) . \tag{4.23}$$

The optimal estimation of $v_{i,j,k}$ is the one that minimizes this error function. This leads to a set of linear equations:

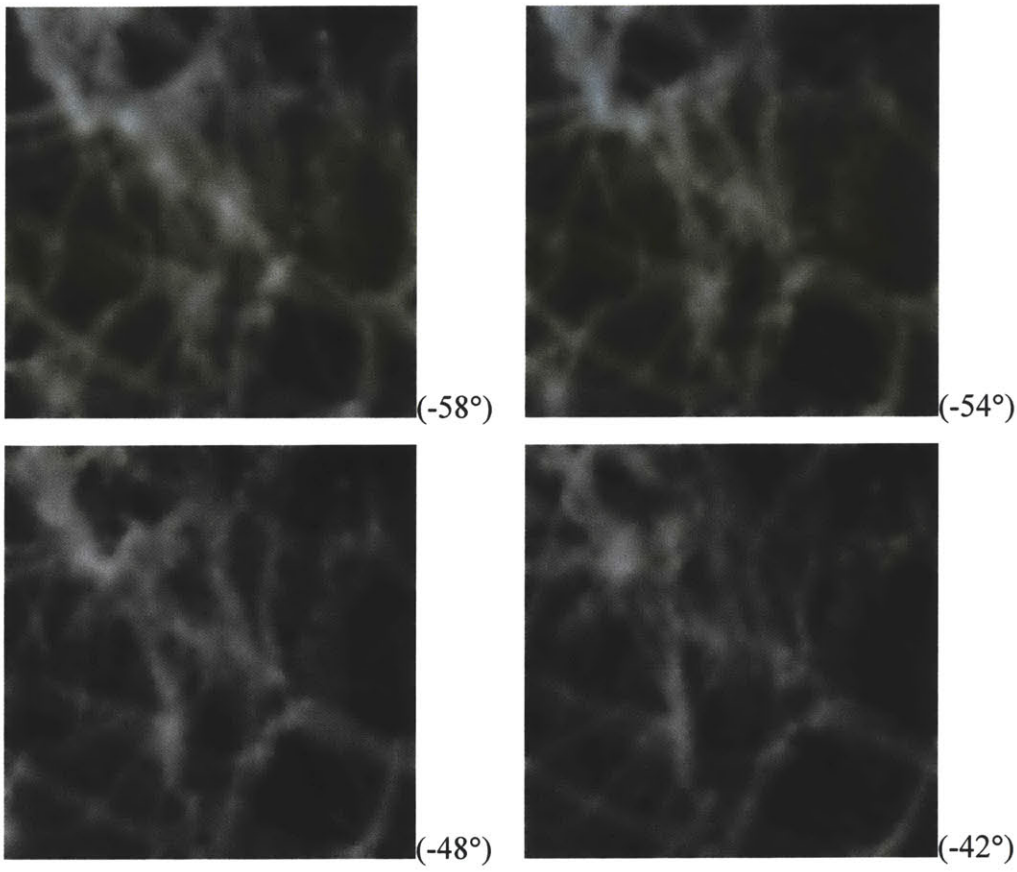
$$\frac{(\sum_s p_{i,j,k}^{(s)} + 6\lambda)}{\lambda} v_{i,j,k} - \sum_{N_{i,j,k}} v_{N_{i,j,k}} = \frac{\sum_s p_{i,j,k}^{(s)} v_{i,j,k}^{(s)}}{\lambda} \quad , \forall i, j, k . \tag{4.24}$$

Equation (4.24) is a sparse system and can be solved by using LAPACK.

4.3 Experiment and Results

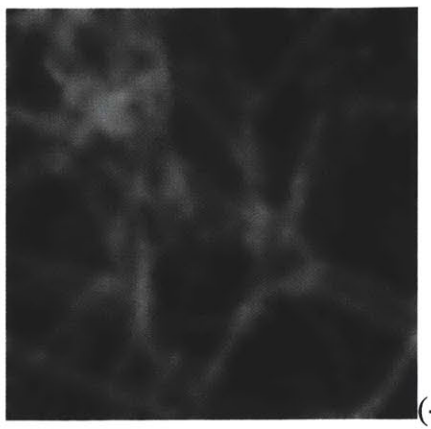
In this section, we apply the proposed algorithm to 3D reconstruction of cell cytoskeleton structures from multiple images. The cell cytoskeleton specimen is prepared by Dr.

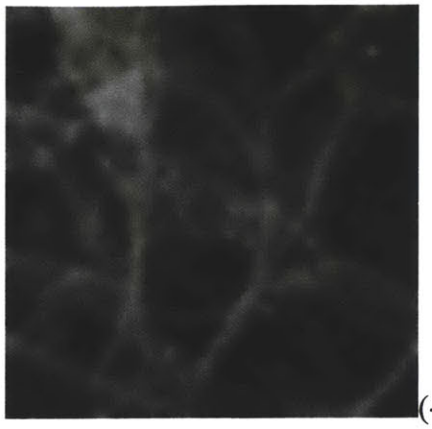
Hartwig (Hartwig and Shevlin 1986; Hartwig 1992). The images are taken on a JEOL JEM4000FX Intermediate-voltage electron microscope (IVEM) at the Wadsworth Center[4.2]. A sequence of images is obtained with the tilt angles ranging between –60° and 60° at an increment of 2.0°. These images are properly aligned with respect to each other by using the algorithm discussed in Chapter 2. An area of 256×256 in each image is selected for reconstruction. The scale in each image is that 1 pixel in the image equals to 1 nm in physical size. Figure 18 exhibits 16 of these images in sequence. The tilt angle is shown next to each image.

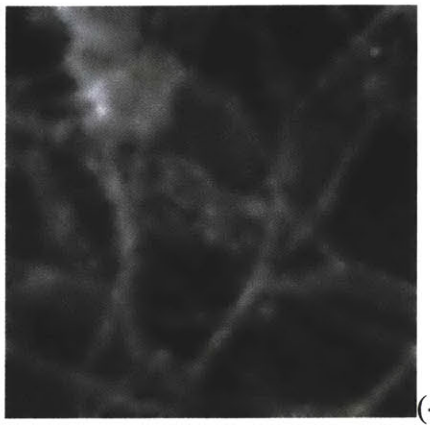


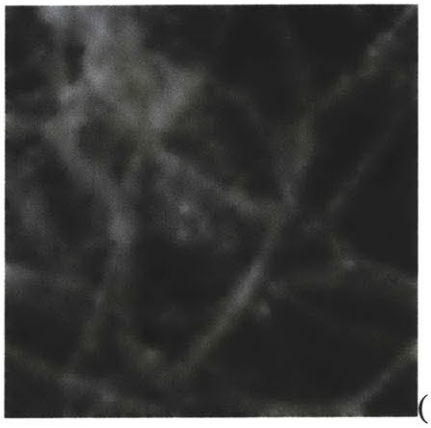
(to be continued in next page)

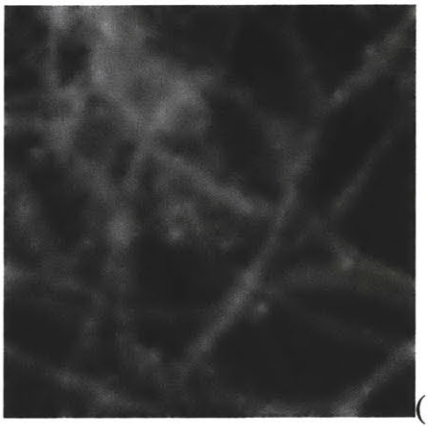
[4.2] The Wadsworth Center, located at Albany, New York, is a NIH National Biotechnological Resource center supported by grant RR 01219 from the National Center for Research Resources (DHHR/PHS). We also would like to thank Dr. Joachim Frank and Mr. Michael Marko for their generous help and support.

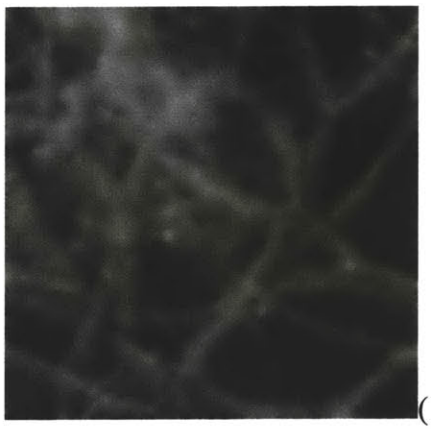

(-34°)


(-26°)


(-16°)


(-4°)


(0°)


(+4°)

(to be continued in next page)

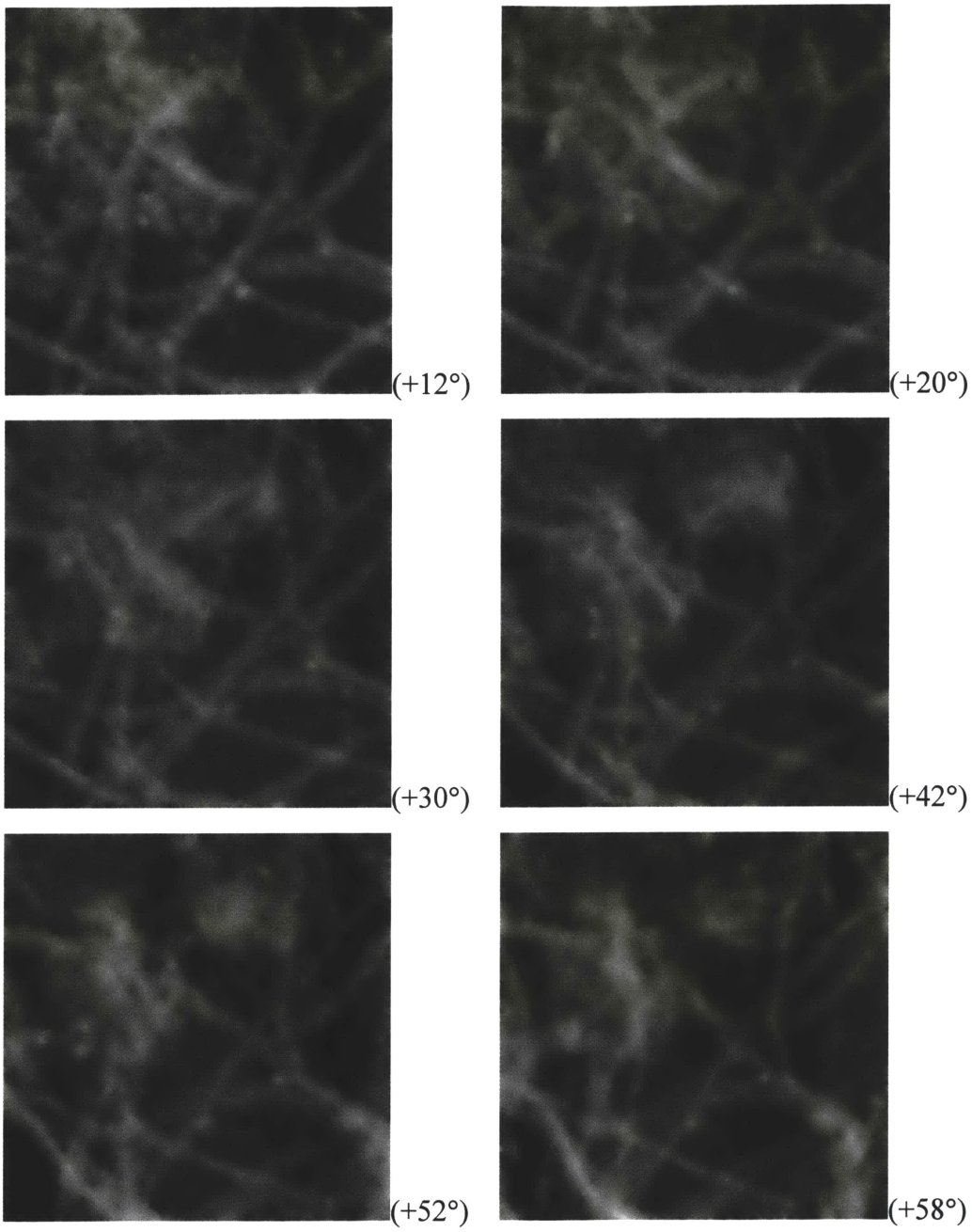
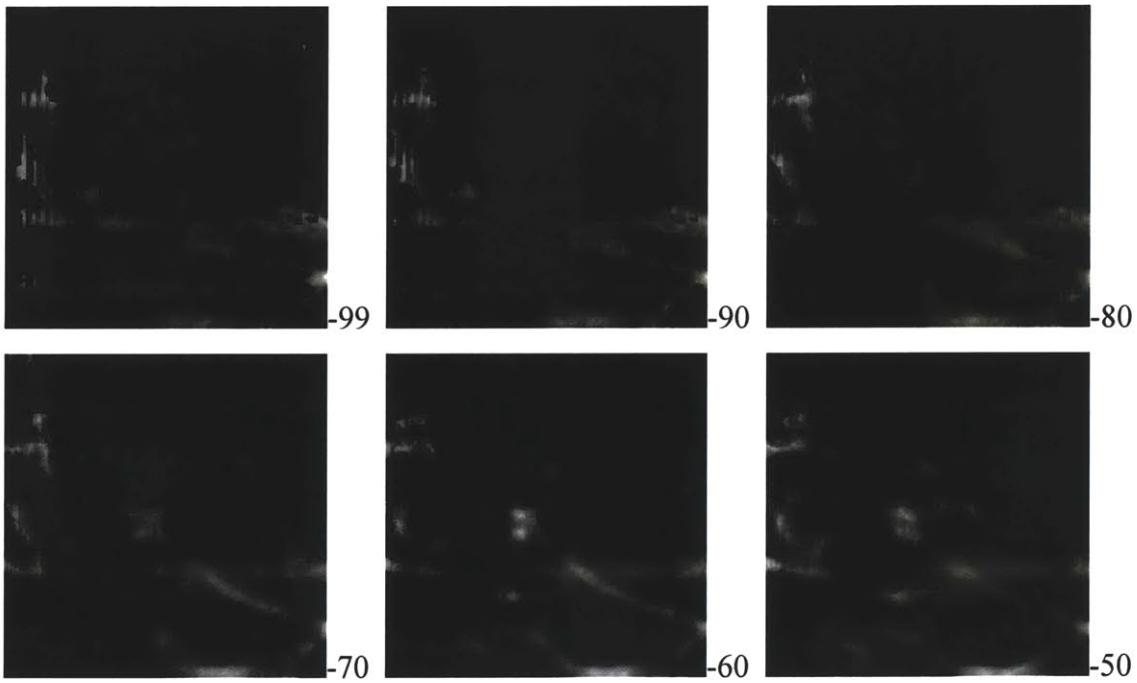


Figure 18 Selected cell cytoskeleton images taken on IVEM

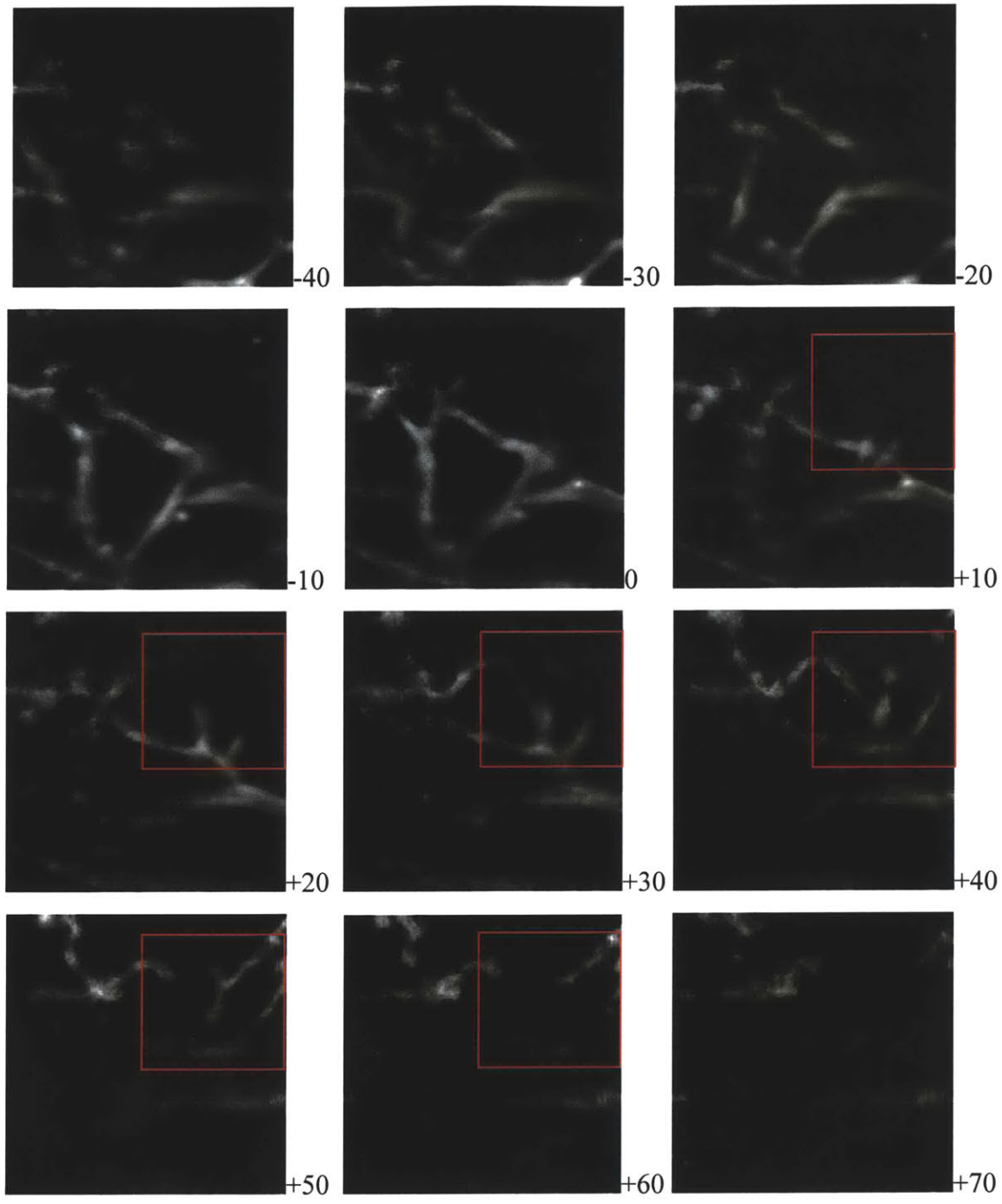
4.3.1 Reconstruction Result

As discussed in Chapter 2, the markers used in alignment gives a rough range of the object space. Based on this, we define a volume that the proposed method attempts to reconstruct. The volume size is chosen to be 201×201×201. Before reconstruction, some

pre-processing operations are employed to enhance contrast and reduce noise in the original image. Such operations are possible because they won't change the geometric relationship between the object and the images. The proposed method is then applied to reconstruct this volume. We conduct moving stereo reconstruction on 21 pairs of stereo images ranging from –48.0° to 46.0°. These pairs are formed from 22 images with tilt angles {-48.0°, -46.0°, -42.0°, -38.0°, -34.0°, -30.0°, -26.0°, -20.0°, -16.0°, -10.0°, -4.0°, 0.0°, 4.0°, 8.0°, 12.0°, 16.0°, 20.0°, 26.0°, 30.0°, 34.0°, 42.0°, 46.0°}. These images are selected from available images. We try to select the ones that have better alignment and image quality. Each pair is immediate-sequential images in the sequence, such as *i*-th image and (*i*+1)-th image. Local interaction is included to encourage local smoothness. Figure 19 shows some slices of the reconstructed volume. Each slice represents a *XY*-plane. The *Z*-coordinate is on the right side of each slice. *Z*-coordinate is measured with respect to the origin which is in the center of the volume.



(to be continued in next page)



(to be continued in next page)

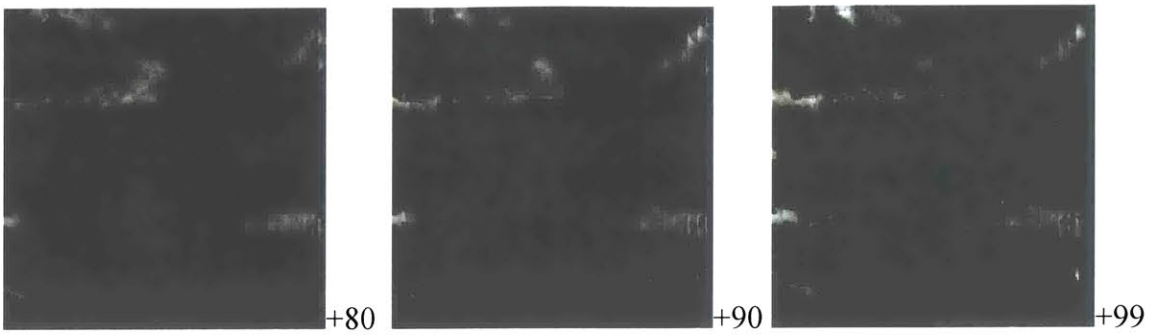


Figure 19 The *Z*-slices of the reconstructed volume with 22 images

The reconstructed volume shown in Figure 19 helps to demystify the 3D structure of the cell cytoskeleton. We can identify and even measure the locations and relationships of the filaments in the volume. For instance, when we look at slices from (+10) to (+60) sequentially, we can clearly see how the structure evolves at the right-top corner of each image (labeled with small red window). The 3D structures and their relationships can also be observed by using visualization techniques. For example, Figure 20 displays a 3D visualization of this volume. The technology used for visualization is ray casting method[4.3].

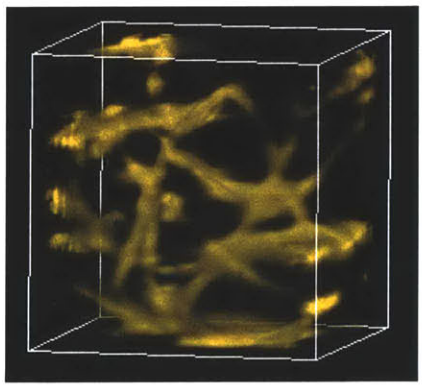


Figure 20 Visualization of the reconstructed volume by ray-casting method

[4.3] The software package we use for visualization is called VolView, developed by Kitware. An evaluation version can be downloaded freely from *http://www.kitware.com/volview.htm*.

In both Figure 19 and Figure 20, we also see some artifacts in some areas. These artifacts are caused because of uncertainty in the reconstruction. There are several sources of uncertainty. One is the poor image quality as observed in Figure 18. Poor imaging conditions, such as poor contrast, noise, inconsistent focus, etc, lead to ambiguity or even incorrect reconstruction. Another source is occlusion. Since some areas may move out the field of view as we rotate the volume, their projections may only exist in some images not in the others. For example in Figure 19, artifacts are more visible in the areas that are close to the boundary of the volume, such as slices from (+90) to (+99), or slices from (-99) to (-90). Objects in these areas have large motions and more likely move out of the image window when we tilt the object volume. One common to alleviate this problem is to conduct reconstruction on an extended volume and then crop the volume to a smaller size (Herman 1980; Suzuki and Yamaguchi 1988).

4.3.2 Number of Images

As presented in this Chapter, reconstruction from multiple images achieves much better precision and accuracy than reconstruction from two images. It is straightforward that the more images we use, the better reconstruction we may obtain. However, the addition of one more image adds additional computation, but may make very little improvement if many images have already been used. Therefore, a realistic question arises: how many images are enough in practice? This section attempts to answer this question in an empirical way.

From the sequence of available images, we select a different number of images to do different experiments of reconstruction. The images used in each experiment cover the same range of tilt angles as in Figure 19. Figure 21 plots the quality of reconstruction with respect to the number of images used. Since we don't know real objects in the volume, RMSE in Figure 21 is actually a relative measure. We use the reconstructed objects from 30 images as reference objects in calculating RMSE. It is equivalent to a shift from zero line, and it won't change the shape of the plot. Figure 21 illustrates that the reconstruction quality improves as the number of images used for reconstruction

increases. It also shows that the improvement rate starts to slow down as more and more images are used, which is quite intuitive. When many images have been used, all information may already have been told by these images. The addition of an extra image may not yield any more new information. In this case, the gain from an extra image may not be worth the extra cost in computation and preparation of images. In practical implementation, it is useful to have an optimal number of images for guidance. Based on the plot in Figure 21, we would suggest the use of 10~20 images for reconstruction.

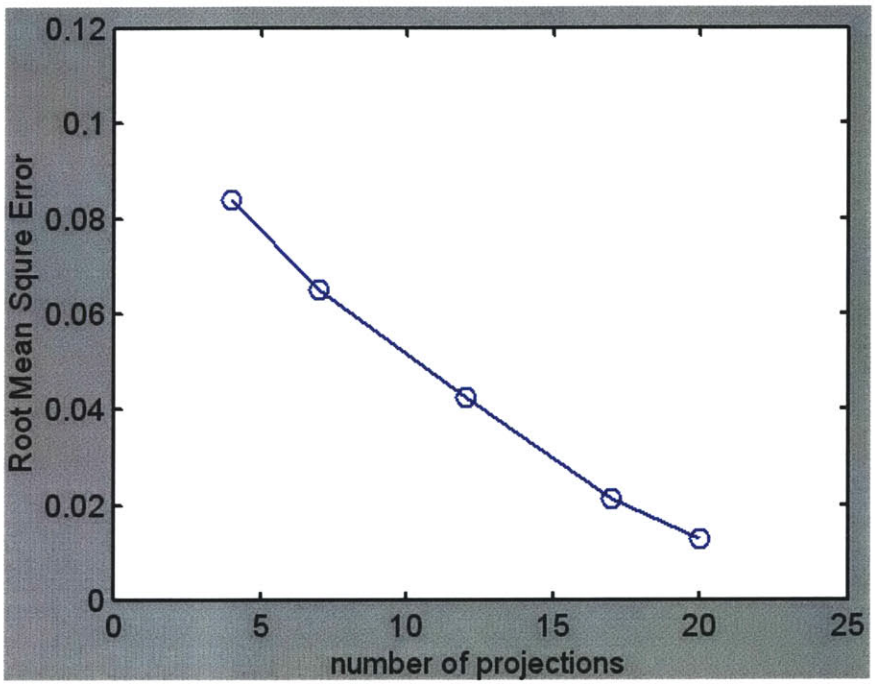


Figure 21 The quality improves as more images are used for reconstruction.

4.4 Discussion and Conclusion

In this Chapter, we presented a reconstruction method from multiple images. The method explored a volumetric reconstruction. Unlike surface reconstruction, the proposed method estimates each object point in the volume on a voxel-by-voxel basis. The overall procedure includes alignment, moving stereo reconstruction and integration. Local interactions, such as local smoothness, can also be incorporated into the integration process. Several integration methods that integrate a sequence of 3D reconstruction from

different stereo pairs are presented. By evaluating certainty values of each stereo reconstruction, each method achieves a globally optimized reconstruction with respect to a different cost function.

One important advantage of the proposed method is flexibility. The method can deal with an arbitrary number of images taken from different tilt angles. The choice for image pair used in each moving stereo reconstruction can be flexible too. The main drawback of the method may be its computation cost. One of future work could be to improve computational efficiency. However, in reality, computation in reconstruction is not the most time-costly step. Steps, including specimen preparation and marker identification for alignment, are more time-consuming procedures. It may be necessary to improve the efficiency of these steps as well.

Finally, we would like to draw some analog between the proposed method and tomography. It is very interesting to explore similarities between them. In fact, the proposed algorithm employed some concepts and ideas developed in tomography. As discussed in Chapter 1, tomography is a technique that reconstructs an image or a volume from its projections. The reconstruction involves two operations: backprojection and filtering. Backprojection simply integrates or sums all the values from each projection as expressed in Equation (1.12). The integration process discussed in section 4.2.2 is actually very similar to the backprojection operation. It also uses sum or weighted sum to obtain the best estimation. The only difference between them is that backprojection in tomography sums the intensity values of each projection while our integration process performs a probability backprojection that sums the probability values from moving stereo reconstruction. Intuitively, this backprojection type operation makes sense. First, it accumulates all the information from each stereo reconstruction. Secondly, it treats each projection equally, which is a very important property. It means that each projection has the same influence in determining the final result. And the result is independent to the order of projections. The other process in tomography is the filtering. In our volumetric reconstruction method, the integration process can be considered a filtering process. As indicated in Equation (4.14) or (4.21), the integration process involves an averaging or

weighted averaging operation. Equivalently, it conducts low-pass filtering which is contrary to high-pass filtering used in tomography. Therefore, our proposed method shares a similar computational architecture to that of tomographic reconstruction. This observation suggests a connection between computer vision based reconstruction and tomographic reconstruction. More connections could be explored as part of future work.

Chapter 5

Discrete Tomography

As discussed in previous chapters, computer vision based reconstruction recovers 3D structure based on the geometric relationships between images. The imaging function doesn't play a role in reconstruction. This type of methods could be a good choice when the imaging relationship between the 3D object and its images is very complicated or the imaging function is unknown. In this Chapter, we turn our attention to tomographic reconstruction in which the imaging function is that the value in each projection is the integral of the object values along projection direction. In transmission electron microscopy, the imaging relationship between projection and object can be approximated to meet this form. Electron tomography has been a fundamental technique that applies tomographic reconstruction methods to study macromolecular or sub-cellular level structures, such as virus or ribosome (Frank 1996).

In this Chapter, we will focus on a special case of tomography, namely discrete tomography. Discrete tomography deals with a set of tomography problems in which the object to be reconstructed is composed of a discrete set of materials each with uniform values. In other words, each point in the object space only takes a finite number of possible values (Herman and Kuba 1999). Such condition or constraint is termed discreteness in this Chapter. In structural biology, there are many cases that meet this condition. For example, the specimen that we study often consists of only a few different objects and each object tends to have one different value. By explicitly modeling this information, discrete tomography aims to achieve better reconstruction quality and accuracy even with a fewer projections.

Not limited to biological structures, this Chapter studies discrete tomography in a rather general format. As described in Chapter 1, 3D tomographic reconstruction can be built upon 2D reconstruction by stacking 2D slices. Similar treatment is applied in discrete tomography. In this Chapter, we will focus only on 2D case of discrete tomography, in which 2D image is reconstructed from its projections. The methods should be easily applicable and extendable to 3D or multi-dimensional reconstruction. To avoid confusion with the term "discrete tomography", this Chapter will use the term "conventional tomography" to refer to the tomography that is commonly known to everyone.

5.1 Introduction and Overview

In inverse reconstruction, there are often cases where the volume (3D case) or image (2D case) to be reconstructed takes only a finite number of possible values. Questions often arise about whether knowing this information helps to make better reconstruction, and if so, how this information can be incorporated into reconstruction algorithms. Discrete tomography is a technique that attempts to address this kind of problems. Recently, it has drawn great attention in many applications, such as medical imaging and electron microscopy (Herman and Kuba 1999) due to its high reconstruction quality and accuracy, even with a small number of projections.

The simplest form of discreteness is the binary field. The reconstruction of binary field from its projections is often called binary tomography. It is probably the most studied form of discrete tomography (although the name of binary tomography came much earlier than that of discrete tomography). Without loss of generality, a binary field can be represented by a (0,1) field, where pixel values of 1 and 0 correspond to presence or absence of the object at a specific location. Projections are simply sums of zeros and ones. For example, Figure 22 depicts a 2D binary field defined on a finite lattice and its two projections. Reconstruction from these projections has been an interesting subject in discrete mathematics and geometry (Ryser 1957; Salzberg 1995; Gardner and Gritzmann 1997; Vardi and Lee 1998; Yagle 1998; Herman and Kuba 1999). It has been

demonstrated that a binary image (or field) can be reconstructed from as few as only two projections (Kuba 1984; Fazekas, Herman et al. 1998; Vardi and Lee 1998). However, theories and approaches developed for binary tomography are hardly applicable to more general discrete tomography problems where the reconstructed objects are composed of multiple gray levels.

2	1 0 1 0 0
2	1 1 0 0 0
3	1 1 0 1 0
1	0 0 0 0 1
	3 2 1 1 1

Figure 22 A binary lattice and its row and column projections

To approach general discrete tomography problem, Frese *et al* (Frese, Bouman et al. 1998) proposed an iterative coordinate descent (ICD) algorithm that deals with multi-level discrete tomography. The reconstruction is formulated as a conventional tomography problem. The discreteness is simply modeled as a constraint to an optimization criterion. ICD is a greedy algorithm that performs a pixel-wise update. At each update, current pixel loops through all possible discrete values and chooses the one that maximizes the optimization criterion with all other pixel values fixed. This process is clearly very expensive in terms of computation. In addition, since the optimization is obtained only at a pixel level at each update, the reconstruction is very likely to be trapped by local maximums, instead of the global maximum. More recently, Chan *et al* (Chan, Herman et al. 1999) developed a two-step reconstruction procedure that is capable of dealing with general discrete tomography. This procedure performs a conventional tomographic reconstruction followed by a classification (or segmentation) process, i.e., discrete tomography is broken into two separate processes: conventional tomography and segmentation. It allows us to take advantage of rich literature in both well-studied fields. However, this procedure also loses an integral view about the problem. One common observation from both papers is that they both treat discrete tomography as a special case of conventional tomography and use existing models of conventional tomography in discrete tomography.

The approach this Chapter will propose, sharing the same theme as (Chan, Herman et al. 1999) and (Frese, Bouman et al. 1998), attempts to develop a framework for general discrete tomography, but with two key contributions. First, a new and integrated model is proposed for discrete tomography. This model incorporates discreteness into a Bayesian formula. A class label variable is defined to denote the probability of each object point taking one particular gray level. The problem of discrete tomography is then converted into a Bayesian labeling process, assigning a unique label to each object point. Second, the relationship between discrete and conventional tomography is derived mathematically, showing that conventional tomography can be treated as a special case of discrete tomography. This is quite contrary to what previous studies suggest.

The algorithm presented here is partially inspired by researches in adaptive segmentation (Wells, Grimson et al. 1996; Kapur, Grimson et al. 1998). In a more general sense, discrete tomography shares the same roots as segmentation. Both problems attempt to recover a discrete-valued field from continuous-valued observations. In adaptive segmentation (Wells, Grimson et al. 1996; Kapur, Grimson et al. 1998), each discrete-valued label class was modeled by a Gaussian function. An expectation-maximization (EM) algorithm was developed that treats the underlying label classes as hidden variables. The computation then alternates between estimating these classes and the inhomogeneous biases accounting for the offset of the real value to the class value. This approach has proven to be very successful (Wells, Grimson et al. 1996; Kapur, Grimson et al. 1998). We take a very similar approach in the development of our EM algorithm for discrete tomography. Our algorithm also uses continuous Gaussian functions to model the discrete-valued label variables. Details will be discussed later in the Chapter.

The EM (Dempster, Laird et al. 1977; Lange and Carson 1984; Green 1990) and other statistical methods (Shepp and Vardi 1982; Lange, Bahn et al. 1987; Mumcuoglu, Leahy et al. 1994) have been used widely and successfully in conventional tomography. The advantages are that they offer a great deal of flexibility in modeling and improve the performance in various situations, such as unusual projection geometry, uneven-sampled

projections, noisy situations, etc. However, the main drawback is that they have a very high computation cost due to their iterative nature. In search of an efficient algorithm, this Chapter introduces some reasonable approximations and relaxations. A linear filter is derived, with which the reconstruction can be computed efficiently. The process resembles efficient reconstruction procedures used in conventional tomography, namely the backprojection and high-pass filtering.

5.2 Review of Statistical Reconstruction

Before proceeding to describe the proposed method for discrete tomography, this section reviews some statistical reconstruction methods used in tomography. Some ideas presented in these methods are adopted in our algorithm for discrete tomography.

In Chapter 1, we have reviewed some commonly used tomographic reconstruction methods, including some efficient algorithms (such as FBP or FABP) and iterative methods using linear algebraic techniques. Statistical reconstruction methods are a different set of iterative methods. They formulate and develop the algorithm by introducing statistical models. Since statistical methods are iterative methods, computation is typically expensive and convergence may be slow. But on the other hand, since noises are properly modeled in the methods, statistical methods reduce statistical noise and artifacts to achieve better reconstruction quality. This property is particularly appealing to applications in PET or SPECT because these modalities generally produce low-quality projections with low spatial resolution. Conventional methods such as FBP don't work well on these situations.

In statistical methods, reconstruction is considered an estimation problem (Shepp and Vardi 1982; Green 1990). The goal is to obtain the best estimate of the object $\boldsymbol{x}$ from its projections $\boldsymbol{y}$. Each projection is modeled by a Poisson process:

$$y_t \sim \text{Poisson}\left(\sum_s a_{ts} x_s \right), \tag{5.1}$$

where i and t index the object and projections. a_{ti} is coefficient representing the emission contribution of pixel x_i on projection point y_t. If the Poisson processes are independent for all i and t. The conditional probability $p(\boldsymbol{y}|\boldsymbol{x})$ is then given by:

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_t \frac{\exp\left(-\sum a_{ts} x_s\right)\left(\sum a_{ts} x_s\right)^{y_t}}{y_t!} \quad . \tag{5.2}$$

If $\boldsymbol{x}$ is considered deterministic but unknown parameters, Equation (5.2) is essentially the likelihood function. As discussed in Chapter 1, we know that the best estimations may be given by the maximum likelihood (ML) method, i.e.,

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{y} \mid \boldsymbol{x}) \quad . \tag{5.3}$$

Solving this equation is not trivial at all because a large number of parameters (all x_s) need to be estimated. The article by Shepp and Vardi (Shepp and Vardi 1982) tackled this problem by using an expectation-maximization (EM) algorithm:

$$z_{ts} = \frac{y_t a_{ts} x_s}{\sum_{s'} a_{ts'} x_{ts'}} \tag{5.4}$$

$$\hat{x}_s = \frac{\sum_t z_{ts}}{\sum_t a_{ts}} \quad . \tag{5.5}$$

In general, the EM algorithm is an approach for maximizing a likelihood or *a posteriori* distribution when some data are "missing" or "hidden" in some sense, and observation of these hidden data would have greatly simplified the estimation of parameters (Green 1990). The EM algorithm is an iterative method, alternating between two steps: E-step and M-step. E-step estimates the expectation of the hidden variables using current model parameters. M-step conducts maximum-likelihood estimation of model parameters given current hidden variables. In the above formula, Z_{ts} is introduced as hidden variables. Equation (5.4) is the E-step and Equation (5.5) is the M-step. The EM algorithm alternatively computes one of the two equations (Equations (5.4) and (5.5)), assuming the other is given. The whole process is often called EM-ML algorithm (Lange and Carson 1984; Lange, Bahn et al. 1987; Green 1990; Ollinger 1994).

If combining Equations (5.4) and (5.5), the estimation of $\boldsymbol{x}$ from $\boldsymbol{y}$ can be obtained by a one-step iterative procedure:

$$\hat{x}_s^{(n)} = \frac{\hat{x}_s^{(n-1)} \sum_t \left(y_t a_{ts} \Big/ \left(\sum_{s'} a_{ts'} \hat{x}_{ts'}^{(n-1)} \right) \right)}{\sum_t a_{ts}} , \tag{5.6}$$

where (n) is the iteration number. $\hat{x}_s^{(n-1)}$ is the previous estimation of x_s, and $\hat{x}_s^{(n)}$ is the updated estimate based on previous estimation. This iteration is repeated until apparent convergence.

When $\boldsymbol{x}$ is considered random variables, the estimation of $\boldsymbol{x}$ can be obtained by maximum *a posteriori* (MAP) method based on Bayesian theorem, i.e.,

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\max_{\boldsymbol{x}} \frac{p(\boldsymbol{y} \mid \boldsymbol{x}) p(\boldsymbol{x})}{p(\boldsymbol{y})} . \tag{5.7}$$

Similar to the EM-ML algorithm, an EM-MAP algorithm can be developed (Levitan and Herman 1987; Chen, Johnson et al. 1990; Green 1990; Herman and Odhner 1991; Chan, Herman et al. 1997). Markov random field (MRF) is often used to model *a priori* probability, $p(\boldsymbol{x})$, in the EM-MAP algorithm.

5.3 Description of Method

In this section, we will describe the proposed method of discrete tomography in details. The method is developed within a Bayesian framework. Some modeling concepts and procedures discussed in previous section are adopted in the method.

5.3.1 Notations

As a convention, a variable name without subscript will denote a vector representing a whole set of the same type of variables if applicable. When a vector is used to describe an image (i.e. 2D matrix), it refers to its row-wise representation. For example, a pixel (i, j)

in an image of size w×h corresponds to a variable at the index (i·w+j) in the vector. An index mapping is given by:

$$(i, j) \leftrightarrow (i \cdot \mathrm{w} + j). \tag{5.8}$$

This representation applies to both object space and projections. Projections are organized in an image format, called sinogram. Each row of sinogram represents a projection at the corresponding direction. If there are h projections and each projection has w samples, sinogram is an image with size w×h.

The notations are defined as follows and will be used throughout this Chapter:

- i *or* j: an index variable used to index the object points in a vector form. i (or j) = 1,2,…,n where n is the total number of points in the object field. For example, if the object image has size of w×h, n = w·h.
- t: an index variable used to index the points in a vector form of the sinogram. t = 1,2,…,m where m is the total number of points in all given projections. If the sinogram has size of w×h, m = w·h.
- u_i: a random variable referring to the i-th object point or its value (i = 1,2,…,n).
- $\boldsymbol{u}$: a vector variable representing the whole set of object points, $\{u_i\}$.
- v_t: a random variable referring to the t-th projection point or its value (t = 1,2,…,m).
- $\boldsymbol{v}$: a vector variable representing the whole set of projection points, $\{v_t\}$, i.e. the sinogram.
- a_{ti}: a coefficient relating the i-th object point to the t-th projection point. It is determined by projection geometry and projection physics. We assume they are known and accurate.
- g: an index variable used to index the class labels. g = 1, 2,…,s indicates there are s discrete levels in the object field to be reconstructed. s is typically known.
- $\mu(g)$: the ideal value of the g-th class (g =1,2,…,s), representing each gray level in the discrete object field. They may not be known but would be very helpful if known.

Γ_{ig}: a class label for the i-th object point and $\Gamma_i = g$ which means i-th object point takes g-th gray level.

Γ: a vector representation of the whole set of class label variables, $[\Gamma_{10}, \ldots, \Gamma_{1s}, \Gamma_{20}, \ldots, \Gamma_{ns}]^T$. The vector has size of n×s.

w_{ig}: *a posteriori* probability of object point i belonging to the class g, given that u_i is known. This is the variable the algorithm attempts to estimate for discrete tomography.

w_{i*}: a vector formed by w_{ig} over all g, i.e., $[w_{i1}, \ldots, w_{is}]^T$.

$P[\ ]$: a probability function of some random variable.

φ_{tg}^2: the variance of i-th object point u_i taking g-th class value.

σ_t^2: the variance of additive random noise at projection point v_t.

As a summary, we have a discrete valued field (the image) with total s discrete levels. The ideal value of each discrete level is denoted by $\mu(g)$, $g = 1,2,\ldots,s$. The goal of discrete tomography is to reconstruct this field from its projections represented as a vector v. We introduce a new set of variables w_{ig}, representing the probability of an object point (e.g., i-th point) belonging to one of the discrete classes (e.g., g-th class). Clearly, the sum of w_{ig} over all g is equal to 1, or

$$\sum_{g=1}^{s} w_{ig} = 1 \ , \qquad \forall i \ . \tag{5.9}$$

In an ideal case, any w_{ig} should be either 0 or 1, and the vector w_{i*} should be drawn from the set $\{[100\ldots0]^T, [010\ldots0]^T, \ldots, [000\ldots1]^T\}$, indicating each object point belongs to only one class exclusively. The reconstruction of discrete tomography is then equivalent to a labeling process, assigning such vector to each object point. However, this ideal case is difficult to approach in practice. In search of a computable algorithm, we relax w_{ig} to be a variable that may vary continuously between 0 and 1, but still satisfy Equation (5.9). In this case, the vector w_{i*} will only imply the most likely class that an object point may belong to. If it is necessary to enforce discreteness, a discrete winner-take-all scheme can be employed (Elfadel 1993). For example, we can simply set the largest w_{ig} to be 1 and all others to be 0's, and the vector w_{i*} will then be one of those unit vectors.

Along with variable w_{ig}, another set of variables, u, is also introduced to describe the object field. u is a random variable which models a noise-added version of discrete object field. Both sets of variables are unknown and need to be estimated from projections v. As described later, we employ an EM algorithm to alternately estimate these two unknowns at the same time. In the EM algorithm, variables w_{ig} are treated as hidden variables. Details of the EM algorithm will be presented in later sections.

We assume that each class value $\mu(g)$, g =1,2,...,s, is known in advance. But in practice, it may not be the case. If so, another level of estimation can be developed and incorporated into the EM algorithm. Frese *et al* (Frese, Bouman et al. 1998) presented a region-based method to estimate the discrete levels concurrently with the reconstruction. The same idea is adopted in our algorithm when class values are not known.

5.3.2 Models Development

The models are built upon a Bayesian framework. The first model is for object field. Let us assume that each point in the object field takes one of s (g=1,2,...s) discrete-valued gray levels, i.e., there are s sets of objects in the field and each set of object has the same value. Due to the stochastic nature of variations, the value at each point is a random variable. Its probability can be modeled as a Poisson distribution (Shepp and Vardi 1982; Lange, Bahn et al. 1987; Mumcuoglu, Leahy et al. 1994):

$$P[u_i \mid \Gamma_i = g] = \frac{\mu(\Gamma_i)}{u_i!} e^{-\mu(\Gamma_i)} \tag{5.10}$$

However, since a Poisson distribution is limited to an integer variable and is impossible to differentiate with respect to the variable u_i, a Gaussian approximation is adopted:

$$P[u_i \mid \Gamma_i = g] = \frac{1}{\sqrt{2\pi\varphi_{ig}^2}} e^{-(u_i-\mu(\Gamma_i))^2/2\varphi_{ig}^2} \quad , \quad \forall i \tag{5.11}$$

where φ_{ig}^2 is the variance of i-th point u_i taking g-th class value. To be comparable to the Poisson model, the variance φ_{ig}^2 can be chosen close to $\mu(\Gamma_i)$. Since Γ_i=g (g=1,2,...s) is

finite, marginal probability can be computed from the sum of conditional probability over all g:

$$P[u_i] = \sum_{g=1}^{s} P[u_i \mid \Gamma_i = g] \cdot P[\Gamma_i = g], \quad \forall i \tag{5.12}$$

where $P[\Gamma_i=g]$ is *a priori* probability distribution of g-th class. Let us first consider $P[\Gamma_i=g]$ to be known and stationary. Later, we will use a Markov random field (MRF) to model a more general case. We further assume that the probability function of each object point is independent to each other. Therefore, the model for the whole set of object points $\{u_i\}$, or $\boldsymbol{u}$, is given by:

$$P[\boldsymbol{u}] = \prod_{i=1}^{n} \left(\sum_{g=1}^{s} P[u_i \mid \Gamma_i = g] \cdot P[\Gamma_i = g] \right) \tag{5.13}$$

The second model is for projections. The projection is simply assumed to be the integral (i.e., the sum) of the emission or transmission along a projection path. It accumulates all contributions from the object points on the path, weighted by coefficient a_{ti}. The projection value is given by:

$$v_t = \sum_{i=1}^{n} a_{ti} \cdot u_i + \beta_t, \quad t = 1, 2, \ldots m \tag{5.14}$$

where β_t is modeled as additive zero-mean Gaussian noise. The coefficient a_{ti} is assumed to be known. The conditional probability of v_t is then derived as:

$$P[v_t \mid \boldsymbol{u}] = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-(v_t - \sum_i a_{ti} \cdot u_i)^2 \,/\, 2\sigma_t^2} \tag{5.15}$$

where σ_t^2 is the variance of β_t. Similarly, we also assume $P[v_t|\boldsymbol{u}]$ is independent for each v_t. The model for whole set of projection points $\{v_t\}$, or $\boldsymbol{v}$, is then expressed as:

$$P[\boldsymbol{v} \mid \boldsymbol{u}] = \prod_{t=1}^{m} P[v_t \mid \boldsymbol{u}] \,. \tag{5.16}$$

5.2.3 EM Algorithm Development

With Equations (5.13) and (5.16), the *a posteriori* probability function is then formulated and written as:

$$P[\boldsymbol{u} \mid \boldsymbol{v}] = \frac{P[\boldsymbol{v}|\boldsymbol{u}] \cdot P[\boldsymbol{u}]}{P[\boldsymbol{v}]} = \frac{\prod_t P[v_t|\boldsymbol{u}] \cdot \prod_i (\sum_g P[u_i|\Gamma_i = g] \cdot P[\Gamma_i = g])}{P[\boldsymbol{v}]} \tag{5.17}$$

where $P[\boldsymbol{v}]$ is independent of u and can simply be considered a normalization constant. The log form of Equation (5.17) becomes

$$\ln P[\boldsymbol{u} \mid \boldsymbol{v}] = \sum_t \ln P[v_t \mid \boldsymbol{u}] + \sum_i \ln(\sum_g P[u_i \mid \Gamma_i = g] \cdot P[\Gamma_i = g]) + const. \tag{5.18}$$

According to Bayesian theorem, the best estimation of u from observation v is the one that maximizes the *a posteriori* function. This maximum-*a-posteriori* (MAP) estimation is formulated by

$$\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} P[\boldsymbol{u} \mid \boldsymbol{v}] \tag{5.19}$$

A necessary condition is that its derivative to any u_i equals zero:

$$\left(\frac{\partial}{\partial u_i} \ln P[\boldsymbol{u}|\boldsymbol{v}]\right)_{\boldsymbol{u}=\hat{\boldsymbol{u}}} = \left(\sum_t \frac{\frac{\partial}{\partial u_i} P[v_t|\boldsymbol{u}]}{P[v_t|\boldsymbol{u}]} + \frac{\sum_g P[\Gamma_i = g] \cdot \frac{\partial}{\partial u_i} P[u_i|\Gamma_i = g]}{\sum_g P[\Gamma_i = g] \cdot P[u_i|\Gamma_i = g]} \right)_{\boldsymbol{u}=\hat{\boldsymbol{u}}} = 0 \qquad \forall i . \tag{5.20}$$

Substitution of Equations (5.11) and (5.15) into Equation (5.20) yields:

$$\left(-\sum_t \frac{\sum_j (a_{tj} \cdot u_j - v_t)}{\sigma_t^2} \cdot a_{ti} + \frac{-\sum_g P[\Gamma_i = g] \cdot P[u_i|\Gamma_i = g] \cdot \frac{u_i - \mu(g)}{\varphi_{ig}^2}}{\sum_g P[\Gamma_i = g] \cdot P[u_i|\Gamma_i = g]} \right)_{\boldsymbol{u}=\hat{\boldsymbol{u}}} = 0 \qquad \forall i . \tag{5.21}$$

Let us define

$$w_{ig} = \frac{P[\Gamma_i = g] \cdot P[u_i|\Gamma_i = g]}{\sum_g P[\Gamma_i = g] \cdot P[u_i|\Gamma_i = g]} . \tag{5.22}$$

It is easy to verify that:

$$\sum_{g=1}^{S} w_{ig} = 1, \qquad \forall i . \tag{5.23}$$

w_{ig} contains unknown variable u_i. If u_i is given, w_{ig} can be computed by Equation (5.22). w_{ig} actually represents *a posteriori* probability of point i belonging to the class g. Using Equation (5.22), Equation (5.21) is rewritten as follows:

$$\sum_g \frac{w_{ig}}{\varphi_{ig}^2} u_i + \sum_j \sum_t \frac{a_{ti}a_{tj}}{\sigma_t^2} u_j = \sum_g \frac{w_{ig}}{\varphi_{ig}^2} \mu(g) + \sum_t \frac{a_{ti}}{\sigma_t^2} v_t \qquad \text{at } u = \hat{u}, \forall i \tag{5.24}$$

Equation (5.24) forms a linear equation about u_i. If w_{ig} is known, u_i can be obtained by solving this linear equation. Therefore, Equations (5.22) and (5.24) constitute an expectation-maximization (EM) algorithm. There are two iterative steps: calculating w_{ig} from Equation (5.22) based on the current u, and computing u from linear Equation (5.24) using current w_{ig}. This iteration eventually converges to a solution that maximizes the *a posteriori* function in Equation (5.19).

In order to exploit more insights from Equation (5.24), let us consider a simple case where we assume all $\sigma_t^2=\sigma^2$ for any t and $\varphi_{tg}^2=\varphi^2$ for any i or g. Thus, Equation (5.24) is simplified as:

$$\frac{1}{\varphi^2} \sum_g w_{ig} u_i + \frac{1}{\sigma^2} \sum_j \sum_t a_{ti}a_{tj}u_j = \frac{1}{\varphi^2} \sum_g w_{ig} \mu(g) + \frac{1}{\sigma^2} \sum_t a_{ti} v_t \qquad \text{at } u = \hat{u}, \forall i \tag{5.25}$$

Let us denote

$$c_i = \sum_g w_{ig} \mu(g), \qquad \forall i . \tag{5.26}$$

Then Equation (5.25) is re-written as:

$$\frac{\sigma^2}{\varphi^2} \boldsymbol{u} + \mathbf{A}^T \mathbf{A} \boldsymbol{u} = \frac{\sigma^2}{\varphi^2} \boldsymbol{c} + \mathbf{A}^T \boldsymbol{v} \tag{5.27}$$

where $\mathbf{A}$ is a m×n coefficient matrix with its element being a_{ti}. $\boldsymbol{u}$ and $\boldsymbol{v}$ are column vectors representing the object field and projections respectively. $\boldsymbol{c}$ is a n×1 column-vector and its element is c_i, i.e., $[c_1, \ldots, c_n]^T$.

It is very interesting to see that conventional tomography is actually embedded in Equation (5.27). If we remove the first terms on both sides of Equation (5.27), it simply becomes the equation for conventional tomography, $\mathbf{A}^T\mathbf{A}\boldsymbol{u} = \mathbf{A}^T\boldsymbol{v}$. It also implies that the essential difference between discrete and conventional tomography lies in those two first

terms which account for the effect of discreteness. This observation is very powerful. In a general sense, Equation (5.27) addresses a generalized tomography problem, explaining the relationship between discrete and conventional tomography. In this format, discrete tomography is a more generalized form of tomography and conventional tomography is merely a special case of such generalization where the discreteness terms are simply neglected. The variances, σ^2 and φ^2, in Equation (5.27) are controllers that can be used to tune the degree of discreteness in a generalized tomography problem, with ideal discrete tomography and conventional tomography at two ends, respectively. Figure 23 demonstrates such relationship.

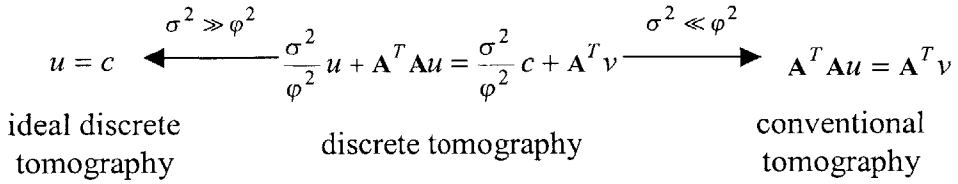


Figure 23 The relationship between discrete and conventional tomography

5.3.4 An Efficient Algorithm for Linear Equation

Equation (5.27) is a $m \times n$ linear equation with respect to unknown variable u. Solving this linear equation is generally very computation-costly. By introducing a computationally efficient filter, this section derives an efficient algorithm to solve Equation (5.27).

As argued in previous sections, conventional tomography is embedded in discrete tomography of Equation (5.27). Therefore, the solution for conventional tomography may help to develop an efficient algorithm for Equation (5.27). In Chapter 1, we reviewed some efficient algorithms such as filtered backprojection (FBP) and filtering after backprojection (FABP). Both FBP and FABP algorithms involve a filtering process and a backprojection process but in an opposite order. We also discussed that there exists a duality between these methods and linear algebraic methods. The duality indicates that matrix $\mathbf{A}$ is equivalent to a projection operator and $\mathbf{A}^T$ is a backprojection operator. In addition, $(\mathbf{A}^T\mathbf{A})^{-1}$ corresponds to the high-pass filter used in FBP method while $(\mathbf{A}\mathbf{A}^T)^{-1}$ is equivalent to the high-pass filter of FABP method.

We utilize these duality properties to develop an efficient algorithm for discrete tomography of Equation (5.27). The derived efficient algorithm is very similar to the efficient algorithm in conventional tomography. By multiplying $(A^TA)^{-1}$ to both sides of Equation (5.27), the solution of Equation (5.27) may be given by:

$$\boldsymbol{u} = \left(\frac{\sigma^2}{\varphi^2}\left(\mathbf{A}^T\mathbf{A}\right)^{-1} + \mathbf{I}\right)^{-1}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\left(\frac{\sigma^2}{\varphi^2}\boldsymbol{c} + \mathbf{A}^T\boldsymbol{v}\right) \tag{5.28}$$

where $\mathbf{I}$ is an identity matrix. This computation can be considered to contain two processes: a quasi-backprojection operation represented in the last term $\left(\frac{\sigma^2}{\varphi^2}\boldsymbol{c} + \mathbf{A}^T\boldsymbol{v}\right)$, and a filtering process represented by the first two terms, $\left(\frac{\sigma^2}{\varphi^2}\left(\mathbf{A}^T\mathbf{A}\right)^{-1} + \mathbf{I}\right)^{-1}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}$.

In FABP method, the high-pass filter represented in Fourier domain is given by (Rowland 1979; Toft 1996):

$$\mathbf{H}(\Omega_x, \Omega_y) = \sqrt{\Omega_x^2 + \Omega_y^2} \tag{5.29}$$

where Ω_x and Ω_y are frequency coordinates in 2D Fourier domain. Therefore, $(A^TA)^{-1}$ in Equation (5.28) essentially performs this high pass filter. In order to applying this filtering to Equation (5.28), row-vector representation of object points has to be reshaped into its 2D form, i.e., $u(x, y)$ where x and y are coordinates in 2D space domain, and with $U(\Omega_x, \Omega_y)$ as its Fourier transform. Substitution of Equation (5.29) into (5.28) gives us:

$$\mathbf{U}(\Omega_x, \Omega_y) = \frac{\mathbf{H}(\Omega_x, \Omega_y)}{\alpha\ \mathbf{H}(\Omega_x, \Omega_y) + 1}\mathbf{B}(\Omega_x, \Omega_y) \tag{5.30}$$

where coefficient $\alpha = \sigma^2/\varphi^2$ is a constant. $B(\Omega_x, \Omega_y)$ is the 2D Fourier transform of the term $\left(\frac{\sigma^2}{\varphi^2}\boldsymbol{c} + \mathbf{A}^T\boldsymbol{v}\right)$. $u(x, y)$ is obtained by taking an inverse Fourier transform of Equation (5.30). It is interesting to note that coefficient α can be used to control the discreteness. As coefficient α vanishes, Equation (5.30) simply becomes conventional tomographic reconstruction by FABP method. This confirms the relationship illustrated in Figure 23.

Using the filtering method discussed above, computation required to solve Equation (5.27) is greatly reduced. However, when implementing high-pass filter of Equation (5.28), some issues arise concerning the quality and performance. To achieve a result comparable to FAB method, certain measures need to be taken when using FABP method (Suzuki and Yamaguchi 1988). For example, experiments show that reconstruction quality improves when filtering is applied on an extrapolated larger size image area. Our algorithm adopts these measures.

5.4 Algorithm Implementation

As described in previous sections, the EM algorithm iteratively alternates estimations of w_{ig} and u, constituting two steps: E-step and M-step. E-step estimates w_{ig} from Equation (5.22) assuming u is known. M-step estimates u from Equation (5.28) given w_{ig}, and may be computed efficiently by Equation (5.30). They are summarized as follows:

E-step:
$$w_{ig} \leftarrow \frac{P[\Gamma_i = g] \cdot P[u_i | \Gamma_i = g]}{\sum_g P[\Gamma_i = g] \cdot P[u_i | \Gamma_i = g]} , \tag{5.31}$$

M-step:
$$\boldsymbol{u} \leftarrow \left(\frac{\sigma^2}{\varphi^2} \left(\mathbf{A}^T \mathbf{A} \right)^{-1} + \mathbf{I} \right)^{-1} \left(\mathbf{A}^T \mathbf{A} \right)^{-1} \left(\frac{\sigma^2}{\varphi^2} \boldsymbol{c} + \mathbf{A}^T \boldsymbol{v} \right) \tag{5.32}$$

5.4.1 A Priori Probability Model

In E-step of Equation (5.31), *a priori* probability $P[\Gamma_i=g]$ is assumed to be stationary. In this section, we employ a Markov random field (MRF) to model non-stationary *a priori* probability. As discussed in Chapter 1, MRF models are often used to enforce local regularity by introducing correlation between a site and its neighbors (Geman and Geman 1984; Li 1995). It has been successfully used in many applications to model various local properties, such as smoothness, continuity, etc. (Geman and Geman 1984; Geiger and Girosi 1991; Kapur 1999) Based on MRF theories, the joint probability function is defined as:

$$P[\boldsymbol{\Gamma}] = \frac{1}{Z} e^{-\frac{1}{T}E(\boldsymbol{\Gamma})} = \frac{1}{Z} e^{-\frac{1}{T}\sum_{c \in \boldsymbol{C}} V_c(\boldsymbol{\Gamma})} \tag{5.33}$$

where T is some constant, called the temperature. $\boldsymbol{\Gamma}$ is the vector representation of all Γ_i, i.e., $[\Gamma_1, \ldots, \Gamma_n]^T$. Z is simply a normalization constant. $E(\boldsymbol{\Gamma})$ is an energy function. $\boldsymbol{C}$ is a collection of all possible cliques defined upon a neighborhood system $\boldsymbol{N}$. $V_c(\boldsymbol{\Gamma})$ is a clique potential function that models the interactions among neighbors.

In our implementation, we limit $\boldsymbol{N}$ to be a first-order neighborhood. In other words, the neighbors of a pixel (i, j) at an image is defined as:

$$\boldsymbol{N}_{i,j} = \{(i+1, j), (i-1, j), (i, j+1), (i, j-1)\}. \tag{5.34}$$

Since we use the vector representation of the image, corresponding neighbors need to be mapped by Equation (5.8). Furthermore, we only consider single-site and pair-site cliques discussed in Chapter 1. We define clique potential function at each site and its neighborhood. The energy function is chosen as:

$$E(\boldsymbol{\Gamma}) = \frac{1}{2} \sum_{i=0}^{n} \sum_{i' \in \boldsymbol{N}_i} \left(\lambda_{gg'} \cdot \left(\mu\left(\Gamma_{ig}\right) - \mu\left(\Gamma_{i'g'}\right) \right)^2 \right), \tag{5.35}$$

where i' is a neighbor of i based on the defined neighborhood $\boldsymbol{N}_i$. g and g' are the class levels that Γ_i and $\Gamma_{i'}$ take. $\lambda_{gg'}$ is called color interaction coefficient. It reflects the interaction effect between neighboring variables. It may encourage or penalize certain combination of neighbors. For instance, if we want to encourage uniform neighborhood and penalize abrupt changes, we may set $\lambda_{gg'}$ be zero when g=g', and set $\lambda_{gg'}$ be positive when $g \neq g'$. Usually the color interaction coefficient is defined as a color interaction matrix (Elfadel 1993). In our implementation, two color interaction matrices are adopted for class level of 2 and 3, respectively, .

$$\mathbf{J}_{2\times 2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \; ; \quad \mathbf{J}_{3\times 3} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}. \tag{5.36}$$

Note that the reason we use the value of class level in Equation (5.35), instead of class level itself, is because we think the value of each class level is more important. There are other approaches to define the energy function, such as the ones used in (Elfadel 1993;

Kapur 1999). In general, one basic rule in defining energy function is that the energy is smaller when neighbors are more alike and vice versa. Under this definition, the MRF model will encourage smoothness and penalize discontinuities in the neighborhood.

Solving MRF is usually not trivial. In implementation, we adopt a Mean-Field (MF) approximation method developed in (Kapur 1999). The basic idea of MF approximation is that *a priori* probability is approximated by its mean field at each iteration. The mean field is estimated from current neighbors' values and their interactions.

5.4.2 Initialization

The EM algorithm is an iterative method. A good initialization not only speeds up convergence but also helps to prevent the maximization step from being trapped at local maximums. As discussed earlier, coefficient α $(=\sigma^2/\varphi^2)$ can be used to control the discreteness. At initialization, we start from $\alpha = 0$, with which M-step of Equation (5.32) essentially computes u as conventional tomographic reconstruction. In E-step, we start with an uniformly equal *a priori* probability $P[\Gamma_i=g]$. w_{ig} is calculated based on u from M-step. As iteration goes on, we gradually increase the coefficient α and estimate *a priori* probability using MF approximation at each iteration.

5.4.3 Estimation of Unknown Class Values

As discussed earlier, the class values may be unknown in advance. In this case, estimation of the class values is necessary. We adopt the region-based method proposed by Frese *et al* (Frese, Bouman et al. 1998), and incorporate this process into our EM algorithm. At each EM iteration, the class values are re-estimated, based on the current reconstruction result, i.e., w_{ig} and u values. The process of estimating class values itself is an iterative process. It resembles an EM-ML procedure discussed earlier in Equation (5.6) (Shepp and Vardi 1982; Green 1990). But in this case, the entire set of points belonging to the same class is treated as one single unit and updated iteratively. Given projections and current w_{ig}, the iterative estimation process can be derived as:

$$\mu_g^{(n+1)} \leftarrow \frac{\mu_g^{(n)} \sum_t \left(v_t \sum_i a_{ti} w_{ig} \Big/ \sum_i a_{ti} \sum_g w_{ig} \mu_g^{(n)} \right)}{\sum_t \sum_i a_{ti} w_{ig}}, \text{ for } g = 1, 2, ..., s. \qquad (5.37)$$

where (n) is the iteration number. The experiment shows that this iterative estimation process converges very quickly. Therefore, only several iterations are taken in our algorithm. The initialization of class values can be chosen from the classification of conventional tomographic reconstruction result.

5.5 Experiment and Results

The experiments are carried out on a synthetic phantom shown in Figure 24. The phantom contains three gray levels including background level, with intensity values μ_g = {0.11, 0.57, 0.99} where g = 1,2,3, respectively. The projections are calculated at evenly spaced angles. We conduct three experiments with different number of projections, i.e., 10, 20 and 30, respectively. The proposed algorithm is applied to obtain discrete tomographic reconstruction. In all experiments, we assume μ_g's are unknown but we know there are three levels of discreteness. In the implementation of the algorithm, we employ a MRF model to enforce local smoothness and use the color interaction matrix of Equation (5.36). Results are compared to reconstructions done by conventional filtered backprojection method. Figure 25 exhibits these reconstructions in parallel.

Figure 26 plots a quantitative measure of the reconstruction quality. It is measured by root mean square error (RMSE) with respect to the original phantom in Figure 24. The comparison demonstrates that reconstruction using our algorithm improves the result greatly, even when a small number of projections are given. In addition, for both cases, the RMSE error decreases considerably as more projections are used for reconstruction. Using our proposed algorithm speeds the error reduction. In other words, our algorithm can achieve even better quality with fewer projections than conventional FBP method with a larger number of projections.

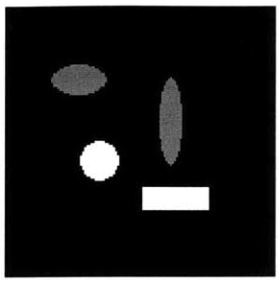


Figure 24 A synthetic discrete-valued phantom

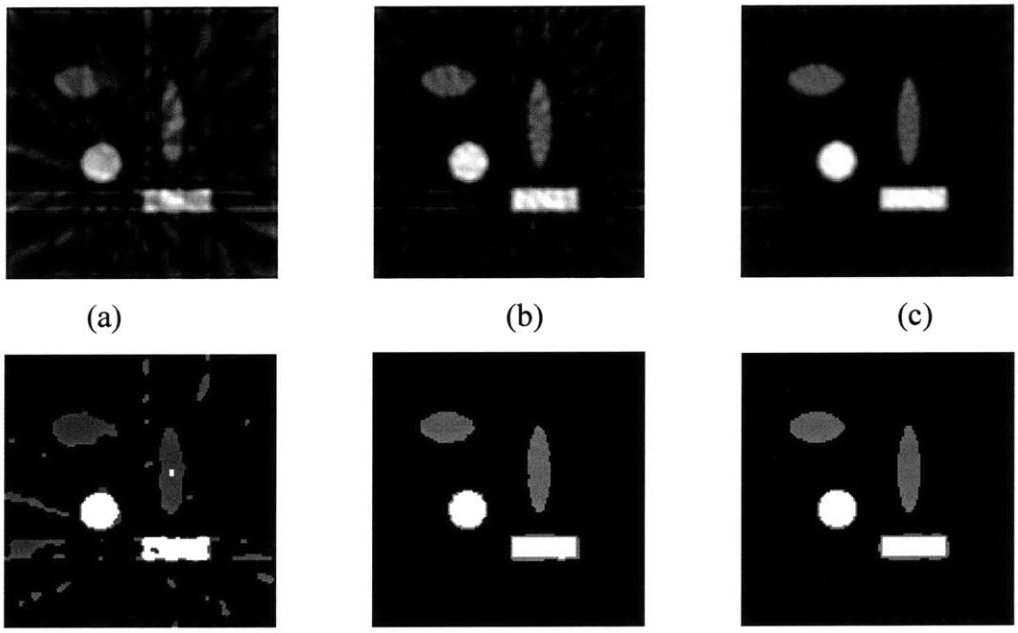


Figure 25 Reconstruction and comparison: (a,b,c) reconstructions using filtered backprojection (FBP) from projections of 10, 20, 30, respectively; (d,e,f) reconstructions done by our algorithm from projections of 10, 20, 30, respectively.

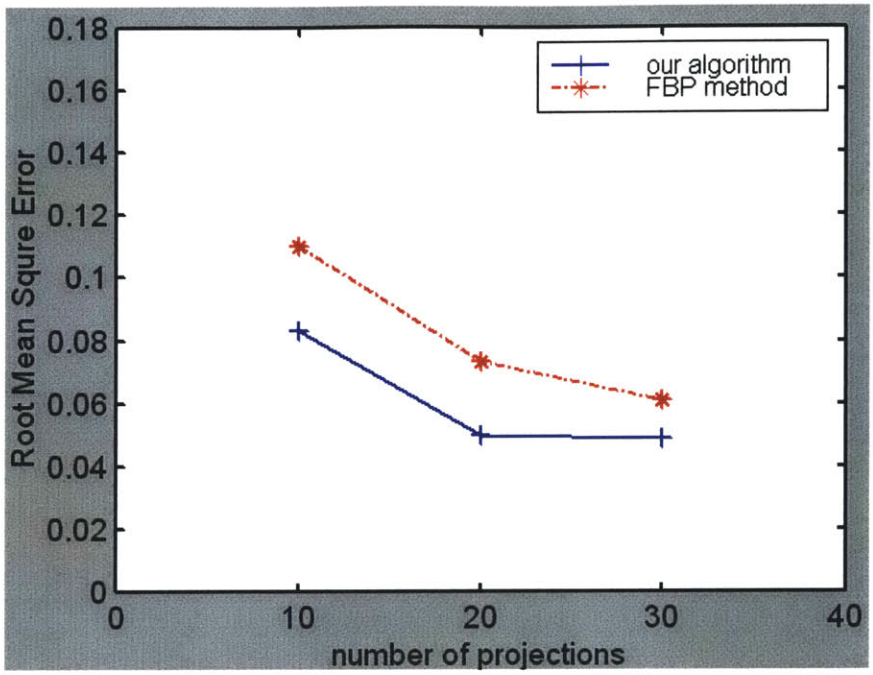


Figure 26 RMSE measure of reconstructions by proposed algorithm and FEB method

5.6 Discussion and Conclusion

We have described an integrated model for discrete tomography. The model incorporates discreteness into conventional tomography. An EM algorithm has been derived to accomplish discrete tomographic reconstruction. The algorithm can computed very efficiently with a derived high-pass filter. In comparison to the results obtained with conventional tomography, our proposed algorithm demonstrates improvement in the quality of the reconstruction even when a small number of projections are given.

In the proposed algorithm, discreteness can be controlled through a single coefficient, i.e., α ($=\sigma^2/\varphi^2$). When it vanishes, the algorithm simply degenerates to conventional tomographic reconstruction. The relationship depicted in Figure 23 has very interesting and powerful implications.

The proposed algorithm uses Markov random field (MRF) model to encourage local smoothness in the object field. Its importance may not be obvious when the object field is

not deteriorated. A MRF regularizer works especially well when noise exists in the field. Besides modeling local regularity, MRF can also be used to model other *a priori* information, such as geometric information, geometric relationship (Kapur 1999), etc. Generally, the more we know about the object field before reconstruction, the better results we can achieve.

The proposed models and algorithms resemble an adaptive segmentation process. In fact, the approach described here can be treated as an integrated version of two separate processes: tomographic reconstruction and segmentation. Such 2-in-1 processes can useful in some medical imaging applications.

Chapter 6

Conclusions and Future Directions

In this thesis, we have studied two different types of 3D reconstruction from 2D images: computer vision based reconstruction and tomographic reconstruction. The main difference between them is the imaging function that relates the intensity value of the projection to the value of the object. In tomographic reconstruction, the imaging function between the object and images is known and its form has to satisfy the projection-slice theorem. The computer vision based reconstruction, on the other hand, deals with cases where imaging function is complicated or unknown. The reconstruction is obtained by using geometric relationships among projections. This thesis has attempted to find both differences and connections between these two types of reconstruction. In Chapter 4, we suggested and explored some connections between them from an algorithm point of view. More studies could be conducted in the future to generalize the relationships.

Different algorithm for each type of reconstruction is developed in the thesis. As demonstrated in the experiments, specific focuses have been given to the application of 3D reconstruction of biological structures. With these experiments, we hope to bring both types of reconstruction to the attention of biological community. At the mean time, we would like to point out that both algorithms developed in the thesis have their generalities. They should be applicable to applications in other areas as well. Future work can extend the proposed algorithms to other applications.

Reconstruction provides accurate 3D structure for researches to understand the structure. It is often the first step in most researches. The next step would be to provide measurements based on the reconstructed 3D structures. It would be useful to be involved

with this part of researches in the future so that we may streamline reconstruction and measurements of the 3D structures.

Finally, we would like to mention that both computer vision reconstruction and discrete tomography remain largely unsolved problems. The intention of the thesis is not to provide perfect solutions for these problems, but rather to illustrate some possible ideas to approach the problem. For these ideas to hold in general applications, a great deal of research is needed not only in proving mathematical formulas but also in applying techniques to solve more real-world problems. These should be one of our main efforts in the future.

Appendix A

Java GUI Program for Image Pair Alignment and Reconstruction

In our cytoskeleton research, we often make measurements from only two images (image pair). Some procedures become very common in the study. This appendix describes some Java-based graphic user interfaces (GUI) that we developed to automate or help these procedures.

A.1 Alignment Program

The alignment method proposed in Chapter 2 requires one manually pick corresponding markers from each image. This process may be very labor-intensive. The Java program is developed to alleviate the workload. The program provides a GUI that allows user to pick markers, modify them, save or load them. The saved marker data is then fed into a matlab program which implements the algorithm discussed in Chapter 2. The matlab program computes the alignment and aligns images accordingly.

Figure A.1 shows a screen shot of this alignment program GUI. Some key features are summarized as follows:

- save and load markers into/from a file with a defined data structure.
- pick markers by mouse click in a matched order between two images.
- modify marker position by mouse drag or keyboard arrows.
- run animation function (under Function menu) that helps users to identify the marker more easily. It pops up a separate window in which the images are cyclically displayed. The scrollbar is used to adjust the speed. Figure A.2 exhibits a screen shot of such animation window.

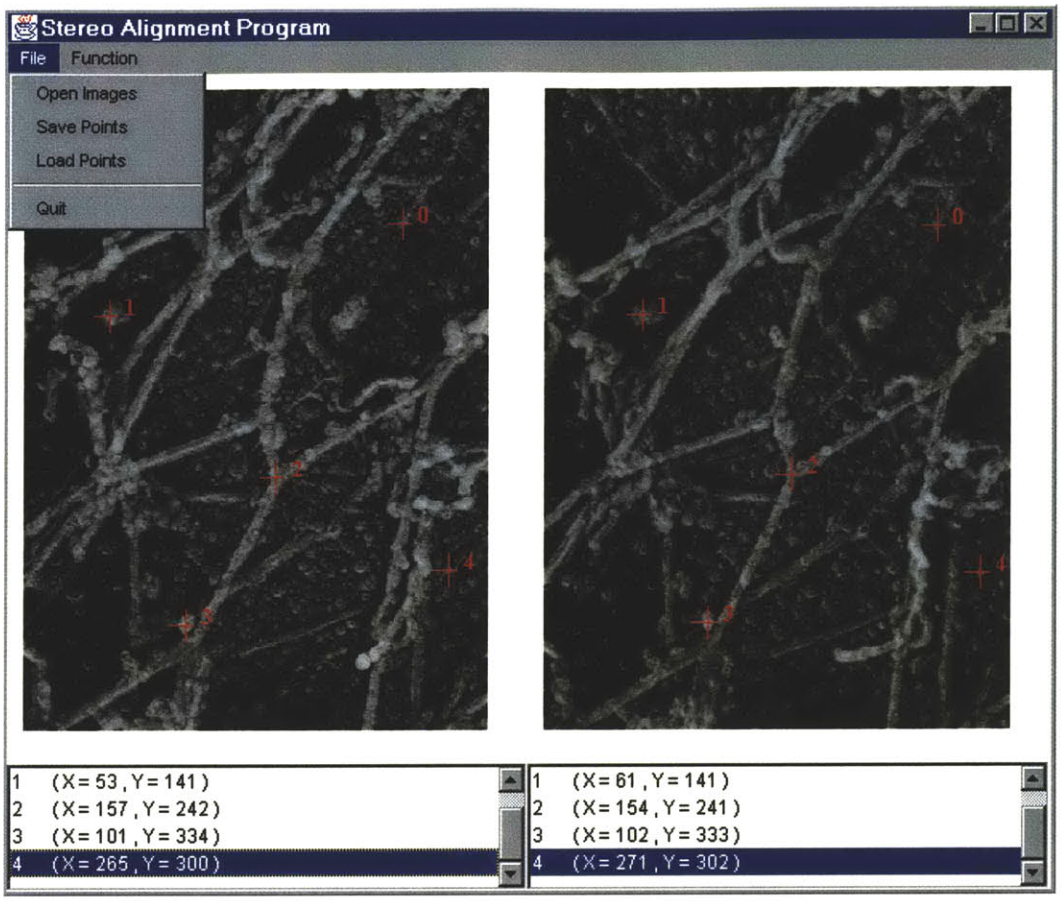


Figure A.1 Screen shot of GUI for alignment program

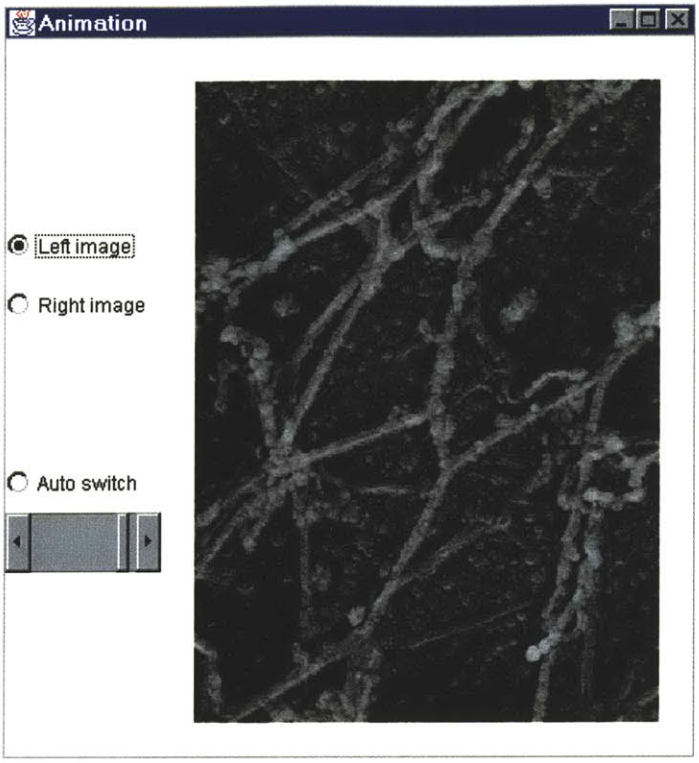


Figure A.2 a screen shot of animation window

A.2 Manual Stereo Reconstruction Program

This program is developed to help users who would like to do reconstruction manually. It was originally developed by Celine Paloc, a visiting student here in 1998. Some new features and functions are added in the new version. Figure A.3 shows a screen shot of the main window of stereo reconstruction program. Figure A.4 displays 3D visualization of the reconstructed filaments. In 3D visualization window, Java3D API is used for graphics.

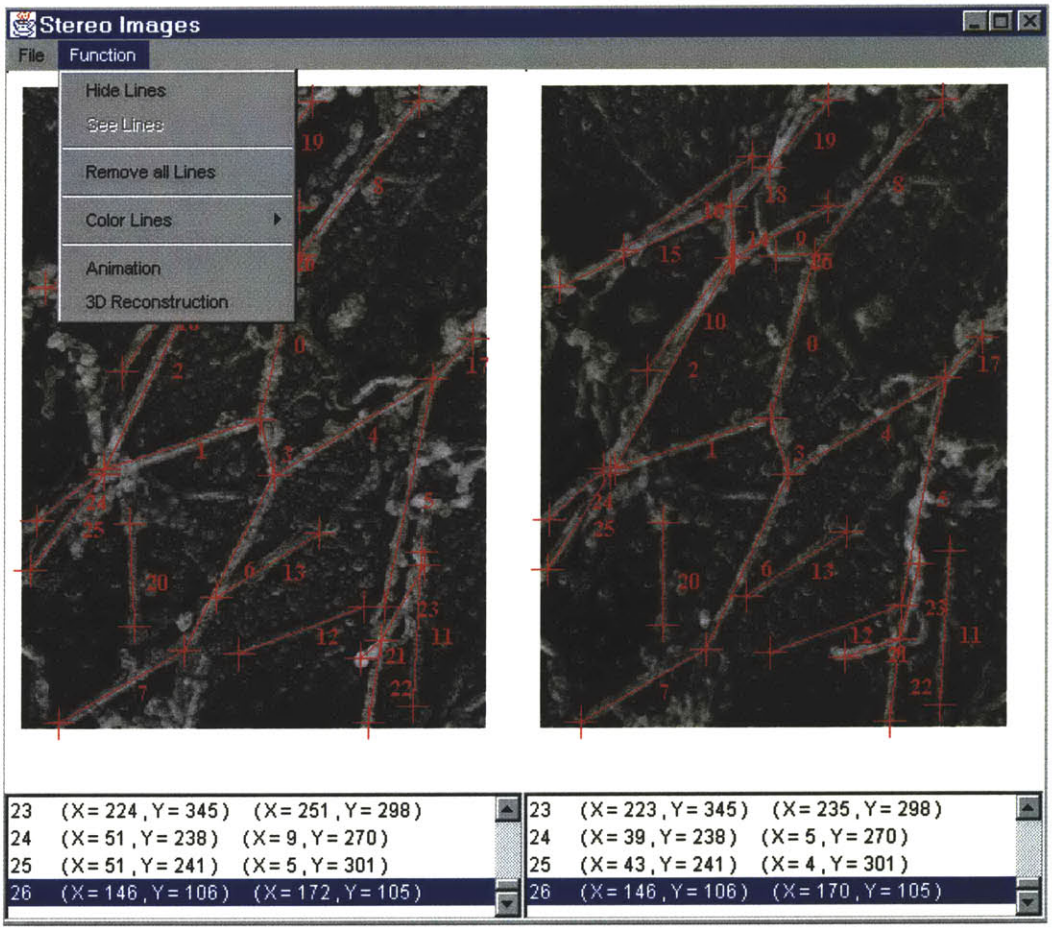


Figure A.3 a screen shot of stereo reconstruction program

Some key features of the program are listed as follows:

- draw matched lines in each image.
- modify line's position by dragging the end point or using keyboard arrows.
- save and reload matched line pairs into/from the file.

- compute reconstruction and save reconstructed lines into a file.
- pop up animation window which is the same as in alignment program.
- pop up 3D visualization window which displays and updates the reconstructed filaments in real time.
- use mouse for rotation, zoom, and translation in 3D visualization window.

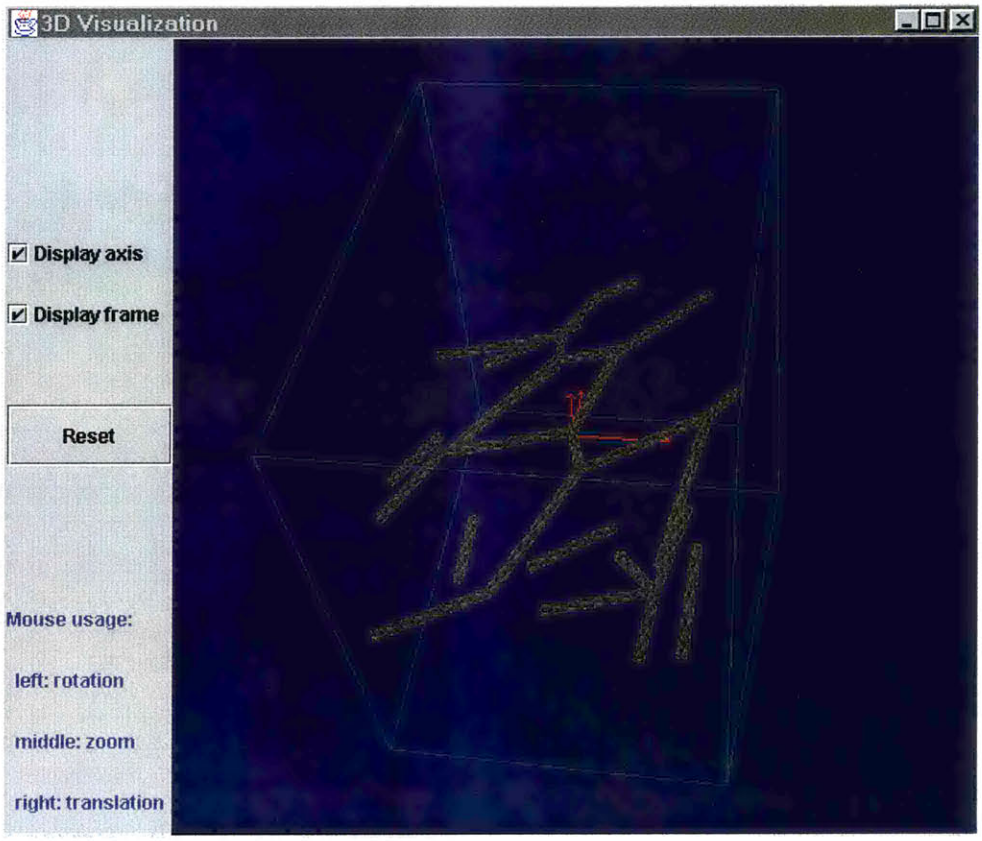


Figure A.4 a screen shot of 3D visualization window

Appendix B

Statistical Measurements of Cells Characteristics

Studies have demonstrated that vascular endothelial cells change their shape when subjected to fluid shear stress (Dewey, Bussolari et al. 1981; Remuzzi, Dewey et al. 1984; Depaola, Gimbrone et al. 1992). This morphological accommodation reflects the ability of cells to respond to external forces. Quantitative measurements of such morphological change are very important for understanding the cell dynamics and mechanics. This Appendix proposes and describes a statistical approach that offers a fast, quantitative measurements of two morphological characteristics: cell alignment direction and eccentricity (elongation ratio).

B.1 Cell Alignment Direction and Eccentricity

The proposed approach assumes that each cell has an ellipsoidal shape and the distribution of the cell sizes is the same anywhere in the region we consider. The direction of the cell alignment is chosen from the probability distribution function of cell directions. The eccentricity (elongation ratio) is obtained by comparing the cell length along the cell alignment direction (*i.e.* major axis) to the length along the perpendicular direction (*i.e.* minor axis).

The directional representation discussed in Chapter 3 is used to obtain cell directions, employing four quadrature filters to extract the orientation of each pixel relative to its neighborhood (local region) (Granlund and Knutsson 1995; Westelius 1995; Cheng, Hartmink et al. 2000). A probability function is then estimated from the data of cell directions. The estimation offers a smooth function so that the maximum can be easily

derived. Details of density estimation will be discussed in next section. In the probability function, we pick the location of the function maximum as the dominant direction of the cell alignment. This direction is also considered the major axis and its perpendicular direction is the minor axis.

Cell lengths are statistically measured from line segments contained between each cell along major and minor axes, respectively. Similarly, a smooth probability function of cell lengths is estimated for each axis from cell length data. The maximums are identified from probability functions and eccentricity (elongation ratio) is then calculated.

B.2 Probability Function Estimation

There are a number of techniques for estimating densities from data. Commonly, they are categorized into two groups: parametric and non-parametric techniques (Duda and Hart 1973). If we have some general knowledge about the probability function and the function can be characterized by a few parameters, the problem of estimating this probability function essentially becomes to estimate the parameter. For instance, a normal density is parameterized by its mean and variance. Parameter estimation is a classical problem in statistics, and it can be approached in several ways. Maximum likelihood (ML) estimation is a widely used method. In ML, the parameters are viewed as fixed but unknown constants. The best estimation is defined to be the one that maximizes the probability of obtaining the observed samples (Duda and Hart 1973). In the cases that these parameters are random variables themselves, Bayesian methods are often used. In a Bayesian framework, the best estimation is the one that maximizes *a posteriori* probability. However, one important limitation of parametric techniques is that the form of the probability function has to be known. In practice, this limitation does constrain the use of parametric methods.

On the other hand, non-parametric approaches provide more flexibility than parametric methods. By assuming a general model of the function, non-parametric methods use

samples to directly define the model (Viola 1995). For example, Parzen-window density estimation approach uses a very general form given by:

$$P(x,a) = \frac{1}{N_a} \sum_{x_a \in a} R(x - x_a) \tag{B.1}$$

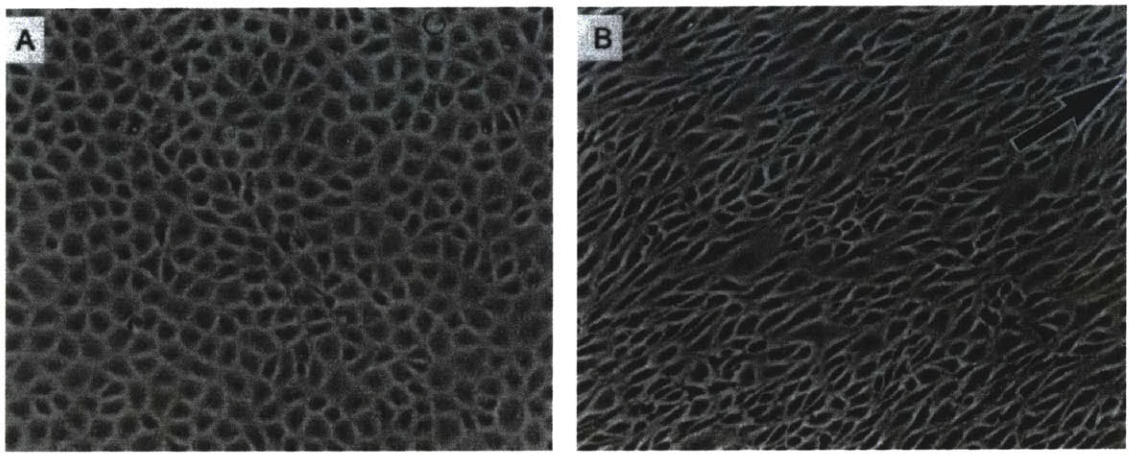
where a is sample data, N_a is the number of sample data, and R is a valid density function. The model described in Equation (B.1) approximates the actually probability function. The Parzen window estimation is a local or windowed estimator. The estimation at a query point, x, is the weighted sum over the samples inside a window. The samples that fall outside of the window do not contribute. The weighting scheme and window size are determined by the window function R. Therefore, the quality of estimation is very dependant on the window function R. A common selection of R is a Gaussian function. According to the model of Equation (B.1), there is one Gaussian centered at each sample. The variance of each Gaussian determines smoothness of the estimated probability function. A thorough study of Parzon window with the Gaussian window function is given in Viola (Viola 1995).

In comparison to parametric estimation, non-parametric methods are computationally expensive. Taking the Parzen window method as an example, the computation of Equation (B.1) is proportional to the size of samples and complexity of the window function R.

B.3 Experiment and Results

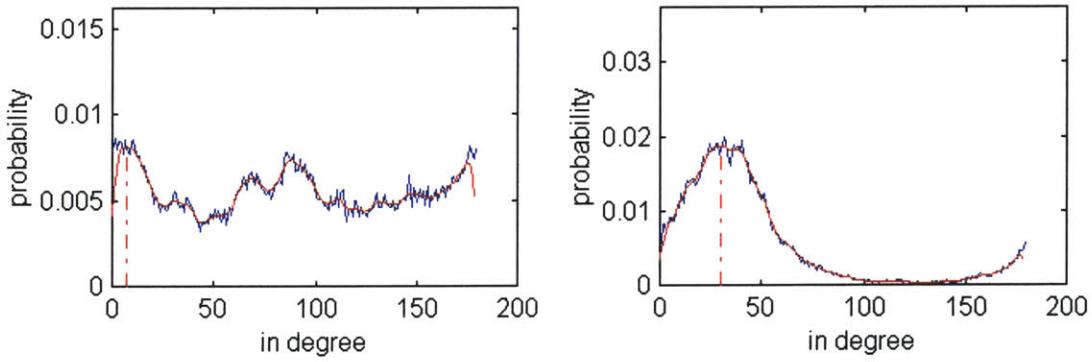
To demonstrate the proposed approach, an example is examined and discussed as follows. Figure B.1 exhibits the images of the cells before and after subjected to fluid shear stress. Figure B.2 shows the corresponding distribution of directions and its estimated probability function (in red) using the Parzen window estimation method. In situation A, the distribution of directions is quite flat, which is true because free cells are oriented randomly. In situation B, cells are shown to align along a dominant direction. Figure B.3 depicts the distribution of cell lengths along the major and minor axes and

their estimated probability function (in red). The eccentricity is defined as the ratio of the length along major axis over the length along minor axis. From situation A to B, the ratio changes from 1.04 to 1.86, which indicates the morphological changes of cell elongation as subjected to fluid shear stress.



(a) (b)

Figure B.1 The cells exhibit the morphological changes under different conditions: (a) situation A without fluid shear stress and (b) situation B with extern fluid shear stress



(a) (b)

Figure B.2 Probability distribution function of cell directions: (a) no dominant direction in situation A; (b) dominant direction is 30.0° in situation B. (red line is estimated probability function.)

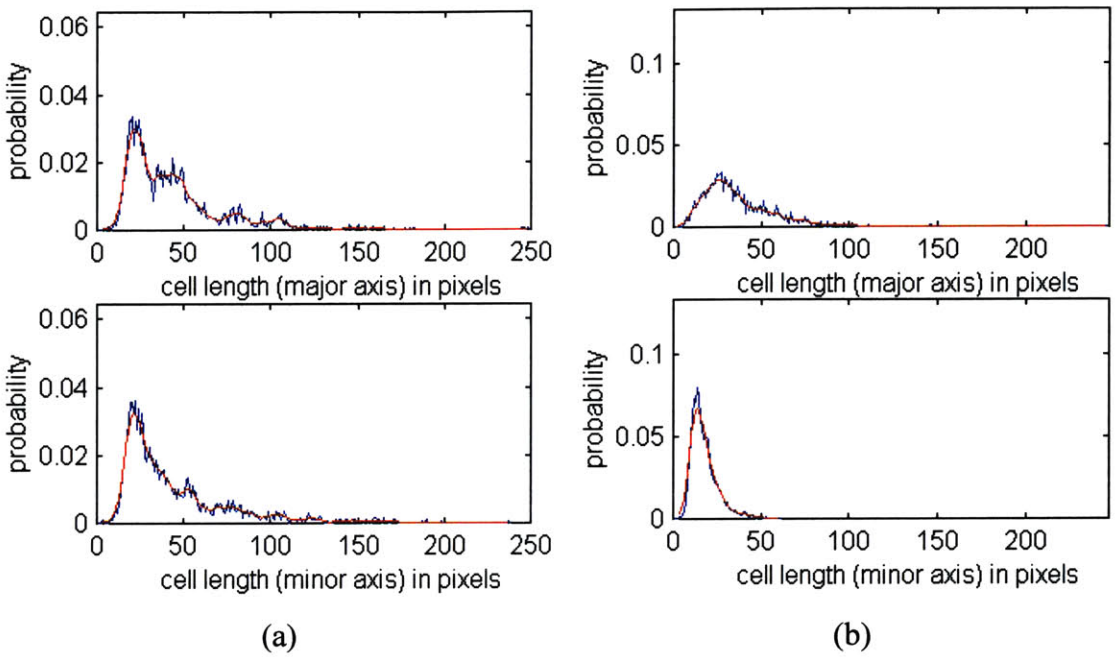


Figure B.3 Probability distribution functions of cell size measured in length of major and minor axes: (a) ratio estimation is about 1.04 in situation A; (b) ratio estimation is about 1.86 in situation B. (red line is estimated probability function.)

B.4 Conclusions

The appendix described a statistical method to offer quantitatively measurements of cell morphological changes. Two important characteristics, alignment direction and elongation, are measured. In comparison to previous methods (Remuzzi, Dewey et al. 1984), our proposed method gives more reliable results and computation is significantly faster.

The success of our statistical method relies on the correctness of statistical assumptions. The measurements obtained only reflect the behavior of an overall population, not individual cells. It assumes that characteristic information builds up over the population and dominate statistically.

References

[1] Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney and D. Sorensen (1999). *LAPACK Users' Guide*. Philadelphia, PA, Society for Industrial and Applied Mathematics.

[2] Barnard, S. T. and M. A. Fischler (1982). "Computational Stereo." *Computing Surveys* **14**(4): 553-572.

[3] Barron, J., D. J. Fleet and S. S. Beauchemin (1994). "Performance of optical flow techniques." *International Journal of Computer Vision* **12**(1): 43-77.

[4] Besag, J. E. (1974). "Spatial interaction and the statistical analysis of lattice systems." *Journal Royal Statistical Society, Series B* **36**: 192-236.

[5] Besag, J. E. (1986). "On the statistical analysis of dirty pictures." *Journal Royal Statistical Society, Series B* **48**: 259-302.

[6] Burt, P. and E. Adelson (1983). "The Laplacian pyramid as a compact image code." *IEEE Transactions on Communication* **31**(4): 532-540.

[7] Chan, M. T., G. T. Herman and E. Levitan (1997). "A Bayesian Approach to PET Reconstruction Using Image-Modeling Gibbs Priors: Implementation and Comparison." *IEEE Transactions on Medical Imaging* **44**(3): 1347-1354.

[8] Chan, M. T., G. T. Herman and E. Levitan (1999). "Probabilistic Modeling of Discrete Images." *Discrete Tomography: Foundation, Algorithms, and Applications*. Ed. by G. T. Herman and A. Kuba. Boston, Birkhauser. **Chapter 9**.

[9] Chen, C. T., V. E. Johnson, W. H. Wong, X. P. Hu and C. E. Mertz (1990). "Bayesian Image Reconstruction in Positron Emission Tomography." *IEEE Transactions on Nuclear Science* **37**(2): 636-641.

[10] Cheng, Y., C. Hartmink, J. Hartwig and C. F. Dewey (2000). "Three-Dimensional Reconstruction of the Actin Cytoskeleton from Stereo Images." *Journal of Biomechanics* **33**(1): 105-113.

[11] Deans, S. R. (1993). *The Radon Transform and Some of Its Applications*. Malabar, Florida, Krieger Publishing Company.

[12] Dempster, A. P., N. M. Laird and D. B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society. Series B* **39**: 1-38.

[13] Depaola, N., M. Gimbrone, Jr., P. Davies and C. F. Dewey, Jr. (1992). "Vascular Endothelium Responds to Fluid Shear Stress Gradients." *Arterio Thromb* **12**: 1254-1257.

[14] DeRosier, D. J. and A. Klug (1968). "Reconstruction of three-dimensional structures from electron micrographs." *Nature* **217**: 130-134.

[15] Dewey, C. F., Jr., S. R. Bussolari, M. A. Gimbrone, Jr. and P. F. Davies (1981). "The Dynamic Response of Vascular Endothelial Cells to Fluid Shear Stress." *Journal of Biomechanical Engineering* **103**: 177-185.

[16] Dhond, U. R. and J. K. Aggarwal (1989). "Structure from Stereo: A Review." *IEEE Transactions on Systems, Man, and Cybernetics* **19**(6): 1489-1510.

[17] Duda, R. and P. Hart (1973). *Pattern Classification and Scene Analysis*, John Wiley and Sons.

[18] Elfadel, I. M. (1993). *From Random Field to Networks*. Ph.D. Thesis, Department of ME, Also as RLE Technical Report, No. 579, Cambridge, MIT.

[19] Faugeras, O. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA, MIT Press.

[20] Fazekas, A., G. T. Herman and A. Matej (1998). "On processing binary pictures via their projections." *International Journal of Imaging Systems and Technology* **9**(2-3): 99-100.

[21] Fleet, D. J., A. D. Jepson and M. Jenkin (1991). "Phase-Based Disparity Measurement." *CVGIP: Image Understanding* **53**(2): 198-210.

[22] Frank, J., Ed. (1992). *Electron Tomography*. New York, Plenum Press.

[23] Frank, J. (1995). "Approaches to large-scale structures." *Current Opinion in Structural Biology* **5**: 194-201.

[24] Frank, J. (1996). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. New York, Academic Press.

[25] Frese, T., C. A. Bouman and K. Sauer (1998). "Discrete Multiscale Bayesian Image Reconstruction." *33rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA.

[26] Fua, P. (1993). "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features." *Machine Vision and Applications* **6**(1): 35-49.

[27] Fua, P. (1997). "From Multiple Stereo Views to Multiple 3-D Surfaces." *International Journal of Computer Vision* **24**(1): 19-35.

[28] Gardner, R. J. and P. Gritzmann (1997). "Discrete Tomography: Determination of Finite Sets by X-rays." *Transactions of the Mathematical Society* **349**(6): 2271-2295.

[29] Geiger, D. and F. Girosi (1991). "Parallel and Deterministic Algorithms from MRF's: Surface Reconstruction." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(5): 401-412.

[30] Geman, S. and D. Geman (1984). "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6): 721-741.

[31] Gordon, R., R. Bender and G. T. Herman (1970). "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and x-ray photography." *Journal of Theoretic Biology* **29**: 471-481.

[32] Granlund, G. H. (1978). "In Search of A General Picture Processing Operator." *Computer Graphics and Image Processing* **8**(2): 155-178.

[33] Granlund, G. H. and H. Knutsson (1995). *Signal Processing for Computer Vision*, Kluwer Academic Publisher.

[34] Green, P. J. (1990). "Bayesian reconstructions from emission tomography data using a modified EM algorithm." *IEEE Transactions on Medical Imaging* **9**(1): 84-93.

[35] Grimson, W. E. (1980). *A Computer Implementation of A Theory of Human Stereo Vision*. AI Memo, 565, Cambridge, MIT.

[36] Grimson, W. E. (1985). "Computational Experiments with a Feature-Based Stereo Algorithm." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7**(1): 17-34.

[37] Hannah, M. J. (1989). "A System for Digital Stereo Image Matching." *Photogrammetric Engineering and Remote Sensing* **55**(12): 1765-1770.

[38] Hartwig, J. (1992). "An ultrastructural approach to understanding the cytoskeleton." *The cytoskeleton: a practical approach*. Ed. by K. Carraway and C. Carraway. Oxford, UK, Oxford University Press: 23.

[39] Hartwig, J. and P. Shevlin (1986). "The Architecture of Actin Filaments and the Ultrastructural Location of Actin-binding Protein in the Periphery of Lung Macrophages." *The Journal of Cell Biology* **103**(3): 1007-1020.

[40] Heel, J. (1991). *Direct estimation of structure and motion from multiple frames*. AI Memo, 1190, Cambridge, MIT.

[41] Heel, J. (1991). *Temporal Surface Reconstruction*. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Cambridge, MIT.

[42] Herman, G. T. (1980). *Image reconstruction from projections: the fundamentals of computerized tomography*. San Francisco, Academic Press.

[43] Herman, G. T. and A. Kuba, Eds. (1999). *Discrete Tomography: Foundations, Algorithms and Applications*. Boston, Birkhauser.

[44] Herman, G. T. and D. Odhner (1991). "Performance evaluation of an iterative image reconstruction algorithm for positron emission tomography." *IEEE Transactions on Medical Imaging* **10**(3): 336-246.

[45] Heuser, J. and L. Evans (1980). "Three-dimensional visualization of coated vesicle formation in fibroblasts." *Journal of Cell Biology* (84): 560-583.

[46] Horaud, R. and T. Skordas (1989). "Stereo Correspondence Through Feature Grouping and Maximal Cliques." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(11): 1168-1180.

[47] Horn, B. K. P. (1986). *Robot Vision*. Cambridge, MIT Press.

[48] Horn, B. K. P. and M. J. Brooks (1989). *Shape from shading*. Cambridge, MA, The MIT Press.

[49] Horn, B. K. P. and E. Weldon (1988). "Direct methods for recovering motion." *International Journal of Computer Vision* **2**: 51-76.

[50] Iu, S. L. and K. Wohn (March, 1989). *Estimation of 3-d motion and structure based on a temporally-oriented approach with the method of regression.* Proceedings Workshop on Visual Motion, Irvine, CA.

[51] Jain, A. K. (1989). *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ, Prentice Hall.

[52] Kanada, T. and M. Okutomi (1994). "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(9): 920-932.

[53] Kapur, T. (1999). *Model Based Three Dimensional Medical Image Segmentation*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Cambridge, MIT.

[54] Kapur, T., W. E. L. Grimson, R. Kikinis and W. M. Wells (1998). *Enhanced spatial priors for segmentation of Magnetic Resonance Imagery.* Medical Image Computing and Computer-Assisted Intervention (MICCAI'98), Cambridge, MA.

[55] Kawata, S. and O. Nalcioglu (1985). "Constrained Iterative Reconstruction by the Conjugate Gradient Method." *IEEE Transactions on Medical Imaging* **MI-4**(2): 65-71.

[56] Kim, N. H., S. J. Aggarwal, A. C. Bovik and K. R. Diller (1990). "3-D Model of Vascular Network in Rat Skin obtained by Stereo Vision Techniques." *Journal of Microscopy* **158**(*Pt* 2): 275-284.

[57] Knutsson, H. and C. F. Westin (1993). *Normalized and Differential Convolution: Methods for Interpolation and Filtering of Incomplete and Uncertain Data.* IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York City, IEEE.

[58] Kuba, A. (1984). "The reconstruction of two-directionally connected binary patterns from their two orthogonal projections." *Computer Vision, Graphics, Image Processing* **27**: 249-265.

[59] Lange, K., M. Bahn and R. Little (1987). "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography." *IEEE Transactions on Medical Imaging* **MI-6**(2): 106-114.

[60] Lange, K. and R. Carson (1984). "EM reconstruction algorithms for emission and transmission tomography." *Journal of Computer Assisted Tomography* **8**(2): 306-316.

[61] LAPACK, *http://netlib2.cs.utk.edu/lapack/*.

[62] Lawrence, M. C. (1992). "Least-Squares Method of Alignment Using Markers." *Electron Tomography*. Ed. by J. Frank. New York, Plenum Press.

[63] Levitan, E. and G. T. Herman (1987). "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography." *IEEE Transactions on Medical Imaging* **MI-6**(9): 185-192.

[64] Li, S. Z. (1995). *Markov Random Field Modeling in Computer Vision*, Springer-Verlag.

[65] Luther, P. K., M. C. Lawrence and R. A. Crowther (1988). "A Method for Monitoring the Collapse of Plastic Sections as A Function of Electron Dose." *Ultramicroscopy* **24**(1): 7-18.

[66] Marapane, S. B. and M. Trivedi (1994). "Multi-Primitive Hierarchical (MPH) Stereo Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(3): 227-240.

[67] Marr, D. and T. Poggio (1979). "A Theory of Human Stereo Vision." *Proc. Roy. Soc. London* **B204**: 301-328.

[68] Matthies, L., R. Szeliski and R. Kanade (1989). "Kalman filter-based algorithms for estimating depth from image sequences." *International Journal of Computer Vision* **3**: 209-238.

[69] McDonnell, M. J. (1981). "Box-filtering techniques." *Computer Graphics and Image Processing* **17**: 65-70.

[70] McIntosh, J. H. and K. M. Mutch (1988). "Matching Straight Lines." *Computer Vision, Graphics, and Image Processing* **43**(3): 386-408.

[71] Mersereau, R. M. (1973). "Recovering Multidimensional Signals from their Projections." *Computer Graphics and Image Processing* **1**(4): 179-195.

[72] Mumcuoglu, E., R. Leahy, S. Cherry and Z. Zhou (1994). "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images." *IEEE Transactions on Medical Imaging* **13**(4): 687-701.

[73] Niederman, R., P. C. Amrein and J. Hartwig (1983). "Three-dimensional Structure of Actin Filaments and of an Actin Gel Made with Actin-binding Protein." *The Journal of Cell Biology* **96**(5): 1400-1413.

[74] Ohta, Y. and T. Kanada (1985). "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7**(2): 139-154.

[75] Okutomi, M. and T. Kanada (1993). "A Multiple-Baseline Stereo." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(4): 353-363.

[76] Ollinger, J. M. (1994). "Maximum-Likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm." *IEEE Transactions on Medical Imaging* **13**(1): 89-101.

[77] Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill.

[78] Penczek, P., M. Marko, K. Buttle and J. Frank (1995). "Double-tilt electron tomography." *Ultramicroscopy* **60**: 393-410.

[79] Ramirez, J. M., S. Mitra and J. Morales (1999). "Visualization of the three-dimensional topography of the optic nerve head through a passive stereo vision model." *Journal of electronic imaging* **8**(1): 92-97.

[80] Remuzzi, A., C. F. Dewey, Jr., P. Davies and M. A. Gimbrone, Jr. (1984). "Orientation of Endothelial Cells in Shear Fields in Vitro." *Biorheology* **21**: 617-630.

[81] Rowland, S. W. (1979). Computer Implementation of Image Reconstruction Formulas. *Image Reconstruction from Porjections: Implementation and Applications*.Ed. by G. T. Herman. Berlin, Springer-Verlag.

[82] Ryser, H. J. (1957). "Combinatorial Properties of Matrices of Zeros and Ones." *Canad. J. Math.* **9**: 371-377.

[83] Salzberg, P. M. (1995). "Binary Tomography on Lattices." *Congressus Numeratium* **111**: 185-192.

[84] Sanger, T. D. (1988). "Stereo Disparity Computation Using Gabor Filters." *Biological Cybernetics* **59**(6): 405-418.

[85] Shashua, A. (1995). "Algebraic Functions For Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **17**(8): 779--789.

[86] Shepp, L. A. and Y. Vardi (1982). "Maximum Likelihood Reconstruction for Emission Tomography." *IEEE Transactions on Medical Imaging* **MI-1**(2): 113-122.

[87] Stein, G. P. (1998). *Geometric and photometric constraints: motion and structure from three views*. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Cambridge, MA, MIT.

[88] Stein, G. P. and Shashua (2000). "Model-based Brightness Constraints: on Direct Estimation of Structure and Motion." *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **22**(9): 992-1015.

[89] Stein, G. P. and A. Shashua (1996). *Direct methods for estimation of structure and motion from three views*. AI Memo, 1594, Cambridge, MA, MIT.

[90] Sun, C. (1997). "A Fast Stereo Matching Method." *Digital Image Computing: Techniques and Applications*, Auckland, New Zealand.

[91] Suzuki, S. and S. Yamaguchi (1988). "Comparison between an image reconstruction method of filtering backprojection and the filtered backprojection method." *Applied Optics* **27**(14): 2867-2870.

[92] Toft, P. (1996). *The Radon Transform: Theory and Implementation*. Ph.D. Thesis, Deptartment of Mathematical Modelling, Technical University of Denmark.

[93] Vardi, Y. and D. Lee (1998). "The discrete Radon transform and its approximate inversion via the EM algorithm." *International Journal of Imaging Systems and Technology* **9**(2-3): 155-173.

[94] Viola, P. (1995). *Alignment by Maximization of Mutual Information*. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Cambridge, MA, MIT.

[95] Vosselman, G. (1992). *Relational Matching*, Springer-Verlag Press.

[96] Wells, W. M., W. E. L. Grimson, R. Kikinis and F. A. Jolesz (1996). "Adaptive segmentation of MRI data." *IEEE Transactions on Medical Imaging* **15**(4): 429-442.

[97] Wells, W. M. (1993), *Statistical Object Recognition*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Cambridge, MA, MIT.

[98] Westelius, C. J. (1995). *Focus of Attention and Gaze Control for Robot Vision.* Ph.D. Thesis, Department of Electrical Engineering, Linköping University, Sweden.

[99] Yagle, A. E. (1998). "An Algebraic Solution to the 3D Discrete Tomography Problem." *IEEE International Conference on Image Processing* **2**: 714-717.