Three Dimensional Reconstruction of the Cell Cytoskeleton from Stereo Images

by

Yuan Cheng

B.S., Precision Instruments and Manufacturing (1994)

Tsinghua University

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Mechanical Engineering

at the

Massachusetts Institute of Technology

June 1998

Signature of Author _____
Department of Mechanical Engineering
May 8, 1998

Certified by _____
C. Forbes Dewey, Jr.
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by _____
Chairman, Department Committee on Graduate Students

Three Dimensional Reconstruction of the Cell Cytoskeleton from Stereo Images

by

Yuan Cheng

Submitted to the Department of Mechanical Engineering
on May 8, 1998 in partial fulfillment of the
requirements for the Degree of
Master of Science in Mechanical Engineering

ABSTRACT

Besides its primary application to robot vision, stereo vision also appears promising in the biomedical field. This study examines 3D reconstruction of the cell cytoskeleton. This application of stereo vision to electron micrographs extracts information about the interior structure of cells at the nanometer scale level. We propose two different types of stereo vision approaches: the line-segment and wavelet multiresolution methods. The former is primitive-based and the latter is a point-based approach.

Structural information is stressed in both methods. Directional representation is employed to provide an ideal description for filament-type structures. In the line-segment method, line-segments are first extracted from directional representation and then matching is conducted between two line-segment sets of stereo images. A new search algorithm, matrix matching, is proposed to determine the matching globally. In the wavelet multiresolution method, a pyramidal architecture is presented. Bottom-up analysis is first performed to form two pyramids, containing wavelet decompositions and directional representations. Subsequently, top-down matching is carried out. Matching at a high level provides guidance and constraints to the matching at a lower level.

Our reconstructed results reveal 3D structure and the relationships of filaments which are otherwise hard to see in the original stereo images. The method is sufficiently robust and accurate to allow the automated analysis of cell structural characteristics from electron microscopy pairs. The method may also have application to a general class of stereo images.

Thesis Supervisor: C. Forbes Dewey

Title: Professor of Mechanical Engineering

Acknowledgements

It is not I alone who have brought this work to this point. There are a number of people who made their contributions in one way or another.

First of all, I would like to thank Professor Forbes Dewey, my advisor, for his inspiration, understanding, trust, and constant support. I sincerely admire him for his enthusiasm, knowledge, and great ideas. Working with him has been a privilege.

I am grateful to Professor Gösta Granlund and Dr. Johan Wiklund of Linköping University for their supplying me with kernel codes and many insightful comments. Their help laid a foundation for this work.

I thank Donna Wilker for her warm heart, her help, and sharing her opinions about Boston, Massachusetts, America, and the world.

Many thanks to Jim McGrath for his picking me up in my first day in America, and his reading my thesis in his last days in MIT. Also, I would like to express my gratitude to my colleagues in the lab for their encouragement, their help, and their friendship. Specially, many thanks to Christopher Hartemink for providing stereo images.

Finally, I would like to thank my wife, Yan Feng, for her caring, her love, and her standing by me throughout these years. Though I may not say it often, I appreciate all the support you have given me both materially and emotionally. In addition, I would like to thank my parents and my brother for their love, their support, and their pride, which have enabled me to become what I am today.

Contents

Chapter 1 Introduction

Stereo vision is usually associated with robot vision and scene reconstruction. The application to the biomedical field is often ignored or underrated. The main purpose of this thesis is to explore the application of stereo vision in the biomedical field in the context of electron microscope stereo pairs.

The thesis topic was introduced during the collaboration with Brigham and Women's Hospital of Harvard Medical School. Related studies have examined the mechanisms of interaction of actin filaments, primary elements of the cell cytoskeleton, with one another. The three-dimensional relationships of actin filaments can be observed in the stereo pair electron micrographs (see Figure 1.1). Three-dimensional information provides evidence and a solid basis to interpret and model the actin-filament networks [Niederman *et al* 1983]. Our goal is to reconstruct the 3D structure from stereo images and provide quantitative measurements to support our modeling and interpretations.

In general, 3D information is always desirable in any field. Stereo images play an important role in obtaining 3D information. Tiny objects, e.g. the objects at 10^{-8}-10^{-6}m scale, can be seen through Electron Microscopes (EM). Neither MRI nor other slicing technology is applicable to reconstruct 3D structure. There is no better way than stereo images to obtain 3D nature of objects at this scale. (Of course, triple images would be better. But essentially, they are the same as stereo images.) This makes our project more important, particularly if we can have computers recreate 3D objects from stereo images automatically and make this procedure routine.

This thesis proposes two approaches to stereo vision. One is the line-segment method and the other is the wavelet multiresolution method. Our experimentation is mainly carried out on EM images of the cell cytoskeleton.

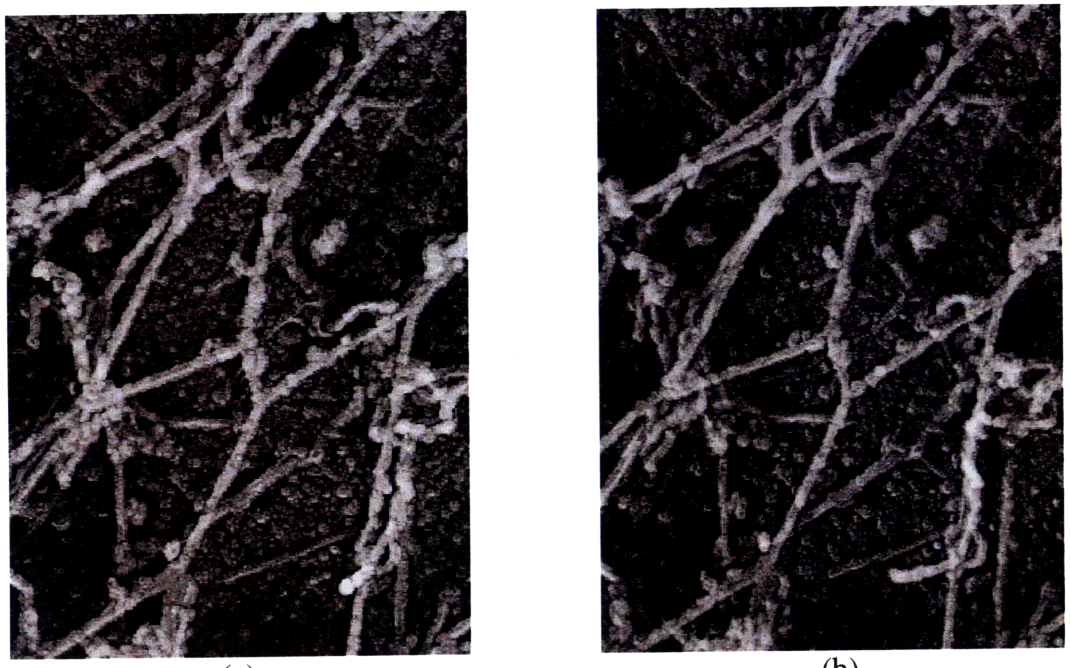
In the rest of this chapter, section 1.1 provides a little more background about endothelial cytoskeleton. Section 1.2 outlines the organization of the thesis.

1.1 Endothelial Cytoskeleton

A monolayer of endothelial cells forms the interface between flowing blood and the vessel walls of the cardiovascular system. This monolayer functions as a barrier to the deposition of lipids in the vascular walls. It has been known that a change in endothelial function induced by the hemodynamic environment is the initial event in forming plagues [Gimbrone 1990], which lead to atherosclerosis and other diseases. This monolayer is subject to dynamic loads in *vivo*, among which the wall shear stress induced by blood flow is the most important one. Dramatic rearrangements in the internal cellular actin network have also been observed after exposure to fluid shear stress both in *vivo* and in *vitro* [DePaola *et al* 1992][Gotlieb *et al* 1991].

Figure 1.1 shows a pair of stereo images of endothelial cell cytoskeleton. (these images will serve as the main example images for experimentation.) The structures in these images have significant directional features. The term "directional feature" refers to both lines and edges. (In image processing field, the line and the edge are different concepts.) A "directional feature" means a filament-type structure.

Measurements of the cytoskeleton's structure are important to the understanding of its role in biology. One and two-dimensional measurements have been proposed and well developed [McGrath 1994]. However, some three-dimensional information is very critical, such as conjunctions, angles of interaction, and the volume of structure, etc. 2D measurements of these types of information are often incorrect or insufficient.



(a) (b)

Figure 1.1 Stereo images of cell cytoskeleton: (a) left image; (b) right image.

1.2 About the Thesis

The thesis intends to emphasize two aspects: stereo vision and its application to cell cytoskeleton reconstruction. Chapter 2 gives a review of stereo vision and summarizes the procedures of stereo vision. Chapter 3 describes stereo geometry. Some important constraints are discussed in detail. Stereo geometry also implies some problems we have to confront, such as the ambiguity issue.

Chapter 4 deals with directional representation theory and its implementation. Directional representation describes structural information which is important for both approaches. An interesting application of directional representation to 2D measurement is proposed as well.

In chapters 5 and 6, the thesis studies stereo vision from two approaches: the line-segment and wavelet multiresolution methods. Both theory and implementation will be presented.

In the line-segment stereo vision method, the line segments are first extracted from the images according to directional information. Then the line segments are approximated by straight lines. Finally, straight lines are matched between the images. A matrix matching algorithm is introduced for global optimal matching. Matching gives 3D structural information directly.

In the wavelet multiresolution method, the images are decomposed into several scales by wavelet transform. Directional representation is also incorporated to provide structural information. The incorporation of structural information improves the performance of wavelet method. In order to be consistent with the multiresolution idea, we subsample the images and obtain the directional information at the different scales. The matching is conducted on a point-by-point basis. A "Coarse-to-fine" analysis strategy is employed. The matching at a high level is used to guide the matching at lower level. Interpolation is used to enhance the final results.

Chapter 7 draws some conclusions, summarizes the thesis, and presents future work.

In the Appendix, some animated demos are shown. The demos are Java applets and available on the ICMIT[1] webpage (see Appendix A).

[1]ICMIT: stands for *International Consortium for Medical Imaging Technology* which is co-directed by Professor C.F.Dewey.

Chapter 2 Stereo Vision

In brief, stereo vision involves using a computer to recover the 3D objects from stereo pictures, which are taken at different angles or positions. For human beings, it is not difficult to fuse the stereo images together and get the sense of three-dimensional objects. (Most of us must have experiences of viewing random-dot stereograms.) But for computers, it is not an easy task.

Stereo vision is the core problem of computer vision, which enables the robots or computers to interact with their environment. In our study, we are exploring a new application of stereo vision to electron microscope images of the cell cytoskeleton.

2.1 Procedures of Stereo Vision

The most successful vision systems are the biological ones. Scientists have gained ideas from the study of human vision and other biological vision systems. In 1977, Marr and Poggio of MIT [Marr and Poggio 1977] proposed a computational framework motivated from human vision. An implementation of their theory was proposed and some experiments were tested by Eric Grimson [Grimson 1981][Grimson 1985]. Their studies paved a new way to model and develop computer vision.

Marr's theories also implied the idea of multiresolution, which is believed to be the possible model of human perception. Zero-crossings were extracted from the different scales and evaluated in pyramids. Granlund extended the multiresolution idea. Two types of hierarchies were proposed [Granlund 1978] [Granlund 1990]: scale hierarchies and abstraction hierarchies. With abstraction hierarchies, images can be viewed as the expansion of image primitives, such as lines, curvatures, objects, and relations, from low level to high level. They are the conceptual or semantic analyses of the images.

From the standpoint of computer implementation, stereo vision consists of three key steps [Marr 1982][Grimson 1993]:

1. Identify a particular point in one image (for example, the left).
2. Find the point in the other image (for example, the right) that is a projection of the same scene point as observed in the first image.
3. Measure the disparity (or difference in projection) between the left and right image points. With the prior knowledge of sensing systems, such as the focal length and camera positions, the actual position of scene point in 3D space can be determined from the disparity.

This scheme of stereo vision algorithm has been commonly accepted. The algorithm is often called the "feature-based" algorithm not only because the images are represented by features but also because the matching is based on the features.

Mathematically, these three steps can be summarized by some more general terms: feature representation, stereo matching, and reconstruction from disparity, respectively. These will be discussed in more detail below.

2.2 Feature Representation

Feature representation involves using features to describe the objects or scenes. In computer vision, objects or scenes are usually represented by one or more features, such as contours, edges, zero-crossings, skeleton lines, regions, etc. Therefore, one of the goals of computer vision is to identify or extract features from the images.

We classify features into two categories: structural features and non-structural features. Structural features, such as edges, lines, curves, regions, etc., are related to some structural information of the object. Non-structural features are similar to zero-crossings, Fourier phases, wavelet coefficients, etc. These features do not contain much structural

information. Intuitively, structural features are more desirable for describing the image. However, in practice, people use non-structural features frequently because non-structural features usually have solid mathematical support, and are easy to analyze and implement. In this thesis, we try to combine both types of features at the same time.

In the following, we review some of the typical features often used in stereo vision.

Skeleton lines are also favorable in the area of pattern recognition. For example, in letter recognition, the first step is commonly to obtain the skeleton line by a thinning algorithm [Zhang and Suen 1984][Pitas 1993]. The biggest difficulty is at cross sections, such as "X" or "Y" intersections. We encounter the same problem in our line-segment method as well.

Edges are of frequent use in stereo vision. Edges are defined as a discontinuity of the image intensity function. They can be obtained by thresholding the derivatives of the intensity function. Many studies in the literature discuss edge detection, and some classical operators are also available, such as Sobel and Prewitt operators. However, the problem is that the edge detector is very sensitive to noise. Preprocessing, such as low-pass filtering, is usually applied before edge detection [Marr and Hildreth 1980]. However, low-pass filtering blurs the edges as well. Some studies have explored the tradeoff between accurate detection and robustness to noise [Canny 1986].

Zero-crossing was introduced as one type of feature describing the objects or scenes in the Marr-Poggio algorithm. The theories developed later by groups at AI Lab of MIT. The algorithm applies a Gaussian operator to the original image first, followed by a Laplacian operator (second-order derivative). The Gaussian operator smoothes the image. The zeros of the Laplacian-Gaussian operator correspond to maximums or minimums of the first order derivatives of the intensity function. The zero-crossings are related to the edges but they are not identical. The operation can be applied at various sizes by varying the spread of the Gaussian function. This is the multiresolution idea. This idea fits the concept of wavelet so well that this operator was later named "Marr's wavelet."

All the features discussed so far are extracted from the spatial domain. Some researchers argued that their accuracy is limited by detectors that are usually spaced one pixel apart. The pixel width therefore determines the disparity resolution of final matching.

"Fourier phase" has been introduced as a type of feature to represent the objects in the images [Sanger 1988][Fleet 1991]. Gabor filters have been a natural choice because they have both limited spatial width and finite bandwidth. There are advantages to use phase information rather than spatial information. One of them is that the phase feature method has sub-pixel accuracy. However, singular points have been the critical obstacle for the phase-based method in the implementation. Westelius proposed a method of reducing the influence of singular points [Westelius 1995].

In the light of Gabor filters, some alternatives were proposed for various reasons, such as quadrature filters [Knutsson 1982] and non-ringing filters. Quadrature filter is complex in the spatial domain, consisting of an even real part and an odd imaginary part. Essentially, the even part is a line filter and the odd part is an edge filter.

Based on this idea, Granlund proposed the concept of general operator descriptor (GOP) [Granlund 1978][Granlund and Knutsson 1995]. The *directional features* are of interest. Successful applications can be found in various areas, such as image restoration by adaptive filtering, texture recognition, and motion estimation, etc. [Knutsson *et al* 1983]. Directional features are the most significant features in the cytoskeleton structure.

Wavelet coefficients are used as features in our wavelet multiresolution approach. Wavelet is an idea tool for multiresolution representation. It provides a compact way to look at signals at different scales or resolutions. Wavelet coefficients tell us feature information of different scales. Surprisingly, there are only a few papers studied stereo vision by explicitly using the wavelet approach [Pan 1996][Kim, Lee, *et al* 1997]. Our studies in this thesis show the success of the wavelet approach with stereo vision. Details will be discussed in Chapter 6.

How and why do we choose one type of feature over another? There are actually certain characteristics that a good feature representation should have. These characteristics form the criteria of selection:

- **Translation (or rotation) invariant:** Usually stereo images are projected from two different angles or two different positions. Stereo matching is based on the feature information of two images. Ideally, feature representation should be the same in two images, except that the positions of features may change.
- **Compatibility and complementarity:** [Granlund 1990] Feature representation should be able to reflect any changes of the objects or scenes. The feature values should be conceptually close if the objects are similar or close, and *vice versa*.
- **Stability over scales:** In multiresolution methods, the stability of feature representation is often required. In coarse-to-fine analysis, the information at a coarse level is commonly used to guide stereo matching or the extraction of information at a finer level. The consistency of features is expected.

2.3 Stereo Matching

Stereo matching is the process of selecting a point or object in one image and then finding its correspondence in the other image. Correspondence means that it is the projection of the same point or object onto the second image.

Stereo matching lies at the heart of stereo vision. It is a very difficult task because the positions of objects might change in two images. Moreover, some information might be lost due to object occlusions, noises or shadows. The major problem in stereo matching is to reduce ambiguities and uncertainties. Ambiguity or uncertainty means that one point in one image may have more than one possible matching point in the other image or even no matching at all. Also, the process is very computationally expensive. In order to reduce the ambiguities and cut down on the computation time, various constraints are imposed

on stereo matching algorithms. Most constraints arise from the imaging geometry, and are discussed below.

Epipolar constraint: the epipolar constraint is one of the most important constraints in stereo matching. It is a geometric property of an imaging system. Briefly speaking, for any point in one image, its possible matching points must lie along a line in the other image. This line is called the epipolar line. Therefore, the searching space of its correspondence is reduced from 2D to 1D. Furthermore, the epipolar line is assumed to be horizontal for simplicity in most cases. The epipolar constraint will be examined in detail in Chapter 3.

Ordering constraint: this constraint requires that, except in the transparent case, right-to-left order should be preserved along an epipolar line in two images. This assumption does not always hold, especially when there are many isolated objects and these objects are not on the same surface. However, the ordering constraint is very reasonable because the human visual system may make this same assumption.

(Figural) continuity constraint: this constraint is based on the assumption that the features describing the surface edges or surface markings should be continuous in some sense. So should their correspondences. Continuity often means that there is only small change in a local area.

Uniqueness constraint: it means that each feature in one image can match only one feature in the other image, and *vice versa*. Sometimes it is very useful to reduce the ambiguity. However, it is not true for some types of feature representation, such as line segments.

Disparity gradient constraint: this constraint requires that the disparity should change very smoothly in the local area. It is widely used in surface reconstruction, which assumes there is no abrupt change in a real surface.

Some or all of the above constraints can be applied at the same time to help to reduce the ambiguities in the stereo matching algorithm. However, ambiguity and uncertainty still exist. Then some questions arise: "what is the best matching?" and "how to determine it?"

Similarity is the fundamental principle in the stereo matching algorithm. The basic assumption is that corresponding features will remain similar in two images. When we match two features, say one in the left image and the other in the right image, if they are evaluated to be very similar under certain constraints, we could say these two features are a possible matching pair.

A match evaluation function is usually defined as comparing the candidates for matching. The similarity function is often used [McIntosh and Mutch 1988][Hellwich and Faig 1994]. Briefly, the similarity function measures the similarity of two features in stereo images. The best matching is the one with largest similarity among all the candidates. Alternatively, a cost function was defined, in [Ohta and Kanade 1985] and [Kim, Lee, *et al* 1997], to determine the matching. The best matching is defined as the one that minimizes the cost function.

In the term of "how to determine the matching," various algorithms have been proposed to optimize the matching process, such as the relaxation labeling technique [Marr and Poggio 1977][Pollard *et al* 1985], graph approaches [Horaud and Skordas 1989], the tree-searching technique, dynamic programming [Ohta and Kanade 1985][Lee and Leou 1994], etc.

Corresponding to the multiresolution in feature representation, the matching can also be carried out in the concept of multiresolution. Matching at a higher level can be used to guide and speed up the matching at a lower level. This is usually referred to as "coarse-to-fine analysis." Grimson [Grimson 1985], Marapane [Marapane and Trivedi 1994], Westelius [Westelius 1995] are such examples.

Another point I would like to mention is that stereo matching is "directed." So called "directed" means that the matching could be performed either from left to right or from right to left. The results from two directions might be different. However, most algorithms assume only the left-to-right matching.

2.4 Reconstruction from Disparity

Disparity is defined as the position difference between two projections in stereo images. Disparities can be obtained directly from the stereo matching results.

In most cases, the disparities are a sparse distribution, especially in the feature-based matching. In order to recover 3D surfaces or objects, an interpolation technique is also employed to generate more disparities. The basic assumption behind interpolation is that the surface is smooth and the local disparity gradient changes smoothly [Faugeras 1993][Wildes 1989]. In our wavelet multiresolution method, linear interpolation is applied to enhance the results. Chapter 6 discusses this.

In order to reconstruct 3D from disparities, we have to have the prior knowledge about the system from which the stereo images were taken. Information, such as focal length, camera positions, rotation axis, etc., must be given. There is a unique relationship between the disparity and the depth of the object. It is discussed in the next Chapter.

Chapter 3 Geometry of Stereo Vision

In this chapter, we deal with some basic but important questions of stereo vision: how will we be able to reconstruct 3D objects from stereo images? What is the correspondence? Is it uniquely determined? The geometry of stereo vision will give us the answers.

3.1. Geometry

Figure 3.1 shows the geometry of stereo vision, which is the general case of stereo vision. Line P1P2 is the object in 3D space (XYZ coordinates). Line p1p2 and p1'p2' are the projections in stereo images, say left and right images. Points c and c' are two focal points of a camera when two images are taken.

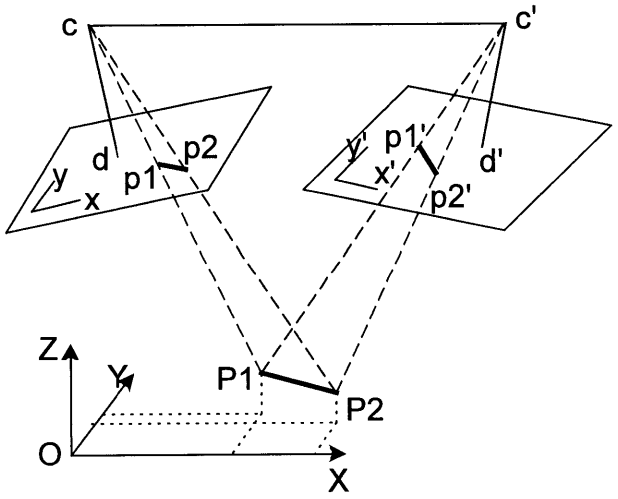


Figure 3.1 Geometry of stereo vision

To find the correspondence pair, such as p1p2 and p1'p2', in left and right images, is called "stereo matching." If we know the positions of p1p2 and p1'p2' in xy and x'y'

coordinates of left and right images, respectively, we will be able to calculate the position of P1P2 on XYZ coordinates. This is called "reconstruction." The system parameters, such as the focal points c and c', the focal length cd or c'd', and the positions of the two image coordinates with respect to the XYZ coordinates, are needed for this reconstruction. These design parameters are referred to as "prior knowledge of the sensing system."

In our system, the object is observed through an electron microscope. The system can be simply illustrated in Figure 3.2. The situation is equivalent to thinking that the object plane is tilted at two angles of $\pm\theta$. The rotation axis is the Y-axis. (The Y-axis is perpendicular to the plane of the figure). Since the projection in our system is parallel, the y coordinates are the same in stereo images. This property makes the computation less complicated.

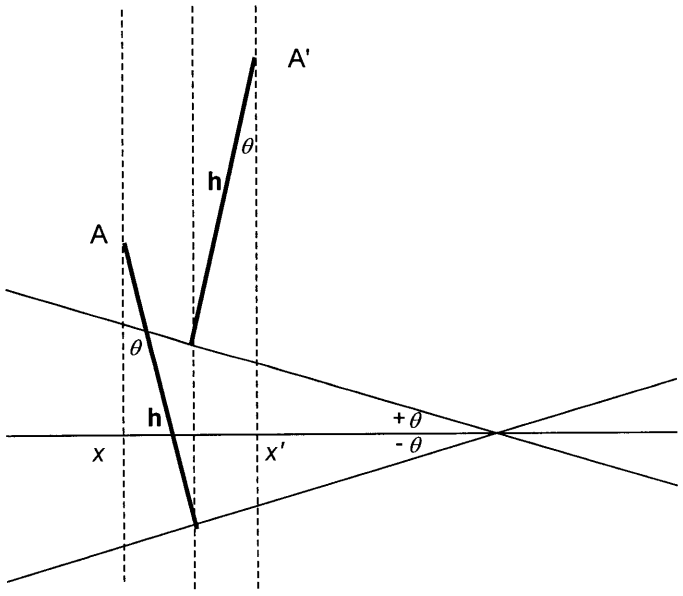


Figure 3.2 Simple geometry of our system

Suppose that the height of point A is h, and its two projections are (x, y) and (x', y'), respectively. The height h can be obtained from the difference of the X-coordinates. The relationship is simply given as follows [Niederman *et al* 1983]:

$$h_{eight} = \frac{(x - x')}{2\sin\theta} \tag{3.1}$$

The height h is often called the depth as well. The depth or height gives the position of the point A in 3D space. The difference between the X-coordinates of the two image planes, $(x\text{-}x')$, is called the disparity. One of the goals of stereo vision is to find the disparities between the points or objects.

3.2. Geometric Constraints

The geometry of stereo vision not only gives the relationship between projections and positions of objects in 3D space but also implies many constraints, such as the epipolar constraint, the ordering constraint, the continuity constraint, etc. These constraints are very important to reduce the ambiguity and computation of image matching.

The epipolar constraint is a very powerful one. We would like to discuss it in detail.

The projection of a point is a point in image plane. For example, in Figure 3.3, point m1 is the projection of point M onto the left image plane. In fact, all the points of the line MC1 have the same projection, the point m1, in the left image plane. (C1 and C2 are the focal points.) However, the projections of all the points on line MC1 onto the right image plane lie on line L_2. This line is called an epipolar line. As a consequence, all the possible matches of point m1 in the right image plane lie on the epipolar line L_2. And *vice versa*, i.e., all the possible matches of m2 in the left image plane are on the epipolar line L_1. This is the epipolar constraint. Therefore, the epipolar constraint helps us to reduce the dimension of search space from two dimensions to one.

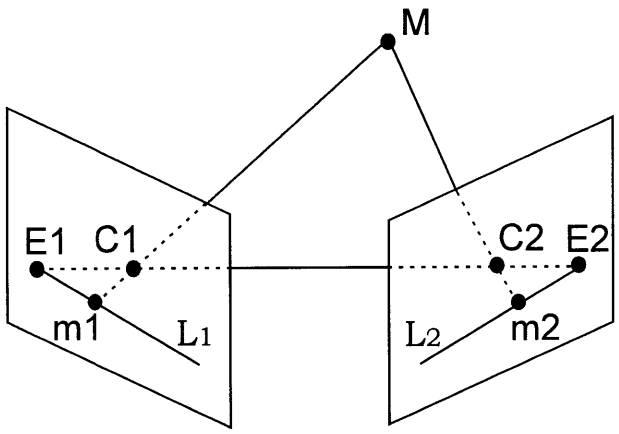


Figure 3.3 The epipolar geometry

It is obvious that the epipolar lines pass through the epipoles, which are the intersection of the line C1C2 with the image planes, such as E1 and E2 in Figure 3.3. If the image planes are parallel to the line C1C2, the epipolar lines in the image planes become parallel. Because of its simplicity, this condition is often assumed in stereo vision. Fortunately, in our system, the epipolar lines are always parallel. We also choose Y-axis as the rotation axis (in Figure 3.2). The epipolar lines in both image planes are parallel to Y-axis. Therefore, we can have the same Y-coordinates for corresponding epipolar lines in stereo images.

With continuity constraint, it is assumed that there are no abrupt changes in features or disparities in the local region. This assumption provides the basis of interpolation, which is important and necessary for surface reconstruction. We also adopt this idea in our second approach in Chapter 6.

The ordering constraint and the uniqueness constraint are often imposed to restrict the matching and reduce the ambiguity as well. However, because of these constraints, some points or features in one image will not find their correspondences in the other image. Consequently, the matching pairs we get are only a subset of matching pairs there should

be in reality. This also results in sparse disparities. Some post-processing, such as interpolation, is often necessary to obtain a full set of matching pairs. The goal of this post-processing is to generate the disparities, based on the local continuity, for those points without correspondences.

Chapter 4 Directional Representation

As discussed in Chapter 2, features can be classified as structural features and non-structural features. Structural features are easily related to the structures that we are interested in, such as edges, lines or regions. However, the non-structural features usually have solid mathematical support, and therefore, analysis and implementation are relatively easier.

Directional representation is a good combination of these two types of features. It describes directional feature, which is the important structural information in our case. Also, directional representation has a well-supported formula for implementation.

Both line-type and edge-type structures are directional features. Because directional representation gives the same results regardless of whether a structure is a line or an edge, it is good for situations in which only structural information is of interest. In our case, the cell cytoskeleton image is characterized by filament-type structures with some width. We allow no difference in feature representation between the edges and the middle of the filaments.

Directional representation is the basis of our line-segment method, in which the line-segments are extracted based on the continuity of directional representation. Directional representation is also incorporated in the wavelet multiresolution method to provide structural information.

Directional representation stems from vector representation or "double angle" representation proposed by Granlund [Granlund 1978][Granlund and Knutsson 1995]. In section 4.1, we present some details of this theory. Section 4.2 gives some interpretations of directional representation and an example of how it works in practice. Section 4.3

examines a simple application of directional representation which greatly simplifies an otherwise computational intensive problem.

4.1 Theory

This section is based on Granlund [Granlund and Knutsson 1995] and Nordberg [Nordberg 1994]. More comprehensive details and proper descriptions can be found in these references.

Section 4.1.1 deals with why the quadrature filter was introduced and how it works in one dimension. Section 4.1.2 focuses on the discussion of orientation estimation by using quadrature filters in two dimensions.

4.1.1 One dimension

As we know, structural detection must give the same results (magnitude values) for either lines, or edges, or their combinations. This implies that the algorithm should contain both line filters and edge filters.

It is also important that both the line filter and the edge filter must be insensitive to the DC component of the image because the flat surfaces are of no interest in the sense of structural detection. It is reasonable to consider the line filter as an even function and edge filter as an odd function (if the origin is set to the center). Figure 4.1 shows a one-dimensional example of the line and edge features. The intensity profiles in Figure (c) and (d) show the symmetric (even) and anti-symmetric (odd) characteristics of the line and edge functions, respectively.

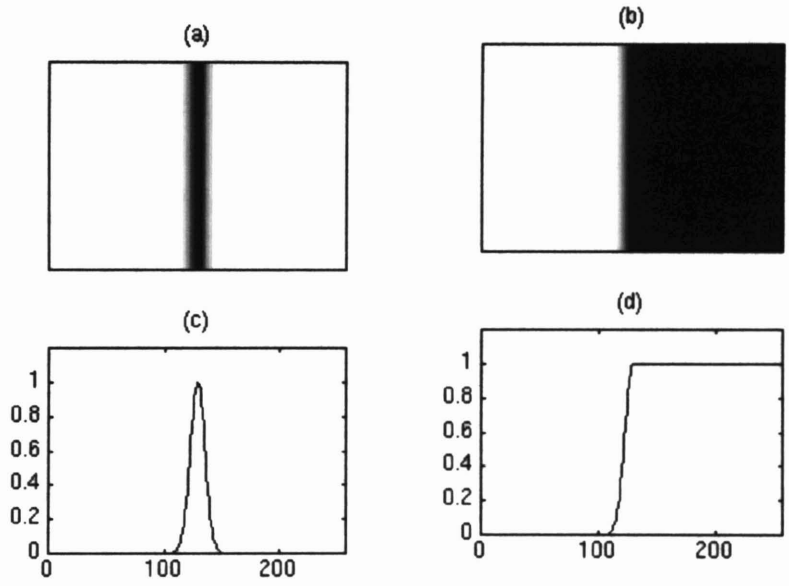


Figure 4.1: (a) 1D example of line; (b) 1D example of edge; (c) intensity profile of a line in(a); (d) intensity profile of an edge in(b).

Assume that a line, $f_{line}(x)$, and an edge, $f_{edge}(x)$, have the exactly the same magnitude function in the Fourier domain, i.e.,

$$||F_{edge}(u)|| = ||F_{line}(u)|| . \tag{4.1}$$

It is required for the line and edge filters to give the identical outputs. Let us assume the line filter, $h_{line}(x)$, and edge filter, $hedge(x)$. In the Fourier domain, we should have relationship:

$$H_{edge}(u)F_{edge}(u) = H_{line}(u)F_{line}(u). \tag{4.2}$$

Substitution of equation (4.1) into (4.2) gives

$$||H_{edge}(u)|| = ||H_{line}(u)|| . \tag{4.3}$$

As discussed earlier, the line filter is even and the edge filter is odd. If both line and edge filters are real, the line filter then has an even and real Fourier transform and the edge filter has an odd and imaginary Fourier transform. Therefore, an edge filter can be related to a line filter by the Hilbert transform,

$$H_{edge}(u) = \begin{cases} -iH_{line}(u), & \text{if } u < 0 \\ iH_{line}(u), & \text{if } u \geq 0. \end{cases} \tag{4.4}$$

If we generate a complex filter with the line filter as its real part and the edge filter as its imaginary part,

$$h(x) = h_{line}(x) - ih_{egde}(x). \tag{4.5}$$

H(u) in the Fourier domain is real and has only the positive part. The negative part is zero due to cancellation of $H_{line}(u)$ and $H_{edge}(u)$.

This type of filter *h(x)* is called a quadrature filter. The output magnitude from a quadrature filter depends only on the signal energy, and not on whether the signal is even (a line), odd (an edge) or a mixture. That is, it will pick up whatever the structure is. The filter phase, on the other hand, indicates the relation between evenness and oddness of the signal relative to the filter center. The phase is not affected by the signal energy. Westelius [Westelius 1995] proposed a stereo vision method based on this phase information.

In general, a quadrature filter is defined as having no support in the left half-plane of the frequency domain, and no DC component. In the spatial (or time) domain, the quadrature filter is a complex function. The imaginary part can be obtained from the real part by the Hilbert transform as in equation (4.4). There are a number of quadrature filters discussed and compared by Westilus [Westelius 1994]. This thesis uses lognorm filters.

Lognorm filters were proposed to estimate the orientation, local phase and local frequency [Knutsson 1989]. The definition is in the frequency domain:

$$F(u) = \begin{cases} e^{-\frac{4}{\log(2)\beta}\log^2(\frac{u}{u_0})} & \text{if } u > 0 \\ 0 & \text{otherwise} \end{cases} \qquad (4.6)$$

The parameters are the center frequency, u_0, and the relative bandwidth, β. It is easy to verify that the lognorm filter has no DC component and no support in the left half plane of frequency domain. In the spatial domain, the real part gives the line filter and the imaginary part gives the edge filter.

4.1.2 Two dimensions

It is not trivial to extend the discussion from one dimension to two dimensions. For example, the concept of the left or right side of a plane makes sense only when a certain direction is considered as the reference.

The quadrature filter in 2D is defined as having a Fourier transform, G, that vanishes on a half-plane in the Fourier domain and is non-zero on the other half-plane (real as well). The half planes are defined by a vector $\mathbf{m}$ that is perpendicular to the border between the half planes and points to the half-plane on which G is non-zero. This quadrature filter is called a directional quadrature filter and its direction is represented by $\mathbf{m}$.

For easy design, a quadrature filter is assumed to be polar separable in Fourier domain,

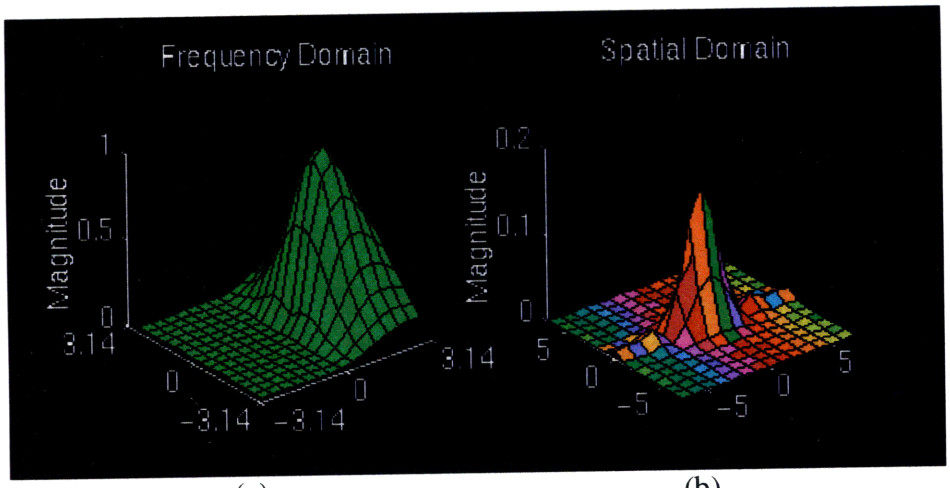
$$G(\mathbf{u}) = G_\rho(|\,\mathbf{u}\,|) \cdot G_{\hat{\mathbf{u}}}(\hat{\mathbf{u}}) \qquad (4.7)$$

where $\mathbf{u} = |\,\mathbf{u}\,|\,\hat{\mathbf{u}}$.

The directional function is given by

$$G_{\hat{\mathbf{u}}}(\hat{\mathbf{u}}) = \begin{cases} (\hat{\mathbf{u}}^*\hat{\mathbf{m}})^2 & \text{if } \hat{\mathbf{u}}^*\hat{\mathbf{m}} > 0, \\ 0 & \text{if } \hat{\mathbf{u}}^*\hat{\mathbf{m}} > 0. \end{cases} \tag{4.8}$$

The radial function $G\rho$ can be chosen arbitrarily. Knutsson [Knutsson 1982] used lognorm functions as the radial part of quadrature filters. The advantages of the lognorm function are similar to those of the 1D lognorm filter. It has no support in one half plane and is non-zero in the other half plane. Figure 4.2 shows an example of lognorm quadrature filter with zero direction (in the spatial domain). In Figure 4.2(b), the color represents the phase of quadrature filter because quadrature filter is complex in the spatial domain. The green corresponds to real value and the red corresponds to imaginary value.



(a) (b)

Figure 4.2 An example of quadrature filter along zero direction (center frequency μ_0: π/2; bandwidth β: 2).

To make the discussion simple, let us consider the situation that the neighborhood contains only one linear structure with a well-defined orientation. The neighborhood is assumed to be constant on parallel lines of a certain direction and have a variation

according to some one-variable function in the direction perpendicular to the line. Let f be a 2D simple function,

$$f(x,y) = h(\mathbf{x}^*\hat{\mathbf{n}}), \qquad \text{where } \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \tag{4.9}$$

and $\hat{\mathbf{n}}$ is a normalized vector which is perpendicular to the lines of constant value. The Fourier transform of f is

$$F(u,v) = 2\pi\, H(\mathbf{u}^*\hat{\mathbf{n}})\, \delta_{line}(\mathbf{u}), \tag{4.10}$$

where H is the one-dimensional Fourier transform of h, and δ_{line} is an impulse function that is zero everywhere except on a line through the origin parallel to $\hat{\mathbf{n}}$. According to the projection theory of the Fourier transform, F also lies on a line in the Fourier domain. The line is perpendicular to the lines of constant value in the spatial domain. Also the projection theory implies that the impulse line of F rotates as spatial function f rotates.

Therefore, given the function f and a quadrature filter g, the filter response s is given by

$$s = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)G(u,v)\,du\,dv = Ce^{i\alpha}(\hat{\mathbf{n}}^*\hat{\mathbf{m}})^2, \tag{4.11}$$

where magnitude C only depends on H and $G\rho$, and where α depends on the directions of function f and filter g. The magnitude of s is given by

$$|s| = C(\hat{\mathbf{n}}^*\hat{\mathbf{m}})^2. \tag{4.12}$$

To generate double angle representation, four quadrature filters of the type as in equation (4.8) are used. The filters, g_k, where k = 0,1,2,3, have the same radial function $G\rho$ but different directions. The directions are denoted by

$$\hat{\mathbf{m}}_k = \begin{pmatrix} \cos(\frac{k\pi}{4}) \\ \sin(\frac{k\pi}{4}) \end{pmatrix}, \quad k = 0,1,2,3. \tag{4.13}$$

Assume the direction of function f is θ,

$$\hat{\mathbf{n}} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \tag{4.14}$$

then the magnitudes of the filter responses are given by

$$|s_k| = C\cos^2(\theta - \frac{k\pi}{4}), \quad k = 0,1,2,3. \tag{4.15}$$

The double angle representation is then constituted by combining these four filter response magnitudes as follows,

$$\begin{aligned} z &= \sum_{k=0}^{3} e^{\frac{k\pi i}{2}} |s_k| = |s_0| + i|s_1| - |s_2| - i|s_3| \\ &= C\left(\cos^2\theta - \cos^2(\theta - \frac{\pi}{2}) + i[\cos^2(\theta - \frac{\pi}{4}) - \cos^2(\theta - \frac{3\pi}{4})] \right) \\ &= C(\cos 2\theta + i\sin 2\theta) = Ce^{2i\theta}. \end{aligned} \tag{4.16}$$

It results in a representation with its phase of 2θ, which is the double of the local direction of function f. Equation (4.16) is correct for any direction θ, i.e., we can have continuous representations of local orientations. The representation is unique as well.

To estimate local orientations in multiple dimensions, tensor representation is necessary. Tensor representation was developed by Granlund and Knutsson [Granlund and Knutsson 1995].

4.2 Interpretations of Directional Representation

The double angle representation describes the orientation of the local linear structure regardless of whether the structure is a line-type or an edge-type. (This property is called "phase invariant.") The local direction θ of the structure can be obtained from the double angle representation in equation (4.16). The magnitude of double angle representation, C,

in equation (4.16), is called the confidence value. It indicates the certainty of the local direction.

In this thesis, we name the representation with the local direction and its confidence value "directional representation." Like a vector, the directional representation contains two components: the local direction and the confidence value.

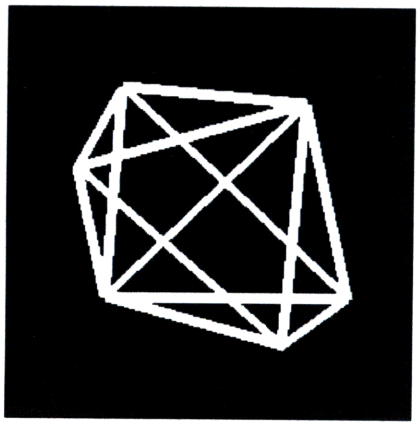
The quadrature filter we use is the lognorm filter. It is generated in the Fourier domain. As Figure 4.2 shows, the filter in the spatial domain is a complex function. Its real part is a line filter and its imaginary part is an edge filter. These filters are "kernels" of image convolution.

Directional representation is performed locally. The kernel size determines the size of the neighborhood. Generally, a larger neighborhood is better because more information is considered. But the larger neighborhood also means less detail information. Therefore, the decision of kernel size is dependent on what kind of information you are interested in. For example, if a small kernel size is chosen, a small object may become large with respect to its neighborhood, and *vice versa*. The typical kernel size we use is 11x11, or 15x15 (pixels), which is suitable for the structures in Figure 1.1.

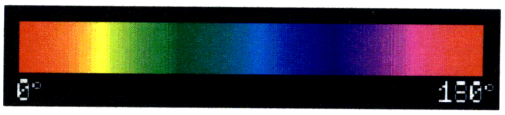
The other parameters that must be considered are the center frequency, u_0, and bandwidth, β, of lognorm filters. Both of these parameters are used to fine-tune the filter to pick up the structures of interest. For example, a large bandwidth is often used to detect sharp directional structures, such as very thin lines.

As an example, Figure 4.3 depicts the directional representation of a pattern image. In the representation images (in Figure 4.3(c) and (d)), the color corresponds to the local direction. The brightness of the color corresponds to the confidence value. The colormap is shown in Figure 4.3(b). Note that the colormap picks up the direction from 0° to 180°. Since the 0° and 180° lines are essentially the same, the corresponding colors should be the same as well. Therefore, the colormap is more like a circle, which well represents the

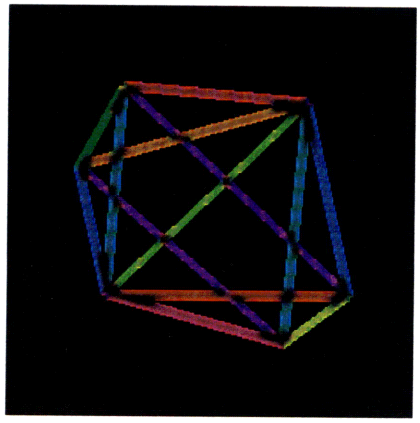
directions we want. (Appendix B gives more details of this colormap. Unless otherwise noted, the whole thesis uses this colormap.)



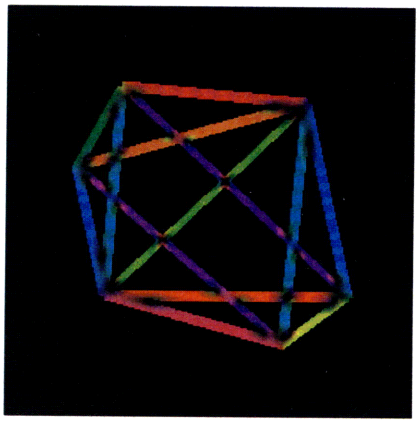
(a)



(b)



(c)



(d)

Figure 4.3 Directional representation of a pattern image: (a) original pattern image; (b) colormap indicating directions; (c) directional representation by 11x11 kernel with $u_0=\pi/4$ and $\beta=2$; (d) directional representation by 15x15 kernel with $u_0=\pi/8$ and $\beta=2$;

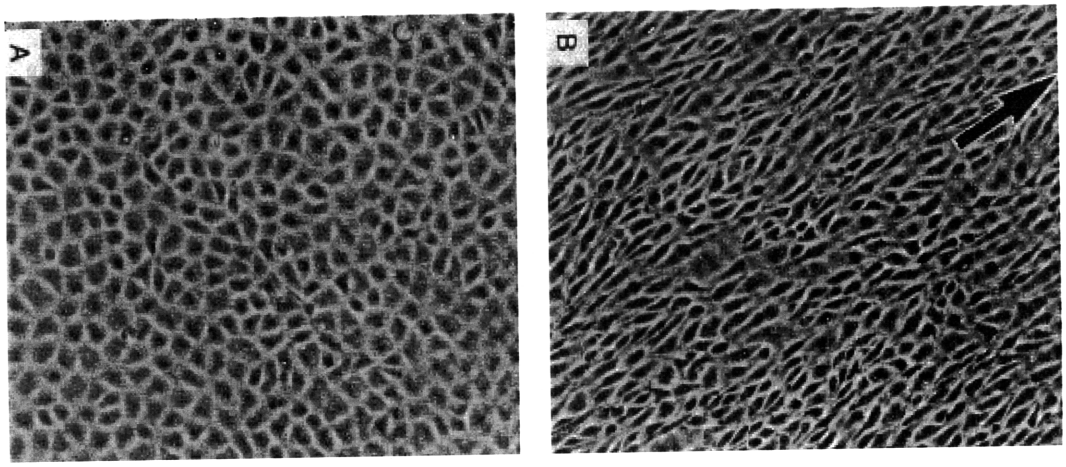
In this example, the filament width of the pattern image is about 7 pixels wide. The kernel sizes of Figure 4.3(c) and (d) are 11x11 and 15x15, respectively. We see that the confidence value of the representation changes in these two cases. This demonstrates that the locality or neighborhood depends on the kernel size.

Some other observations deserve mention. An important one is the continuity of the direction along the filament. This continuity appears as the same color in the representation image. It is a fundamental principle of our line-segment method. However, the discontinuity often takes place around the regions like corners or intersections. The directions at these regions are usually not unique because several lines meet. The directions we get from directional representation are often incorrect or meaningless, which leads to the discontinuity of line extraction. Fortunately, confidence values are usually very low at these regions. This observation is very helpful to track along the filament based on directional continuity. Chapter 5 discusses this.

4.3 A Direct Application

As introduced in Chapter 1, the endothelial cells undergo the structural as well as functional changes when subjected to the fluid shear stress. Eventually, the cells realign themselves along the shear direction, as Figure 4.4 shows. Figure 4.4(a) is the image of non-sheared cells. Figure 4.4(b) is the image of cells after they are subjected to the fluid shear [Davies *et al* 1986].

It is crucial to evaluate this process quantitatively. Scheneider [Schneider 1986] and Remuzzi [Remuzzi 1984] proposed an ellipse method to evaluate the changes of cell shapes quantitatively. In their method, each individual cell was represented by an ellipse that best fits the boundary of the cell. The direction of the major axis and the value of the eccentricity of the ellipse (ratio of major axis to minor axis) were considered to be the primary measure of the cell realignment.

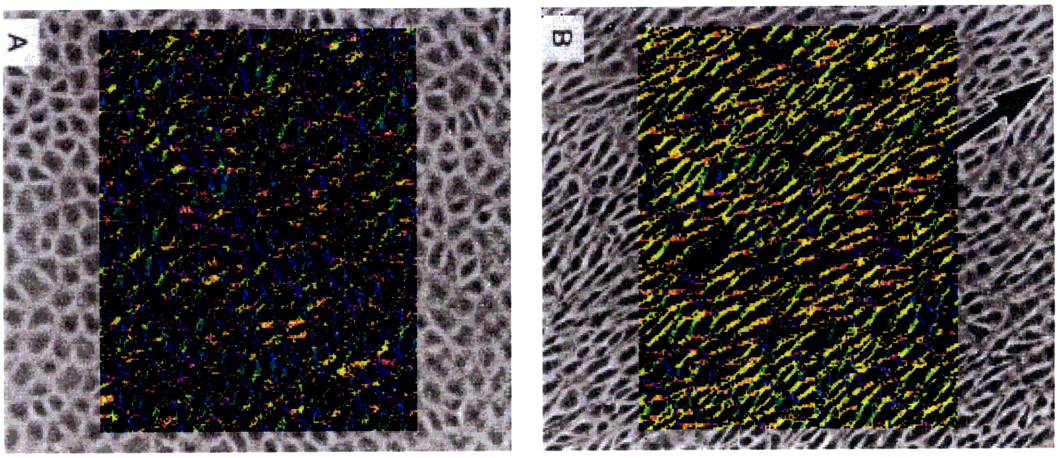


(a) (b)

Figure 4.4 (a) Endothelial cells before they are subjected to fluid shear stress; (b) endothelial cells after they are subjected to fluid shear stress.

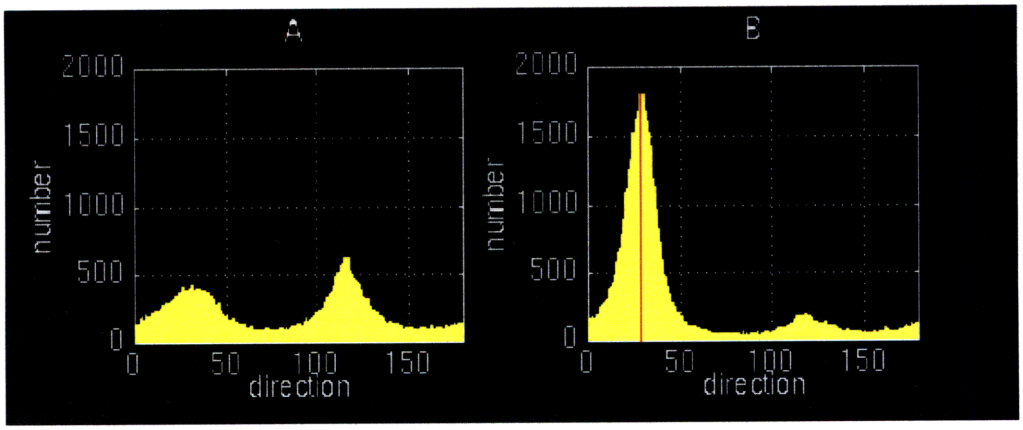
Directional representation provides an alternative way to evaluate this process statistically. The measurement giving the same indication can be carried out simply.

Figure 4.5 exhibits the directional representations of the images in Figure 4.4(a) and (b), respectively. Only a certain region is considered in order to eliminate the boundary effects. Again, the color in Figure 4.5 represents the local directions of that point. A threshold is predefined so that only the directions with the confidence value higher than a certain value are examined. Figure 4.6 plots the histograms of direction distributions in two cases of Figure 4.5. We can see the changes right away. In Figure 4.6(a), we could say the directions are almost uniformly distributed. There is no big difference between 500 and 200. But the difference is dramatic between 1700 and 200 as in Figure 4.6(b). This change indicates the cell realignment. We can even estimate the direction of the realignment by picking up the peak value. For example, in this case, it is about 30° with respect to the horizontal axis. (It is indicated in red.) This result is consistent with that of the ellipse method [Schneider 1986], but our method is more robust and computation is much less.



(a) (b)

Figure 4.5 (a) Directional representation of image in Figure 4.4(a); (b) directional representation of image in Figure 4.4(b).



(a) (b)

Figure 4.6 (a) Histogram of direction distribution in Figure 4.5 (a); (b) histogram of direction distribution in Figure 4.5(b).

Chapter 5 Line-Segment Method

The existing approaches to stereo vision can be grouped into two categories: the point-based method and the primitive-based method. The point-based method uses non-structural features, such as intensities, Fourier phases or wavelet coefficients, to describe the image or object. These features are extracted locally at each point and the matching is based on point-to-point matching. For example, the correlation stereo vision method [Gennery 1980][Fua 1991] was based on the intensity value and the variance in the neighborhood. Fleet [Fleet *et al* 1989] proposed a method based on the phase behavior of bandpass Gabor filters.

In the primitive-based method, structural features, such as edges, lines, curves, or regions, are often used. These features describe the image or object concisely. Because of their conciseness, the computation could be reduced in comparison with the point-based method. Structural information and the relationships between objects could be obtained directly from the matching results [Marapane and Trivedi 1996]. The difficult part of the primitive-based method is the feature representation.

Some arguments exist about which method is better. In fact, it depends on the specific applications. In this thesis, we study both methods. In this Chapter, we employ the primitive-based method with the line-segments as primitives. The wavelet multiresolution method in Chapter 6 is the point-based method.

In this chapter, section 5.1 gives an overview of the line-segment method. Section 5.2 deals with line-segment extraction from directional representation. Section 5.3 discusses some issues of matching and proposes a new matching idea — matrix matching. In section 5.4, a pattern example is presented to illustrate the procedure. The discussion and a brief summary of the line-segment method appear in section 5.5.

5.1 Architecture

It is very reasonable to represent filaments in cytoskeleton images as line-segments. Alternatively, it is also possible to use edges as features to describe the images. Edges are often used in scene reconstruction to simplify the objects. However, edges are not appropriate for our case because each thin filament-type structure would produce two-sided edges. In addition, edge detection cannot cross intersection regions so that this leads to the distortions in those regions. Edge detection is also very sensitive to noises.

Our line-segment method is based on directional representation. Line-segments are extracted according to the continuity of direction representation along each filament. After extraction, matching is carried out between two line-segment sets of stereo images. Finally, the three-dimensional structure is reconstructed from matching results. Figure 5.1 depicts the architecture of implementation.

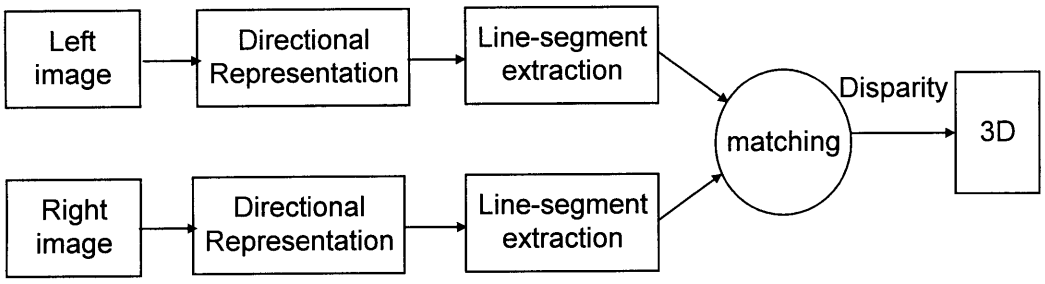


Figure 5.1 Architecture of line-segment method

5.2 Line-Segment Extraction

As discussed in Chapter 4, directional representation is performed locally and contains two components: the local direction and the confidence value. The confidence value indicates the certainty of the local direction. The continuity of the local direction along the filament is the basis of line-segment extraction.

In practice, it is not easy to pick up the whole filament directly. First of all, the filaments we are dealing with are not one pixel wide. Instead, the width in our case is in the range between 7 to 10 pixels wide. This makes our line extraction more like segmentation. Instead of a line, we actually pick up the region of a single filament. Second, the discontinuities of directional representation break filament into several pieces. In Figure 5.2, the discontinuity takes place around the corner or intersection regions, as Chapter 4 points out. "Line linking" at these regions is necessary.

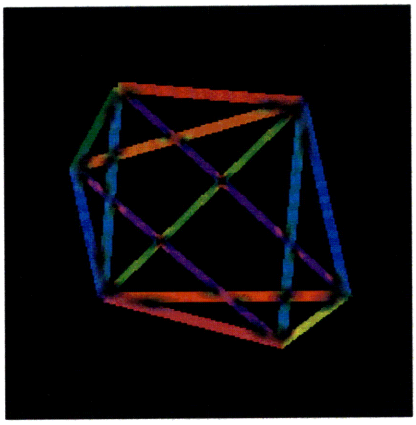


Figure 5.2 Directional representation of a pattern image of Figure 4.3

As in Figure 5.2, we want to extract a single filament with the same direction (the color) each time. After we obtain this filament, we can use a straight line to approximate this filament. The approximation is done by the least square method. The straight line-segment is then used for stereo matching.

5.2.1 Filament segmentation

Segmentation is based on some criteria which only an object of interest satisfies (or not satisfies). It identifies the object and distinguishes it from the rest of image. In our case, we want to pick up each filament each time. The unique feature of each filament is its

direction. The continuity of the directions follows along the filament, except the points near the corners or intersections.

The confidence values of directional representation would also be helpful. The confidence values near the corners or intersections are usually low even though the local direction might be incorrect or meaningless. This observation is incorporated into our algorithm. We assume that it is an intersection or corner point and is accepted for line tracking, if its confidence value is lower than a small threshold. This assumption helps the filament tracking to cross the intersection regions.

From the standpoint of programming, the criteria for filament segmentation are summarized as follows:

- The local direction of each point along the filament should be similar and the confidence value should be higher than a predefined threshold.
- If the confidence value is very low, (lower than another predefined threshold), we treat it as the intersection or corner point and assume that the continuity is satisfied.

A seed-growing technique is employed for segmentation. The search for the candidates is based on the breadth-first tree-searching algorithm [Vosselman 1992]. The following pseudocode describes the procedures of this filament segmentation algorithm:

1. Pick a seed with sufficiently high confidence value and set up criteria according to the directional representation of the seed.
2. Search in 3x3 neighborhood of the seed and evaluate each neighboring point according to the criteria. Label the point if it satisfies the criteria.
3. Search in 3x3 neighborhood of just-labeled points and evaluate each neighboring point. And again, label each point that satisfies the criteria.
4. Iterate the step (3).
5. Segmentation stops when no more points satisfy the criteria.
6. Repeat step(1)-(5) to segment another filament.

There are many practical issues we have to consider in the implementation. For example, it is very important to set up various thresholds to ensure a good filament extraction, although the choice of a proper threshold is often a trial-and-error process. Another important issue arises during breadth-first searching. It is difficult to avoid picking up two connected filaments at the same time, especially when they have very close directions. Therefore, more constraints have to be imposed on the evaluation of neighboring points.

5.2.2 Line linking

Confidence values help the filament tracking process to cross the intersections or identify corners. However, the filament may still be broken into several pieces. In cases where the discontinuity of directional representation is very severe, line linking is applied to establish filament continuity.

Line linking bridges a small gap and connects broken line segments together. The lines that can be linked must satisfy certain compatibility criteria. In practice, the compatibility criteria are implemented as follows:

- Only small gap is allowed.
- The two line segments, e.g. L_1 and L_2 in Figure 5.3, must have very close directions.
- The linking segment, e.g. L_{12} in Figure 5.3, also must have a direction very similar to the two line segments.
- The final integrated line must keep the original direction as closely as possible.

The last two conditions just insure that the two line segments are truly broken pieces of a single filament. After linking, the segmented filament is represented by a straight line. The approximation is performed by the least square method [Melter *et al* 1993].

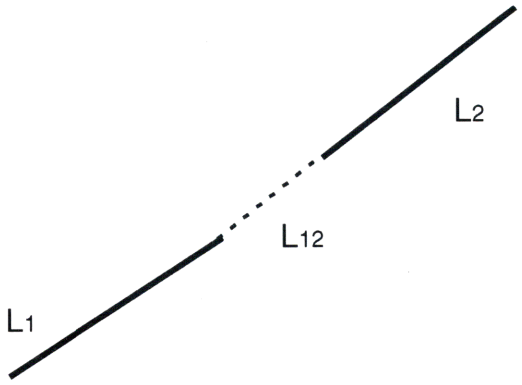


Figure 5.3 Line linking

Figure 5.4 shows the line-segments of a pattern image, extracted from the directional representation in Figure 5.2. The colors again correspond to directions. The highlighted line at the middle of each filament is the straight-line segment that represents the filament. Each straight-line segment is obtained by the least square method and characterized by its beginning, ending points and its direction.

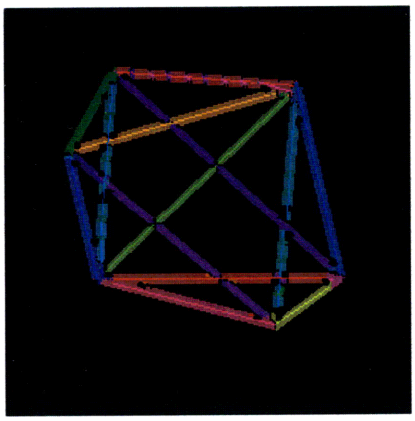


Figure 5.4 Line-segments extracted from Figure 5.2

At the early stage of our study, the thinning technology [Zhang and Suen 1984][Yu and Tsai 1990] was considered to skeletonize the structure first. Line-segments were then

extracted from skeletons (which are only one pixel wide). Nevertheless, it didn't work very well because the thinning algorithm distorted the intersections and corners. From our results in Figure 5.4, it shows that our line-extraction performs acceptably. There are almost no distortions at intersections and corners. The line extraction crosses the intersections very well in this synthetic and idealized test image.

5.3 Matching

Matching is the process of finding the correspondence in the other image. For example, if we have a line-segment in the left image, matching is to find a proper line-segment in the right image and then these two line-segments are considered a matching pair.

As discussed in Chapter 2, similarity is a fundamental principle in stereo matching. The basic assumption is that features are similar in the stereo-paired images. For example, there is a line-segment with direction θ in the left image. The corresponding line-segment in the right image should have its direction very close to θ. Similarity indicates how close two features are.

Therefore, the first task of matching is to compare two features of stereo images. In the implementation, each feature is often represented by its attributes and then the attributes are compared. For example, a line-segment could be represented by its attributes like length, direction, position, intensity, etc. The goal of evaluating attributes is to describe and quantify the feature for comparisons.

Ideally, the features are uniquely one-to-one matched in the left and right images. In practice, due to occlusion, noises, or other facts, the ambiguity always exists in image matching. Many constraints are imposed on image matching to reduce the ambiguity.

5.3.1 Similarity function

In our study, a similarity function is used to provide comparisons among all the candidates. The similarity of two features is defined as the weighted average of similarity values of all attributes:

$$S_{similarity} = \frac{\sum_{i} (w_i \bullet v_i)}{\sum_{i} w_i} \tag{5.1}$$

where v_i is the similarity value of *i-th* attribute, and w_i is the weight of *i-th* attribute. The similarity is normalized to be [0, 1]. The weight of an attribute is set according to the importance of the attribute. For example, the direction of a line-segment is more important than its intensity. Therefore, the weight of direction is larger than that of intensity.

Usually, the similarity value of *i-th* attribute is computed according to the nature of the specific attribute. In most cases, it can be defined as

$$v_i = \frac{\min(A_L(i),\ A_R(i))}{\max(A_L(i),\ A_R(i))} \tag{5.2}$$

where $A_L(i)$ is the *i-th* attribute value in left image and $A_R(i)$ is the *i-th* attribute value in right image. All the similarity values of the attribute are between 0 and 1.

In our study, line-segments are the features used for image matching. We consider five attributes of each line-segment. They are *line direction, line length, line intensity, line position* and *the overlap ratio of two line-segments*. Some researchers have argued that some relational information, such as parallelism or collinearity in neighborhood, should be considered as well. We think our five attributes are sufficient for our case.

The similarity values of the line length and intensity can be calculated by equation (5.2). But equation (5.2) is not suitable for others.

For the line direction, we only care about the difference of directions, not the absolute values. Therefore, the difference of directions is compared to the expected maximum difference. The maximum difference is a predefined threshold value. It indicates that the current candidate is not valid if the direction difference is greater than the maximum difference. The similarity value of the line direction [McIntosh and Mutch 1988] is given by:

$$V_{direction} = \frac{\text{MaxDirectionDifference} - |\,\text{Direction(L)} - \text{Direction(R)}\,|}{\text{MaxDifferenceDifference}} \quad . \quad (5.3)$$

The situation of the position of the line-segment is the same as that of the line direction. We only care about the position change of the line-segments, not the position itself. Similarly, the maximum disparity is predefined. In the implementation, we set the similarity value to 1 if the position change is less than the maximum disparity. Otherwise, it is set to 0. This means that we treat all possible disparities equally as long as they are below the maximum value. From this point, this attribute is more like a constraint condition.

The overlap ratio [Wu and Leou 1995] is defined as

$$V_{overlap} = \begin{cases} y_{overlap} / \min(\Delta y_L, \Delta y_R), & \text{if } y_{overlap} > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (5.4)$$

where $y_{overlap}$ is the overlap of y-coordinates, as shown in Figure 5.5. L_i and R_j in Figure 5.5 are two line-segments in left and right images, respectively. Δy_L,, Δy_R are the differences of y-coordinates of the ending points of the line-segments. Note that in our stereo system discussed in Chapter 3, the y-coordinates of two matching points are assumed to be the same. Ideally, the line-segments of a matching pair should have identical y-coordinates in two images, and the overlap ratio is then equal to 1. However, due to object occlusion and the approximation during line-segment extraction, the y-coordinates of two matched line-segments may be slightly different. The overlap ratio will be less than 1. Additionally, definition of equation (5.4) is invalid for horizontal or

near-horizontal line-segments. In those situations, we simply assume the overlap ratio is 1 in our algorithm.

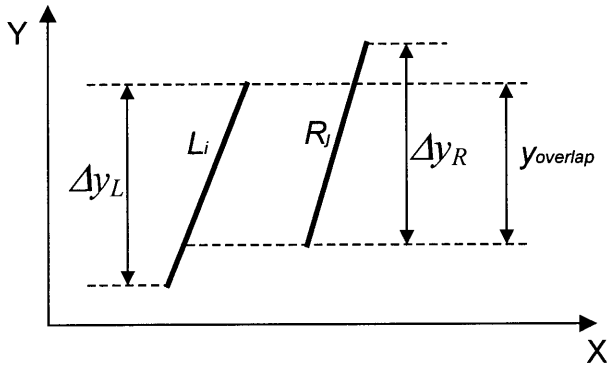


Figure 5.5 The overlap ratio between two line-segments

5.3.2 Matrix matching algorithm

Constraints are imposed on image matching to reduce the ambiguity. In the line-segment method, *epipolar constraint, disparity constraint* and *ordering constraint* are used to help to determine the matching. A horizontal epipolar line is assumed in both stereo images. (This is actually true for our stereo system.) That is, the matching pairs are assumed to appear at the same horizontal scanline in both images. This assumption reduces a lot of unnecessary computation. The disparity constraint means that the large disparity is not allowed. In our algorithm, we predefine a maximum disparity value. This constraint is consistent with the similarity value of the position attribute of the line-segment.

Despite these constraints, ambiguity is not totally eliminated. For a line-segment in the left image, for example, there might still be a certain number of line-segments in the right image, satisfying all the constraints. The similarities of these candidates can be calculated by equation (5.1). The question is how to choose the best one from all possibilities.

Another problem is the conflict. For example, each line-segment in the left image may have several candidates in the right image, and *vice versa*. Conflicts may exist between candidates. Therefore, the global consideration is necessary to optimize the whole matching set. It is not easy to determine the best matching globally. Some graph-based algorithms are often used for global matching because the graph method seems ideal for dealing with global relationships of elements [Horaud and Skordas 1989][Ohta and Kanade 1985].

We propose a new method, matrix matching algorithm, to tackle this problem. It provides a new way to look at the globe matching. The matching problem turns out to be very clear in this matrix form. The matrix is constituted as follows: each column j of the matrix corresponds to each line-segment of the right image; each row i of the matrix corresponds to each line-segment of the left image; the element, (i,j), of the matrix is the similarity value of the *i-th* line-segment in the left image and the *j-th* line-segment in the right image; the element is set to zero if there is no similarity value available for two corresponding line-segments. Figure 5.6 gives an example of such a matrix.

$$
\begin{array}{c}
 \\
(i)
\end{array}
\begin{array}{c}
(j) \\
\begin{bmatrix}
0.9740 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7542 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.7854 & 0 & 0 & 0 \\
0 & 0 & 0.6750 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8979 \\
0 & 0 & 0 & 0 & 0.8755 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.8798 & 0 & 0 & 0 & 0 & 0 & 0.5587 & 0 \\
0.8078 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7127 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.8767 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.6947 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.7488 & 0 & 0 & 0 & 0.9432 & 0 & 0 \\
0 & 0.6576 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9008 & 0 \\
0 & 0.9056 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

Figure 5.6 An example of matrix

From this matrix, we can see obvious conflicts. For example, in column (1),(8) and row (1),(6),(10) in Figure 5.6, if a uniqueness constraint is imposed, we can only have 0.9740

and 0.9432 chosen. However, if there is no uniqueness constraint, we could have 0.9740, 0.8078, 0.7488 and 0.9432. These four are the best ones in either their column or their row. Therefore, it is easy to understand that the best matching actually depends on how we define the best matching. The best matching set is only a subset of all elements in the matrix.

For some simple cases, global matching is very easy to implement using matrix form. For example, we have four simple choices to define the best global matching set: to select the largest similarities (1) based on each column; (2) based on each row; (3) in both column and row; (4) in either column or row. Actually, most literature employs one of these definitions. In our algorithm, we pick the third definition.

For complex cases, we may have to resort to linear algebra. Our matrix is usually sparse. All the elements in the matrix are positive and less than 1. These are the observations about the matrix. Linear algebra may tell us more, such as by matrix decomposition.

5.4 Reconstruction

Matching results provide us disparity information. The depth can be computed according to the relationship of equation (3.1). Since matching is based on line-segments, reconstruction gives the 3D positions of straight-line segments. In order to make the reconstruction look better, we assign the straight-line segments certain thickness. (Alternatively, the straight-line segment can also be assumed as the rotation axis of a cylinder.)

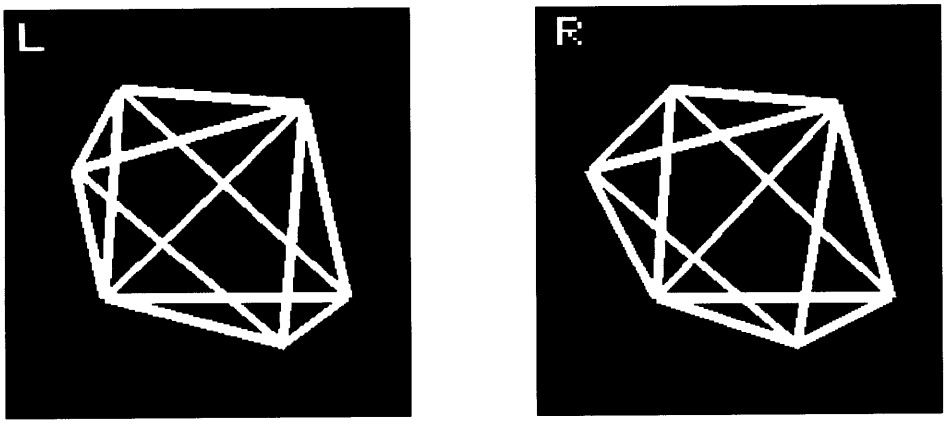
Figure 5.7 exhibits the whole procedure of the line-segment method on the pattern images. The procedure is consistent with the diagram in Figure 5.1.

Figure 5.7(a, b) are stereo pattern images. Since they are generated from a given set of 3D data, we will be able to compare our reconstruction result with the real 3D data.

Figure 5.7(c) and (d) are directional representations of the left and right pattern images. Figure 5.7(e) and (f) are the line-segments extracted, based on their directional representations. Again, in all these images, the colors correspond to the local directions. The matching is then conducted between these two sets of line-segments.

Figure 5.7(g) shows the reconstructed results. Note that the image is actually a 3D image. We relate the intensity to the depth. The lower the intensity, the larger the depth. From this 3D image, we can see relationships between the filaments, which are otherwise hard to see in the original stereo images. For example, some filaments appear to intersect with each other in 2D images. However, in reconstructed 3D image, they do not intersect at all.

Figure 5.7(h) depicts the real 3D image created directly from the 3D data given. In comparison with our reconstructed one, we can only find some minor difference.



(a) (b)

Figure 5.7 Line-segment method on stereo pattern images (a-b)

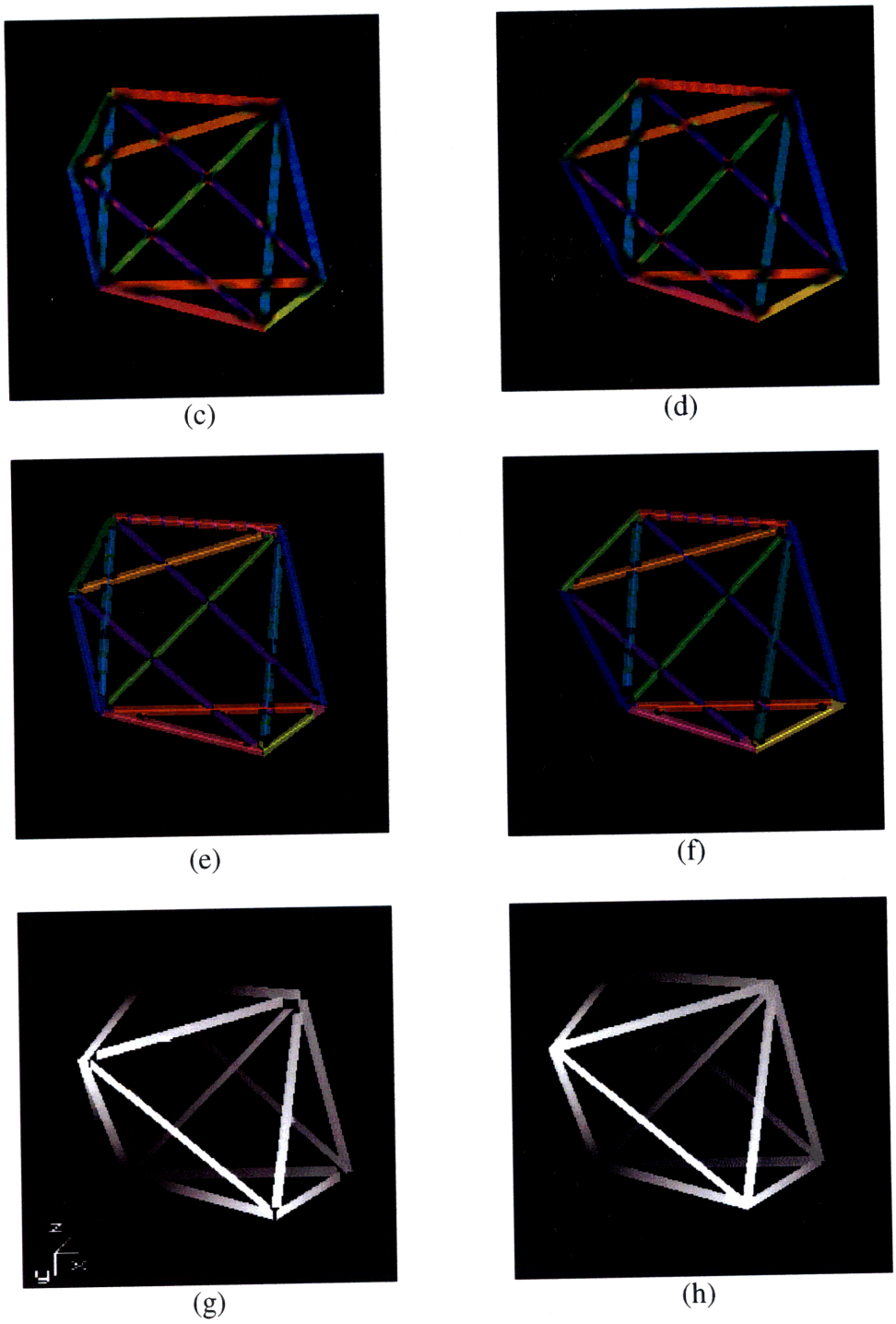


Figure 5.7 Line-segment method on stereo pattern images (e-h)

5.5 Discussions

The line-segment method gives 3D structural information directly, such as the 3D lines. This is one of advantages of the line-segment method.

Unfortunately, when we apply the line-segment method to the real case images such as the images in Figure 1.1, the results are poor. Figure 5.8 shows the reconstructed result of the stereo images in Figure 1.1. We cannot see the structure clearly. (Note that the intensity in the image of Figure 5.8 corresponds to the depth as well. Each filament is represented by a cylinder.)

The following factors could explain why.

1. Discontinuities in directional representation are much more severe than those in the pattern images.
2. Straight-line approximations distort filaments differently in the two images, resulting in incorrect reconstruction.
3. Background noise affects the results.

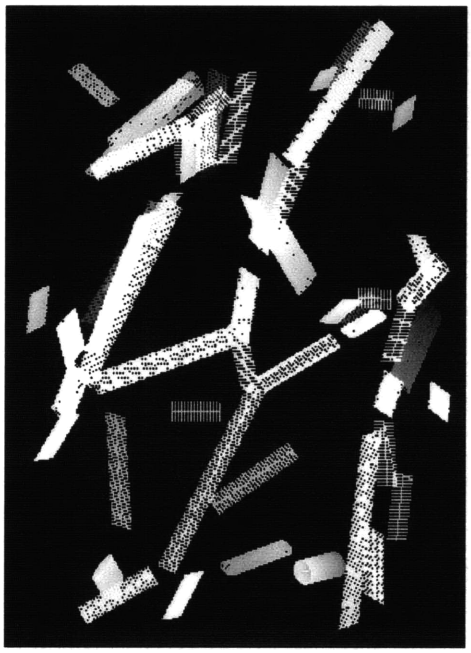


Figure 5.8 Reconstructed image of cell cytoskeleton stereo images in Figure 1.1

Chapter 6 Wavelet Multiresolution Method

The concept of multiresolution is a philosophy to think and a strategy to approach a problem. In stereo vision, the idea of multiresolution appears as several forms, such as coarse-to-fine analysis [Marr and Poggio 1977][Grimson 1985] or hierarchical approach [Marapane and Trivedi 1996]. The strategy based on multiresolution helps to reduce ambiguity and enhance stereo matching. It is considered one of the most efficient and reasonable ways for stereo vision.

Wavelet has been an ideal tool for multiresolution analysis since its invention in the 1980's. Unlike Fourier analysis, which works in the frequency domain and the time (or spatial) domain, wavelet analysis deals with the scale domain and the time (or spatial) domain. "Scale" refers to different resolutions. Wavelet decomposes a signal or data into various components at different scales. More importantly, wavelet may offer perfect reconstruction. Perfect reconstruction implies no loss of information during wavelet decomposition.

There are other benefits of using wavelet multiresolution. Normally, a Gaussian or Laplacian operator is used in multiresolution analysis because these filters are easily adjustable in both the time (or spatial) domain and the frequency domain. However, they lose much information during filtering because they do not have perfect reconstruction. Furthermore, the data at separate levels are correlated and redundant in some senses [Kim, Lee, *et al* 1997]. These characteristics may lead to inaccurate matching. The wavelet multiresolution approach does not have these weaknesses. It can provide pyramidal representation without duplication or loss of information. Therefore, the wavelet multiresolution method is more attractive.

Little work has been reported on the application of wavelet to stereo vision [Pan 1996][Kim, Lee, *et al* 1997][Zhou and Dorrere 1994]. The results of existing methods

have not been very successful. Part of the reason is that in their methods only wavelet decompositions were considered for stereo matching. As discussed in Chapter 2, because the wavelet coefficients are non-structural features, they are not very effective for stereo matching.

In this chapter, we propose a wavelet multiresolution method for stereo vision, incorporating structural information. We argue that structural information is important for image matching. Our structural information is directional feature, which is given by directional representation. To be consistent with the multiresolution idea, directional representation is extracted at different scales. Therefore, two pyramids are formed: the wavelet pyramid and the directional representation pyramid. In our algorithm, the wavelet is the perfect reconstruction type, i.e., all the information will contribute to the stereo matching.

In the following, section 6.1 provides an overview of wavelet theory. Section 6.2 describes the implementation of our wavelet multiresolution method in detail. In section 6.3, we will discuss the pros and cons of this method.

6.1 Wavelet Theory

Wavelet theory is a relatively recent development even though some ideas and notations appeared early in this century. The early studies aimed at finding a function basis alternative to the Fourier basis. Many studies were going on independently in various fields. A synthesis took place during 1980's, and the formal mathematical theory was finally worked out with contributions from Grossman and Morlet (1980), Mayer (1986), Daubechies (1988), Mallat (1989), etc. Since then, wavelet applications have been explored at an explosive rate. A historical review can be found in Graps [Graps 1995] and Meyer [Meyer 1993].

Briefly, wavelet offers a new way to represent a signal or data in scales. The wavelet transform cuts up a signal or data into different frequency components, and then studies each component with a resolution matched to its scale [Daubechies 1992].

This section is mainly based on Strang [Strang and Nyuyen 1996] and Daubechies [Daubechies 1992]. The discussion is limited to discrete wavelet transform, which is a dyadic wavelet form.

6.1.1 One-dimensional DWT and multiresolution

Wavelet is closely related to filter banks. The scaling function $\phi(t)$ is defined as

$$\phi(t) = 2\sum_{k=0}^{N} h_0[k]\phi(2t-k) \tag{6.1}$$

where $h_0[n]$ is a low pass filter with length N. Equation (6.1) is often called the "dilation equation." The dilation equation couples the representations of scaling function $\phi(t)$ at two scales. The scaling function $\phi(t)$ is determined by a discrete-time filter $h_0[n]$. Similarly, the wavelet function $w(t)$ is defined as

$$w(t) = 2\sum_{k=0}^{N} h_1[k]\phi(2t-k) \tag{6.2}$$

where $h_1[n]$ is a high pass filter. The characteristics of h_0, h_1 determine the characteristics of $\phi(t)$ and $w(t)$. For example, if $h_0[n]$ and $h_1[n]$ are finite, the scaling function $\phi(t)$ and the wavelet function $w(t)$ are compact supported. Compact support is one of the most important properties that make wavelet analysis more favorable than Fourier analysis in some cases.

Perfect reconstruction (PR) is another important property. Let us look at perfect reconstruction of a filter bank. Figure 6.1 describes a two-channel filter bank with H_0 and H_1 in the analysis side and F_0 and F_1 in the synthesis side. Perfect reconstruction means that the output $x_2(n)$ at the synthesis side is only the shift version of the original signal

$x_1(n)$. Four filters H_0 , H_1 , F_0 , F_1 are related to each other for perfect reconstruction. Therefore, the scaling function $\phi(t)$ and the wavelet function $w(t)$ are related as well.

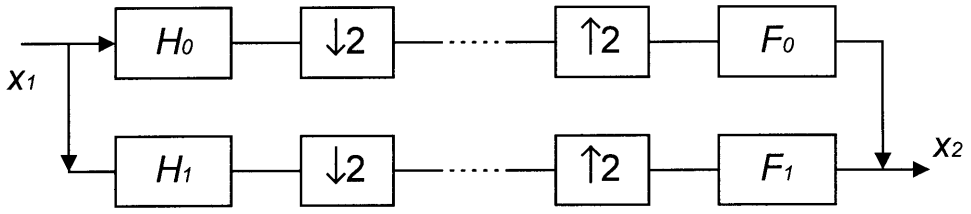


Figure 6.1 Two-channel filter bank

For multiresolution analysis, the signal is broken into different scales by a series of downsampling. Figure 6.2 shows the filter bank decomposition. The decomposition is performed at a logarithmic tree structure, with lowpass iteration only. At each decomposition, H_0 gives "approximation", such as a_1, a_2, and H_1 gives "detail", such as d_1, d_2, at corresponding scales. The "detail" component can also be treated as the difference between two consecutive "approximations".

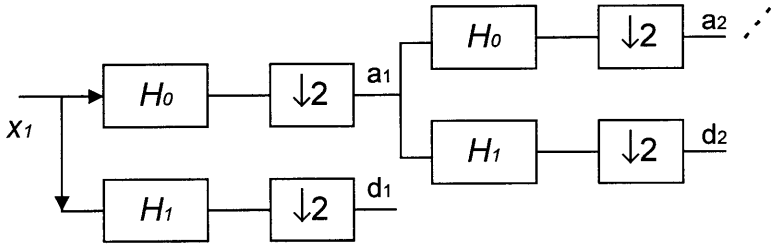


Figure 6.2 Logarithmic tree decomposition

Wavelet decomposition is conducted in the same way. Let us introduce two notations:

$$\phi_{l,k}(t) = 2^{l/2}\phi(2^l t - k) \tag{6.3}$$

$$w_{l,k}(t) = 2^{l/2} w(2^l t - k) \tag{6.4}$$

where j is scale parameter and k is shift parameter. Equation (6.3) gives all basis functions of dilated and translated versions of the scaling function $\phi(t)$, and equation (6.4) gives all basis functions of dilated and translated versions of the wavelet function $w(t)$.

Two function spaces, Vj and Wj, are defined as

$$\mathbf{V}_j = \{ \textit{all combinations } \sum_k a_j[k]\phi_{j,k}(t) \textit{ at scale } j \} \tag{6.5}$$

$$\mathbf{W}_j = \{ \textit{all combinations } \sum_k b_j[k]w_{j,k}(t) \textit{ at scale } j \} \tag{6.6}$$

$\mathbf{V}_j$ forms embedded subspaces, which mean

$$\ldots \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \ldots \subset \mathbf{V}_j \subset \mathbf{V}_{j+1} \subset \ldots$$

i.e., each $\mathbf{V}_j$ is contained in the next space $\mathbf{V}_{j+1}$. All $\mathbf{V}_j$, $j \in (-\infty, \infty)$, form a complete set of basis functions, which correspond to the "approximations" at all scales. The scaling coefficients, a_j, of a function $f(t)$ are the projections of $f(t)$ to $\mathbf{V}_j$, which are given by

$$a_{j,k} = \langle f, \phi_{j,k}(t) \rangle = \int_{-\infty}^{+\infty} f(t)\phi_{j,k}(t)dt . \tag{6.7}$$

Spaces $\mathbf{W}_j$ are the complementary space of $\mathbf{V}_j$, i.e.

$$\mathbf{V}_j \oplus \mathbf{W}_j = \mathbf{V}_{j+1} \tag{6.8}$$

where $\oplus$ is direct sum. The spaces $\mathbf{W}_j$ are the differences between $\mathbf{V}_j$ and $\mathbf{V}_{j+1}$. It is also easy to show that spaces $\mathbf{V}_j$ are the sum of all $\mathbf{W}_j$.

$$\begin{aligned} \mathbf{V}_J &= \mathbf{W}_{J-1} \oplus \mathbf{V}_{J-1} = \mathbf{W}_{J-1} \oplus \mathbf{W}_{J-2} \oplus \mathbf{V}_{J-2} = \ldots \\ &= \sum_0^{J-1} \mathbf{W}_j \oplus \mathbf{V}_0 \end{aligned} \tag{6.9}$$

Therefore, all $\mathbf{W}_j$, $j \in (-\infty, \infty)$, also provide a complete set of basis functions, which correspond to the "details" at all scales. The wavelet coefficients, b_j, of a function $f(t)$ are the projections of $f(t)$ to $\mathbf{W}_j$:

$$b_{l,k} = \langle f, w_{l,k}(t) \rangle = \int_{-\infty}^{+\infty} f(t) w_{l,k}(t) dt, \qquad (6.10)$$

and $\{ a_0, b_0, b_1, ..., b_j \}$ are all coefficients of function $f(t)$ at j-th decomposition. From equation (6.9), we can also obtain the approximation coefficients, a_j, at all scales (0~j-th).

In the implementation, a fast algorithm is available based on equation (6.9). This pyramid algorithm is called the Mallat algorithm. The transform is the fast discrete wavelet transform (FWT), similar to the fast Fourier transform. The important point of this pyramid is that any wavelet coefficient, $b_{j,k}$, at level j is the difference between two consecutive scaling coefficients $a_{j+1,k}$ and $a_{j,k}$ (Figure 6.3). This is exactly like the filter bank case in Figure 6.2.

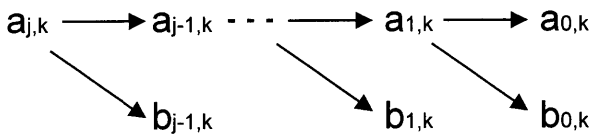


Figure 6.3 Mallat algorithm for wavelet coefficients

6.1.2 Two-dimensional decomposition

It is much more complicated to find a genuine 2D wavelet transform for a 2D case, such as image analysis. For simplicity, a 2D function $f(x,y)$ is often assumed to be separable. Then, the analysis is treated as two one-dimensional cases in series. A 1D Wavelet transform is performed in each row, then in each column. There are three wavelet functions associated with one scaling function. They are as follows:

$$\begin{aligned} \phi(x,y) &= \phi(x)\phi(y) \\ w_1(x,y) &= \phi(x)w(y) \\ w_2(x,y) &= w(x)\phi(y) \\ w_3(x,y) &= w(x)w(y) \quad . \end{aligned} \qquad (6.11)$$

Therefore, there are four coefficients at each scale, instead of two in the 1D case. Among these coefficients, one is the approximation and the other three are different details.

Multiresolution analysis can be conducted in a similar way to the 1D case. The pyramidal structure for the 2D case is shown in Figure 6.4, where A_i represents approximation component at *i-th* scale, and H_i, V_i, D_i are three detail components at *i-th* scale. Also, only approximation A_i decomposes further at each scale.

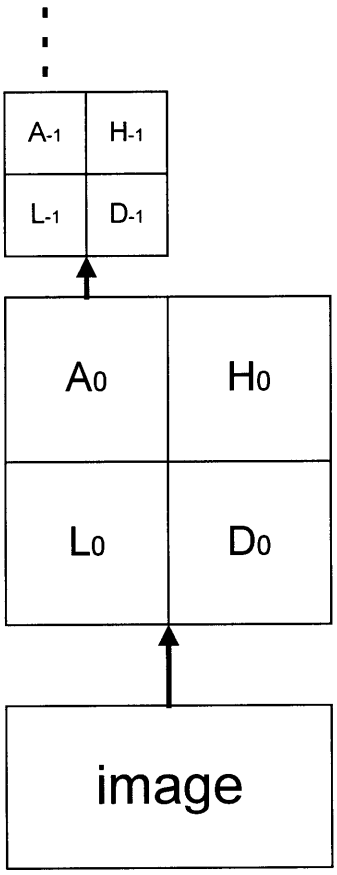


Figure 6.4 Wavelet pyramid structure for image

Mathematically, perfect reconstruction (PR) is biorthogonal because $\mathbf{F} = \mathbf{H}^{-1}$ is required in Figure 6.1. If $\mathbf{F} = \mathbf{H}^{-1} = \mathbf{H}^{T}$, it is then orthogonal. Wavelet with PR is always desirable because it means there is no information loss or distortion. Among various wavelet families, the Daubechies wavelet family is very popular because it not only offers perfect reconstruction but also it is compact supported (compact support means that the filters

$h_0[n]$ and $h_1[n]$ are finite). Both biorthogonal wavelets and orthogonal wavelets are provided. Nevertheless, biorthogonal wavelets are symmetric while orthogonal wavelets are unsymmetrical. For image analysis, symmetric wavelets are more favorable because unsymmetrical wavelets induce undesirable shift. In our study, Daubechies biorthogonal wavelets are used.

6.2 Implementation of Wavelet Multiresolution Method

As stated earlier, the wavelet coefficients are non-structural features and not effective for image matching. In our method, we take directional representation into account because directional representation provides us with the structural information. Therefore, two pyramids are formed for each image: one is the wavelet decomposition, and the other is the directional representation at several scales.

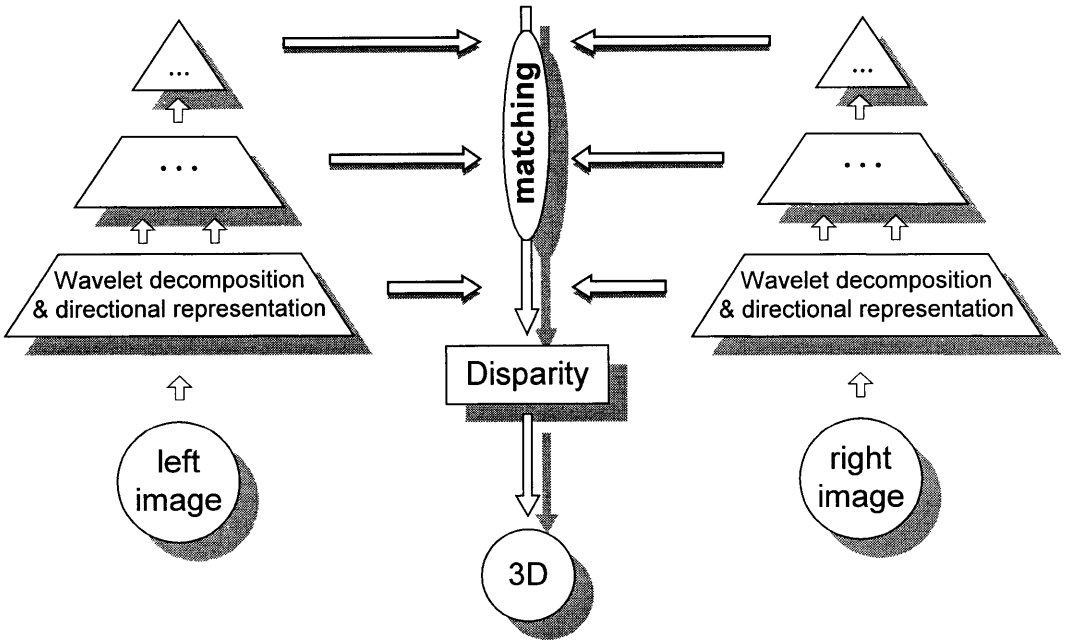


Figure 6.3 architecture of wavelet multiresolution method

Figure 6.3 exhibits the architecture of implementation. The pyramidal architecture contains two parts: an analysis part and a matching part. Both parts are conducted in a

multiresolution strategy. The analysis part goes from bottom to top and the matching part goes from top to bottom. Matching at a high level is carried out first between stereo images. Matching at a lower level is guided and constrained by the matching results at the high level.

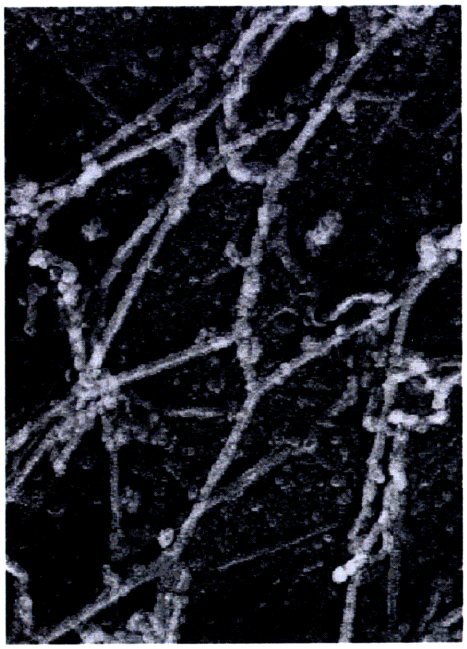
6.2.1 Bottom-up analysis

The pyramid for wavelet coefficients is formed by the discrete wavelet transform. The wavelet family we chose is "biorthogonal wavelet family". There are certain good properties associated with biorthogonal wavelets:

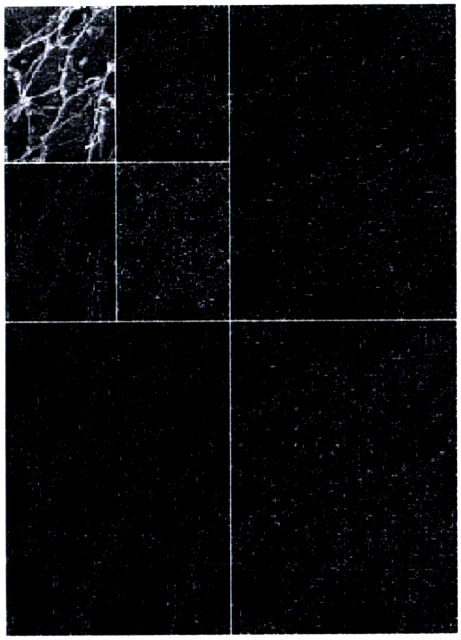
- they are compact supported;
- they offer perfect reconstruction;
- they are symmetric so that no shift is introduced during decomposition;
- a fast algorithm, which is important for computation, is available for both decomposition and reconstruction.

As for the pyramid of directional representations, we subsample the image and then obtain directional representation at different scales. We use "subsample" to distinguish from "downsample" used in wavelet theory. "Downsample" refers to keeping only odd-numbered elements and dumping even-numbered ones. "Subsample" means to apply a low-pass filter in a 2x2 neighborhood. In our algorithm, we average each 2x2 neighborhood so that the image is subsampled with half size. Then directional representation is obtained on the subsampled image. This constitutes one level of directional representation in the pyramid. We can repeat the process to obtain various levels in the pyramid.

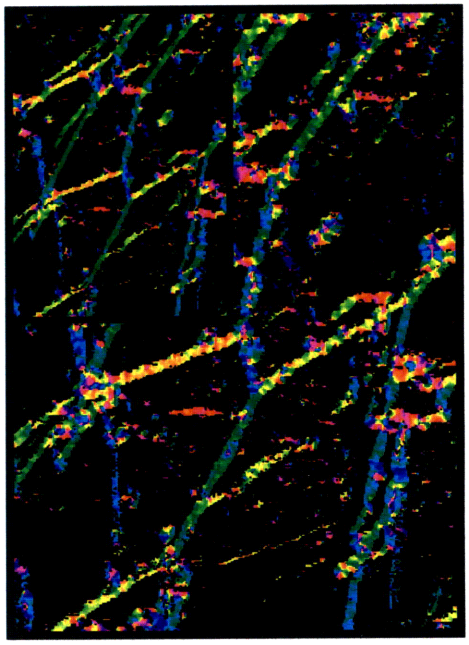
Figure 6.4 depicts an example of the bottom-up analysis. Figure 6.4(a) is the cell cytoskeleton image as in Figure 1.1(a). Figure 6.4(b) and (c) exhibit two pyramids, respectively. Both of them are illustrated in two scales.



(a)



(b)



(c)

Figure 6.4 Bottom-up analysis: (a) original image; (b) wavelet pyramid; (c) directional representation pyramid.

6.2.2 Top-down matching

Stereo matching is conducted in a coarse-to-fine strategy. The matching starts at the coarsest resolution, i.e. the highest level in the pyramid. The matching at a high level puts limitations on the matching at a lower level. On the other hand, the matching at the low level refines the matching at the high level.

The matching is based on point-to-point matching. As one can anticipate, ambiguity and computation are major problems. The pyramidal strategy does help a lot to reduce the ambiguity and computation time. However, a number of constraints still have to be imposed on image matching. In fact, these constraints are enforced in our algorithm. Our principle is that we would rather have no matching than have wrong matching.

The constraints that we employ include the epipolar constraint, disparity constraint, ordering constraint and uniqueness constraint. The search is only carried out at the horizontal scan-line because of horizontal epipolar line. A maximum disparity is predefined to restrict the search space further so that the search space is only limited to a segment of the scan-line. The ordering constraint and uniqueness constraint are used to keep the search from going wrong. As a consequence, the disparity distribution is very sparse because of the strict constraints. An interpolation technique is employed to make it dense. We will discuss it later.

A similarity function is also defined to compare the matching among all the candidates. The calculation is almost the same as in section 5.3.1. The only difference is that we are dealing with is the point feature, instead of the line-segment in Chapter 5.

Each point feature contains wavelet coefficients and directional representations at various scales. Wavelet coefficients at each scale have four components: (*A, H, V, D*). The attributes of each point feature are listed as follows:

$$\{ A(i),\ H(i),\ V(i),\ D(i),\ D_{irection}(i), C_{onfidence}(i),\ I_{ntensity} \mid i = 0,1. \}$$

where i means the *i-th* level in pyramid, $D_{irection}(i)$ means the local direction from the *i-th* directional representation, $C_{onfidence}(i)$ is the corresponding confidence value of the *i-th* directional representation, and $I_{ntensity}$ is the intensity value of the point in the original image. Only two levels are assumed in our case, i.e., two-level wavelet decompostion is performed. As a consequence, there are a total of 13 attributes for each point feature.

The calculation of similarity values of the attributes $\{D_{irection}(i), C_{onfidence}(i), I_{ntensity}\}$ are the same as in Chapter 5. The similarity values of the attributes $\{A(i), H(i), V(i), D(i)\}$ are computed according to the equation (5.2). The weight selection is based on trial-and-error as well. Usually, the weights of $(H(i), V(i), D(i))$ are set to be relatively smaller than those of $A(i)$ and $D_{irection}(i)$.

Our coarse-to-fine analysis makes great use of these attributes. At each horizontal scan-line, the points on this line should have their corresponding points on the same y-coordinate scan-line in the other image. A dynamic programming method is used to determine the best global matching. In fact, "global" only means at each scan-line in our algorithm, i.e., dynamic programming is performed each time at each scan-line.

Dynamic programming is typically applied to optimization problems. It solves problems by combining the solutions to subproblems [Cormen 1990]. Successful attempts at using dynamic programming for solving stereo matching problems are those of Baker [Baker 1981] and Ohta [Ohta and Kanade 1985]. In both cases, they chose edges as the basic primitives. The stereo matching was determined by finding an optimal path in the graph.

Instead of edge primitives, points are considered in our algorithm. At each scanline, a two-dimensional grid is formed, which is very similar to the matrix in Chapter 5. Each point (i, j) of the grid represents a pair with i corresponding to the *i-th* point in the scanline of the left image and j corresponding to the *j-th* point in the scanline of the right image. The point (i, j) is weighted by the similarity between the *i-th* and *j-th* points in two images. The stereo matching is therefore considered as finding an optimal weighted path in the grid as shown in Figure 6.5. Furthermore, the ordering constraint is imposed on the

path so that all admissible paths have to be monotonous [Faugeras 1993]. In other words, if $m=(i, j)$ and $m'=(i', j')$ are two successive points on the path, then they must satisfy $m<m'$ which is defined as:

$$m(i,j) < m'(i,j) \text{ iff } \begin{cases} i < i' \text{ and } j < j' \\ \text{or, } i = i' \text{ and } j < j' \\ \text{or, } j = j' \text{ and } i < i'. \end{cases} \tag{6.12}$$

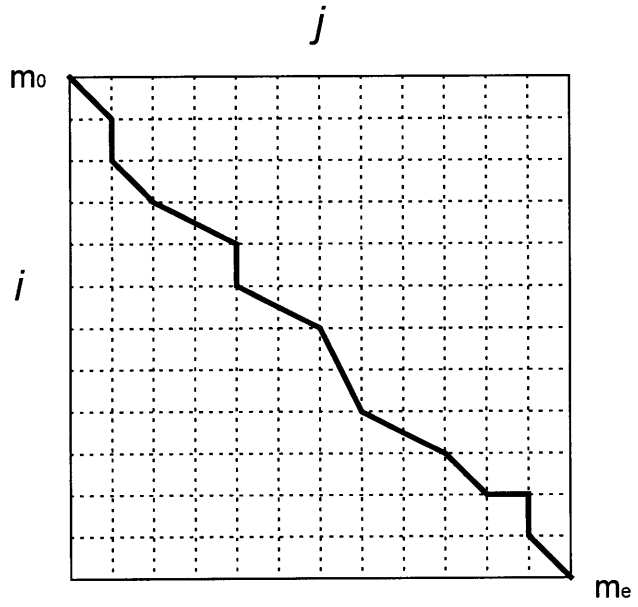


Figure 6.5 The stereo matching is similar to find an optimal path in the grid

The optimal path from m_0 to m_e is defined as the path which maximizes an evaluation function $\boldsymbol{S}(m)$ in our algorithm. It is recursively given by

$$\boldsymbol{S}(m) = \max_{p \text{ in } V_m}(s(m) + \boldsymbol{S}(p)) \tag{6.13}$$

where V_m is the set of nearest neighbors of m with $p<m$, and s(m) is the similarity value of the matching pair $m(i,j)$. The optimal path can be found by backtracking. For example, if the last matching pair (a point on grid) is m_e, we find p_e according to equation (6.13), and then the search becomes to find the maximum of $\boldsymbol{S}(p_e)$. We continue this backtracking until we reach m_0.

In our algorithm, each scanline is processed independently. In fact, the correspondences in one scanline have a strong dependency on the correspondences in the neighboring scanlines because the structure might cross all these scanlines. Ideally, the inter-scanline consistency should also be considered. Ohta and Kanade [Ohta and Kanade 1985] proposed an intra- and inter-scanline search algorithm for edge primitives, in which the search space was extended to the three dimensions. However, for points as in our case, it is almost impossible to realize because of heavy computation.

6.2.3 Results of wavelet multiresolution method

A picture is worth a thousand words. Let us look at an example of the wavelet multiresolution method. Figure 6.5(a) and (b) are stereo images of the cell cytoskeleton. After going through bottom-up analysis and top-down matching for both images, we then recreate a 3D image from disparities. The reconstructed 3D image is shown in Figure 6.5(c), where the intensity represents the depth.

As discussed earlier, we impose very strict constraints and criteria on image matching. The good side is that we do not have many wrong matching pairs and the ambiguity is also reduced greatly. The downside is that disparity distribution is very sparse. So is the reconstructed 3D image as Figure 6.5(c) shows.

Disparity interpolation will help to make the disparity dense. The basic assumption behind the interpolation is the continuity. That means, the disparity of each point should be consistent with the disparities of the neighboring points. Interpolation is used to generate the disparities of the points that are failed in the stereo matching. In our algorithm, linear interpolation is applied because we assume the linear continuity. For a 2D case, the linear interpolation is not easy to compute. In practice, we interpolate the average for simplicity. Figure 6.6 depicts the new 3D image after linear interpolation. We are able to see some 3D structures from this reconstructed 3D image, which is not

revealed in 2D images. For example, some filaments are actually afloat above the other filaments. However, we see them as if they were connected with each other in the 2D images.

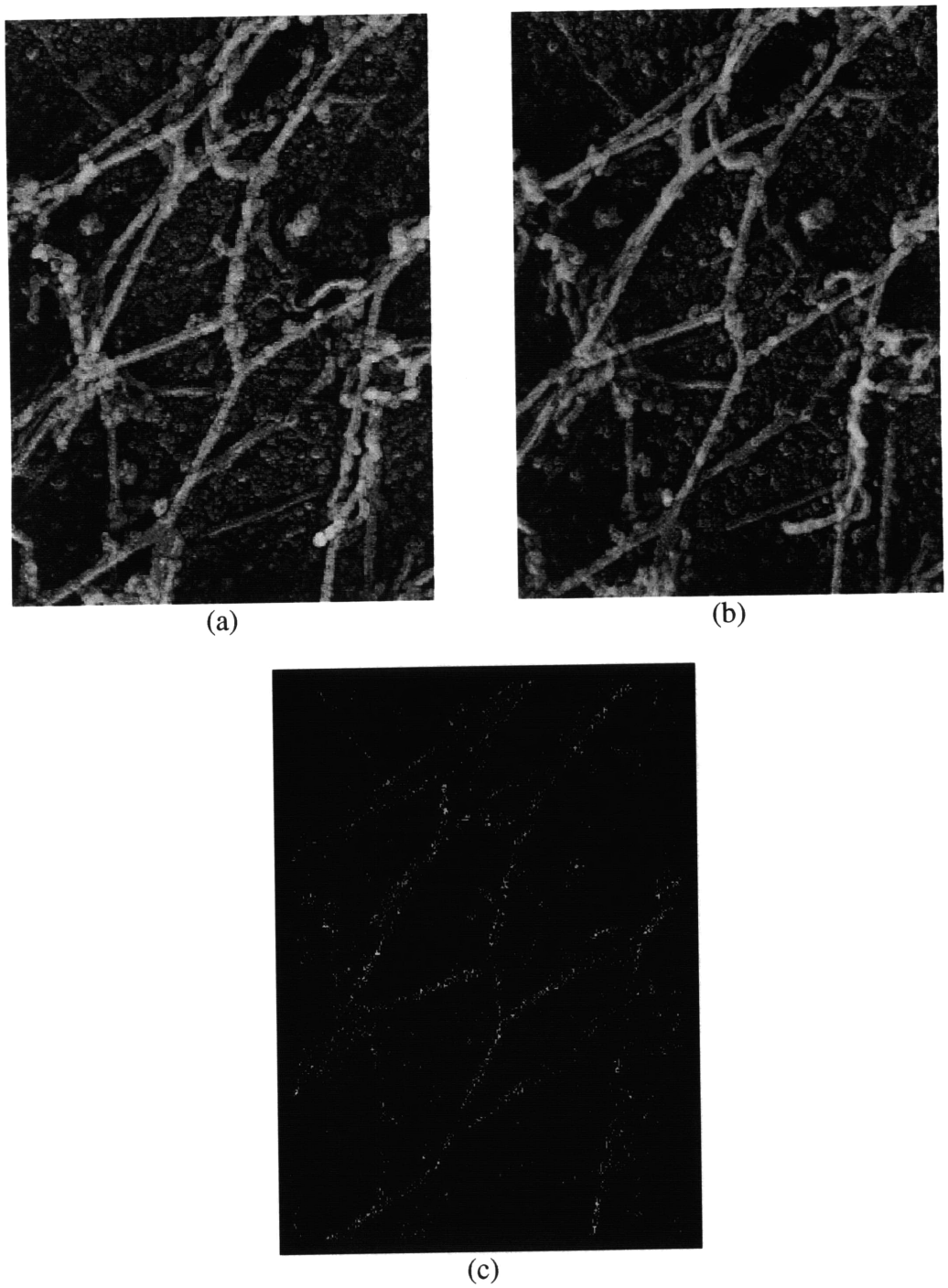


Figure 6.5 Result of cell cytoskeleton stereo images: (a) left image; (b) right image; (c) reconstructed 3D image.

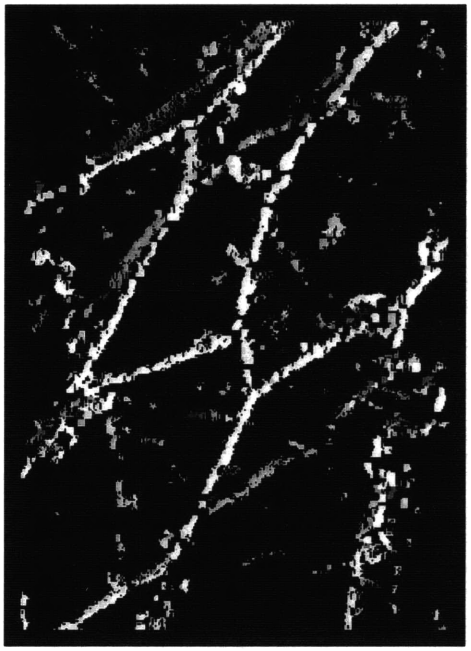
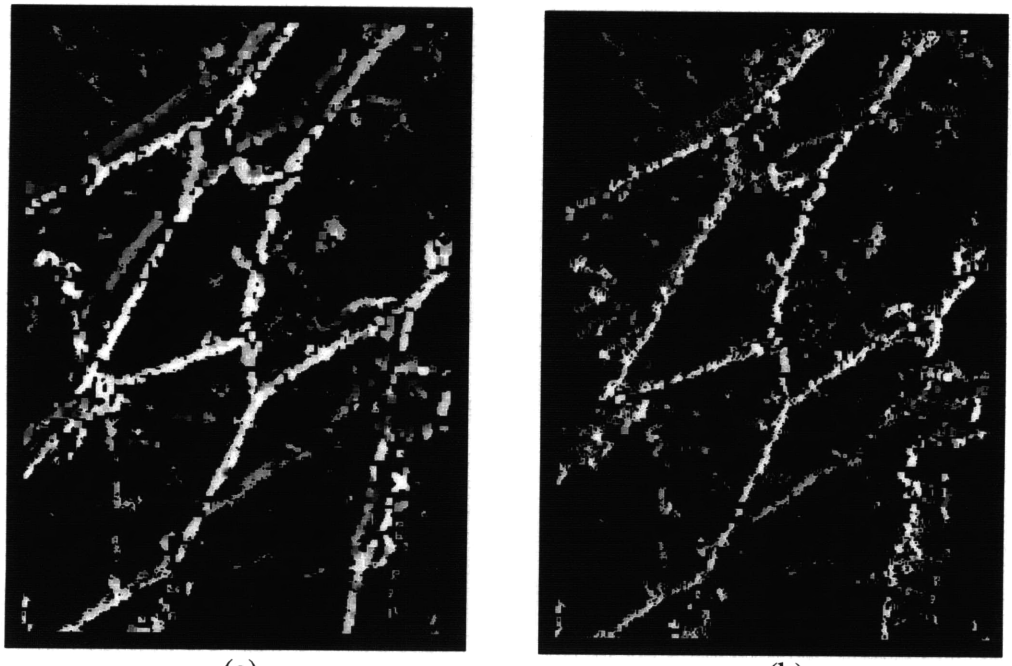


Figure 6.6 New 3D image after interpolation

Note that the 3D image is created from the 3D data calculated. Therefore, we could do any transformation on the whole data set, such as rotations and translations. For comparison, Figure 6.7(a) and (b) exhibit two views of the 3D image with the same view angles as when we took stereo images through the electron microscope. They are $\pm 10^\circ$ in our case. (Figure 6.6 actually corresponds to the view at 0°.)

We have also developed an animated demo to show the 3D structures. The demo is a Java applet, which is available from the ICMIT webpage. (See Appendix A.)



(a) (b)

Figure 6.7 (a) 3D image at view angle of 10°; (b) 3D image at view angle of -10°.

6.3 Discussions

Our results in Figures 6.6 and 6.7 show satisfactory 3D reconstruction from stereo images. More important is that the examples are real images of the cell cytoskeleton. Therefore, the wavelet multiresolution method outperforms the line-segment method.

The only weakness is that we only obtain 3D points from stereo matching. We still have to work on them to get 3D structural information, such as 3D lines. The operation has to be carried out in 3D space. However, this is only a computation problem.

There is also a theoretical pitfall in general wavelet approaches. Wavelet transform does not ensure translation-invariance. Therefore, it seems not quite ideal to choose wavelet coefficients as features for stereo matching. Remember that translation-invariance is a

major property that a good feature representation should have in stereo vision (see Chapter 2). However, this translation-variance generally happens to the detail-components of the wavelet decomposition. In our implementation, we assign very low weights to detail-components so that they only contribute very little to the matching decision. Their effects are also attenuated due to the incorporation of other information, such as directional representations in our method.

In summary, our wavelet multiresolution method not only provides the possibility and feasibility of reconstruction of 3D structure, but also shows the capability to extract and measure structural information in 3D space. This method is also very robust. It is applicable to more complicated situations.

Chapter 7 Conclusions and Future Work

In this thesis, we studied stereo vision from two different approaches. Our results show not only the feasibility but also the capability of 3D reconstruction from stereo images. This is very important for EM images, which are the main applications of our study.

Cell cytoskeleton images were studied extensively as primary examples in this thesis. The reconstructed structure reveals important 3D information which is very difficult to see in the original images. This will help us to understand the relationship of actin filaments. The results also provide the possibility of making measurements based on 3D information. For the cell cytoskeleton, 3D measurements have rarely been exploited previously.

From the standpoint of vision stereo, two approaches were studied on different bases. The line-segment method is primitive-based and the wavelet-multiresolution method is point-based. The selection of features is strongly related to these methods. Our classification of features indicates an explicit relationship. The features are classified into two categories: structural features, related to the primitive-based method, and non-structural features, which are related to the point-based approach.

Different computation frameworks, which corresponded to the two approaches, were proposed. Both performances are very case-dependant. In our cell cytoskeleton examples, the wavelet multiresolution method has proved to be more satisfactory. Part of the credit can be attributed to its pyramidal computation strategy. Table 7.1 lists comparisons of two approaches based on their performances. The main advantage of the line-segment method is that it gives 3D structural results directly, but it is only suitable for some regular pattern images (based on our study). Instead, the wavelet multiresolution method works reasonably well for real case images. However, only 3D points are obtained from

stereo matching. It does not give 3D structural information. Information about 3D structure has to be extracted by 3D operations.

	Structural Information	***Applicability***	***Resistance to noise***
Line-segment Method	yes	only for pattern images	no
Wavelet Multiresolution Method	not directly	real images	to some degree

Table 7.1 comparison of two stereo vision approaches

Structural information has been stressed throughout the thesis. We argue that it is very important for both analysis and stereo matching. In this thesis, structural information is mainly given by directional representation. Directional representation gives a very good description of local structure, which contains two components of the feature value and the confidence value. It is essentially the basis of the line-segment method. In the second approach, the incorporation of structural information with wavelet analysis overcomes some of the weaknesses of the other wavelet methods.

Our study also verifies that directional representation is ideal for filament-type structures. Granlund and Knutsson [Granlund and Knutsson 1995] used similar representation for adaptive filtering and object identification.

Of course, there are still some issues that need improvement. The matching algorithm is one of the major concerns. Even though our matrix matching method provides a clear view of global matching, the problem is not complete. More studies are required to prove this method theoretically and verify it experimentally. Besides, it is crucial to make great use of constraints and find a good search algorithm.

Another issue is the improvement of the wavelet method. The translation-variance of wavelet coefficients is not ideal for the stereo vision. Theoretically, it might be possible to find a translation-invariant wavelet. Practically, we can find some way to reduce the effects of this unfavorable property on the final results.

7.1 Future Work

There are a number of potential directions for further studies. Some of them are already on their way. We briefly consider three areas: 3D measurements of the cell cytoskeleton, stereo vision algorithm and implementation, as well as other applications.

3D measurements of cell cytoskeleton: This will be our main direction of study. This thesis studied reconstruction from stereo vision. 3D data were obtained from reconstruction. The next step concerns the extraction 3D information.

Our further study may focus on the wavelet-multiresolution method since it has proved to be more satisfactory. The wavelet-multiresolution method gives 3D points but no 3D structural information. 3D operation is necessary to work on 3D data. Similar to directional representation in 2D, a tensor operator will be an optimal candidate for multi-dimension representation [Granlund and Knutsson 1995].

Some statistical measurements will be conducted on 3D structural extraction. For example, we are going to measure the length distribution of filaments, the angle distribution of filaments, the typical distance between two filaments, and the volume that filament structure forms, etc. These measurements are important for us to understand and model the mechanism showing how the cell cytoskeleton works. Relaxation measurements will be necessary to reveal the dynamic process of cell cytoskeleton rearrangement as a function of time [Remuzzi *et al* 1984].

Analysis can also be extended to multiple images from different angles. Our algorithm is applied to every two images, and then a synthesis process combines all the results to give a united result. Certainly, it will provide more accurate and detailed information.

Stereo vision algorithm and implementation: Stereo vision is the foundation of 3D measurements. Improving the stereo vision algorithm will be a parallel process to our study on 3D measurements.

Wavelet analysis to stereo vision itself is worth further study. We think wavelet analysis is ideal for "coarse-to-fine" stereo matching. Incorporation of structural information improves the performance of wavelet method. More experiments will be carried out on a wide array of images to verify its applicability and robustness. For the different applications, different structural information can be used with wavelet analysis.

Other applications: There are many applications of stereo vision. Scene reconstruction will be of interest since stereo vision is one of the major topics in robot vision. Besides typical applications, the application of stereo vision to biomedical fields also appears promising. For instance, similar to cell cytoskeleton structure, blood vessels [Kim, Aggarwal, *et al* 1990] and DNA chains [Margalef *et al* 1996] are also filament-type structures. It will be very easy for us to modify our approaches to these cases.

Clearly, most of above applications are speculative, and further work is required to verify and realize them.

Appendix A: Demos

There are several animated demos available on ICMIT webpage. Each demo is a Java applet, which can run over the network. (http://icmit.mit.edu/projects/pia/research.html).

Stereo demo

This demo exhibits original stereo images. Two images are flipped back and forth and give a sense of 3D structure. From this demo, we will be able to check the rotation axis. Figure A.1 is a screenshot of stereo demo. Clicking "left image" or "right image" displays the corresponding still image. Click "Auto switch", and the animation is activated. The scrollbar adjusts the speed of flipping.

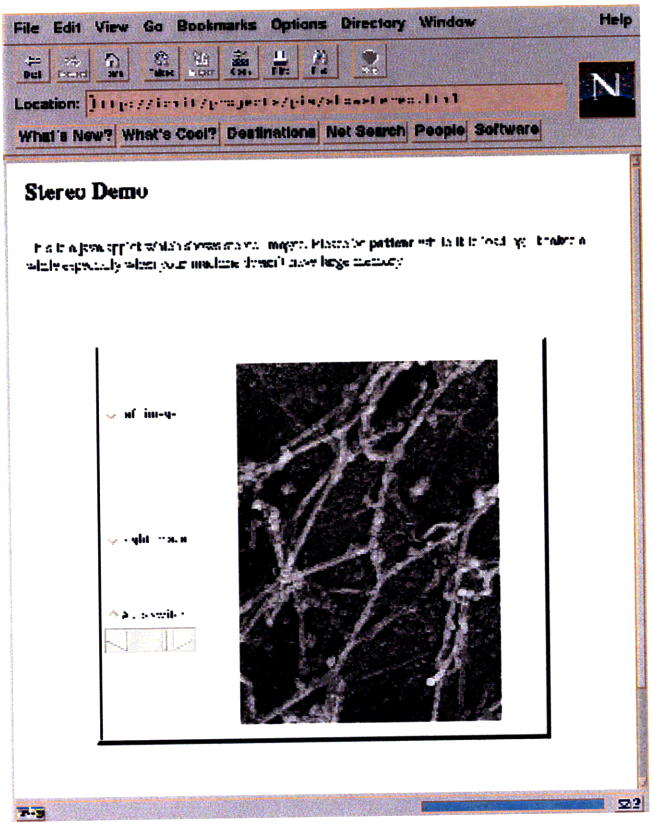


Figure A.1 Screenshot of stereo demo

Reconstructed stereo demo

This demo demonstrates reconstructed 3D images obtained by wavelet multiresolution method. Three images are flipped back and forth (corresponding to the view angle of –10°, 0°, and 10°, respectively.) The intensity in the images is related to the depth (Z-coordinate). The lower the intensity, the larger the depth. The operations are the same as in stereo demo.

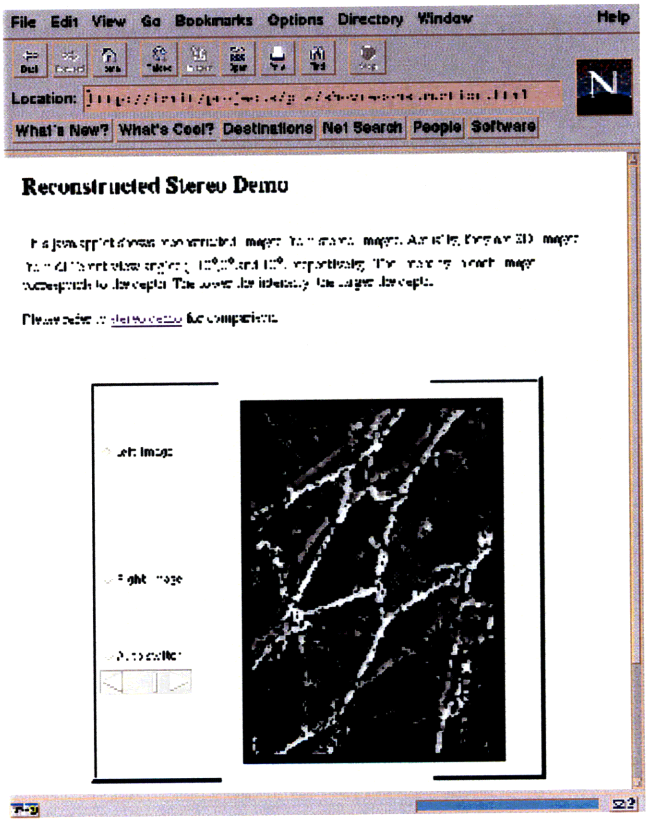


Figure A.2 Screenshot of reconstructed stereo demo

Appendix B: Colormap of Directional Representation

The colormap we use to display the directional representation is adopted from HSV (Hue-Saturation-Value) colormap. The colors begin with red, pass through yellow, green, cyan, blue, magenta, and return to red. This map is particularly useful for displaying periodic functions. In our case, the colors correspond to the directions between [0°, 180°]. Since the line with 0° direction is the same as the line with 180° direction, the corresponding colors must be same as well. This colormap is ideal for our purpose.

Figure B.1 Colormap

In the implementation, we create a look-up table (LUT) to map each direction to each color. Only 60 colors are actually used in our algorithm. Every three directions map to one color. The mapping is listed as follows:

Directions	*R*	*G*	*B*
0°~3°	1.0000	0	0
3°~6°	1.0000	0.1000	0
6°~9°	1.0000	0.2000	0
9°~12°	1.0000	0.3000	0
12°~15°	1.0000	0.4000	0
15°~18°	1.0000	0.5000	0
18°~21°	1.0000	0.6000	0
21°~24°	1.0000	0.7000	0
24°~27°	1.0000	0.8000	0
27°~30°	1.0000	0.9000	0
30°~33°	1.0000	1.0000	0
33°~36°	0.9000	1.0000	0
36°~39°	0.8000	1.0000	0
49°~42°	0.7000	1.0000	0
42°~45°	0.6000	1.0000	0
45°~48°	0.5000	1.0000	0

48°~51°	0.4000	1.0000	0
51°~54°	0.3000	1.0000	0
54°~57°	0.2000	1.0000	0
57°~60°	0.1000	1.0000	0
60°~63°	0	1.0000	0
63°~66°	0	1.0000	0.1000
66°~69°	0	1.0000	0.2000
69°~72°	0	1.0000	0.3000
72°~75°	0	1.0000	0.4000
75°~78°	0	1.0000	0.5000
78°~81°	0	1.0000	0.6000
81°~84°	0	1.0000	0.7000
84°~87°	0	1.0000	0.8000
87°~90°	0	1.0000	0.9000
90°~93°	0	1.0000	1.0000
93°~96°	0	0.9000	1.0000
96°~99°	0	0.8000	1.0000
99°~102°	0	0.7000	1.0000
102°~105°	0	0.6000	1.0000
105°~108°	0	0.5000	1.0000
108°~111°	0	0.4000	1.0000
111°~114°	0	0.3000	1.0000
114°~117°	0	0.2000	1.0000
117°~120°	0	0.1000	1.0000
120°~123°	0	0	1.0000
123°~126°	0.1000	0	1.0000
126°~129°	0.2000	0	1.0000
129°~132°	0.3000	0	1.0000
132°~135°	0.4000	0	1.0000
135°~138°	0.5000	0	1.0000
138°~141°	0.6000	0	1.0000
141°~144°	0.7000	0	1.0000
144°~147°	0.8000	0	1.0000
147°~150°	0.9000	0	1.0000
150°~153°	1.0000	0	1.0000
153°~156°	1.0000	0	0.9000
156°~159°	1.0000	0	0.8000
159°~162°	1.0000	0	0.7000
162°~165°	1.0000	0	0.6000
165°~168°	1.0000	0	0.5000
168°~171°	1.0000	0	0.4000
171°~174°	1.0000	0	0.3000
174°~177°	1.0000	0	0.2000
177°~180°	1.0000	0	0.1000

References

1. Baker, H.H., "Depth from Edge- and Intensity-Based Stereo", Ph.D. Thesis, University of Illinois, 1981.

2. Canny, J.F., "A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8: 769-798, November 1986.

3. Corman, T., "Introduction to Algorithms", MIT press, 1990.

4. Daubechies, I., "Ten lectures in Wavelets", Society for Industrial and Applied Mathematics, Philadelphia, 1992.

5. Davies, P., Remuzzi, A., Gordon, E., Dewey, C. Jr., and Gimbrone, M., "Turbulent Fluid Shear Stress Induces Vascular Endothelial Turnover *in vitro*", *Proc. Natl. Acad. Sci., U.S.A.*, 83: 2114-2117, 1986.

6. Depaola, N., Grimbrone Jr., M., Davis, P., and Dewey, C. Jr., "Vascular Endothelium Responds to Fluid Shear Stress Gradients", *Arterio. Thromb.*, 12: 1254-1257, 1992.

7. Faugeras, O., "Three-Dimensional Computer Vision — A Geometric Viewpoint", The MIT Press, Cambridge, MA, 1993.

8. Fleet, D.J., Jepson, A.D., and Jenkin, M., "Phase-Based Disparity Measurement", *CVGIP: Image Understanding*, Vol.53, No.2, pp.198-210, March 1991.

9. Fua, P., "Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities", *In Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pp.1292-1298, August 1991.

10. Gennery, D., "Modelling the Environment of an Exploring Vehicle by Means of Stereo Vision", Ph.D. thesis, Stanford University, June 1980.

11. Gimbrone, M., "Endothelial Dysfunciton and Atherosclerosis", *J. Cardiac Surgery*, 4(2): 180-183, 1989.

12. Gotlieb, A., Langille, B., Wong, M., and Kim, D., "Biology of Disease: Structure and Function of the Endothelial Cytoskeleton", *Lab. Invest.*, 65: 123, 1991.

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

14. Granlund, G.H., "Hierarchical computer vision", *In Proceedings of EUSIPCO 90*, Barcelona, Spain, September 1990.

15. Granlund, G.H., Knutsson, H., "Signal Processing for Computer Vision", Kluwer Academin Publisher, 1995.

16. Graps, A., "An Introduction to Wavelets", *IEEE Computational Science and Engineering*, Vol. 2, No. 2, 1995.

17. Grimson, W.E., "A Computer Implementation of A Theory of Human Stereo Vision", *Phil. Tran. Roy. Soc. London*, vol. B 292, pp.217-253, 1981; see also, "A Computer Implementation of A Theory of Human Stereo Vision", AI Lab., Mass. Inst. Technol., Cambridge, MA, Memo 565, 1980.

18. Grimson, W.E., "Computational Experiments with a Feature-Based Stereo Algorithm", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 7, No.1, pp.17-34, January 1985.

19. Grimson, W.E., "Why Stereo Vision is Not Always About 3D Reconstruction", AI Lab., Mass. Inst. Technol., Cambridge MA, Memo 1435, 1993.

20. Hartwig, J., and Shevlin, P., "The Architecture of Actin Filaments and the Ultrastructural Location of Actin-binding Protein in the Periphery of Lung Macrophages", *The Journal of Cell Biology*, Volume 103, 1007-1020, Sept. 1986.

21. Hellwich, O., and Faig, W., "Graph-Based Feature Matching Using Descriptive and Relational Parameters", *Photogrammetric Engineering & Remote Sensing*, Vol. 60, No. 4, pp.443-450, April 1994.

22. Horaud, R., and Skordas, T., "Stereo Correspondence Through Feature Grouping and Maximal Cliques", *IEEE Trans. Pattern Analysis Mach. Intell.*, PAMI-11(11), 1168-1180, 1989.

23. Knutsson, H., "Filtering and Reconstruction in Image Processing", PhD Thesis, Linköping University, Sweden, Diss. No. 88, 1982.

24. Knutsson, H., Wilson, R., and Granlund, G.H., "Anisotropic Nonstationary Image Estimation and Its Application: Part I — Restoration of Noisy Images", *IEEE Transactions on Communications*, Vol. COM-31, No. 3, pp.388-397, 1983.

25. Lee, S., and Leou, J., "A Dynamic Programming Approach to Line Segment Matching in Stereo Vision", *Pattern Recognition*, Vol. 27, No. 8, pp. 961-986, 1994.

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

27. Kim, Y.S., Lee, J.J., and Ha, Y.H., "Stereo Matching Algorithm Based on Modified Wavelet Decomposition Process", *Pattern Recognition*, Vol. 30, No.6, pp.929-952, 1997.

28. Marapane, S.B., and Trivedi, M., "Multi-Primitive Hierarchical(MPH) Stereo Analysis", *IEEE Trans. Pattern Analysis Mach. Intell.*, PAMI-16(3), 227-240, 1994.

29. Margalef, F.P., Furrer, P., Kunt, M., and Dubochet, J., "The Flying Cylinder: A new Algorithm for Filament Recognition in Noise Stereo Images", *Journal of Structural Biology*, 116, pp.25-29, 1996.

30. Marr, D., and Poggio, T., "A Theory of Human Stereo Vision", *Proc. Roy. Soc. London*, vol. B 204, pp.301-328, 1979; see also, " A Theory of Human Stereo Vision", AI Lab., Mass. Inst. Technol., Cambridge, MA, Memo 451, 1977.

31. Marr, D., and Hildreth, E., "Theory of Edge Detection", *Proc. Roy. Soc. London*, vol. B 207, pp.187-217, 1980.

32. Marr, D., "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information", W.H. Freeman and Company, San Francisco, 1982.

33. Meyer, Y., "Wavelets: Algorithms and Applications", Society for Industrial and Applied Mathematics, Philadelphia 1993.

34. McGrath, J.L, "Measuring Actin Dynamics in Endothelium", Master Thesis, MIT, 1994.

35. McIntosh, J.H., and Mutch, K.M., "Matching Straight Lines", *Computer Vision, Graphics, and Image Processing*, Vol. 43, No. 3, pp. 386-408, 1988.

36. Melter, R., Stojmenović, I., and Žunić, J., "A new Characterization of Digital Lines by Least Square Fits", *Pattern Recognition Letters*, 14, pp. 83-85, 1993.

37. Niederman, R., Amrein, P.C., and Hartwig, J., "Three-dimensional Structure of Actin Filaments and of an Actin Gel Made with Actin-binding Protein", *The Journal of Cell Biology*, Volume 96, 1400-1413, May 1983.

38. Nordberg, K., "Signal Representation and Processing Using Operator Groups", PhD Thesis, Linköping University, Sweden, Diss. No. 366, 1994.

39. Ohta, Y., and Kanade, T., "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming", *IEEE Trans. Pattern Analysis Mach. Intell.*, PAMI-7(2), 139-154, 1985.

40. Pan, H.P., "General Stereo Image Matching using Symmetric Complex Wavelets", *SPIE Conference: Wavelet Applications in Signal and Image Processing*. Denver, August 1996. Published in SPIE Proceedings Vol. 2825.

41. Pitas, I., "Digital Image Processing Algorithms", Prentice Hall, 1993.

42. Pollard, S.B., Mayhew, J.E., and Frisby, J.P., "A Stereo Correspondence Algorithm Using A Disparity Gradient Constraint", *Perception*, 14: 449-470, 1985.

43. Remuzzi, A., Dewey, C. Jr., Davies, P., and Gimbrone, M., "Orientation of Endothelial Cells in Shear Fields in Vitro", *Biorheology*, 21, pp.617-630, 1984.

44. Sanger, T.D., "Stereo Disparity Computation Using Gabor Filter", *Biological Cybernetics*, 59, 405-418, 1988.

45. Strang, G., and Nguyen, T., "Wavelets and Filter Banks", Wellesley-Cambridge Press, 1996.

46. Vosselman, G., "Relational Matching", *Lecture Notes in Computer Science*, No. 628, Springer-Verlag press, 1992.

47. Westelius, C.J., "Focus of Attention and Gaze Control for Robot Vision", Ph.D. Thesis, Linkoping University, Sweden, 1995.

48. Wildes, R., "On Interpreting Stereo Disparity", Ph.D. Thesis, MIT, Cambridge, MA, 1989; see also, "On Interpreting Stereo Disparity", AI Lab., MIT, Memo 1112, 1989.

49. Wu, M.S., and Leou, J.J., "A Bipartite Matching Approach to Feature Correspondence in Stereo Vision", Pattern Recognition letters, 16, pp.23-31, January 1995.

50. Yu, S.S., and Tsai, W.H., "A New Thinning Algorithm for Gray-Scale Images by the Relaxation Technique", *Pattern Recognition*, Vol. 23, No. 10, pp.1067-1076, 1990.

51. Zhang, T.Y., Suen, C.Y., "A Fast Parallel Algorithm for Thinning Digital Patterns", *Communications of the ACM*, Vol. 27, No.3, pp.236-239, 1984

52. Zhou, X., and Dorrer, E., "Automatic Image-Matching Algorithm Based on Wavelet Decompostition", *IAPRS*, Vol. 30, pp.951-960, *ISPRS Comm. III Symposium: Spatial Information from Digital Photogrammetry and Computer Vision*, September 1994, Munich.