Multiple Layer Image Analysis for Video Microscopy

Brian S. Eastwood

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2009

Approved by:

Russell M. Taylor II, Advisor

Marc Niethammer, Reader

Gary Bishop, Reader

Leonard McMillan, Committee Member

David Hill, Committee Member

Abstract

Brian S. Eastwood: Multiple Layer Image Analysis for Video Microscopy. (Under the direction of Russell M. Taylor II.)

Motion analysis is a fundamental problem that serves as the basis for many other image analysis tasks, such as structure estimation and object segmentation. Many motion analysis techniques assume that objects are opaque and non-reflective, asserting that a single pixel is an observation of a single scene object. This assumption breaks down when observing semitransparent objects—a single pixel is an observation of the object and whatever lies behind it. This dissertation is concerned with methods for analyzing multiple layer motion in microscopy, a domain where most objects are semitransparent.

I present a novel approach to estimating the transmission of light through stationary, semitransparent objects by estimating the gradient of the constant transmission observed over all frames in a video. This enables removing the non-moving elements from the video, providing an enhanced view of the moving elements.

I present a novel structured illumination technique that introduces a semitransparent pattern layer to microscopy, enabling microscope stage tracking even in the presence of stationary, sparse, or moving specimens. Magnitude comparisons at the frequencies present in the pattern layer provide estimates of pattern orientation and focal depth. Two pattern tracking techniques are examined, one based on phase correlation at pattern frequencies, and one based on spatial correlation using a model of pattern layer appearance based on microscopy image formation.

Finally, I present a method for designing optimal structured illumination patterns tuned for constraints imposed by specific microscopy experiments. This approach is based on analysis of the microscope's optical transfer function at different focal depths.

For Lisa.

Acknowledgments

There are many people to whom I am grateful for their help in completing this dissertation. I thank my advisor, Russell Taylor, for working closely with me over the past six years and for providing encouragement and clarity when I needed it. I will miss our many hours of brain hurting math.

I consider myself very fortunate to have had the support of my committee members, Marc Niethammer, Gary Bishop, David Hill, and Leonard McMillan. They have provided a wealth of guidance and insight throughout my graduate career. I thank Richard Superfine for—along with Russell—fearlessly leading CISMM, for helping me understand Fourier optics, and for providing very helpful and constructive criticism. My thanks go to Mary Whitton for being a strong mentor to me, helping me with my writing, and providing guidance in teaching. I also extend special thanks to Lamar Mair for creating the test square grid patterns used throughout this work.

This work was funded in part by the National Institute of Biomedical Imaging and Bioengineering (NIBIB), grant number P41-EB002025.

The staff of the Computer Science department at UNC are second to none. I thank them all for their help over the years.

I have many friends, collaborators, and colleagues at UNC and beyond who have influenced my work and provided encouragement. They are Jill Aldredge, Rob Aldredge, Brian Austin, Greg Baker, Amanda Baltzley, Michael Baltzley, Eric Bennett, Aaron Block, Kerry Bloom, David Borland, Fred Brooks, Sabrina Burmeister, Samuel Caro, Brianna Carstens, Mukta Chakraborty, Diana Clark, Jeremy Cribb, Kyla Davidoff, Edward Dale, Ben Evans, Mike Falvo, David Feng, Yoni Friedman, Matt Fuxjager, Nico Galoppo, Guido Gerig, Casey Goodlett, Nathan Hudson, Melanie Kessels, Bill Kier, Anselmo Lastra, Christina Lebonville, Steven Lebonville, Cathy Lohmann, Kenneth Lohmann, David Marshburn, Bruce Maxwell, Justin McAlister, Suzanne McAlister, Daniel Millard, Tim O'Brien, Angela Oyeno, Ted Oyeno, Justin Peters, Stephen Pizer, Marc Pollefeys, Rebecca Price, Cory Quammen, Danielle Racke, Jenny

Rice, Ted Salmon, Katrina Salvante, Jason Sewall, Kendra Sewall, Ryan Schubert, Adam Shields, Montek Singh, Dale Skrien, Jack Snoeyink, Keith Sockman, Ricky Spero, David Stotts, Vinay Swaminathan, Stephanie Taylor, Scott Taylor, Jur van den Berg, Leandra Vicci, Sean Washburn, Stephen Weiss, Jeremy Wendt, Buddy Whitman, Brian Wilson, Shun Xu, and many others.

I am eternally grateful to my parents, Abe Eastwood and Judy Brook, who have instilled in me the passion for lifelong learning and provided me with many opportunities and experiences. My father taught me how to write my first computer program. My mother spent summers teaching me about the natural sciences. I also thank Sutton Monro, Jo Ann DiMartini, and Anne and Vinnie Mangiamele, for their love and support. I thank my brother, Chuck, for being the big brother I always look up to. I thank Suzanne and Carl Eastwood for keeping me smiling.

Finally, I thank my love, my life, my wife, Lisa Mangiamele. She is the reason I am who I am and I do what I do.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1D . . . . . one-dimensional

2D . . . . . two-dimensional

3D . . . . . three-dimensional

ADC . . . . . analog-to-digital converter

CCD . . . . . charge-coupled device

CISMM . . . . . Computer Integrated Systems for Microscopy and Manipulation

DC . . . . . direct current, mean amplitude

DFT . . . . . discrete Fourier transform

DIC . . . . . differential interference contrast

EM . . . . . electromagnetic

GLT . . . . . gradient logarithm transmission

GUI . . . . . graphical user interface

ITK . . . . . Insight Toolkit

LTM . . . . . local transmission metric

MCL . . . . . Mad City Labs

MSC . . . . . model-based spatial correlation

NA . . . . . numerical aperture

NCC . . . . . normalized cross-correlation

NIH . . . . . the National Institutes of Health

OTF . . . . . optical transfer function

OTL . . . . . optical tube length

PSF . . . . . point-spread function

PSNR . . . . . peak signal-to-noise ratio

RANSAC . . . . . random sample consensus

RMS . . . . . root mean squared

SNR . . . . . signal-to-noise ratio

TEM . . . . . transmission electron microscope

UNC . . . . . the University of North Carolina at Chapel Hill

USD . . . . . United States Dollars

VTK . . . . . Visualization Toolkit

WPC . . . . . weighted phase correlation

List of Symbols

cm . . . . . centimeter, 10^{-2} meters

Cr . . . . . chromium

dB . . . . . decibel

$^{\circ}$ . . . . . degrees

fps . . . . . frames per second

Au . . . . . gold

μm . . . . . micrometer or micron, 10^{-6} meters

μs . . . . . microsecond, 10^{-6} seconds

mm . . . . . millimeter, 10^{-3} meters

nm . . . . . nanometer, 10^{-9} meters

pixel . . . . . a picture element

Chapter 1

Introduction

This dissertation considers analysis of images formed of scenes containing multiple semitransparent objects. The goal of image analysis is to extract information from images that provide insight about the scene, such as the size, shape, and motion of objects in it. Models of image formation drive the image analysis process by providing constraints that inform the researcher about how an object will appear when imaged.

A great deal of image analysis research considers objects to be opaque and nonreflective, such as a piece of clay. In an image of such an object, the information at a single image location comes from a single point on the object. The situation changes, however, when imaging semitransparent objects, such as frosted glass. In an image of such an object, a single image location contains information from the semitransparent object *and* whatever lies behind it. Under these conditions, different models of image formation are required to form accurate insight about the scene being observed.

Multiple layer imaging presents a combination of challenges and opportunities. Many image analysis techniques established for images of opaque objects break down in the mixture of information from multiple layers. The benefit of semitransparency, however, is that an object is visible at all times, even when hidden behind another object. This means there is information about the object in the image—the trick is getting at it.

One domain in which multiple layer images abound is microscopy. In observations

of biological specimens at the microscopic scale, transparency is the norm—small fish, worms, parameciums, cell membranes, organelles, and cellular scaffolding are all components of a semitransparent world of great interest to science.

My thesis statement is:

> Tracking multiple, semitransparent, moving layers in microscopy videos requires image analysis techniques that are different from those used in tracking opaque objects. Median gradient estimation of log-intensity images enables the accurate removal of the stationary component from videos containing semitransparent moving objects. Harmonic analysis of structured illumination pattern frequencies and model-based spatial correlation enable three dimensional stage tracking for microscopy. Semitransparent patterns composed of optimally-selected sinusoids enable tracking with accuracy below the Abbe resolution limit.

The thesis statement is examined in the context of the following novel results of this research:

- I present a novel technique for recovering a model of stationary objects in multiple layer images [ET07].
  - Median gradient estimation of log-intensity images yields an estimate of the constant gradient of light transmitted through each location in a series of bright-field microscopy images. A Fourier transform-based gradient integration constructs a model of the stationary light transmission field.
  - This model provides a specimen-specific field correction that removes stationary components from the images.
  - This method has been shown to enhance motion in microscopy images, and serves as a preprocessing step that improves motion estimation.

- I present a novel structured illumination technique that introduces a semitransparent image layer to microscopy images to provide stage tracking information even in the presence of stationary and moving specimens.
  - Magnitude comparisons at the frequencies present in the pattern layer provide estimates of pattern orientation and focus depth.
  - Two pattern-layer tracking techniques are examined, one based on phase correlation and one on spatial correlation.
  - The phase correlation approach could be fast enough for online tracking, and is accurate to within 0.01 pixel for unoccluded light paths and within 0.2 pixel in the presence of moving semitransparent specimens with sparsely-distributed contrast at focal distances up to 6 times the depth of field of the objective lens.
  - The spatial correlation approach is accurate to within 0.5 pixel in the presence of a moving semitransparent specimen that has densely-distributed contrast.
  - Analysis of the microscope objective optical transfer function (OTF) enables optimal pattern design of sinusoidal patterns tuned for specific microscopy experiments.

1.1 Microscopy Applications

The techniques developed in this research are applicable to a broad range of microscopy imaging applications. The following represents a brief overview of active research that may benefit from my research.

Cilia-driven flow

Ciliated human epithelial lung cells are grown from tissue culture in glass chambers to study the dynamics of cilia-driven mucus flow [BB08]. In experiments performed by some of my collaborators at the University of North Carolina at Chapel Hill (UNC), an inverted light microscope focuses on the cilia or into the mucus layer at the top surface of the cells, through the bottom of the container, the cell substrate, and the cell bodies. The thickness of the cell layer is approximately $10\,\mu$m, the cilia layer is $7\,\mu$m, and the mucus layer is up to $50\,\mu$m. A long working distance lens is required to focus this far into the specimen, and subsequently the image formed has a large depth of field. Visible within one image are the fixed cell structures, the beating cilia, and particles in the flowing mucus layer.

Stationary occlusion removal, discussed in Chapter 4, processes such videos to exclude the non-moving cell layer, enhancing the motion at the cilia and mucus layers. Ongoing research is concerned with observing mucus flow over regions spanning multiple fields of view. Structured illumination microscopy would provide a method to track the stage motion, establishing the relative position of different observations with the cell culture.

Vesicle transport

Intracellular vesicle transport occurs along microtubules—cell scaffolding structures—driven by molecular motor proteins such as kinesin and dynein [HPBH04]. Transport along $100 - 200\,\mu$m neurites that are fixed to the top surface of a cover slip are imaged with a 60X, 1.0NA water immersion lens and images are acquired with a $55 \times 55\,\mu\text{m}^2$ field of view. Vesicle motion up to $15\,\mu$m is tracked using differential interference contrast (DIC) microscopy with background subtraction from a computed sliding mean of images.

My occlusion removal method may enhance the view of vesicle transport in bright-field microscopy and structured illumination may enable tracking vesicle transport over

longer ranges, across multiple fields of view or even along pathways oblique to the image plane.

Bead diffusion

Diffusion experiments investigate the motion of small particles—for example, polystyrene microbeads—through liquids of different concentrations (such as sucrose), semipermeable membranes (such as cell and nucleus membranes), and meshes (such as fibrin protein clots). Current experiments are constrained to 200 μm fields of view, observed with a 40X lens (Tim O'Brien, personal correspondence). Structured illumination microscopy may enable observations in such experiments over long ranges with an understanding of how far from the seed location beads have diffused.

Cell motility

Cell motility research seeks to understand the mechanics of cell motion. Observed over a long period of time, some cells migrate long distances, alternately extending filopodia and contracting the cell body. Typical cell bodies are approximately 10 μm thick (Tim O'Brien, personal correspondence). Structured illumination microscopy may provide a way to track individual cells over many fields of view, maintaining accurate information about total cell motion.

Sea urchin larvae development

Research comparing the evolution of larvae development in related populations of Pacific and Caribbean sea urchins involves making three-dimensional (3D) measurements of larva arms using bright-field microscopy [McA08]. In this research, lateral (x and y) positions are recorded using a *camera lucida*, which enables simultaneously viewing the microscope field and a digital drawing tablet. The axial (z) positions are obtained

through coupling the microscope's fine focus knob to an optical encoder. Measuring distances of 240 μm in 3D is typical in this research. Structured illumination may provide a method to measure these 3D distances more directly.

Nematode tracking

Nematodes are small, semitransparent worms used in research of chemical, mechanical, and thermal sensing and motion regulation. Recent research has concentrated on effective methods to track the sinusoidal motion of these specimens over long ranges, but this is often constrained to the field of view of the imaging system [TT07]. Low magnifications are used to track populations of the worms within a wide field [HS06], and motorized microscope stages provide tracking of individual animals at higher resolutions [GCB+04]. Structured illumination microscopy may provide tracking at high magnifications without the use of motorized stages.

Combined fluorescent and bright-field microscopy

Fluorescent dyes stain cellular structures, such as actin or tubulin, so that they are visible with a fluorescent microscope, but not with transmitted light illumination (bright-field). Most fluorescent microscopes also have bright-field capabilities, and some microscopes enable switching quickly between the two modes [SSW+03]. With such a setup, moving structures could be observed in fluorescence and tracked using structured illumination microscopy.

1.2 Outline

The remainder of this dissertation is organized as follows. Chapter 2 provides an introduction to microscopy image formation and digital image acquisition. Chapter 3 surveys prior image analysis research related to my area of study. Chapter 4 discusses a tech-

nique for removing stationary, semitransparent image layers from microscopy videos. Chapter 5 discusses two techniques for determining the lateral motion of semitransparent patterns and the application to tracking a microscope stage. Chapter 6 discusses the design of patterns optimized for structured illumination microscopy tracking, and extends stage tracking to three dimensions. Chapter 7 summarizes the dissertation and presents avenues for future research in this field.

Chapter 2

Microscopy Image Formation and Acquisition

As mentioned in Chapter 1, this work is concerned with analyzing information from multiple mixed image layers. Multiple layer images arise often in biological microscopy where the objects under study are semi-transparent and a non-zero depth of field means images are formed of objects from multiple levels in the specimen. Before examining how to extract layer information from microscopy images, it is necessary to understand how microscopy images are formed. In this chapter I will discuss the optics of microscopy image formation, digital image acquisition, and image sensor noise characteristics.

The fundamental concern in image acquisition is the behavior of light. Visible light is a thin spectrum of *electromagnetic (EM) radiation*. EM radiation is a transverse wave phenomenon—the radiation propagates in one direction and a pair of orthogonal electric and magnetic fields oscillate tangential to the propagation direction [Mur01, Ch. 2]. The EM waves in visible light have wavelengths between 400 and 700 nm. The speed of light propagation is constant for all wavelengths, but depends on the medium through which the light is passing:

$$v = \lambda f, \tag{2.1}$$

where v is the speed of light for a particular medium and λ and f are the wavelength

and frequency of the light wave, respectively.

Like many natural phenomena, several models exist to explain how light behaves, and each model is appropriately applied at different scales of observation. In free space, light waves travel in a straight line and parallel light waves remain parallel, like waves across the open ocean. At large scales, light can be modeled as traveling in straight rays. This is the domain of geometrical optics analysis, which can be used to describe image magnification in lenses. In computer graphics, ray tracing is an application of geometrical optics analysis.

When light is obstructed by a barrier, the light waves bend—*diffract*—spread out on the other side of the barrier, like ocean waves passing through a breakwater in a harbor. Though light diffracts in all optical systems, the diffraction effect is more apparent for smaller apertures, especially for apertures smaller than several wavelengths of the light. This is the domain of Fourier optics, which considers the propagation and interference of EM waves. Macroscale photography can often accurately model image formation by considering only the ray behavior of light, but microscopy imaging requires considering the effects of diffraction.

The energy in EM radiation is carried in quantized packets called *photons*. When light is incident on a surface, such as certain metals, the photons can interact with electrons in the material. This is the domain of quantum physics, which considers the particle nature of light. This atomic-scale phenomenon is fundamental to the operation of digital image sensors.

The quantum theory of light—specifically quantum electrodynamics—is consistent in describing the behavior of light at the diffraction and macroscopic scales; Fourier optics theory is consistent in describing behavior at the macroscopic scale. Each larger scale theory, however, offers simplifications that make analysis easier at each of the different scales. In the sections that follow, I will rely on each of the light models to build a complete explanation of how images are formed and captured in digital video

microscopy.

2.1 Microscope Optics

In bright-field microscopy, light passing through a specimen is collected and focused by a series of lenses to form an image of the specimen at the image sensor plane. Figure 2.1 diagrams the major components of a bright-field microscope. Bright-field microscopes share a common optimal lens alignment scheme—known as *Köhler* [1] *alignment*—with several types of light microscopes, *e.g.* dark-field, phase contrast, differential interference contrast, and fluorescence [Mur01, Ch. 1]. Each of these imaging modalities relies on a different property of light to generate contrast, and a single microscope can often switch among these different modes.

Understanding how images are formed in a microscope requires knowledge of its optical alignment. This provides a basis for understanding what effect different elements in the microscope light path have on the images formed. Köhler alignment, introduced by August Köhler in 1893, specifies how to position the lenses in a microscope in order to create two conjugate sets of image planes. *Field planes* are locations where the specimen is in focus. *Aperture planes* are locations where the illuminating lamp is in focus. Light from the lamp is completely out of focus at the specimen plane, providing a uniform illumination field. Köhler alignment provides two major benefits:

1. Uniform specimen illumination. In an alternative illumination scheme—one used on microscopes before 1893—an image of the light source is focused on the specimen plane, confounding understanding of the specimen structure.

2. Independent adjustment of numerical aperture (NA) and field of view. The NA controls the quantity of light collected by the objective lens and determines the

[1] Köhler frequently appears written Koehler.

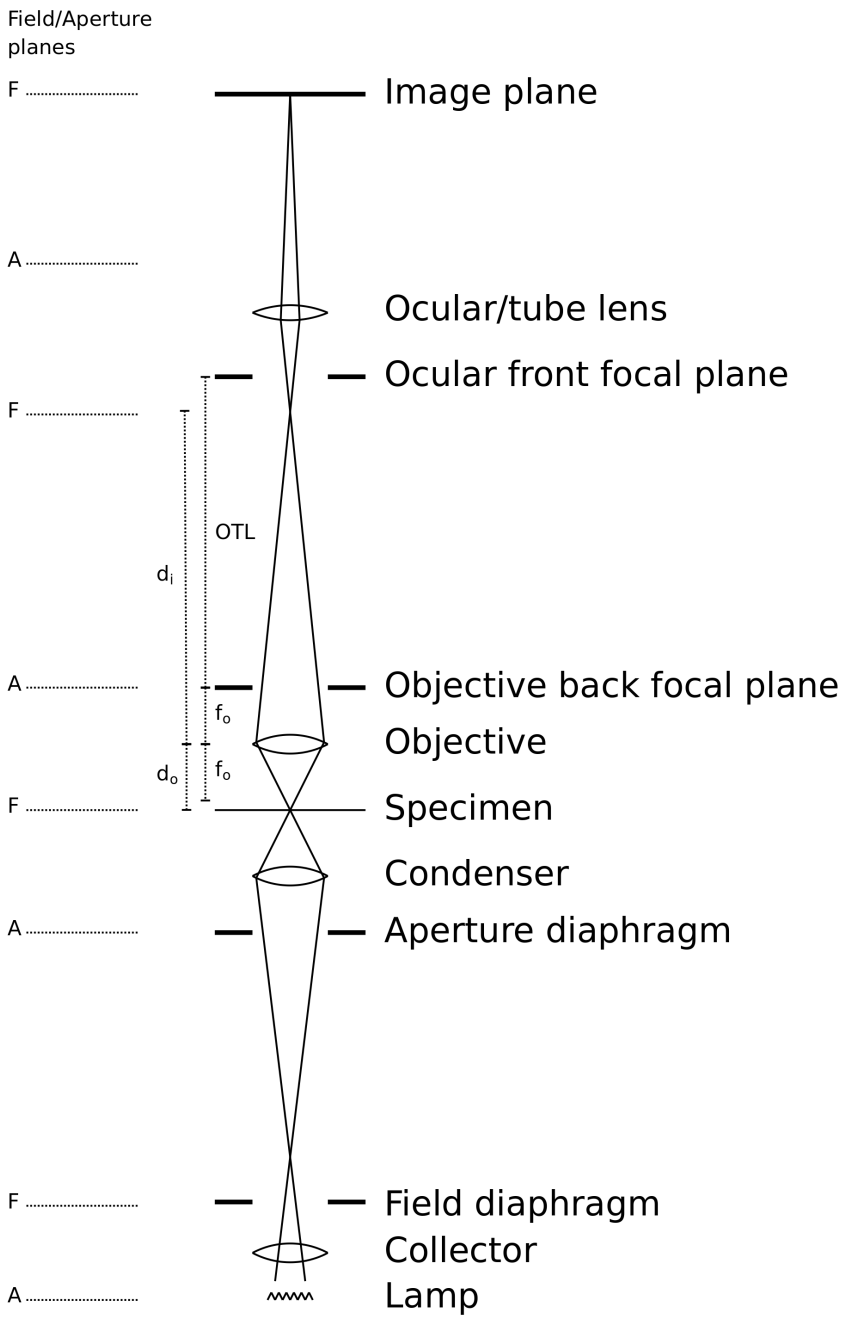


Figure 2.1: The optical path of a bright-field microscope. Light rays (straight lines) intersect at the field planes, where the specimen is in focus. The specimen is placed slightly in front of the front focal plane of the objective lens, which projects a magnified intermediate image in the microscope top tube. The ocular lens further magnifies this image to form an image at the image observation plane.

resolution limit of the microscope. The field of view determines the area of the specimen that is illuminated.

The aperture light path forms focused images of the illumination lamp at several points in the microscope's optical path. The collector lens focuses light from the lamp onto the aperture diaphragm. Adjusting the aperture diaphragm changes the angle of the cone of light that illuminates the specimen, which affects contrast and resolution. In practice, the aperture diaphragm is usually left fully open. The condenser lens defocuses the lamp image at the aperture diaphragm into parallel rays at the specimen plane, providing uniform illumination across the field of view (the image of every point on the lamp is spread out over the specimen plane). The objective lens creates a focused image of the lamp at the back focal plane of the objective lens. This image is observed during the alignment procedure by removing the ocular lens. The ocular lens (eyepiece) creates another image of the lamp at the back focal plane of the ocular, just in front of the observer's eye.

The field light path forms focused images of the specimen at several points in the microscope's optical path. The condenser lens focuses an image of the field diaphragm onto the specimen plane. (There is therefore also a focused image of the specimen at the field diaphragm.) Adjusting the opening of the field diaphragm changes the field of view. The objective lens forms a focused image of the specimen at an intermediate image plane near the front focal plane of the ocular. (This field plane is less common in modern microscopes. Instead, the specimen image is "infinity focused" through this section of the microscope to facilitate introduction of light conditioning filters, such as polarizers and phase prisms.) If the image is being viewed by a human observer, the ocular lens and the observer's eye lens act together to form a virtual image of the specimen on the retina. Oculars are designed to create this image 250 mm in front of the eye, which is a comfortable viewing distance for the human visual system to examine an object. If the image is being recorded by a digital sensor, a tube lens forms a real image of the

specimen on the image sensor.

2.1.1 Geometrical Optics Analysis

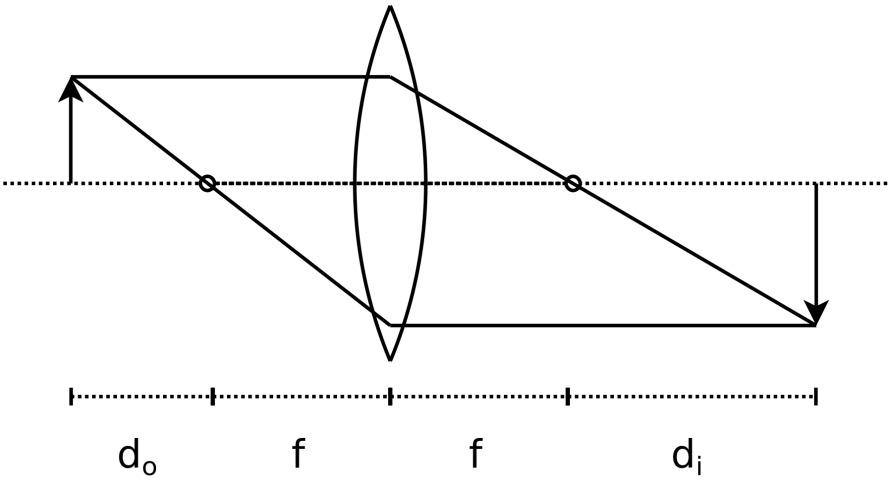


Figure 2.2: Image formation by an ideal thin lens using geometrical optics analysis. Rays parallel to the optical axis of a lens converge at the focal point on the opposite side of the lens. Tracing two rays from a point on an object placed in front of a lens reveals where the image of that object will be formed behind the lens.

Geometrical optics analysis—in which light is modeled as straight, non-diverging rays—provides a framework to discuss the magnification in an imaging system. An ideal, convex thin lens causes light rays parallel to the optical axis (infinity-focused light) incident on one side of the lens to converge at a single focal point that lies on the optical axis on the other side of the lens. The axial distance from the center of the lens to the focal point is known as the focal distance, f, of the lens. When an object is placed in front of a convex lens, an image of the object is formed behind the lens, as seen in Figure 2.2. The thin-lens equation relates the axial distance from an object to a lens, d_o, to the image of the object formed by the lens, d_i, as

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}. \tag{2.2}$$

The linear magnification provided by the lens is the ratio of the image distance to the

object distance,

$$M = \frac{d_i}{d_o}. \tag{2.3}$$

Microscopes are compound systems of lenses, and the basic principles of their imaging can be understood with repeated application of the thin-lens equation. In the microscope diagram of Figure 2.1, the specimen is placed slightly in front of the front focal point of the objective lens. The magnified intermediate image is formed at the front focal plane of the tube lens. The total magnification provided by this compound lens system at the observation image plane is the product of the magnifications from the objective and tube lenses. Under a strict geometrical analysis, then, the image formed of a semitransparent planar object is

$$I(x, y) = \frac{1}{M} LT\left(\frac{x}{M}, \frac{y}{M}\right), \tag{2.4}$$

where M is the magnification of the system, L is the uniform illumination intensity, and $T(x, y)$ describes the *optical transmission* of the object—the fraction of light passed by each point in the object.

Most modern microscopes have interchangeable objective lenses of different focal lengths that offer different magnifications in the microscope system. In a *parfocal* system, the matched objectives for a particular microscope form the intermediate image at the same location in order to minimize the need to refocus after changing objectives [FHC03]. The distance between the focal points of the objective and tube lenses stays fixed—this is known as the optical tube length (OTL).

The numerical aperture (NA) of the objective lens measures the light-gathering ability of a microscope, which impacts its resolution limit. Let θ_2 be the solid half angle of light admitted into the objective from a point in the center of the specimen plane, as depicted in Figure 2.3. NA is defined as:

$$\text{NA} = n_2 \sin \theta_2, \tag{2.5}$$

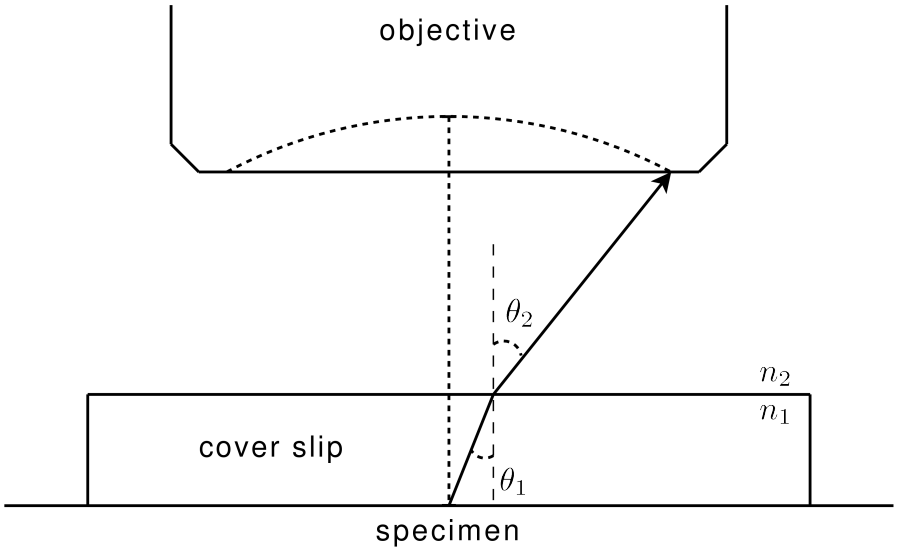


Figure 2.3: The numerical aperture of an objective lens is proportional to the sine of the half-angle of light admitted by the lens. The index of refraction of the medium between the cover slip and objective lens determines the maximum obtainable NA.

where n_2 is the index of refraction of the medium between the specimen and objective lens. The range of this refractive index is $n_2 \in [1.0 \ldots 1.5]$ with commonly used media including air ($n_2 = 1.00$), water ($n_2 = 1.33$) and oil ($n_2 = 1.52$) [Mur01].

Observing specimens through a glass cover slip places a constraint on the maximum obtainable NA. Light rays traveling through the cover slip are refracted at the boundary of the cover slip, as see in Figure 2.3. The angle of incidence and refraction for two mediums with refractive indices (n_1, n_2) are related by

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{2.6}$$

Total internal reflection occurs when the angle of refraction $\theta_2 \geq \frac{\pi}{2}$—all light is reflected into the cover slip. The critical angle of incidence above which total internal reflection

occurs is therefore

$$\theta_{c1} = \arcsin\left(\frac{n_2}{n_1}\right), \tag{2.7}$$

which is defined when $n_1 \geq n_2$. From Equations 2.6 and 2.5, the theoretical maximum NA is equal to the index of refraction of the medium between the cover slip and lens. For example, air and cover slip glasses have refractive indices of $n_a = 1.00$ and $n_c = 1.52$, respectively. The critical angle for the cover slip-air interface is therefore $\theta_{c1} \approx 41.8°$. The practical value is slightly less than this, so dry objectives are limited to NAs of about 0.95. NAs of greater than 1 are obtainable using an immersion medium of water ($n_2 = 1.33$) or oil ($n_2 = 1.52$) between the cover slip and objective lens.

The impact of NA on the microscope resolution limit is discussed in Section 2.1.3, after a more formal discussion of image formation.

2.1.2 Fourier Optics Analysis

As seen in Equation 2.4, geometrical optics analysis predicts that images formed by lenses are perfect magnified representations of objects. In fact, diffraction and lens aberrations cause the image of a point to spread out over a small area in the image plane. The blurred image of a point on the observed object is defined by the optical system's point-spread function (PSF). Even for an object that is in perfect focus, light rays reflected or transmitted from a point are not all projected to a single point in the image plane.

The image analysis literature provides several models of point spread functions for optical systems. A geometrical optics approach obtains a PSF by projecting the aperture of the optical system onto the image plane. Watanabe and Nayar use such an approach in obtaining shape information by analyzing the defocus of textured surfaces [WN97, WN98]. The projection of a circular aperture yields a "pillbox" function, where the light from a single point source is uniformly distributed over a circular region. Because

this model does not consider diffraction effects, it is appropriate for macroscale imaging applications, but not for microscopy.

Cheezum *et al.* use the following radially symmetric PSF to simulate images of particles in fluorescence microscopy [CWG01], derived from Born and Wolf [BW75]:

$$\mathrm{PSF}(r) = \left(\frac{2J_1(ra)}{r}\right)^2 \tag{2.8}$$

$$a = \frac{2\pi \mathrm{NA}}{\lambda}, \tag{2.9}$$

where r is the radial distance from the center of a point in the geometrical optics projection, λ is the wavelength of light emitted by the fluorescing particle, and J_1 is the first order Bessel function. As shown below, this equation describes the PSF for a microscope considering diffraction effects but only when the object is in perfect focus. The work presented here involves both diffraction and focus effects, so a more complete model of image formation is required.

Fourier optics analysis, which considers the diffraction and interference of light waves, provides a complete method for understanding how light propagates through space to form images in a microscope. Joseph Goodman's *Introduction to Fourier Optics* provides a thorough discussion of the image formed by placing a semitransparent planar object in front of a convex lens [Goo68, Ch. 5 & 6]. A brief overview follows.

Fourier optics analysis is based on the propagation of complex-valued EM radiation fields through space. Let $\mathbf{U}_o(x_o, y_o)$ be the complex EM field immediately in front of a semitransparent planar object that has been illuminated by a planar EM wave. When placed in front of a lens, the field formed at a point in the image plane, $\mathbf{U}_i(x_i, y_i)$, is the integral of the shifted product of the object field with the PSF, $\mathbf{h}(x_i, y_i; x_o, y_o)$:

$$\mathbf{U}_i(x_i, y_i) = \iint_{-\infty}^{\infty} \mathbf{h}(x_i, y_i; x_o, y_o)\mathbf{U}_o(x_o, y_o)dx_o dy_o. \tag{2.10}$$

That is, the PSF provides the optical system output at a location (x_i, y_i) in the image plane due to an impulse in the object plane at (x_o, y_o). Although the PSF in Equation 2.10 is dependent on both object plane and image plane coordinates, this is purely a matter of formality at this point—the PSF will be shown to be *shift invariant*, dependent only on the difference in coordinates, $(x_i - x_o, y_i - y_o)$. With a shift-invariant PSF, Equation 2.10 represents a convolution.

The task remains to determine how the EM field propagates from a point on the object field through a lens and onto an image plane. The EM field radiates spherically from the point source, and this spherical wave is incident on the lens. The lens aperture admits a portion of the wave front which diffracts in the region beyond the aperture. It is the finite aperture, not the lens itself, that gives rise to diffraction in an optical system—light diffracts after passing through an aperture whether or not a lens is present.

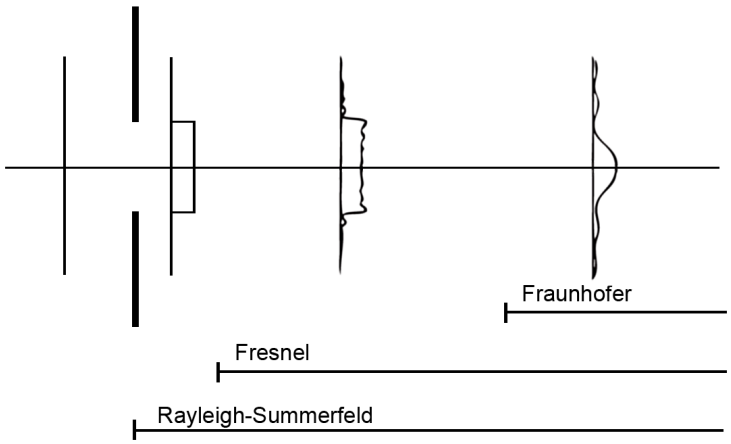


Figure 2.4: The profiles in this figure represent the intensity of a planar wave front propagating from left to right and passing through an aperture. The diffraction of light admitted through an aperture is described by the Rayleigh-Sommerfeld equation at all distances from the aperture. The Fresnel and Fraunhofer equations become valid approximations to the Rayleigh-Sommerfeld equation at distances farther from the aperture [Gas78].

Several equations describe the diffraction of light admitted by an aperture, the most general of which is the Rayleigh-Sommerfeld diffraction equation that holds at any

distance from the aperture [Gas78, Ch. 10]. As depicted in Figure 2.4, two other equations are approximations to the Rayleigh-Sommerfeld equation that are valid in regions some distance away from the aperture. In microscopy, the image plane is very far from the objective lens aperture where the far-field approximations provided by the Fraunhofer equations are sufficient. (Note that Figure 2.4 serves only as an illustration of diffraction regions—it depicts a planar wave incident on an aperture, whereas the light from a point source incident on an objective lens is spherical. It is the aperture, not the shape of the wave front nor the presence of the lens, that causes diffraction.)

Goodman employs the Fraunhofer diffraction equation and the assumption that the imaging application is concerned with measuring the light intensity (not phase) of the image formed—which is the case in microscopy—to arrive at the following approximation of the PSF:

$$\begin{aligned}\mathbf{h}(x_i, y_i; x_o, y_o) \approx & \frac{1}{\lambda^2 d_o d_i} \iint_{-\infty}^{\infty} P(x,y) e^{i\frac{k}{2}\left(\frac{1}{d_o}+\frac{1}{d_i}-\frac{1}{f}\right)(x^2+y^2)} \\ & \times e^{-ik\left[\left(\frac{x_o}{d_o}+\frac{x_i}{d_i}\right)x+\left(\frac{y_o}{d_o}+\frac{y_i}{d_i}\right)y\right]} dxdy, \end{aligned} \tag{2.11}$$

where λ is the monochromatic frequency of light emitted by the point source, d_o and d_i are the axial distances from the lens to the object and image planes, respectively, $k = \frac{2\pi}{\lambda}$ is a wave number, and $P(x, y)$ is the lens aperture function.

For example, the lens aperture function for a circular aperture of radius a is

$$P(x,y) = \begin{cases} 1 & \text{if } \sqrt{x^2+y^2} \le a \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

Here, the physical lens aperture radius a expressed in terms of NA is

$$a = d_o \sin\theta = \frac{d_o \mathrm{NA}}{n}, \tag{2.13}$$

not $d_o \tan\theta$, as might be assumed from Figure 2.3 [Paw06, Ch. 11].

Returning to Equation 2.11, making the following substitutions

$$M = \frac{d_i}{d_o} \tag{2.14}$$

$$x'_o = -Mx_o \qquad y'_o = -My_o \tag{2.15}$$

yields

$$\begin{aligned}\mathbf{h}(x_i - x'_o, y_i - y'_o) =& \frac{1}{\lambda^2 d_o d_i} \iint_{-\infty}^{\infty} P(x,y) e^{i\frac{\pi}{\lambda}\left(\frac{1}{d_o}+\frac{1}{d_i}-\frac{1}{f}\right)(x^2+y^2)} \\ & \times e^{-i\frac{2\pi}{\lambda d_i}\left[(x_i - x'_o)x + (x_i - x'_o)y\right]} dxdy. \end{aligned} \tag{2.16}$$

This equation represents a comprehensive, shift-invariant PSF model that considers both diffraction and defocus effects. That is, evaluating Equation 2.16 for a fixed object distance d_o provides a slice of the 3D point spread function of the microscope. Note that the integral is evaluated over the domain of the aperture function while the PSF is defined over the coordinate differences between the image and object planes $(x_i - x'_o, y_i - y'_o)$.

When the term $\frac{1}{d_o} + \frac{1}{d_i} - \frac{1}{f} = 0$ in Equation 2.16, the thin lens Equation 2.2 is satisfied. Under this condition the object is in focus—in fact, this is where the thin lens equation is derived from. Goodman notes that at optimal focus the PSF is the Fraunhofer diffraction pattern of the lens aperture,

$$\mathbf{h}(x_i - x'_o, y_i - y'_o) = \frac{1}{\lambda^2 d_o d_i} \iint_{-\infty}^{\infty} P(x,y) e^{-i\frac{2\pi}{\lambda d_i}\left[(x_i - x'_o)x + (x_i - x'_o)y\right]} dxdy. \tag{2.17}$$

The Fraunhofer diffraction of a circular aperture is known as the *Airy disk*, as seen in Figure 2.5. The analytical solution to Equation 2.17 is a first-order Bessel function—this

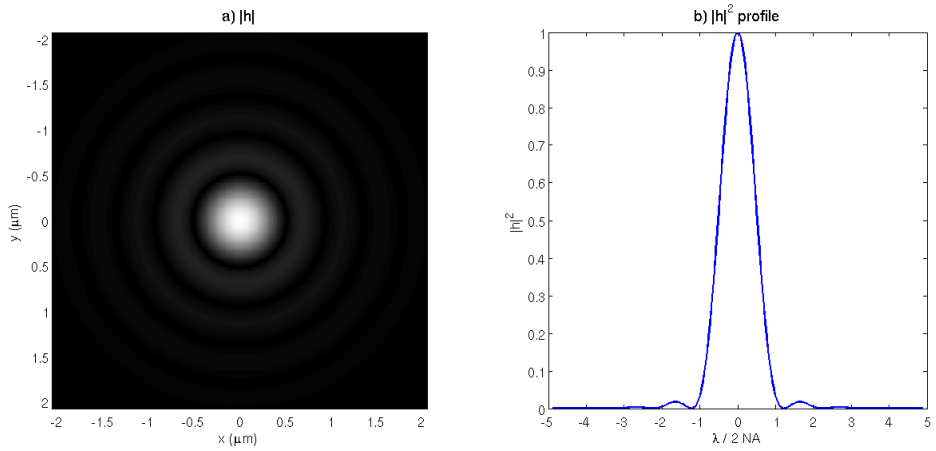


Figure 2.5: A simulated point spread function for a 40X dry objective lens with circular aperture and NA = 0.65. a) The image of the PSF, with size measured in object plane coordinates. b) The profile of the normalized PSF, with units $\frac{\lambda}{2\text{NA}}$. The first minimum occurs at $r = \frac{1.22\lambda}{2\text{NA}}$.

solution gives rise to the PSF model used by Cheezum *et al.* described by Equation 2.8.

At optimal focus, the image formed by the objective lens most closely resembles the object placed in front of the lens. The distortion in the image is a result of the lens aperture selecting a fraction of the light diffracted from from the object. Because light that is diffracted more is excluded, and this diffracted light carries high frequency information, the lens behaves as a non-ideal low-pass filter. A section of the microscope PSF through the XZ plane, seen in Figure 2.6, demonstrates how the more general Equation 2.16 behaves when the object is out of focus.

Microscopy imaging is not concerned with the complex electromagnetic field formed at the image plane, but rather the intensity of that field. Expressing Equation 2.10 in terms of image intensity yields

$$I_i(x_i, y_i) = \iint_{-\infty}^{\infty} |\mathbf{h}(x_i - x'_o, y_i - x'_o)|^2 \, I_g(x'_o, y'_o) dx'_o dy'_o, \tag{2.18}$$

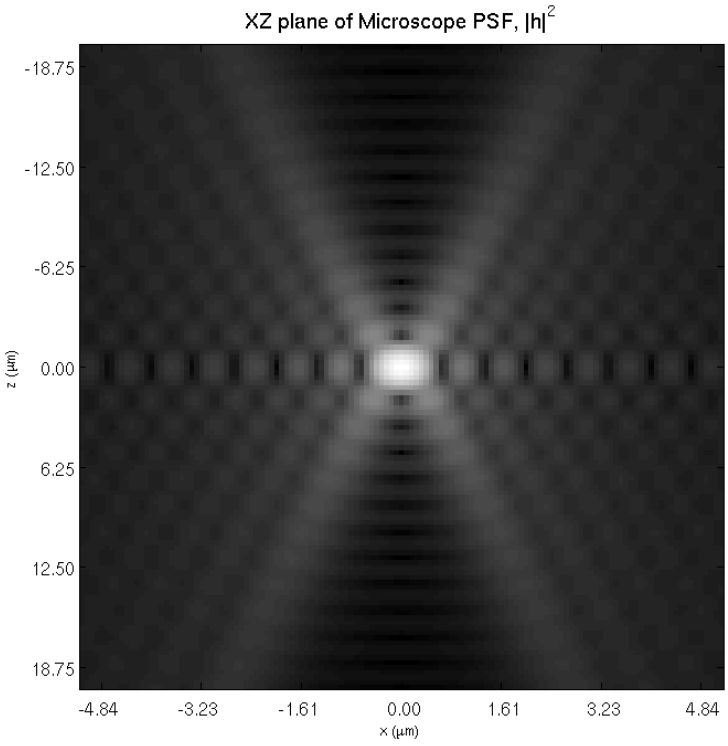


Figure 2.6: The XZ plane of the microscope PSF shows the effect of focus at different distances from the optimal focus plane ($z = 0$). Low intensity values are emphasized in this image using $I_o = \sqrt[4]{I_i}$.

where $I_g(x'_o, y'_o)$ is the image predicted by geometric optics,

$$I_g(x'_o, y'_o) = \left| \frac{1}{M} \mathbf{U}_o \left(\frac{x'_o}{M}, \frac{y'_o}{M} \right) \right|^2, \tag{2.19}$$

which is the complex field form of Equation 2.4. The normalized frequency representation of $|\mathbf{h}|^2$ is known as the optical transfer function (OTF) of the system

$$\mathfrak{H}(f_X, f_Y) = \frac{\mathcal{F}\{|\mathbf{h}(x_i, y_i)|^2\}}{\iint_{-\infty}^{\infty} |\mathbf{h}(x_i, y_i)|^2 dx_i dy_i}, \tag{2.20}$$

where $\mathcal{F}\{\mathbf{h}\}$ is the Fourier transform of $\mathbf{h}$. Combined with the convolution theorem [FvDFH97, Ch. 14], the OTF provides a convenient way to express Equation 2.18 as a product of Fourier transforms

$$\mathcal{I}_i(f_X, f_Y) = \mathfrak{H}(f_X, f_Y) \mathcal{I}_g(f_X, f_Y), \tag{2.21}$$

where $\mathcal{I} = \mathcal{F}\{I\}$, the Fourier transform of I.

In summary, Fourier optics analysis provides the tools necessary to predict the image of an object formed in a bright-field microscope. For a semitransparent object, the process requires finding the PSF of the objective lens for the distance from the object to the lens. Convolution of the magnified image of the object predicted by geometrical optics with the PSF yields the image produced by the optical system.

2.1.3 Resolution

Several empirical conventions are used to specify the resolution limit of a light microscope. The Rayleigh criterion specifies the minimum distance between two point sources that can be distinctly identified as

$$r = \frac{1.22\lambda}{2\mathrm{NA}}, \tag{2.22}$$

where λ is the wavelength of light used for the observation. This criterion follows directly from Equation 2.17, the focused PSF. As shown in Figure 2.5, at this distance, the peak at the center of one point's Airy disk coincides with the first minimum of the other point's Airy disk. For example a microscope equipped with an oil immersion objective lens with NA = 1.4 using monochromatic illumination with green light (λ = 550 nm) has a theoretical minimum resolvable distance between two points of

$$r = \frac{1.22 * 550 \text{ nm}}{2 * 1.4} \approx 240 \text{ nm}. \tag{2.23}$$

The Abbe limit is another convention that pertains to the ability to resolve repeating patterns in a specimen, such as a diffraction grating or muscle striations. Light waves admitted through neighboring gaps in the specimen diffracts and interferes, leading to interference patterns in the resulting wave front. In order to resolve a repeating structure, the undiffracted and 1^{st} order interference patterns must be collected by the objective lens. The distance between the orders of interference patterns depends on the pattern spacing; the smallest pattern that can be resolved is

$$d = \frac{\lambda}{2\text{NA}}. \tag{2.24}$$

Though the constant terms in Equations 2.22 and 2.24 are different, under either criterion the resolution limit is defined by the wavelength of light and the objective NA. Bounds on the resolution limit of a light microscope are placed by the wavelength of observable (visible or otherwise detectable) light and the index of refraction of practically useful immersion media. For visible light microscopy, one cannot improve significantly on the values used in Equation 2.23. Electron microscopy leverages the shorter wavelength behavior of electrons to examine specimens at a higher resolution. Structured illumination techniques, such as those discussed in Chapter 3, use multiple images with specific illumination patterns to computationally extend the resolution limit of fluores-

cence microscopy.

The NA also determines the depth of field in a microscope image. The sharpest image is formed of objects in the focal plane; image sharpness decreases as objects move away from the focal plane. Objects within a small distance of the focal plane can still be said to be in focus if their image is reasonably sharp. Smaller depths of field are desirable for fine optical sectioning of a specimen. With larger depths of field, more of the specimen contributes to the image, though not all components are equally focused. Like the resolution limit, defining the depth of field is, to a large degree, a matter of convention—for an image sensor with a minimum resolvable distance of e, one empirical convention specifies the depth of field as

$$d = \frac{\lambda n}{\mathrm{NA}^2} + \frac{ne}{M\mathrm{NA}}. \tag{2.25}$$

Intuitively, the angle of the cone of light admitted into the objective lens (as determined by the NA) corresponds to a cone of light that is focused on the image plane for each point on the object. As the object moves further from focus, a wider disk is projected onto the image plane. As higher NAs admit broader cones of light, this disk size is larger at a fixed distance from focus for higher NAs. The second term in Equation 2.25 depends on the resolvable distance of the sensor, e. For a regular grid array—as used in digital image acquisition, described next—this is empirically set to be the distance spanned by three sensor elements. Two objects can be distinguished if they illuminate two sensor elements with an non-illuminated element between them.

2.2 Image Acquisition

Section 2.1 discussed how bright-field microscopes form images; this section discusses how these images are digitized for computer storage and analysis. Fourier optics explains that the image formed by a microscope is a complex EM field. When projected onto a

surface and observed, a microscope image is a continuous function of *irradiance*—the power of light energy incident on a surface area [FP03, Ch. 4]. A charge-coupled device (CCD) image sensor is commonly used in image acquisition to sample the continuous distribution of light to produce a digital image.

A digital image is a collection of data organized in a structured, regular grid, for example a two-dimensional array of intensity values [FvDFH97, Ch. 14 & 17]. A digital image is composed of picture elements—*pixels*—each of which has a coordinate position and data value defined at that coordinate location. A digital image is therefore a discrete function over the domain of grid locations.

In the convention adapted for this work, pixel coordinates specify the center of a pixel and the pixel value is strictly defined only at the center of pixels. This work is concerned with the analysis of time sequences of two-dimensional (2D) images. Other types of images may be generated through this analysis, for example 2D vector images that describe motion fields.

2.2.1 Charge-Coupled Devices

The image formed by a CCD is a sampled representation of a continuous distribution of light. A CCD is composed of a regular grid of photosensitive elements that operate under the principles of the photoelectric effect. When light is incident on a metal, photons may interact with electrons in the metal. A photon can transfer its energy to an electron, possibly ejecting the electron from the atom to which it is bound. A freed photoelectron can be trapped in a potential energy well associated with a photosensor site on a CCD.

A CCD captures freed photoelectrons over the exposure time period [HK94]. At the end of the exposure, the electrons in each potential well are transfered sequentially to a readout register where a voltage signal proportional to the number of photoelectrons captured is created. The analog sequence of voltage signals is converted to a digital

stream by an analog-to-digital converter (ADC). This stream of digital pixel values is transferred from the camera to a storage device or computer for storage and analysis.

If the purpose of a CCD is to measure photoelectrons created by a distribution of light intensity on a surface, the ideal value, $I(x, y)$, recorded by a camera for an individual sensor element would be proportional to the number of photons incident on the sensor integrated over the exposure period, $E(x, y)$:

$$I(x, y) = kE(x, y), \tag{2.26}$$

where k is the device-specific constant of proportionality. Of course, it is impossible to accurately measure the incident intensity function $E(x, y)$ exactly—several types of noise corrupt the image acquisition process [HK94, TRK01, FP03]. Understanding the character of these noise sources is important for accurate image analysis and simulation tasks, such as those described in this dissertation. Below I construct the CCD imaging model used by Healey and Kondepudy while explaining each source of noise.

The arrival of photons at a sensor site is a Poisson-distributed random process [HK94] that leads to the random effect known as *shot noise*. The photoelectron count at a site is therefore a sample of a Poisson random variable, N_s with variance equal to the ideal number of photoelectrons collected:

$$I(x, y) = N_s(kE(x, y)). \tag{2.27}$$

Healey and Kondepudy represent shot noise with a Poisson distribution that is shifted to have zero mean, implicitly making the following manipulation:

$$\begin{aligned} I(x, y) &= kE(x, y) + (N_s(kE(x, y)) - kE(x, y)) && (2.28) \\ &= kE(x, y) + N'_s. && (2.29) \end{aligned}$$

Shifting the shot noise to have zero mean affords a convenience to the analysis that follows: when all noise sources are zero-mean and additive, the expected value of a pixel's intensity can be obtained by averaging many observations. For this reason, I adopt this notation here.

Fixed pattern noise arises from different physical sensors having slightly different sizes and quantum efficiencies, therefore counting a different number of photons for the same incident light intensity. The fixed pattern noise can be incorporated by the scaling factor k in Equation 2.26; $k(x, y)$ has a normal distribution with a mean of 1 and variance σ_k^2:

$$I(x, y) = k(x, y)E(x, y) + N'_s(x, y). \tag{2.30}$$

Thermal energy in the image sensor can also free electrons that become trapped in the potential energy wells associated with a CCD sensor element. These *dark current* electrons are indistinguishable from photoelectrons and so contribute an additional source of noise in the image measurement:

$$I(x, y) = k(x, y)E(x, y) + N'_s(x, y) + N_t(x, y). \tag{2.31}$$

The dark current increases with exposure time and sensor temperature.

The circuitry that reads out the potentials stored at each sensor site contributes a small, zero-mean, Gaussian-distributed noise, modeled by N_r:

$$I(x, y) = k(x, y)E(x, y) + N'_s(x, y) + N_t(x, y) + N_r(x, y). \tag{2.32}$$

Equation 2.32 represents the number of electrons collected as the input signal to the ADC process. The ADC applies a gain, A, that converts this electron count in to an intensity level in arbitrary units, often called "counts". Quantization of the signal by

the ADC involves an additional noise term, N_q:

$$I(x,y) = (k(x,y)E(x,y) + N'_s(x,y) + N_t(x,y) + N_r(x,y))A + N_q(x,y). \tag{2.33}$$

Healey and Kondepudy argue that N_q is approximately zero-mean and uniformly distributed over $[-\frac{q}{2}, \frac{q}{2}]$, where q is the quantization step size. This equation represents the actual image $I(x,y)$ recorded by a CCD in response to an ideal photoelectron count function $E(x,y)$. Rearranging a few terms yields a convenient form:

$$\begin{aligned} I(x,y) &= \mu(x,y) + N(x,y), \\ \mu(x,y) &= k(x,y)E(x,y)A + \mu_t(x,y)A, \\ N(x,y) &= N'_s(x,y)A + N_r(x,y)A + N_q(x,y). \end{aligned} \tag{2.34}$$

Here, $\mu(x,y)$ is the expected value of $I(x,y)$, $\mu_t(x,y)$ is the expected value of the dark current, $N_t(x,y)$, which is the only non-zero-mean noise source, and $N(x,y)$ is a zero-mean random variable that encapsulates all other temporal noise sources.

In summary, the formation of digital bright-field microscopy images consists of the following key components:

1. Köhler illumination ensures that a uniform light field illuminates the specimen.

2. The microscope's objective gathers light transmitted through the specimen, forming an image of the specimen convolved with the microscope's PSF on the image sensor.

3. The image sensor integrates the light irradiance distribution function over an exposure time by counting photoelectrons collected in a regular grid of potential energy wells.

4. The camera scans the charges collected at each sensor site and digitizes these

sensor values.

5. A computer saves the collected digital image for further image processing.

2.3 Camera Calibration

Working with the CCD image acquisition model of Equation 2.33, it is possible to characterize and calibrate the image sensor in a camera attached to a bright-field microscope. I close this chapter with a practical camera calibration procedure, based on the combined wisdom provided by several camera calibration procedures [HK94, TV98, TRK01, MMG05]. To illustrate the procedure, I present calibration results obtained from a Pulnix TM-6710CL progressive scan, monochrome, 8-bit camera, which records 648×484 pixel^2 images at a maximum rate of 120 fps. The camera is attached to an inverted Nikon Eclipse TE2000-E microscope with an optical train modified to accommodate other optical components such that strict Köhler illumination is not possible: the microscope is missing a collector lens and aperture diaphragm.

2.3.1 Dark Current Estimation

The thermal noise in an acquired image introduces a positive offset at each pixel. Taking a series of dark images enables estimating the expected value of this offset. A series of dark images is obtained by turning off the microscope lamp and covering the camera aperture to block all light from the image sensor. If the camera aperture cannot be blocked, the room lights can be turned off to make the acquisition environment as dark as possible. A series of n_d images are then acquired using the exposure settings that will be used in future image acquisition.

From the model of Equation 2.34, each dark image is

$$I_d(x, y) = \mu_t(x, y)A + N(x, y), \tag{2.35}$$

and the mean of n_d dark images has expected value $\mu_t(x,y)A$ and variance $\frac{\sigma_N^2}{n_d}$:

$$\hat{I}_d(x,y) = \frac{1}{n_d}\sum_{i=1}^{n_d} I_{di}(x,y) = \mu_t(x,y)A. \tag{2.36}$$

The mean dark image, $\hat{I}_d(x,y)$, can be subtracted from any image taken with the same exposure settings to obtain a dark-current-corrected image. For the Pulnix camera operating at 120 fps, $\hat{I}_d(x,y)$ has a mean value of 3.252×10^{-6} counts and a mean variance, $\sigma_N^2 = 3.267 \times 10^{-6}$ counts^2. The dark current offset for this camera can therefore reasonably be neglected.

2.3.2 Flat-field and Fixed-pattern Noise Estimation

Despite the best efforts to align the optical elements in a microscope, the illumination field provided even by Köhler alignment is never truly uniform. Each sensor in the CCD receives a different intensity of light, so the effect of nonuniform illumination is a per-pixel scaling of the photoelectron count. In the CCD model of Equation 2.33, the parameter $k(x,y)$ can therefore incorporate both fixed-pattern noise and nonuniform illumination effects. Accounting for nonuniform illumination is known as *flat-field correction*. To calibrate a flat-field image, the microscope is adjusted for uniform Köhler illumination, and a clean, blank slide and cover slip are placed on the stage. A series of n_f images of this approximately uniform field is acquired at a single lamp intensity. The mean of these images, $\hat{I}_f(x,y)$, has the expected value

$$\hat{I}_f(x,y) = k(x,y)E(x,y)A + \mu_t(x,y)A. \tag{2.37}$$

The term $\mu_t(x,y)A$ is estimated by dark current calibration as described in Section 2.3.1, and can be subtracted to provide

$$\hat{I}_f(x,y) - \hat{I}_d(x,y) = k(x,y)E(x,y)A. \tag{2.38}$$

Because $k(x,y)$ has expected value 1, the mean value of this resulting image is an estimate of EA, the signal that an ideal sensor would register everywhere for a uniform illumination. Thus,

$$\begin{aligned} \mu_f &= \frac{1}{nm}\sum_{x=1}^{n}\sum_{y=1}^{m}\hat{I}_f - \hat{I}_d = EA, \\ F(x,y) &= \frac{\hat{I}_f(x,y) - \hat{I}_d(x,y)}{\mu_f} \\ &= \frac{k(x,y)E(x,y)A}{\mu_f} = k(x,y). \end{aligned} \tag{2.39}$$

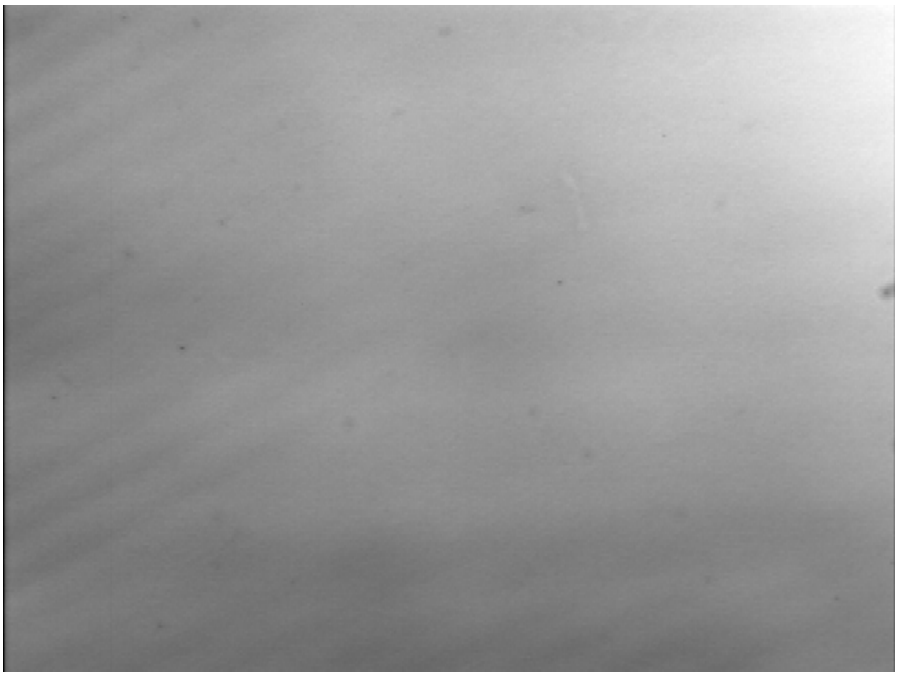


Figure 2.7: A flat-field calibration image obtained from the Zeiss microscope and Pulnix camera.

Figure 2.7 shows a flat-field calibration image, $F(x,y)$, obtained from 600 blank

field images captured at 120 fps. The flat-field image values fall within the range $[0.574 \ldots 1.288]$ with variance 6.84×10^{-3}. This image reveals that the lamp is not properly aligned with the center of the camera. The diagonal banding lines are fixed-pattern aberrations probably arising from electrical interference from some other component in the microscope system. The dark splotches in the image are from dust on the image sensor or dirt on the slide. There are several columns on the left side of the image where the signal drops considerably; this indicates that the image sensor may not be properly seated in the camera enclosure.

The calibration images $F(x, y)$ and $\hat{I}_d(x, y)$ provide a means to correct any image, $I(x, y)$, taken under the same conditions as the calibration images, including frame rate, lamp level, optical alignment, and slide and cover slip thickness. From Equations 2.33, 2.36, and 2.39, a corrected image is

$$\begin{aligned} I_c(x, y) &= \frac{I(x, y) - \hat{I}_d(x, y)}{F(x, y)} \\ &= E(x, y)A + N_s(x, y)A + N_r(x, y)A + N_q(x, y). \end{aligned} \tag{2.40}$$

2.3.3 ADC Gain Estimation

The gain applied by the ADC can be approximated using a number of images of a uniform field taken at different illuminations. The variation in uniform field images (or small regions of interest) should be due to the noise introduced by the CCD image acquisition and not from illumination inconsistencies. It is not necessary to use flat-field corrected images for this process provided that a reasonably uniform region can be found within the calibration images.

To obtain calibration images for gain estimation, the microscope is adjusted for uniform Köhler illumination, and a clean, blank slide and cover slip are placed on the stage. The lamp intensity is set to a high level that does not cause clipping (clamping to the maximum intensity output by the camera) of the acquired images. Pairs of

images are then captured at different illumination levels, ideally by inserting several neutral density filters into the optical train. Alternatively, a large number of images are acquired while the lamp is slowly dimmed, and any consecutively captured pair of images is considered to be obtained at the same intensity setting. This approach is not as robust as using neutral density filters because dimming the lamp changes the spectrum of light emitted, and the CCD sensor response depends on the frequency of incident light.

Given a pair of images $I_1(x, y)$ and $I_2(x, y)$ captured at the same intensity, the summed image, $I_{\Sigma 12}(x, y) = I_1(x, y) + I_2(x, y)$, has expected value

$$\begin{aligned} \mu_{\Sigma 12} &= \frac{1}{nm} \sum_{x=1}^{n} \sum_{y=1}^{m} I_{\Sigma 12}(x, y) \\ &= 2\mu(x, y) = 2(k(x, y)E(x, y)A + \mu_t(x, y)A), \end{aligned} \tag{2.41}$$

where each image has dimensions $n \times m$. The difference image, $I_{\Delta 12}(x, y) = I_1(x, y) - I_2(x, y)$, has variance

$$\text{var}[I_{\Delta 12}] = 2\sigma_N^2. \tag{2.42}$$

At a high intensity level, Healey and Kondepudy show that

$$\sigma_N^2 = A\mu + \sigma_c^2, \tag{2.43}$$

where σ_c^2 is the variance of $N_r A + N_q$, the temporal noise that does not depend on the electron count. Given n_i pairs of images taken at multiple intensity levels, linear least-squares regression can be used to find maximum likelihood estimates for the gain, A and constant noise variance, σ_c^2.

Figure 2.8 shows the results of gain calibration on the Pulnix camera capturing images at 120 fps. Images are obtained using the lamp dimming approach because the microscope is not equipped with neutral density filters. With the high frame rate of

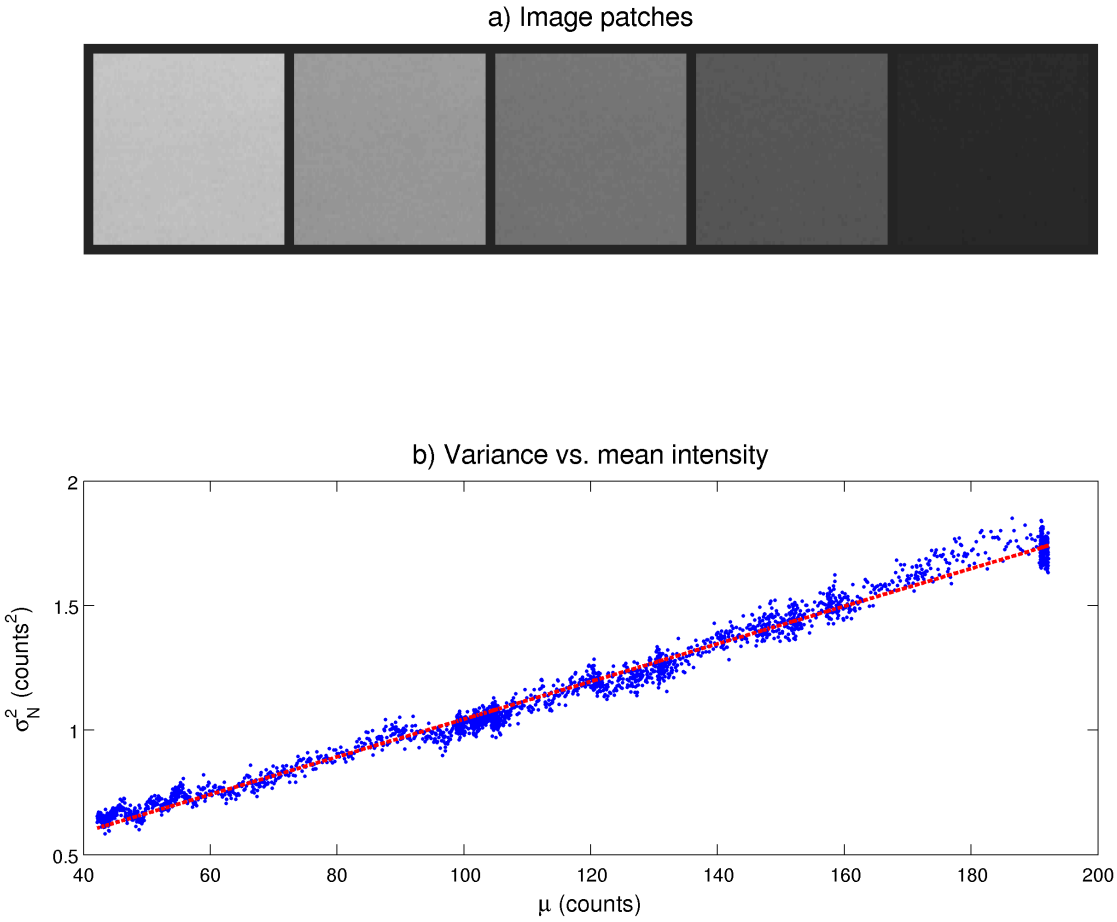


Figure 2.8: Gain calibration for a CCD image sensor involves acquiring a series of paired images at multiple intensity settings. a) A selection uniformly illuminated image patches obtained by slowly dimming the lamp while capturing images at 120 fps. b) The gain calibration obtained by plotting σ_N^2 vs. μ for all consecutive image pairs. The dashed line shows the fit of Equation 2.43 computed with linear least-squares regression.

the Pulnix, this approach produces calibration estimates based on a large number of samples, 2399 image pairs in this case. The mean value and variance of the sum and difference images are computed from a 100×100 pixel^2 region of interest that is selected for having approximately uniform illumination and no fixed aberrations (*e.g.* dirt on the image sensor or slide). Linear regression reveals a sensor gain of $A = 7.56 \times 10^{-3}$ counts per electron (130 electrons / count) and a constant noise variance of $\sigma_c^2 = 0.289$ counts^2. This gain estimate provides a basis for the noise model used in simulating microscopy images in Chapters 5 and 6.

Chapter 3

Related Work

This work is concerned with forming models of motion and structure in multiple layer imaging scenarios. Biological microscopy provides numerous examples of multiple layer imaging. Biological specimens are usually semitransparent, and the microscope optics impose a non-zero depth of field. For this reason, the images formed in a microscope are a composition of multiple scene elements.

The unconstrained motion problem—determining the motions that explain intensity differences between images—is ill-posed, as any mapping that assigns all pixels in one frame with a particular intensity to a single pixel in the next frame with the same intensity is "valid." All motion analysis research is therefore based on some set of assumptions about the imaging system and the observed scenes. The computation of motion for rigid and deformable opaque objects is well studied. The scenarios for which computing multiple layer motion has been solved remains more limited. The human visual system, however, is able to comprehend multiple layer motions in many more situations—although it is prone to optical illusions caused by balanced opposing motion signals [QAA94]. The human visual system provides a proof by example that computational analysis of multiple layer motion remains a worthy area of active research.

3.1 Computing Motion from Images

Determining motion from images is a fundamental problem in image analysis, and serves as the basis of many other image analysis tasks, such as structure estimation, object segmentation, and video compression [TV98, Ch. 8]. A significant portion of motion analysis research relies on a brightness constancy assumption: the image of a scene object maintains constant intensity over time, at least for small time steps. This assumption has the following properties:

- Objects that emit light do not change their emittance; objects that reflect light do not change their reflectance; objects do not appear or disappear.
- The illumination of the scene remains constant.
- Surfaces do not occlude each other.
- Surfaces are Lambertian—radiance is independent of viewing direction.

In 1981, Lucas and Kanade proposed an image registration method that is the basis for many feature-based motion techniques [LK81]. This method relates the motion in the direction of the image gradient to the temporal difference between images. The method can be applied to images as a whole or to image patches locally, with the assertion in either case being that motion over the observation window is constant and can be described with a parametric model. Applied to small image patches, this technique provides an estimate of *optical flow*, the projection of scene motion onto the image plane. The computation is often constrained to features with high gradients in both directions, as these are locations where the local tracking is likely to succeed [ST94, SMB00]. Applied over the whole image plane, this technique provides a single parametric model of image motion for all image locations.

Horn and Schunck use the same image gradient constraint in an energy minimization problem and introduce a regularization term that enforces smoothness of the computed

velocity field [HS81]. The result is a technique that provides a dense estimate of optical flow through an iterative solution. Bruhn *et al.* embed the local motion estimate of Lucas and Kandade into the energy minimization function of Horn and Schunck to provide a combined local and global optical flow computation [BWS05]. This approach provides a balance between using local motion information where it is available and filling in smooth flow fields in areas of low image contrast. An implementation of this combined local and global optical flow computation appears in ImageTracker, a software package described in Appendix A.

In multiple layer imaging, gradient-based techniques fail because image gradients are not preserved in overlapping transparency layers. In other words, an edge appearing in one image can be due to a feature in any component layer or the alignment of features in multiple layers. As features move in each layer independently, edges in the resultant image will appear and disappear. The two remaining popular motion computation techniques are explored in this dissertation—frequency-based methods and correlation-based methods [BB95]. These techniques share a low-dimensional motion model that is robust to information from problematic local areas.

Phase correlation is a frequency-based motion computation that leverages the *Fourier shift property*—translation in the spatial domain is equivalent to a phase shift in the frequency domain. Phase correlation readily provides whole-pixel translations between two images, and much recent research on phase correlation has addressed extending the technique beyond translation transforms and to subpixel accuracy. Reddy and Chatterji explain how to extend phase correlation to recover rotation and scaling in addition to translation using a coordinate transform [RC96]. Foroosh *et al.* show how to obtain accurate subpixel translations from downsampled images [FZB02]. Hoge provided another subpixel registration technique that uses singular value decomposition to separate the phase estimation onto the coordinate axes [Hog03]. Takita *et al.* use low pass frequency domain weighting function to restrict the spectrum of frequencies used in phase correla-

tion [TAS+03]. They further fit the peak indicating translation in the spatial domain to the analytical functions predicted by the weighting. Recently, Ho and Goecke proposed a phase correlation method that computes parametric scale, rotation, and translation transforms within image blocks using a subpixel function fitting [HG08]. They further smooth the flow fields across image blocks using a weighted vector averaging. In Chapter 5, I propose a method to determine subpixel phase correlation based on phase estimation of individual frequencies known to be dominant in the images.

Correlation-based methods compute parametric image motion by finding the transform parameters that maximize a spatial comparison metric of the intensities in two images [BB95]. Like gradient-based image registration, this matching approach can be applied to small image patches or the entire image frame. Computing the correlation for all possible transforms between images is computationally expensive, so practical methods depend on coarse-to-fine refinements and optimization strategies to drive the parameter search towards the optimal solution [ISNC05]. In Chapter 5, I propose a method to perform correlation-based tracking using a pattern known to exist in an image layer and using a model of microscope image formation to predict layer appearance.

3.2 Computing Transparency from Images

Another active area of image analysis research involves estimating the transparency of objects from images. This problem is the inverse of creating composite images from individual image layers, which is used often in image editing and video production. Porter and Duff present a mathematical framework for creating composite images from distinct image elements [PD84]. An image element is an independently-rendered image with an associated coverage mask, or matte, that determines where that image element should appear in the composite image. The matte information is specified as an *alpha channel* value that determines the weight the image element should contribute at each pixel.

The authors demonstrate that the alpha value for an image element can equivalently represent either the transparency of a solid object covering an entire pixel or the partial pixel coverage of an opaque object. The framework proposed in this paper is used throughout the transparency analysis literature. The operator relevant to most multiple layer imaging in microscopy is the *over* operator, which specifies a pixel value as a linear blending between foreground and background layers.

Smith and Blinn demonstrate that the transmission matte of a foreground object viewed in front of a flat background can be determined using two images of the same scene with different backgrounds, but not using a single background, as often used in the film industry [SB96]. Recovering the transparency of image layers in settings where swapping out backgrounds is not possible, then, depends on other assumptions about the scene composition. Ruzon and Tomasi and Chuang *et al.* present methods for segmenting a foreground object with complex boundaries based on statistical distributions of pixels known with certainty (based on user interaction) to belong to the foreground and background layers [RT00, CCSS01]. Chuang *et al.* extend these techniques to moving scenes using optical flow to estimate the locations of each region over the video given user-specified regions at key frames [CAC+02]. Yeung *et al.* present a method to extract smooth transparent foreground layers from backgrounds that have sufficient gradients [YWT08]

In Chapter 4, I present a technique that estimates the transparency of a non-moving layer in a video composed of an arbitrary number of moving layers using no user interaction or optical flow information. The key assumptions I rely on are that the background layer remains stationary and the foreground layers have sufficient motion over the video.

3.3 Multiple Layer Motion Analysis

Multiple layer motion analysis combines the domains of motion and transparency estimation. One aspect of this problem is to determine multiple motion models for each image location that explain the optical flow of independent image layers. Another aspect of this problem is to determine the composition of image layers in the presence of mixed motion information from multiple layers. As always, each contribution makes assumptions about the scene and imaging that enable extracting information from the images.

Shizawa and Mase present a method for estimating multiple motions using frequency-domain analysis [SM90]. Their approach is an adaptation of phase correlation in which they find two planes through frequency space that represent the motion of two rigid image layers undergoing translations. As with phase correlation, this approach requires broad spectral support for each of the image layers.

Irani *et al.* describe a method for computing the motion of multiple occluding or transparent rigid objects [IRP94]. This approach first finds a dominant moving object and its region of support using multiscale registration followed by moving object segmentation. The dominant component of motion is removed from the video and the process is repeated to find the second component of motion. This technique relies on wide spatial motion support and rigidity of the moving objects.

Black and Anandan incorporate a robust estimation framework into optical flow computations in order to handle gross errors introduced by multiple motion conditions [BA96]. The authors demonstrate that three common optical flow approaches (spatial correlation and the local and global gradient techniques) are least-squares minimization problems. The problem with least-squares estimation is that outliers are weighted heavily by the penalty function. Least-squares is robust when the data follows a normal distribution, but when outliers do not fit a normal distribution the model becomes skewed to fit these gross outliers. In multiple motion situations, there may be

a dominant motion present, but any data from a secondary motion would manifest as gross outliers. Black and Anandan replace the quadratic penalty functions with more general ρ-functions to apply robust estimation techniques to motion computation. The goal of choosing a robust estimation penalty function is to reduce the effect that gross outliers have in model fitting.

Szeliski *et al.* address the problem of recovering image layer composition from a sequence using a constrained least-squares solution assuming the transforms of each layer are known [SAA00]. When the transforms are not known *a priori*, the authors propose an iterative solution to recovering both the motion and layer composition by first recovering the dominant motion. An image mosaic is formed using this motion information, which provides an upper bound of intensities for the dominant layer and a lower bound of intensities for the remaining motion layers. An iterative refinement improves the estimates of layer composition until the constrained least-squares approach can be applied.

Sarel and Irani present a method to separate two layers of motion from a video where one motion is repetitive [SI05]. Considering a video as a spatio-temporal volume, they formulate a registration process to bring into alignment two arbitrary lengths of video. A coarse registration is used to find the periodicity within the image sequence by comparing a spatio-temporal volume with all segments in the video. This leads to alignment of all occurrences of the repeated motion, followed by a local registration refinement. The repetitive motion layer is separated from the non-repetitive layer by median-filtering the spatio-temporal derivatives of the registered video volume. The second motion layer is extracted from the original video by finding the blending coefficients that minimize the correlation between the first and second motion layer.

Gai *et al.* present a method for determining the motion and composition of multiple rigid image layers undergoing translation using gradient information [GSZ08]. The authors assume that the edge information in each of the layers is sparse, the information

in each of the layers is statistically independent, and each layer is non-periodic. These three assumptions together enable the shifts for all layers to be found by correlation of gradient magnitude images.

3.4 Structured Illumination

Structured illumination is an image analysis technique in which a scene is illuminated by a nonuniform light pattern of known structure. The appearance of the illumination pattern in images of the scene can be used to discern more information about the scene—often geometry. The general technique relies on using calibrated projector and camera systems to illuminate and image objects in a working volume. Correspondences between points in the images and points in the projected pattern provide scene geometry via triangulation.

Vuylsteke and Oosterlinck propose an binary encoding pattern that provides unique spatial locations from observations of a small image window [VO90]. This enables recovering range maps from single images, which is a requirement for scenes containing moving objects. Gärtner *et al.* propose an efficient encoding pattern that enables determining pattern locations from multiple observations of a static scene [GLT96]. In this case, the intensity encoding that identifies a location is time-modulated. Kang *et al.* use frequency-modulated sinusoidal patterns to enhance the discrimination between points identified in images from multiple cameras, improving the stereo correspondence task [KWZK95]. Noguchi and Nayar and Nayar *et al.* use a checkerboard pattern projected through a microscope onto a surface to determine shape from focus of the pattern [NN94, NWN95]. In their work, a collection of images taken at different z depths is analyzed to find the best focus at each image location based on response to a Laplacian focus operator.

In Chapter 5, I propose a new structured illumination technique that provides lateral

tracking information for microscopy. The structured illumination pattern I use introduces a semitransparent layer to the acquired images. The fixed pattern is inserted into the light path of the bright-field microscope and remains fixed to the reference frame of the microscope stage. In Chapter 6, I demonstrate how to obtain depth information from analysis of the pattern's defocus.

3.5 Computational Microscopy

Another active field of research is computational microscopy. The unifying premise of research in this field is that if microscopy images are to be analyzed by a computer, the optimal images that enable this analysis may differ significantly from optimal images for analysis by humans.

A flurry of papers have been published recently on using structured illumination to enhance the lateral resolution of fluorescence microscopy [GAS00, FKS00, RHM+03, HB06, Car08, TRL+08]. These all rely on the same principles, with the clearest explanation provided by Carlton. The approach leverages the *convolution theorem*—convolution in the spatial domain is equivalent to multiplication in the frequency domain—and the fact that lenses perform Fourier transforms [Goo68, BW75]. A structured illumination pattern is used to excite fluorophores such that an individual frequency-domain image of the emitted light contains the sum of magnitudes from frequencies that fall within the Abbe limit, but also from frequencies that exceed the Abbe limit. Shifting the phase of the structured illumination pattern to capture multiple images enables recovering the individual magnitudes at all frequencies. The Fourier transform of the reconstructed frequencies provides an image with resolution beyond the Abbe limit.

Levoy *et al.* propose another computational microscopy technique that trades lateral resolution for obtaining depth information from a single image [LNA+06]. The technique records a light field by placing a lens array at the intermediate image plane which forms

multiple images of the specimen from slightly different perspectives. Combining the information in these subimages enables constructing multiple perspective views of the specimen or constructing a sequence of images focused at different depths through the specimen, all from a single image.

The structured illumination microscopy technique I present in Chapters 5 and 6 contributes to the field of computational microscopy. The images acquired have nonuniform illumination, making them non-optimal for human observers. However, the extra information in the images enables obtaining 3D tracking information through computer analysis. The occlusion removal technique presented in Chapter 4 provides a method to remove the structured pattern layer from the images.

Chapter 4

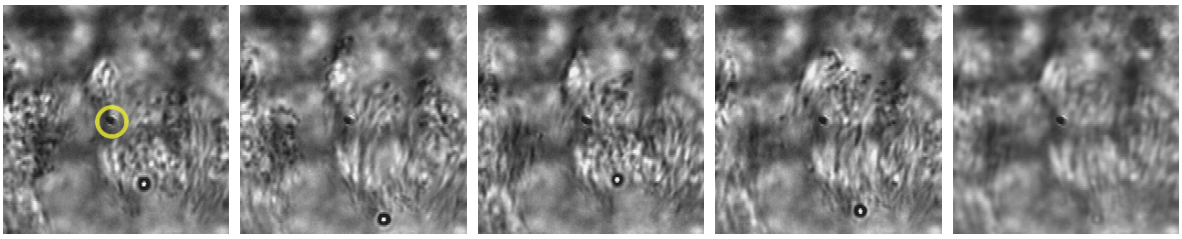
Removal of Stationary Semitransparent Image Layers

A significant component of the work in this chapter was presented at the 12^{th} International Conference on Computer Analysis of Images and Patterns (CAIP) in 2007 [ET07].

Video microscopy figures prominently in many fields of research, such as biology, pathology, materials science, and physics. In some situations, an experiment imposes constraints that introduce undesirable artifacts into the captured images. One such scenario arises when using long working distance lenses in microscopy of moving specimens. A long working distance lens, which has a smaller numerical aperture (NA) than a standard lens at a particular magnification, enables focusing further into a specimen, but also increases the depth of field. Consequently, a thicker slab of the specimen is visible in the final image.

Figure 4.1 shows several frames from a video of beating cilia on epithelial lung cells. A microbead in the bottom third of the image frame is attached to one clump of beating cilia. The inverted microscope used to create these images focuses through the substrate and cell bodies onto the cilia layer, and these components modulate the image of the moving cilia. This appears as a blurry constant background behind the beating cilia.

Other artifacts, such as debris on the image sensor (also seen in Figure 4.1), slide, or cover slip, also contribute to the final image. Microscope cleaning and alignment are



(a) (b) (c) (d) (e)

Figure 4.1: Ciliated epithelial lung cells move a bead in bright-field microscopy observed with a long working distance lens. The spot circled in the first frame is from dirt on the image sensor. a-d) Frames from the original captured video. e) The mean of 61 frames from the video. Original images from David Hill.

crucial steps for obtaining the clearest images. The reality, however, is that a great deal of microscopy data is collected with suboptimal alignment and cleaning. The data collected is interesting, but corrupted by optical impurities.

With these considerations, a typical microscopy video consists of one or more moving "foreground" specimen objects and static "background" impurities. (The use of foreground and background does not imply which object is physically in front of another; I am choosing the convention that assigns the object of interest, the specimen, to the foreground.) The background impurities will obscure the foreground specimen as the specimen moves across the image plane. An occluding object and specimen become darker where they overlap, but rarely does an object absorb all light.

In this chapter, I present two methods for repairing video microscopy data that suffers from stationary partial occlusions. These approaches use the simplified geometric model of image formation in the bright-field microscope from Chapter 2, and rely on motion to enable separating the video into moving and stationary components. In addition to improving the visual clarity of microscopy videos, these techniques are useful as a preprocessing step for other image analysis tasks, such as optical flow computation. Compared to background subtraction, these approaches have the important characteristic of preserving the relative intensity relationships between image sensor locations, which is necessary for photometry. The techniques presented here each have strengths

and weaknesses; I will compare the effectiveness of these methods, and describe which method suits different scenarios.

Section 4.1 discusses prior research related to occlusion removal; Section 4.2 discusses the principles behind the general occlusion removal technique; Section 4.3 discusses algorithm implementation details; Section 4.4 quantitatively and qualitatively evaluates the repair methods; Section 4.5 summarizes the findings presented here.

4.1 Related Work

As discussed in Section 2.2, the details of image acquisition, such as noise and sensor calibration, are concerns in any imaging application [HK94, TRK01]. Sensor calibration considers image formation from the image sensor onwards, this work concentrates on image formation in the bright-field microscope before the image sensor. These methods, therefore, are complementary. Specifically, the linear response of the sensor should be validated before applying these techniques, as outlined in Section 2.3. The image sensors usually used on research microscopes have a linear—or nearly linear—response.

Background removal techniques for microscopy include background subtraction and flat-fielding [IS97, Ch. 12]. These approaches use specimen-free calibration images to remove artifacts due to illumination variations by frame subtraction or division. The approach presented here depends on the presence of a moving specimen but works on existing microscopy data without calibration. Hill *et al.* use a moving average image for background subtraction to improve the visual clarity of DIC microscopy images [HPBH04]. The methods presented here perform this function and at the same time preserve the image intensity relationships required for photometry.

Schechner and Nayar compute the transmission map of a filter with spatially-varying transmittance [SN03]. Their initial estimate is based on image statistics, and Section 4.2.2 demonstrates how to improve this estimate. Their robust estimation relies on

image registration with a rigid-body constraint, a condition that is rarely satisfied in biological microscopy. Shizawa's multiple optical flow estimation technique also relies on rigid-body assumptions [SM90].

A significant body of computer vision research focuses on macroscopic occlusion detection and processing [SLKS05, JWTT04]. In the macroscopic case, occlusion is typically complete. Complete occlusion constitutes a nonlinearity in image formation that is manifested by a complete loss of signal in the captured image. In microscopy, however, occluding objects do not block light completely. In the multiplicative image model, each object absorbs a constant proportion of light, and the signal from the occluded object transmits through the occlusion. Throughout this chapter the term "occlusion" refers to a partial occlusion as encountered in bright-field microscopy, unless otherwise specified.

4.2 Theory

In a bright-field microscope, a light source is tuned to uniformly illuminate a specimen across the field of view, as discussed in Section 2.1. A series of lens assemblies (*e.g.* condenser, objective, and ocular) focus the light transmitted through the specimen onto an image sensor [IS97, Ch. 2].

Light interacting with any object in the optical path is either transmitted, absorbed, or diffracted. The first-order image formation model assumed here ignores diffraction effects, the consequences of which are discussed below. Recall from Chapter 2 that the image of a planar object predicted by geometrical optics is given by Equation 2.4:

$$I(x, y) = \frac{1}{M} LT\left(\frac{x}{M}, \frac{y}{M}\right). \tag{4.1}$$

The magnification of the objective lens, M, specifies a scaling of the image and object coordinates, and can be disregarded in this analysis. The lamp intensity, L, however,

is considered to vary across the image plane. The image formed on the image sensor, $I(x, y)$ is therefore simply the product of the light field incident on the specimen, $L(x, y)$, and the transmission through the specimen, $T(x, y)$:

$$I(x, y) = L(x, y)T(x, y), \qquad T(x, y) := \{0 \leq T(x, y) \leq 1, \forall(x, y)\}. \tag{4.2}$$

Equation 4.2 constitutes a multiplicative light model for the bright-field microscope. For brevity in the analysis that follows, I replace the coordinates (x, y) with vector notation, $\mathbf{x}$, to specify 2D positions.

Considered over time, the transmission map, $T(\mathbf{x}, t)$, is a composite of transmission coefficients from stationary and moving objects. The time varying transmission map can be decomposed into a stationary transmission map, $T_c(\mathbf{x})$, and a time-varying specimen transmission, $T_s(\mathbf{x}, t)$. The image formed on the sensor is then

$$I(\mathbf{x}, t) = L(\mathbf{x})T_c(\mathbf{x})T_s(\mathbf{x}, t) \ . \tag{4.3}$$

Given the stationary transmission map, a repaired video, without the presence of stationary occlusions, is obtained by

$$I_r(\mathbf{x}, t) = \frac{I(\mathbf{x}, t)}{T_c(\mathbf{x})} = L(\mathbf{x})T_s(\mathbf{x}, t) \ . \tag{4.4}$$

The image repair problem therefore reduces to finding an accurate estimate of the stationary transmission map, $T_c(\mathbf{x})$. The following sections present two methods for estimating T_c.

The image sequence $I(\mathbf{x}, t)$ contains noise in every intensity measurement, and the stationary transmission map amplifies this noise as well as the signal in Equation 4.4. The transmission map, therefore, also provides a measure of relative error in the repaired intensity values—the intensity value of a pixel that has high gain from the transmission

map will be less reliable than one with low gain.

As a consequence of ignoring diffraction effects, occlusions on the image sensor are modeled better than occlusions elsewhere in the optical path. Debris on lens elements are completely out of focus in a well-aligned microscope, so their effects are distributed across the image plane. Light diffracted from partially-focused, stationary debris, however, may add intensity to a local image region. Experimental results in Section 4.4.2 demonstrate that this simplification yields useful information in a broad range of microscopy data, despite assuming a diffraction-free image model.

4.2.1 Approach 1: A local transmission metric

Consider the mean intensity image from all frames in a microscopy video, as seen in Figure 4.1e. The mean intensity of a video with sufficient motion has low spatial variance, except in regions that contain stationary occlusions. The effect of motion on a mean intensity image is similar to the effect of a mean filter on a single image: the intensity from a single moving object is spread over many pixels in the mean image. Given sufficient uncorrelated motion, each point in the scene is imaged by all pixels within some neighborhood. The result is that the mean intensity image is approximately constant within local neighborhoods. For example, the predominant spatial variance in Figure 4.1e is due to stationary background objects. This observation leads to a method for estimating the stationary transmission map.

Assume that for a sequence of N images the mean intensity,

$$\overline{I}(\mathbf{x}) = \frac{1}{N} \sum_{t=1}^{N} I(\mathbf{x}, t), \tag{4.5}$$

is locally constant except where occluded by stationary background objects. If the occlusion regions are relatively small compared to the image plane, each occluded pixel will have within a small neighborhood a number of unoccluded pixels. The mean inten-

sity in this neighborhood will be greater than the mean intensity at the occluded pixel. Comparing the mean intensity of an occluded pixel to that of its unoccluded neighbors determines the amplification necessary to bring the occluded pixel in line with its neighbors. In a sense, the occluded pixel appears as impulse noise, where the value that should have appeared at that pixel has been replaced by another value due to the occlusion. One common method used to detect impulse noise is to compare a pixel value to the median of neighboring pixel values [GHCH05]. For a neighborhood Ω,

$$\overline{I_\Omega}(\mathbf{x}) = \text{Median}\left[\,\overline{I}(\mathbf{x}_i)\right], \qquad \{\mathbf{x}_i \,|\, \mathbf{x}_i \in \Omega\}. \tag{4.6}$$

The per-pixel mean intensity of Equation 4.5 provides a measure of how much light was transmitted to an image location while the local median of Equation 4.6 provides an estimate of how much light was transmitted to the pixels within a neighborhood. Knowing the expected size of occlusions in the image plane enables choosing a neighborhood size such that all neighborhoods have fewer than half of their pixels occluded. For this condition to be met for all neighborhoods, the occlusions present must be relatively small and sparsely distributed across the image plane. When such a neighborhood size can be chosen, however, an estimate of the stationary transmission map is

$$T_c(\mathbf{x}) = \frac{\overline{I}(\mathbf{x})}{\overline{I_\Omega}(\mathbf{x})}. \tag{4.7}$$

Equation 4.7 ensures that if a pixel has an intensity that is always lower relative to its neighbors, its intensity will be amplified after applying Equation 4.4. Similarly, if a pixel has an intensity that is always higher than its neighbors, its intensity will be diminished. Note that this latter case exposes a weakness in this approach applied to bright-field microscopy: an occluding object would never *amplify* the intensity at a pixel, only diminish it.

I will refer to the method of estimating the stationary transmission map using Equa-

tion 4.7 as the local transmission metric (LTM).

4.2.2 Approach 2: A global transmission metric

I next describe a framework for image intensity comparisons and use it to formulate an alternative estimate of the stationary transmission map. A stationary occlusion transmits a constant proportion of light at an image location. Foreground objects move over the occlusion throughout a video, so the intensity at that location changes over time. What remains constant, however, is the ratio of light transmitted at an occluded pixel compared to its neighbors. The transmission map estimate is constructed by forming a model of the stationary gradient of transmission at each pixel in the image plane.

The transmission image formation Equation 4.3, $I(\mathbf{x}, t) = L(\mathbf{x})T_c(\mathbf{x})T_s(\mathbf{x}, t)$, contains two time-independent terms, $L(\mathbf{x})$ being the spatially-varying illumination of the lamp and $T_c(\mathbf{x})$ being the spatially-varying transmission from stationary components in the specimen. The nonuniform illumination can be calibrated and corrected using the flat-fielding techniques described in Section 2.3. In the absence of flat-field calibration, however, these two terms are indistinguishable—that is, both terms result in a stationary, nonuniform illumination at the image plane. The two terms, therefore, can be combined:

$$I(\mathbf{x}, t) = T_c(\mathbf{x})T_s(\mathbf{x}, t). \tag{4.8}$$

As noted by Shizawa, the logarithm transforms this multiplicative imaging model into a linear imaging model [SM90]:

$$\log I(\mathbf{x}, t) = \log T_c(\mathbf{x}) + \log T_s(\mathbf{x}, t). \tag{4.9}$$

Taking the gradient of both sides of this equation yields

$$\nabla \log I(\mathbf{x}, t) = \nabla \log T_c(\mathbf{x}) + \nabla \log T_s(\mathbf{x}, t). \tag{4.10}$$

The benefit of transforming the imaging equation in this way is the following: the moving components in the specimen effect the constant components only at the edges of the moving components. Said another way, if the moving edges move sufficiently, there should be some estimator for which the contribution of the moving edges over all images is zero,

$$E[\nabla \log T_s(\mathbf{x}, t_i)] = 0, \qquad \{i \,:\, i = 1 \ldots N\}. \tag{4.11}$$

Applying such an estimator, E, to Equation 4.10 yields

$$E[\nabla \log I(\mathbf{x}, t_i)] = E[\nabla \log T_c(\mathbf{x})], \qquad \{i \,:\, i = 1 \ldots N\}. \tag{4.12}$$

The task remains to find a suitable estimator that distinguishes between elements from stationary occlusions and the moving specimen. Because the measurements of constant transmission at a single pixel are corrupted with gross errors from moving transmission components, robust estimation suggests itself [PTVF07, Ch. 15]. In practice, the median performs well in a number of cases, with the understanding that the median will provide a reliable gradient estimate at a pixel that is covered by moving foreground edges no more than half the time. Notice that the use of the gradient in Equation 4.12 means that a large moving object may completely cover a region of the image plane in all frames, but as long as its edges are moving the constant transmission within this region can nevertheless be recovered.

The logarithm of the constant transmission map is computed by integrating the gradient estimate of Equation 4.12 over the image plane

$$\log T_c(\mathbf{x}) = \int_C E[\nabla \log I(\mathbf{x}, t_i)] ds, \qquad \{i \,:\, i = 1 \ldots N\}, \tag{4.13}$$

where C denotes any path over the image plane from the origin to $\mathbf{x}$. Converting back

to intensity space reveals the estimate of constant transmission,

$$T_c(\mathbf{x}) = \exp\left(\int_C E[\nabla \log I(\mathbf{x}, t_i)]ds\right), \qquad \{i \,:\, i = 1 \ldots N\}. \tag{4.14}$$

I will refer the method of estimating the stationary transmission map from Equation 4.14 as the gradient logarithm transmission (GLT) method and its variants due to estimator selection in Equation 4.11 as GLT-mean and GLT-median.

Note that if the estimator, E, is linear, the estimation and gradient commute, simplifying Equation 4.14 to

$$T_c(\mathbf{x}) = \exp\left(E[\log I(\mathbf{x}, t_i)]\right), \qquad \{i \,:\, i = 1 \ldots N]\}. \tag{4.15}$$

For the mean estimator, Equation 4.15 reduces to the geometric mean of all frames. (As a result, the GLT-mean method does not actually involve a gradient computation in its implementation; I nevertheless preserve its acronym as a reference to its derivation.) No such simplification is possible for the median estimator.

4.3 Implementation

To evaluate different stationary component removal algorithms, I implemented both of the techniques outlined in Section 4.2 as well as mean-image background subtraction in C++ using the Insight Toolkit [ISNC05] (ITK) as a framework. Some implementation details deserve mention.

4.3.1 Enforcing Integrability

The GLT method requires an integration of the stationary gradient estimate in Equation 4.13. The gradient estimates, however, are not guaranteed to be integrable. That is, choosing a different order of integration will result in a different transmission map.

Frankot and Chellappa investigated this problem in research on determining the geometry of a surface from shading cues (shape from shading) [FC88]. I adopt this approach to construct a transmission map, and therefore briefly summarized the method here.

The problem is to compute the 2D surface $z(x, y)$ given estimates $z_x(x, y) = \frac{\partial z(x,y)}{\partial x}$ and $z_y(x, y) = \frac{\partial z(x,y)}{\partial y}$ of the derivatives of the surface z. The image derivatives have discrete Fourier transform coefficients $\mathcal{C}_x$ and $\mathcal{C}_y$, such that

$$z_x(x, y) = \sum_{\omega_x, \omega_y \in \Omega} \mathcal{C}_x(\omega_x, \omega_y) e^{i(\omega_x x + \omega_y y)} \tag{4.16}$$

$$z_y(x, y) = \sum_{\omega_x, \omega_y \in \Omega} \mathcal{C}_y(\omega_x, \omega_y) e^{i(\omega_x x + \omega_y y)}, \tag{4.17}$$

where Ω is the range of frequencies (ω_x, ω_y) used in the Fourier transform. Presuming that the derivatives are estimated by a discrete central difference,

$$z_x(x, y) = \frac{1}{2} \left[z(x + 1, y) - z(x - 1, y) \right], \tag{4.18}$$

Frankot and Chellappa show that the *periodic* surface that has *integrable derivatives* closest to the estimated z_x and z_y is given by the Fourier coefficients

$$\mathcal{C}(\omega_x, \omega_y) = \frac{-i \sin(\omega_x) C_x(\omega_x, \omega_y) - i \sin(\omega_y) C_y(\omega_x, \omega_y)}{|\sin(w_x)|^2 + |\sin(w_y)|^2}. \tag{4.19}$$

The elegance provided by this solution is that a simple combination of Fourier coefficients from the derivative estimates has simultaneously projected the surface derivatives into the nearest integrable subspace *and* performed the integration.

Note, however, that this Fourier-based integration finds a *periodic* surface as an estimate of the stationary transmission map. This presents a problem if the actual transmission map is non-periodic, as is the case for most natural images. To improve the behavior of the GLT method, I pad each log-intensity image with a border of half the image size in each direction and apply a Hanning window ([Har78]) that apodizes

intensity values in this padded region to smoothly transition between the pixel value at the image boundary and the mean image value at the border boundary. This embeds the image data within a signal that smoothly transitions at boundaries, which eliminates discontinuities in the infinite repetition of the signal considered by the Fourier transform. After Fourier-based integration, I crop the resulting surface estimate to the original image region. This windowing technique applies only to the GLT-median implementation; the GLT-mean implementation uses the simplification of Equation 4.15.

4.3.2 Handling Zero Intensity

Because log(0) is undefined, zero intensity values require special attention when computing the logarithm in Equation 4.9. At low intensities, the measure from an image sensor is dominated by noise [HK94]. Because the intensity estimate is less accurate at the low intensity range, it is least harmful to treat zero-valued and one-valued pixels the same. I therefore clip the minimum input image intensity to one. The alternative approach—offsetting all intensity values by one—would destroy the relative intensity ratios for all pixels.

If a pixel records zero intensity values for all frames of a video, there is total occlusion of light at that pixel. The transmission map would require infinite gain to repair the intensities at that pixel, and the repaired image intensity would have infinite error. In other words, the repaired image intensities should not be trusted at all for pixels that record low intensity values over the entire video. Because the gain scales noise along with image intensity, the transmission map functions as both a transmission gain and a confidence measure—intensity values with high transmission map gain should be trusted less than intensity values with low transmission map gain. It may be desirable to mask out any pixels whose transmission values fall below a certain threshold, or pass along the transmission map as a confidence measure for further image analysis processes.

At times a sharp intensity fall-off occurs along one edge of the image plane; this

appears, for example, in the flat-field calibration of Section 2.3, Figure 2.7. This is due to misalignment of the image sensor in the camera housing, and this situation should be apparent during camera calibration. Any affected pixels should be cropped before image analysis.

4.3.3 Transmission Map Scaling

The integration in Equation 4.13 introduces an integration constant to the estimate of the logarithm of stationary transmission. Exponentiation in Equation 4.14 transforms this free parameter into a scale factor. So, the GLT method recovers the constant transmission up to a scale factor, and the task remains to select an appropriate value for that scale factor.

An unoccluded image location should have a transmission factor of one. It is unknown *a priori* that any particular location is unoccluded—in fact, the whole image plane could be partially occluded. But such a situation is equivalent and indistinguishable from inserting a neutral density filter into the light path or reducing the intensity of the microscope lamp. Intuitively, the image location that receives the most light over the course of the video would be an obvious choice for an unoccluded location. The pixel with the maximum transmission metric, then, is a good candidate, and this choice reinforces the assumption that objects in the light path only absorb light. The computed stationary transmission map is therefore be rescaled to have a maximum of 1.

Highlights in the image due to diffraction, such as at the center of beads, may set this maximum transmission value too high, overly dimming the output images. In this case one may select some other reference value, such as the 90^{th} percentile of the transmission map, to scale to 1.

Another possible solution scales the mean transmission to one. Although violating the assumption that light is only absorbed, this choice maintains the mean intensity between input and output images, and preserves visual consistency especially when

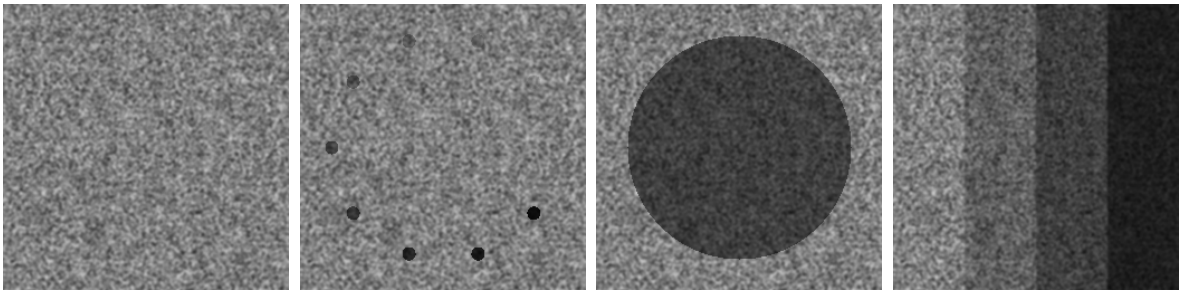
occlusions cover a small portion of the image plane.

A similar situation occurs in background subtraction. Mean image subtraction involves a shifting of intensity values. As this process does not preserve the intensity ratios between pixels anyway, an arbitrary intensity shift can be applied to put the output image intensities into some desired range. In my background subtraction implementation, I maintain the mean image intensity between input and output image sequences to preserve visual consistency.

4.4 Evaluation

I evaluated the microscopy stationary object removal algorithms using both synthetic and real data. Synthetic data provides a known ground truth by which to quantitatively assess the techniques. Real microscopy data must be qualitatively assessed.

4.4.1 Simulated Data



(a) (b) (c) (d)

Figure 4.2: Sample test data of transforming noise fields: a) no occlusion, b) small occlusions with varying transmission, c) large occlusion, and d) step function occlusion.

The simulated data is generated from a normally-distributed noise field undergoing rigid transformations. Each test video consists of 180 frames of 256×256 pixels. To simulate microscopy image formation, we modulate these image sequences with fixed

patterns—discs and bands of various sizes and transmissions—as seen in Figure 4.2.

We apply each of our video repair algorithms and compute the peak signal-to-noise ratio (PSNR) over all frames as a comparison metric [Bov05]. The PSNR, measured in decibels (dB), for two $m \times n$ images $I_1(x, y)$ and $I_2(x, y)$ is

$$\begin{aligned} \mathrm{PSNR}(I_1, I_2) &= 10 \log_{10} \left(\frac{\max(I_1)^2}{E(I_1, I_2)} \right), \\ E(I_1, I_2) &= \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(I_1(i, j) - I_2(i, j) \right)^2 . \end{aligned} \tag{4.20}$$

Higher PSNR values indicate a more faithful reconstruction of an image, and PSNR values between 30 and 40 dB are typical in the image processing literature [ABMD92, SP96].

Table 4.1 shows the reconstruction results for a subset of test cases. Because PSNR is sensitive to gain, and the GLT methods have a free scale parameter, as described in Section 4.3, I choose to scale the GLT transmission maps to maximize PSNR.

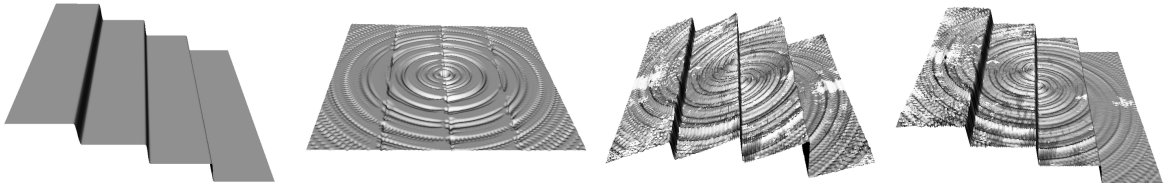
Table 4.1: Mean PSNR (dB) measurements for repairing simulated microscopy data; r is half the width of the square neighborhood used in the LTM method; *window* refers to padding with a Hanning window in the GLT-median method. The best result for each data set is indicated in bold face.

Data set	Noise							
Transform type	Rotation				Translation			
Occlusion type	none	small	large	step	none	small	large	step
Background subtraction	42.12	38.98	16.04	12.20	38.98	18.32	14.51	11.77
LTM, r=5	**49.66**	28.97	12.97	10.35	**46.21**	29.18	13.27	10.64
LTM, r=15	46.12	**46.16**	13.04	10.37	43.36	**43.44**	13.33	10.66
LTM, r=100	—	—	41.65	10.38	—	—	38.47	10.68
GLT-mean	41.83	41.83	**41.83**	**41.79**	39.87	39.88	**39.85**	**39.61**
GLT-median	39.66	39.65	39.66	12.10	36.62	36.61	36.62	12.70
GLT-median window	35.19	35.20	35.15	35.20	33.39	33.39	33.37	33.28

Using the case of a rotating noise field as an example, the data set with no occlusion indicates the quality of reconstruction obtainable by each method. Except for LTM

with small radius, all methods match performance in the unoccluded case when small occlusions of varying modulation are added to the video. The LTM method performs best in this example, followed closely by the GLT-mean.

The presence of a large occlusion exposes the weakness in the LTM method. Large occlusions require the use of large neighborhoods (Ω in Equation 4.7) to ensure an accurate estimate of the unobstructed intensity for each neighborhood. A neighborhood with a radius of 100 pixels was required to repair the test case with large occlusion seen in Figure 4.2. The running time of the LTM method scales quadratically with the scale parameter. For example, a desktop computer equipped with an AMD Athlon 64 3700+ processor and 1 GB of RAM, the LTM computation with $r = 100$ took over an hour, an unacceptable penalty in most situations. (The LTM method with $r = 100$ was not computed for the data sets with no occlusion and small occlusions because a smaller radius was sufficient to achieve accurate image recovery.) The GLT methods do not suffer from this scale parameter dependency.



(a) (b) (c) (d)

Figure 4.3: Transmission maps recovered from rotating noise field image sequence with step function occlusions: a) ground-truth step function; b) transmission map recovered with LTM; c) transmission map recovered with GLT-median without windowing; d) transmission map recovered with GLT-median with Hanning window applied.

The LTM and non-windowed GLT-median methods do not recover an accurate representation of the transmission map for image sequences modulated by a step function. Figure 4.3 displays the computed transmission maps displayed as height fields to provide insight into this result. Figure 4.3a shows the ground-truth transmission map used to

generate the data. The transmission map in Figure 4.3b reveals that the size and shape of the occluded regions make it impossible for the LTM method to determine intensity values from unoccluded pixels for all neighborhoods, as only pixels in the left-most strip can be considered unoccluded.

The transmission map in Figure 4.3c demonstrates that the Fourier-based integration enforces a periodic transmission map in the GLT-median method without windowing. Padding and windowing, as discussed in Section 4.3, is necessary to recover non-periodic transmission maps using the GLT-median method, as seen in Figure 4.3d.

4.4.2 Microscopy Data

Figure 4.4 shows the results of applying stationary occlusion removal to the video of beating cilia seen in Figure 4.1. Background subtraction does a reasonable job of removing stationary objects, but suffers from ghosting of the moving bead in the bottom center of the image. The LTM method succeeds in removing most of the dust occlusion in the center of the frame; a faint diffraction halo remains visible around the dust spot because the light model only handles occlusions that absorb light. The GLT methods suffer less from this artifact. Diffraction effects are also visible around the moving bead, but these foreground components have no effect on any of the stationary occlusion removal methods.

The LTM method does the worst job of removing the large stationary occlusions of the cell culture. In this example, the local neighborhood is constrained to a 30 pixel^2 area, but the cell culture occlusions are large and irregular enough that a larger neighborhood would not likely improve the result. The GLT methods do not suffer from these limitations because they form a complete model of all non-moving components of the video. Removing the stationary cell culture components accentuates the motion of the cilia.

The GLT-median is the only method that does not suffer from ghosting of the moving

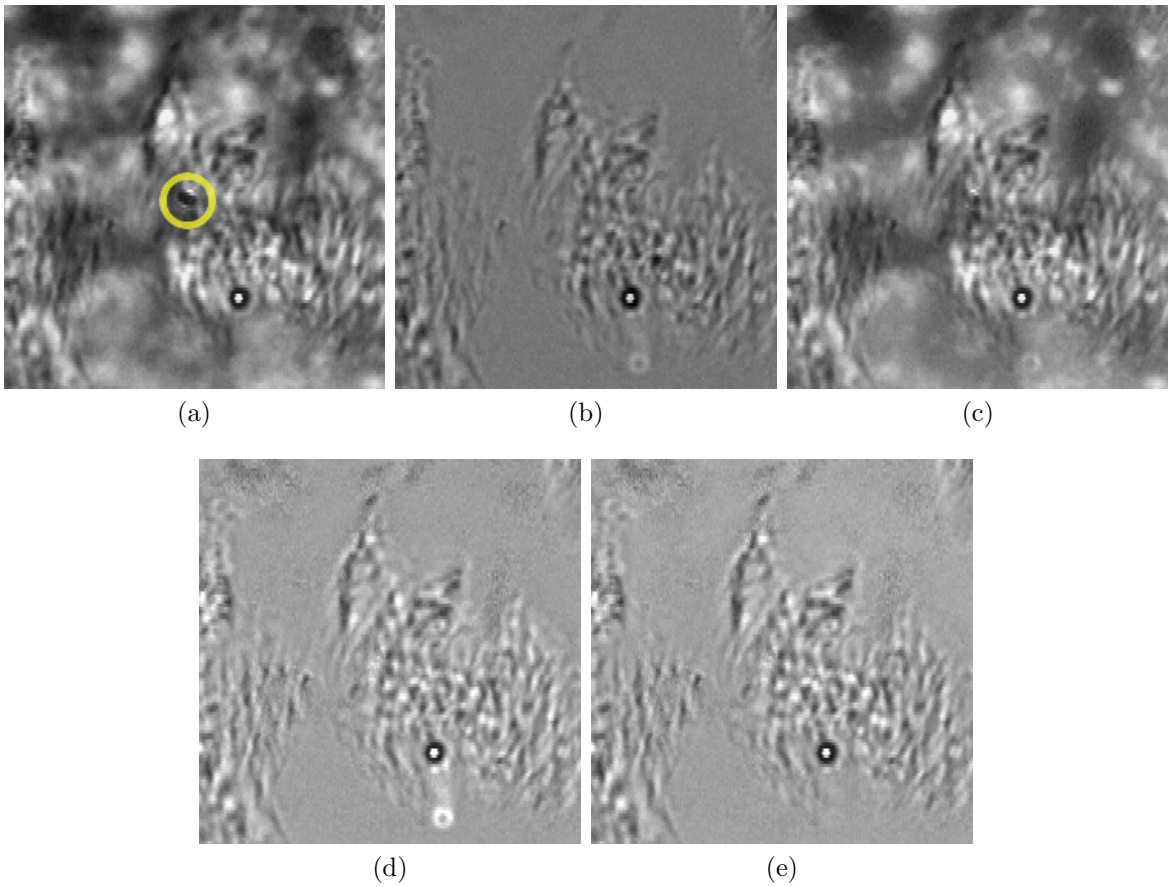


Figure 4.4: Stationary occlusion removal applied to a microscopy video of beating cilia, 248×250 pixels, 61 frames: a) one frame from the original sequence with a circle surrounding a dust spot on the image sensor, b) the same frame recovered with mean background subtraction, c) LTM ($r = 15$), d) GLT-mean, and e) GLT-median.

bead in the bottom center of the images. This example highlights the strength of the GLT-median approach—when the edge of a foreground object covers an image location for fewer than half the frames in a video, its effect does not perturb the background of the repaired video.

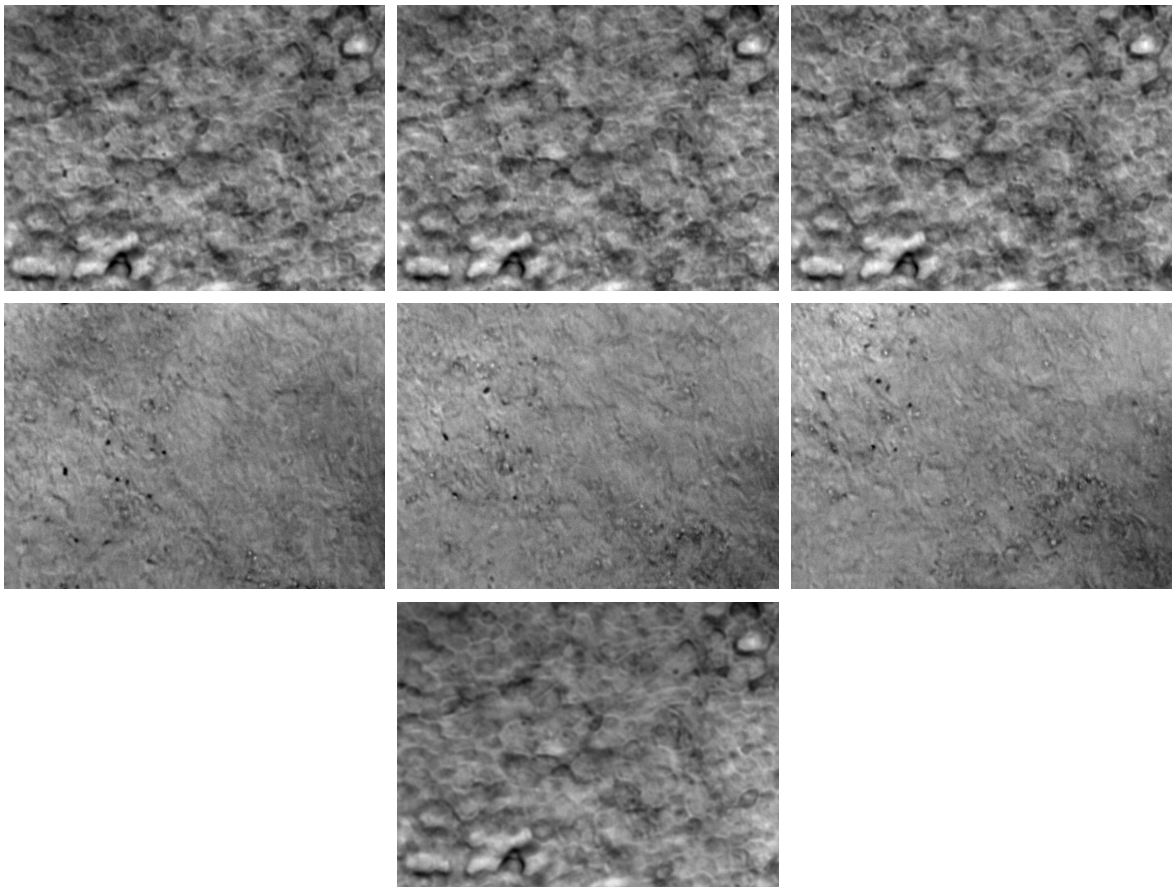


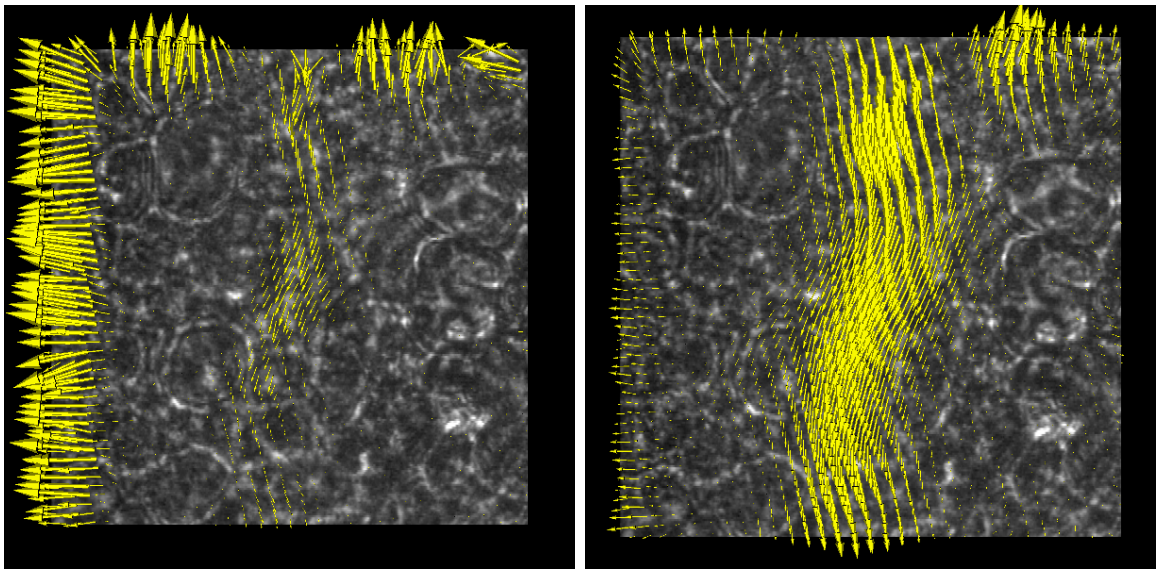
Figure 4.5: GLT-median stationary occlusion removal applied to cilia-driven mucus flow video, 648×484 pixel, 60 frames. Top: cilia on human airway epithelium cells drive mucus flow towards the upper left corner of the frame. Middle: stationary occlusion removal applied to this video makes it easier to discern moving cilia, mucus, and particles trapped in the mucus. Bottom: the stationary component recovered from this video displays the cell layer. Original images from David Hill.

Having demonstrated that GLT-median most effectively removes the visible artifacts from stationary occlusions in cilia videos, Figure 4.5 shows the method applied to another data set. This example displays a much larger portion of a ciliated cell culture. The selection of a long working distance objective lens required to focus through the specimen

means that the cell culture, cilia, and mucus layers are visible. Although motion—especially of particles in the mucus layer—is visible in the video, the cell culture is the predominant visible feature. After stationary occlusion removal, however, the moving mucus layers are featured prominently, and it is easy to discern beating cilia structures in some regions. The stationary component of the video, shown on the bottom of Figure 4.5, contains a clear view of the epithelial cell culture.

The occlusion removal example from Figure 4.5 used 60 frames from a 600 frame video to construct a model of the stationary cell layer. Using the same transmission mask to repair the entire video reveals that the stationary cell layer moves slowly over the entire video, leading to ghosting artifacts in later frames. This suggests that a sliding window of frames should be used to compute a transmission map for long videos with slowly moving background components. Hill *et al.* used such a sliding window approach to enhance motion in DIC videos of vesicle transport [HPBH04]. An interesting avenue of future research is to extend the single layer transmission estimate to multiple layers. In this example, a stationary transmission map would be computed for the entire video, and a time-varying transmission map would account for the slow motion of the cell layer. A third transmission map may account for mucus flow over the cilia.

Figure 4.6 shows one step along the path towards estimating motion and composition of multiple deforming, semitransparent image layers. The image in the background of Figure 4.6a is a single frame from a video of cilia-driven mucus flow. In the original video, the flow is visible as a subtle layer over the top of the cells, with the predominant flow downwards in a "channel" in the center of the field. The arrows in Figure 4.6a represent the net optical flow computed from the original video. Flow computation is performed with an implementation of the combined local and global gradient-based technique of Bruhn *et al.* [BWS05], as discussed in Section 3.1. The magnitude of an arrow indicates the total flow starting at a single pixel from the beginning of the video. In this computation, the high-contrast, stationary boundaries of the cells corrupt the



(a) (b)

Figure 4.6: Flow computation for cilia-driven mucus flow. The gray scale image is a single frame from a bright-field microscopy video of cilia-driven mucus flow. Arrows show net computed optical flow from the beginning of the video (150 frames). a) Flow computed without stationary occlusion removal. b) Flow computed after applying GLT-median stationary occlusion removal. Original images from Jeremy Cribb.

flow estimation—the dominant computed flow is at the left edge of the image frame instead of the middle. Additionally, this computed flow does not preserve the volume of fluid moving through the channel—there is more flow in the middle of the frame than towards the bottom.

Figure 4.6b shows the result of optical flow computation at the same frame for the video preprocessed with stationary occlusion removal using the GLT-median metric. Here, the dominant motion in the video is correctly attributed to the channel in the middle of the image frame. The volume of mucus flow through the middle appears to be conserved. The cluster of arrows pointed upwards in the top right of the field of view indicate motion from a beating cluster of cilia.

4.5 Discussion

This chapter presented two methods for removing occlusions from microscopy videos that exhibit advantages over background subtraction. Each method and variation has strengths which lead it to outperform the others in different scenarios. The LTM method performs best on videos with small, bounded occlusions, but handles large occlusions and diffraction effects poorly. The GLT methods work equally well on any size occlusion and at least visually improves diffraction effects. The GLT-median estimation outperforms the GLT-mean estimation when a moving scene object covers an image region for less than half of the video.

As mentioned in Sections 2.3 and 4.1, flat-fielding is a commonly-used technique in bright-field microscopy that counters the effect of non-uniformities in Köhler illumination. Flat-fielding uses a specimen-free calibration image—or, better, a set of images—to create a map of how much light reaches each pixel in the image sensor. A typical flat-field image is brightest at the center and falls off towards the edges. Flat-field correction involves dividing any specimen images by the flat-field calibration image, as in Equa-

tion 4.4.

The transmission map computed in stationary occlusion removal is similar to the flat-field calibration image. In fact, because no distinction is made between light loss due to nonuniform illumination and light loss due to fixed occlusions, stationary occlusion removal also performs flat-fielding. Given enough microscopy images in which no part of the specimen is stationary, stationary occlusion removal will recover the flat-field image for nonuniform illumination. Therefore, there are two major advantages to stationary occlusion removal:

1. stationary occlusion removal can be applied as a post-process in the absence of calibration data, and

2. stationary occlusion removal recovers a specimen-specific calibration image that "flat-fields" non-uniformities from stationary parts of the specimen.

Chapter 5

Lateral Tracking with Structured Illumination Microscopy

Structured illumination is a computer vision technique for obtaining information about a scene. In this technique, a structured light pattern illuminates a scene in a controlled manner, and the effect of this illumination on images of the scene is used to determine scene properties such as geometry [VO90, NN94, RBY+99]. This chapter explores a new application of structured illumination to determine the 3D position of a microscope stage.

The microscope stage is the platform on which a specimen sits during microscopy observation. Determining the precise position of the stage is important in many microscopy applications. Stage tracking enables monitoring motion over large ranges. For example, when tracking mucus transport across cell cultures or investigating cell motility, the specimen can move further than one field of view will allow observation. The microscope stage can be repositioned to track the specimen, but the motion of the stage needs to be known independent of the specimen motion to discern the complete motion of the specimen over the entire experiment.

Given stage position information, one can also make measurements spanning multiple fields of view, for example to determine object size or measure the distance between landmarks. Tracking the X, Y, and Z position of the stage extends this capability to

3D.

Another aspect of microscopy imaging is the trade-off between magnification and field of view. High-magnification images reveal fine structural details about the specimen while wide field of view images provide the overall context how the specimen is organized. Microscopists are often interested in both of these aspects simultaneously, for example to understand muscle morphology [KS07]. Stage tracking provides information about where in a specimen an observation is being made. Given a collection of high magnification images and accurate stage positions, one can assemble a microscopy mosaic image that presents both fine structural details and broad context.

In the structured illumination approach discussed here, a semitransparent pattern on the cover slip or slide modulates the light transmitted through the bright field microscope. This pattern acts as a transmission mask, partially attenuating the light illuminating the specimen in a structured manner. Analysis of the pattern appearance enables tracking the stage in three dimensions—laterally in X and Y and axially in Z.

Because the tracking is performed in software, this technique does not rely on any microscope instrumentation aside from patterned cover slips and a digital camera. This technique can therefore be used to obtain stage positions on standard research microscopes with hand-driven micrometer stages.

The remainder of this chapter is organized as follows. Section 5.1 explains the theory behind lateral pattern tracking methods. Sections 5.3 and 5.4 examine the performance of this imaging system using simulated and real data, respectively. Section 5.5 discusses using this tracking system in a microscopy mosaicking application. Chapter 6 discusses extending the pattern analysis to include axial tracking and evaluates the technique for moving specimens.

5.1 Theory

In a bright-field microscope, light from a uniform light source is transmitted through a specimen and collected by a series of lenses to form an image on a camera sensor. Image contrast is generated by the transmission and diffraction of light through the specimen. As discussed in Section 2.1.2, if the image of the specimen predicted by geometrical optics can be modeled as a thin, planar, transmission mask $S(x, y)$, the image formed by the microscope is:

$$I(x, y) = L(x, y) \left[h(x, y) \otimes S(x, y) \right], \tag{5.1}$$

where $L(x, y)$ is the illumination at the focal plane, $h(x, y)$ is the microscope's point-spread function (PSF) at the focal plane, and $\otimes$ is the convolution operator.

Assume a small pattern, a *micropattern* $P(x, y)$, is inserted into the microscope's light path on the cover slip, slightly above the specimen. The micropattern acts as a transmission mask that attenuates the illumination field, but its image is convolved with a different part of the microscope's 3D PSF. The image formed on the camera sensor becomes the multiplication of images of each plane

$$I(x, y) = L(x, y) \left[h(x, y, z_s) \otimes S(x, y) \right] \left[h(x, y, z_p) \otimes P(x, y) \right], \tag{5.2}$$

where z_s and z_p are the axial positions of the specimen and pattern respectively. Note that this model does not precisely account for the diffraction of light between the micropattern and the specimen. However, because the distance separating these planes is much smaller (*e.g.* as little as $5\,\mu$m) compared to the distance the light diffracts within the microscope (*e.g.* 160 cm), this approximation is reasonable. This approximation gains the additional benefit that the ordering of the micropattern and specimen layers is unimportant.

The position of the micropattern is fixed with respect to the slide, which remains fixed to the microscope stage. Introducing the micropattern, then, transforms the stage

tracking problem into a pattern tracking problem. Obtaining accurate stage tracking requires that the pattern be tracked independently from the specimen, and a successful tracking method will assume little about the specimen while obtaining accurate pattern tracking. Specifically, pattern tracking holds an advantage over any tracking method that relies on specimen information because it applies whether or not the specimen has sufficient contrast and whether or not the specimen is moving. Because the pattern is semitransparent and has a known appearance, the effect of the pattern can be removed using an extension of the stationary occlusion removal techniques described in Chapter 4.

The pattern tracking problem can be broken down into these major components:

1. Find the orientation and size of the pattern in all input images.
2. Determine the axial position (focus) of the pattern with respect to the specimen.
3. Determine the lateral position of the pattern along each pattern axis independently or simultaneously.
4. Repeat steps 1-3 with the additional information gained at each step.

In this chapter, I discuss orientation and lateral tracking steps assuming the axial position is known. Focus determination is reserved for Chapter 6.

5.1.1 Determining Orientation

Consider the image (and its Fourier transform) of a regular pattern composed of two identical, orthogonal square patterns depicted in Figure 5.1. Note that this pattern's Fourier transform has a fundamental frequency determined by the spacing between square centers plus harmonics along each of the pattern's coordinate axes. The harmonics occur at integer multiples of the fundamental frequency. The first step in tracking is to determine the orientation of the pattern's coordinate axes so that tracking can proceed along each axis independently.

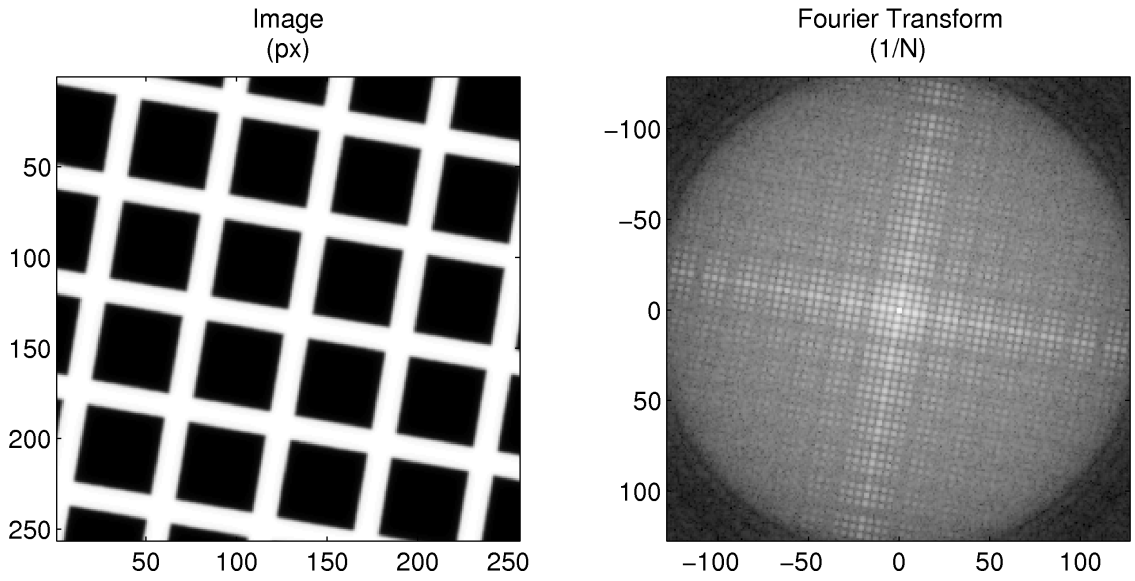


Figure 5.1: An image of a regular grid pattern and its Fourier transform.

An initial estimate of the pattern orientation is determined with a search in frequency space. Rotation in the spatial domain corresponds to rotation in the frequency domain, so the frequency components from the pattern fundamental and harmonics lie a set distance from the center of frequency space. A search for maximum magnitude along the frequency bands of the fundamental and harmonics—marked as rings for the square wave pattern in Figure 5.2a—yields four peaks for each frequency, 90° apart. The orientation estimate can be refined through a one-dimensional (1D) optimization along the frequency band. The mean of the peak orientations for all target frequencies provides an orientation estimate. This search process can be adapted for any other pattern, given the frequencies and orientations present in the pattern.

If the pattern spacing is only roughly known, an orthogonal search provides a refined estimate for pattern spacing given an initial orientation estimate. In this case, the 1D optimization operates along the pattern axis to find the frequency with maximum magnitude. Figures 5.2c and d show the magnitude along each axis, with vertical lines marking maxima. The orientation and frequency searches alternate until convergence.

The accuracy of this orientation estimation technique is discussed in Section 5.3.2.

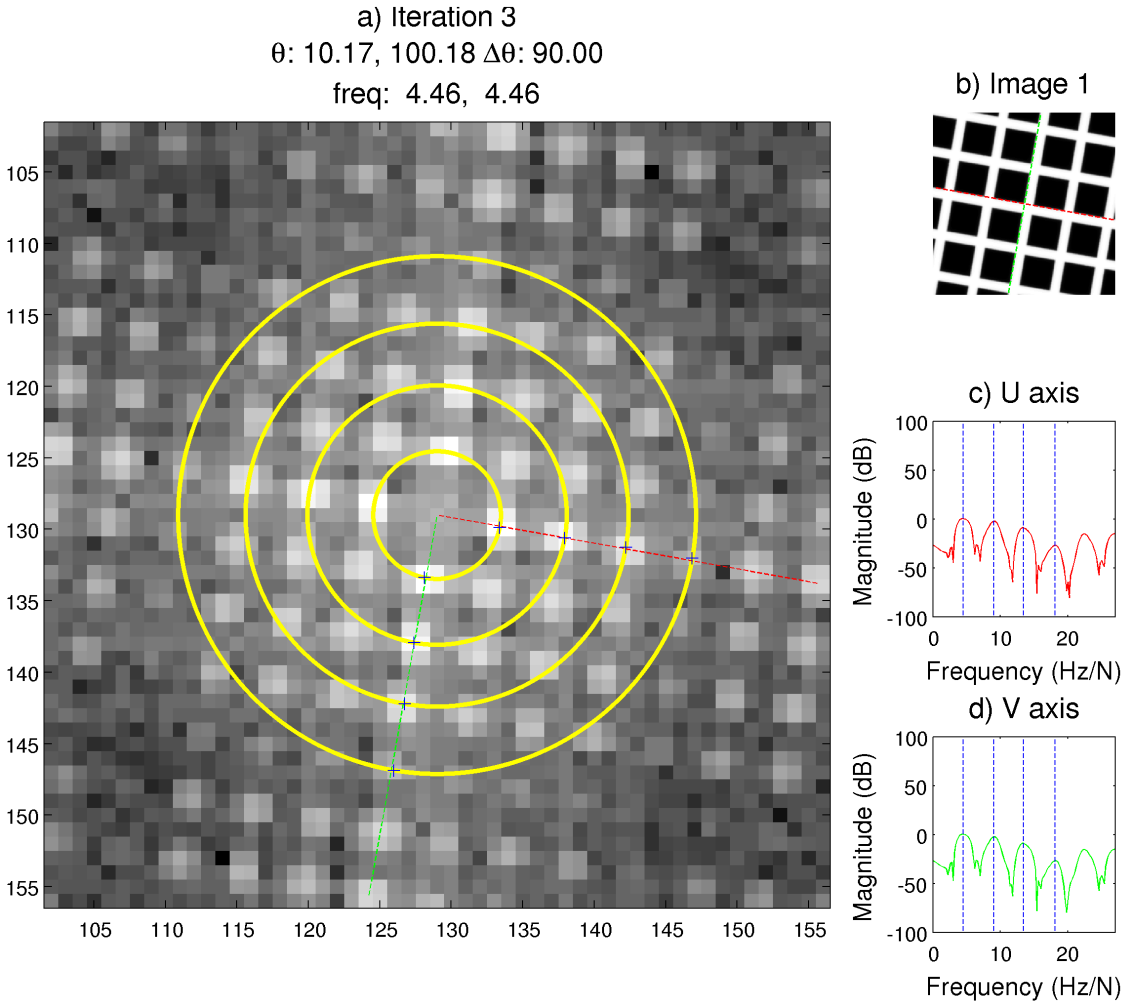


Figure 5.2: Finding the orientation of a regular pattern involves searching for peaks in the Fourier transform at harmonic frequencies in the pattern. a) A subset of the Fourier transform magnitude of the input image (b). Pattern is rotated 10°; the mean image intensity was subtracted to set the DC component to zero. Circles indicate the frequency bands at which the search is conducted; dashed lines show the computed orientation axes. c and d) The frequency magnitude along the pattern axes; dashed vertical lines at the maxima indicate optimized frequencies.

5.1.2 Pattern Tracking

Because the slide orientation stays fixed during stage motion, one can use multiple images to estimate the pattern orientation. Once the orientation is known, tracking requires recovering the translation of the pattern from frame to frame. Several image analysis techniques recover translation from images, and obvious choices for pattern tracking include feature tracking, spatial correlation-based matching, and phase correlation [BB95]. The fundamental image analysis task in tracking the pattern is to separate image information in the pattern layer from the specimen layer. Feature tracking depends on identifying local image features that can be tracked across multiple frames. This class of tracking method is not considered here because it would require feature detection and tracking that is robust to interference from the specimen layer. Although this is a very interesting area of research, the fact that the pattern is rigid and covers the entire image plane provides strong incentive to investigate methods that consider information about the pattern over the whole image plane. For this reason, spatial correlation and phase correlation are considered in detail.

Model-based Spatial Correlation

Image registration is a class of image analysis techniques for determining motion from images. Assume two $n \times m$ images, $I_1(x, y)$ and $I_2(x, y)$, are taken of a planar object that is oriented perpendicular to the optical axis. If there is only translation of the object (or the camera) perpendicular to the optical axis between the capture time of the two images, the images are related by $I_1(x, y) = I_2(x - u, y - v)$. This suggests that finding the translation (u, v) that describes the translation between the two images amounts to minimizing an energy function, such as the ℓ^2 distance:

$$E(u, v) = \sum_{x,y} \left[I_1(x, y) - I_2(x - u, y - v) \right]^2 . \tag{5.3}$$

Lewis demonstrates that the image comparison metric known as normalized cross-correlation (NCC) can be derived from the expansion of the image comparison term in Equation 5.3 [Lew95]. For two image patches $I_1(x, y)$ and $I_2(x, y)$,

$$d^2 = \sum_{x,y} \left(I_1(x, y)^2 + I_2(x, y)^2 - 2I_1(x, y)I_2(x, y) \right). \tag{5.4}$$

Assuming the terms $\sum_{x,y} I_1(x, y)^2$ and $\sum_{x,y} I_2(x, y)^2$ are approximately constant and normalizing the remaining image comparison term yields,

$$\mathrm{NCC}(I_1, I_2) = \frac{\sum \tilde{I}_1(x, y)\tilde{I}_2(x, y)}{\sqrt{\sum \tilde{I}_1(x, y)^2 \sum \tilde{I}_2(x, y)^2}}, \tag{5.5}$$

where $\tilde{I}(x, y) = I(x, y) - \frac{1}{nm} \sum I(x, y)$, and all sums are over the image plane. Because NCC normalizes the distribution of intensities in the compared images, it compares the *landscape* of the two images and is robust to intensity shifts. Note that finding the translation between two images requires finding the alignment that *maximizes* Equation 5.5.

Image registration provides another method for tracking the pattern layer in structured illumination microscopy. A model of the pattern's appearance is used to generate simulated pattern images. An optimization procedure finds the model parameters that maximize the NCC metric between a microscopy image and the simulated pattern images. That is, given a model that produces a pattern image, $M(u, v) \rightarrow P_{u,v}(x, y)$, find the translation parameters (u, v) that minimize

$$E(u, v) = -\mathrm{NCC}(I(x, y), P_{u,v}(x, y)). \tag{5.6}$$

This method will be referred to as model-based spatial correlation (MSC).

The image formation model needs to account for the pattern design (the ideal image predicted by geometric optics), orientation, translation, and the effect of defocus. For

a fixed orientation and focal depth, finding the translation of the pattern in an image involves a search over the parameter space of all possible translations. An exhaustive search that considers sub-pixel displacements is slow, so a practical approach relies on maximizing the NCC objective function using an optimization strategy. The shape of the objective function depends on the pattern and the specimen; the influence of the pattern will be discussed in greater detail in Chapter 6.

The following optimization routine arrives at a reasonable result in all cases I have tried. The optimization method involves a coarse-resolution search through parameter space to initialize a Nelder-Mead simplex-based optimization [NM65]. For the 2D optimization at hand, this technique maintains a set of three connected points in parameter space—the simplex—and the values of the objective function at those points. The simplex is updated by evaluating the objective function near the simplex, moving the high-valued simplex vertex to lower valued locations in an attempt to "walk downhill" towards the minimum. This approach, of course, depends on having a good initial guess to initialize the simplex.

The initial search proceeds as follows:

1. The initial parameter range is chosen to allow for the largest translation of the pattern between frames that can be distinguished—half the largest wavelength in the pattern in either direction.

2. The target microscopy image is smoothed with a Gaussian filter and down-sampled to a coarse resolution to make the image simulation and NCC comparisons fast. A down-sampling factor that is a power of two is chosen to ensure that a single pattern repetition covers at least an 8×8 pixel region in the down-sampled image. The standard deviation of the Gaussian filter is chosen to be the square root of the down-sampling factor to avoid aliasing in the down-sampled image.

3. An initial translation step size is chosen to ensure that each pixel within a down-

sampled pattern region is sampled at least once.

4. Coarse pattern images are simulated for each translation using large virtual CCD sensors to account for the reduced resolution.

5. The NCC metric is computed for each translation to assemble a correlation map over the parameter space.

6. The parameter range and step size are contracted around the maximum in the correlation map by a factor of two.

7. Steps 3–5 are repeated twice at the same resolution.

8. Steps 2–6 are repeated using a higher-resolution target image and smaller CCD sensors (each by a factor of 2) in the simulation.

9. When this search has completed on an image resolution half the size of the original image, the current translation and range initialize a modified Nelder-Mead simplex-based optimization.

The Nelder-Mead optimization procedure usually conducts an unbounded search in the parameter space, but setting the objective function to a very poor value outside the range determined in the multiple resolution search constrains the parameter space. The initialization method arrives at a translation that is no more than 1 pixel away from the final optimized translation. This spatial correlation method is evaluated for structured illumination microscopy patterns in Sections 5.3 and 5.4.

Note that a similar model-based correlation approach can improve the orientation estimate provided by frequency analysis. Given an initial orientation and position estimate, pattern images are simulated at different orientations to find the one that maximizes the NCC metric. This 1D parameter search with an initial estimate close to the optimum is readily implemented with a golden section optimization [PTVF07, Ch. 10].

I now turn to the other major pattern tracking method, phase correlation. The tracking methods are compared in Section 5.3.

Phase Correlation

Phase correlation provides an efficient registration method for recovering translation transforms, such as those introduced by moving a microscope stage [FZB02]. Phase correlation depends on the Fourier shift property: a translation in the data domain is equivalent to a phase shift in the frequency domain. For two images I_1 and I_2, where

$$I_2(x, y) = I_1(x - x_0, y - y_0), \tag{5.7}$$

the cross power spectrum yields

$$\mathcal{C}(u, v) = \frac{\mathcal{I}_2(u, v)\mathcal{I}_1(u, v)^*}{|\mathcal{I}_1(u, v)\mathcal{I}_1(u, v)^*|} = e^{-i(ux_0+vy_0)}, \tag{5.8}$$

where $\mathcal{I}(u, v)$ is the Fourier transform of $I(x, y)$ and $\mathcal{I}^*$ is the complex conjugate of $\mathcal{I}$. The inverse of the cross power spectrum is a Dirac delta function centered at the translation coordinates (x_0, y_0), which can be found by locating the maximum in $C(x, y)$, the Fourier transform of $\mathcal{C}(u, v)$.

Efficiency is one of the major advantages to phase correlation. The method requires three Fourier transforms, two pixel-wise image products, a pixel-wise image division, and a search for a maximum. The computational complexity of phase correlation is therefore limited by the Fourier transform, which for a pair of images with N pixels has $O(N \log N)$ complexity [CT65]. In contrast, spatial correlation has $O(N^2)$ complexity—at every possible translation, a pixel-wise image product and summation is performed.

There are two barriers to using phase correlation to track a regular micropattern. First, because phase correlation treats all frequencies as equal contributors to the registration, this method requires broad spectrum support to recover a translation [TAS+03].

The micropatterns I propose, however, contain narrow bandwidth information only. Considering all frequencies in an image of a micropattern and specimen also includes information from the specimen, which may or may not move in concert with the micropattern.

Second, phase correlation requires applying a windowing function to the image data to attenuate discontinuities at image edges. This weights information in the center of the image more heavily than information at the edges of the image. Though this is less of a problem when considering rigid translations of objects covering the whole image plane, in micropattern analysis it is desirable to treat all valid pattern data as equal contributors to registration. Specifically, it is unknown where in the image the pattern information might dominate the specimen information; if this occurs only near the borders of the image, windowing would discount this important tracking data.

The phase estimate for a particular pattern frequency needs to be very accurate to obtain a reasonable tracking accuracy. For a phase error of E_ϕ measured in degrees, the pixel tracking error is

$$E_{px} = \frac{E_\phi}{360f}, \tag{5.9}$$

where f is the frequency of the component used for tracking. For a fixed phase error, tracking accuracy improves with higher frequency, shorter wavelength pattern components. The range of displacements between frames that can be unambiguously determined with a single component, however, is limited to half a wavelength in each direction. This suggests using large wavelengths to enable large displacements between frames along with short wavelengths to provide accuracy.

The following modifications to phase-based tracking ultimately do not provide a solution to pattern tracking in the presence of an arbitrary moving specimen. However, there are specific imaging applications discussed in Section 5.6 where tracking involves only the pattern. For this reason it is useful to describe the current state of this line of research. A review of the Fourier transform and discrete Fourier transform (DFT)

provides context for discussing attempted approaches to overcome the barriers to pattern tracking with phase correlation.

Discrete Fourier Transform

The Fourier transform is a representation of a signal as a collection of frequency components. For a continuous, infinite 1D signal, $f(x)$, the component of the Fourier transform for a frequency u is given by

$$\mathcal{F}(u) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi ux} dx, \tag{5.10}$$

where $i = \sqrt{-1}$. The Fourier transform is a change in data representation between the *spatial domain* (all spatial positions x) and the *frequency domain* (all frequencies u). Each component of the Fourier transform has a *magnitude* that indicates the energy of the frequency component contained in the signal and a *phase* that indicates the relative position of the frequency component in the signal. For a complex component, u, the magnitude and phase are

$$\begin{aligned} M(u) &= |\mathcal{F}(u)| = \sqrt{\Re(\mathcal{F}(u))^2 + \Im(\mathcal{F}(u))^2}, \\ \Phi(u) &= \arctan\left(\frac{\Im(\mathcal{F}(u))}{\Re(\mathcal{F}(u))}\right), \end{aligned} \tag{5.11}$$

where $\Re$ and $\Im$ represent the real and imaginary components of a complex value, respectively. The magnitude of a signal is often reported in *decibels*, a unitless measure that compares a component magnitude M to some reference magnitude M_0 on a logarithmic scale. For images, the scale convention is

$$M(dB) = 20 \log_{10} \frac{M}{M_0}. \tag{5.12}$$

The discrete Fourier transform (DFT) is a discrete approximation to the Fourier

transform. For a 1D discrete signal, $f(x)$, defined over the domain $x = [0 \ldots N-1]$, the k^{th} component of the DFT is given by

$$\mathcal{F}(k) = \sum_{x=0}^{N-1} f(x) e^{-i\frac{2\pi kx}{N}} \qquad k = 0 \ldots N-1. \tag{5.13}$$

The two major differences between Equations 5.10 and 5.13 are (1) the input signal has finite extent and (2) the transform is defined for a discrete set of uniformly-spaced frequencies $\{u \,|\, u = \frac{k}{N},\, k = 0 \ldots N-1\}$.

Each component of the DFT is the correlation of the input signal with a sinusoidal basis function (from Euler's formula, $e^{-i\theta} = \cos\theta - i\sin\theta$). Because the basis functions are periodic and have infinite extent, the DFT treats the input signal as periodic as well. In other words, the DFT of a signal computed over a the finite signal domain is equivalent to the DFT of an infinite number of copies of the signal computed over an infinite domain.

Another way of thinking about this is to consider the finite signal as a truncated portion of an infinite signal. The finite signal is created by multiplying the infinite signal by a rectangular *windowing function* that has value one within a finite domain and zero everywhere else. In fact, there are many forms of windowing functions used in Fourier analysis. Generally defined, a windowing function has value zero everywhere outside a finite domain. Every DFT computation on measured data involves windowing because every observable signal is finite in extent.

From the convolution theorem, multiplication in the spatial domain is equivalent to convolution in the frequency domain [FvDFH97, Ch. 14]. The effect of multiplying a signal by a windowing function in computing the DFT is convolution in the frequency domain by the Fourier transform of the windowing function. The result of this convolution is that a single frequency component includes contributions from neighboring frequencies, an effect known as *leakage*. Selecting a windowing function involves a trade

off between the width and fall-off of the windowing function's Fourier transform [Har78]. These parameters affect how many neighboring frequencies contribute to a single component and the magnitude of each contribution, respectively. An ideal windowing function would have a frequency domain representation of a single point, but this would require a windowing function of infinite extent in the spatial domain.

Harmonic analysis—determining the magnitude and phase of the frequencies in a signal—is an *estimation problem* [Har78]. Figure 5.3 illustrates the effect of windowing on parameter estimation of sinusoidal signals of the form

$$f(x) = A\cos(2\pi ux + \theta) \tag{5.14}$$

In Figure 5.3a, $A = 10$, $u = \frac{5}{N}$, and $\theta = 90°$. This signal is composed of a single component that has an integral number of periods over the data domain $x = [0 \ldots N]$. The dotted lines in the signal plot show how the signal repeats outside the window extent, as considered by the DFT. The magnitude is plotted in *decibels*, with the reference magnitude for all plots in Figure 5.3 the maximum magnitude from the signal in Figure 5.3a. Because there is only a single frequency ($u = \frac{k}{N}$) in the signal and all of its energy falls into a single DFT bin, the magnitude plot for this signal exhibits two sharp spikes corresponding to frequencies $\pm\frac{k}{N}$. The magnitude and phase computed for this DFT component are $A = 10.0$ and $\theta = 90.0°$, matching the input signal precisely.

Figure 5.3b presents the result obtained when a window is applied to the signal from Figure 5.3a. Here, a Blackman-Harris window is used because it provides a good balance between the width and fall-off of the frequency domain representation [Har78]. The effect of spectral leakage, though is evident in this case—energy from the single frequency component in the signal spreads into neighboring frequencies. The signal parameters estimated in this case are $A = 10.0$ (when rescaled to account for windowing) and $\theta = 90.0°$, matching the input signal precisely.

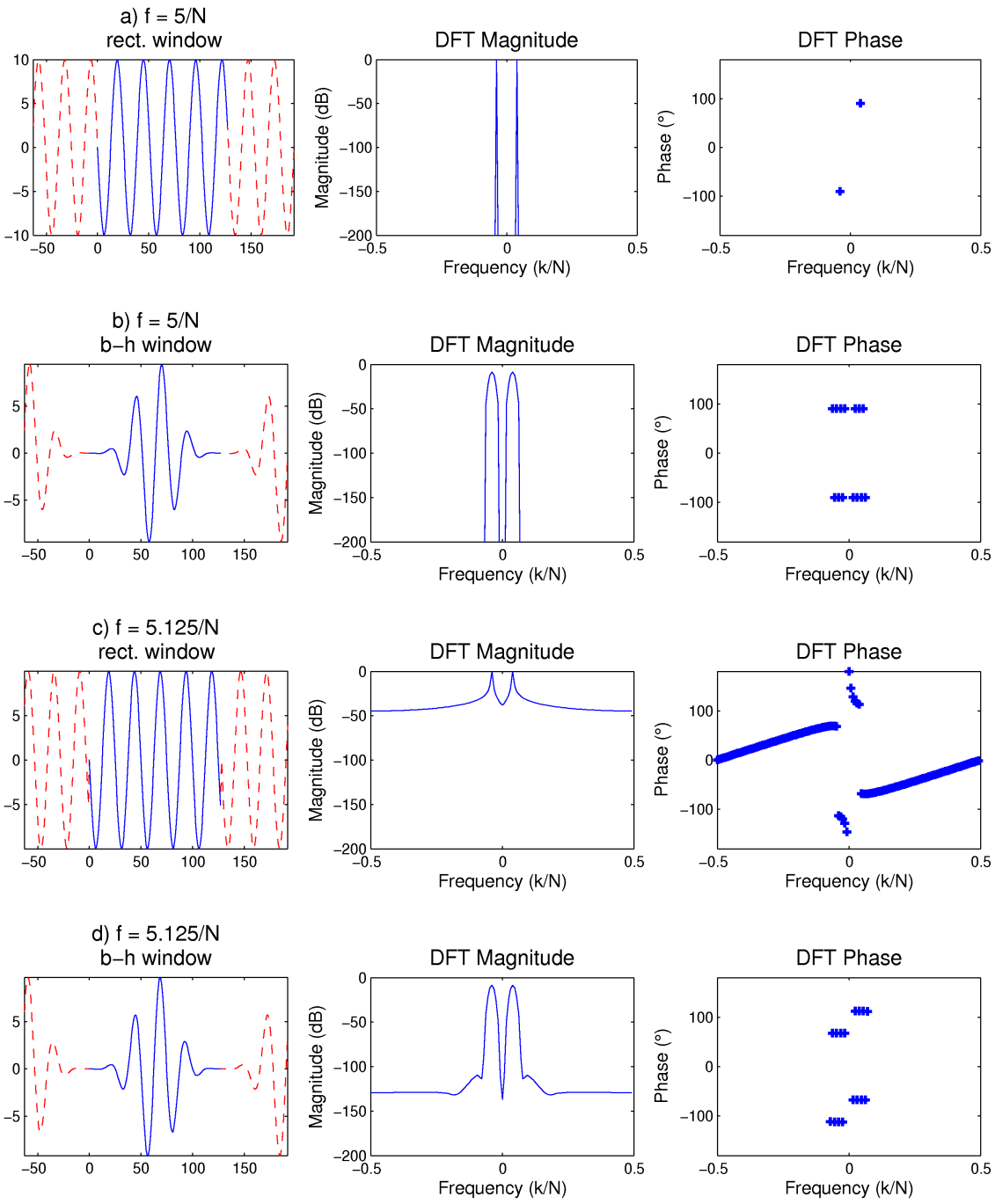


Figure 5.3: Signals containing single frequency components and their computed DFTs. The magnitude scale for each DFT is normalized to set the maximum response from the signal in (a) at 0 dB. Phases are not displayed where the magnitude drops below −100 dB because any phase estimate based on a low signal is unreliable. a) The signal has an integral number of periods in the data domain. b) The signal from (a) is multiplied by a Blackman-Harris window. c) The signal has non-integral number of periods in the data domain. d) The signal from (c) is multiplied by a Blackman-Harris window.

Figure 5.3c presents a signal composed of a single component that has a non-integral number of periods over the data domain—the function has the same form as that of Figure 5.3a with $u = \frac{5.125}{N}$. The dotted lines show how the signal repeats outside the window extent, as considered by the DFT. Although the frequency present in this signal differs only slightly from that of Figure 5.3a, the non-periodicity at the data boundaries introduces a step edge to the signal, which has energy at *all* frequencies, as seen in the magnitude plot. The magnitude plot shows that the peak amplitude is decreased slightly ($A = 9.64$), but energy from the single frequency component present in the signal leaks energy into several neighboring DFT bins. The most important consequence for phase correlation, however, is that the phase shift measured at the peak amplitude is grossly mis-estimated at $\theta = 112.7°$.

Figure 5.3d presents the result obtained by applying a Blackman-Harris window to the signal from Figure 5.3c. In the frequency domain, the windowing function attenuates the signal to be continuous at the boundaries, eliminating the step edge seen in Figure 5.3c. The peak magnitude and phase are estimated to be $A = 9.94$ and $\theta = 112.5°$, respectively. Although the windowing has removed the discontinuity at the data edges, the phase estimate remains inaccurate because the input signal does not exactly match one of the DFT bins.

Weighted Phase Correlation

As discussed by Harris, the use of different windowing functions improves the signal detection task—determining whether a signal is present [Har78]. But windowing does not necessarily improve the phase estimation task. Because phase estimation is crucial in pattern analysis, I present a frequency-specific phase analysis technique, weighted phase correlation (WPC), that computes individual components of the Fourier transform for the frequencies present in a pattern.

Recall that the discrete Fourier transform (DFT) is the correlation of a signal with a

set of sinusoidal basis functions. Equation 5.13 provides the DFT for a 1D signal $I(x)$,

$$\mathcal{I}(k) = \sum_{x=0}^{N-1} I(x) e^{-i\frac{2\pi kx}{N}} \qquad k = 0 \ldots N-1. \tag{5.15}$$

Let f_u be a target frequency defined by a wave number u such that $f_u = \frac{u}{N}$. For now, let the wavelength, $\lambda_u = \frac{1}{f_u}$, be an integer ($\lambda_u \in \mathbb{Z}$); this constraint is later relaxed. Let $I(x)$ be a signal containing only frequency f_u. The DFT $\mathcal{I}(k)$ has a single magnitude spike at $k = u$ if f_u cycles an integral number of times in the data domain (*i.e.* $u \in \mathbb{Z}$). Even if this is not the case, however, the phase at frequency f_u in the signal can be computed through a reweighting of the input signal—this is the goal of WPC.

Because $I(x)$ is periodic at intervals of λ_u, the partial summation

$$P_\lambda(x) = \sum_{j=0}^{j \leq \frac{N}{\lambda_u}} P(x + j\lambda_u), \qquad x = 0 \ldots \lambda_u - 1, \tag{5.16}$$

aligns all components of $I(x)$ that come from the same position along the wavelength and would be multiplied by the same part of the Fourier basis function in Equation 5.15. The number of terms from the input signal that contribute to each point on $P_\lambda(x)$ is given by

$$W_\lambda(x) = \sum_{j=0}^{j \leq \frac{N}{\lambda_u}} 1, \qquad x = 0 \ldots \lambda_u - 1. \tag{5.17}$$

A reweighted signal,

$$I_\lambda(x) = \frac{P_\lambda(x)}{V_\lambda(x)}, \tag{5.18}$$

is an appropriately scaled representation of $I(x)$ over a single cycle. At this point, a windowing function can be applied to the reweighted signal. The single frequency component at the target frequency f_u can be computed as

$$\mathcal{I}'(u) = \sum_{x=0}^{\lambda_u - 1} I_\lambda(x) e^{-i\frac{2\pi ux}{N}}. \tag{5.19}$$

The phase component of $\mathcal{I}'(u)$ is the same as the phase of f_u in the input signal. The magnitude component differs by a scaling factor dependent on the number of repetitions of the signal over the original data domain. Figure 5.4 depicts WPC for a 1D signal with a single frequency component $f_u = \frac{4.5}{N}$.

This formulation has assumed that the wavelength of the target frequency is integral so that the correlations can proceed as component-wise multiplication and summation. If λ_u is non-integral, the input data can be resampled to enforce this condition. The proper resampling technique chooses a resampling interval and determines precisely how much energy from each initial signal sample falls within each resampling bin. A simple resampling approach chooses a resampling interval that is slightly larger than the original data interval so that each input sample point maps into no more than two resampling bins, followed by reconstruction of the signal data with a triangular kernel (linear interpolation). Comparisons of both approaches on a 2D implementation of WPC showed no practical difference between the estimated phases.

Extending WPC to 2D requires accounting for the orientation of the target frequencies in the image, $(f_u = \frac{u}{N}, f_v = \frac{v}{N})$. For an axis-aligned pattern, each axis can be treated independently; that is, $v = 0$ for $u \neq 0$ and $u = 0$ for $v \neq 0$. An example of 2D WPC appears in Figure 5.5. The first step in 2D WPC is to axis-align the pattern by rotating the image. The rotated image contains regions of invalid pixels, where the image data is unknown; these pixels are discounted from the weighting function (these regions are visible, for example at the top and bottom of Figure 5.5d). Any pixels that receive no contributions during the partial summation are discounted from the reweighting step because reweighting is undefined at those locations. Otherwise, the 2D computation follows very closely from the 1D computation.

Note, however, that the reweighting in WPC does not avoid windowing. Although it address the problem of the target frequency not falling within a single DFT bin, it does not address the problem of *other* frequency components in the input signal that repeat

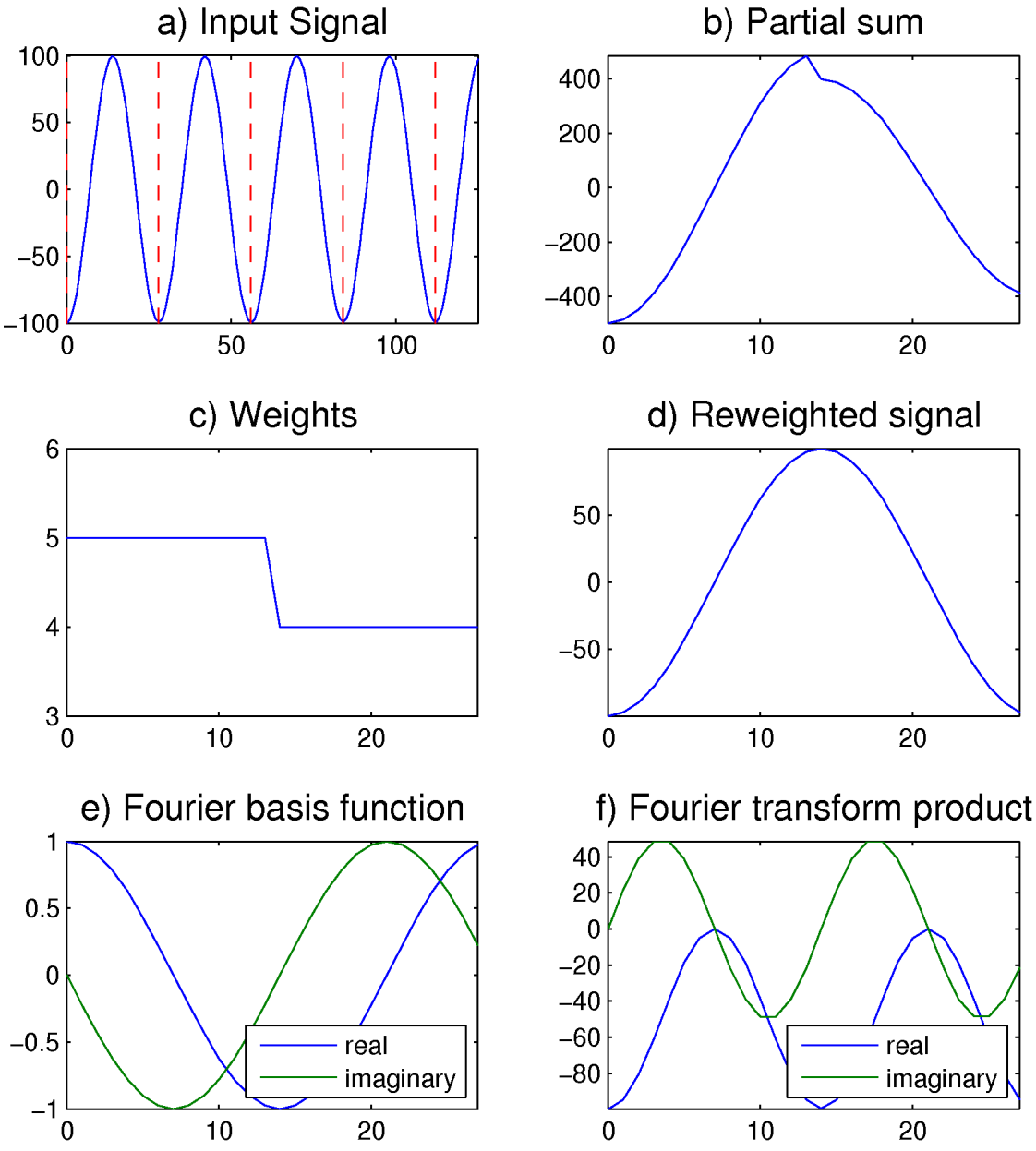


Figure 5.4: Weighted phase correlation (WPC) illustrated in 1D. a) Original input signal. Target wavelength is marked with dashed vertical lines. b) Partial sum of (a) by target wavelength. c) Weights showing the number of contributions for each element of (b). d) Partial sum (b) reweighed by (c). e) Fourier basis function with target wavelength. f) Weighted signal (d) multiplied by Fourier basis function (e).

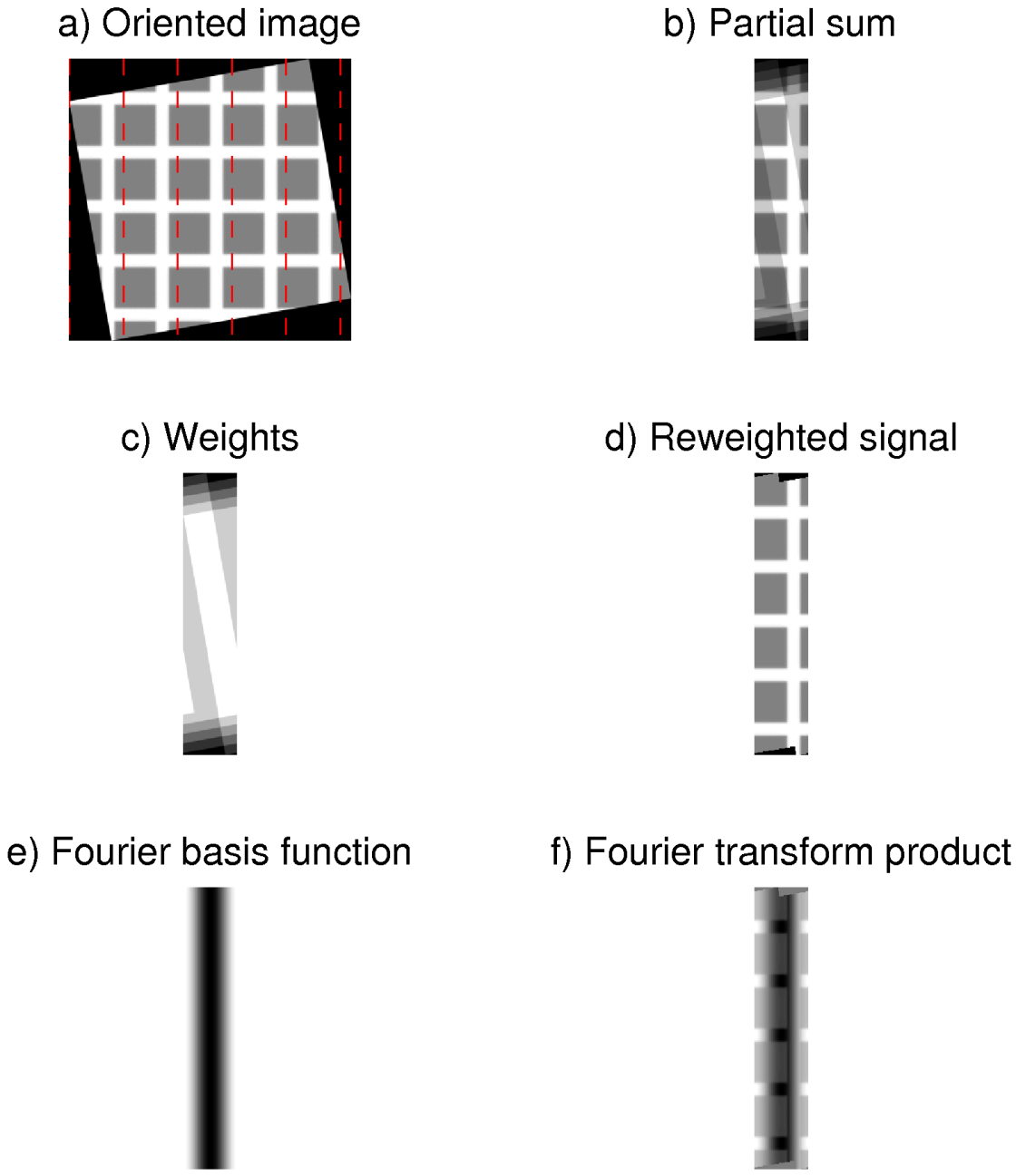


Figure 5.5: Weighted phase correlation (WPC) illustrated in 2D. a) Axis-aligned input image. Target wavelength is marked with dashed vertical lines. b) Partial sum of (a) by target wavelength. c) Weights showing the number of contributions for each element of (b). d) Partial sum (b) reweighed by (c). e) Fourier basis function with target wavelength, real component. f) Weighted signal (d) multiplied by Fourier basis function (e), real component.

a non-integral number of times in the windowed domain. The data has been aligned and reweighted according to a particular frequency that is present in the signal, but there are many other frequencies present which still produce leakage in the frequency domain. A windowing function can be employed as usual to attenuate the effect of discontinuities at the data boundaries. But, any fractional wavelength segment of the input signal may introduce discontinuities *in the middle* of the data domain. This is particularly bad for rotated images, which have discontinuities along the diagonal boundary between the signal domain and the unknown data domain.

Comparing the phases for pattern frequencies in two images yields a phase shift corresponding to the translation of the pattern in the images. The phase shift can be measured for any frequency present in the pattern (*e.g.* the fundamental frequency and harmonics in the square grid) along each axis. Phase shifts from multiple frequencies can be combined to improve the translation estimate, with proper consideration that the phase shifts from higher frequencies may indicate multiple cycles over the translation. Computing the translation between a set of images and a reference image provides the translations that register all images to a common reference frame.

In summary, the process described is similar to the computation of particular DFT components. However, the computation is carried out for a specific set of frequencies with potentially non-integral wave number, and a reweighting is applied to account for this and for missing data in the input image. The result of this discrete, frequency-specific, weighted correlation can be used to compute the phase for any frequencies that provide information about the pattern in the input image.

WPC is evaluated in terms of phase estimation in 1D in Section 5.2 and for tracking structured illumination microscopy patterns in Section 5.3.

5.2 Comparing DFT and WPC in 1D

Estimating phase parameters using general phase correlation is not usually a problem for wide bandwidth signals, such as when a complete image undergoes translation. In this case, phase correlation relies on corroborating measures from many frequencies in the full DFT to arrive at an average phase shift. For narrow bandwidth applications, however, the phase shift estimation is performed on a small number of samples in frequency space, and for each sample there is significant prior knowledge about the signal—*i.e.* its periodicity is known. The WPC method relies on this prior knowledge and removes the requirement that the signal has an integral number of repeats in the image.

Comparative evaluation of the phase estimation task for 1D signals of different frequencies using DFT and WPC provides further justification for the WPC approach. Following the algorithm evaluation approach outlined by Trucco and Verri in [TV98, A1], in this analysis a set of signals representing a range of test scenarios are generated. Each signal contains a single frequency component with a known amplitude and phase shift, with the same form as Equation 5.14:

$$f(x) = A\cos(2\pi ux + \theta). \tag{5.20}$$

For each test scenario, 20 instances of the signal are corrupted with zero-mean, Gaussian-distributed noise with standard deviation 10% of the signal amplitude. The phase shift is estimated using both the DFT and WPC techniques, and the mean phase estimate is computed from the 20 samples for each test scenario. The signal frequency is varied over the range $\{u \mid u = \frac{1}{\lambda},\ \lambda = [5, 6, \ldots N]\}$ and phase is varied over $\theta = [-\pi \ldots \pi]$. From all results, the root mean squared (RMS) error and standard error are computed

to summarize the behavior of each method. The RMS error is

$$\begin{aligned} E_i &= x_i - x_{T_i} \\ \text{RMSE} &= \sqrt{\frac{1}{N}\sum_{i=1}^{N} E_i^2}, \end{aligned} \tag{5.21}$$

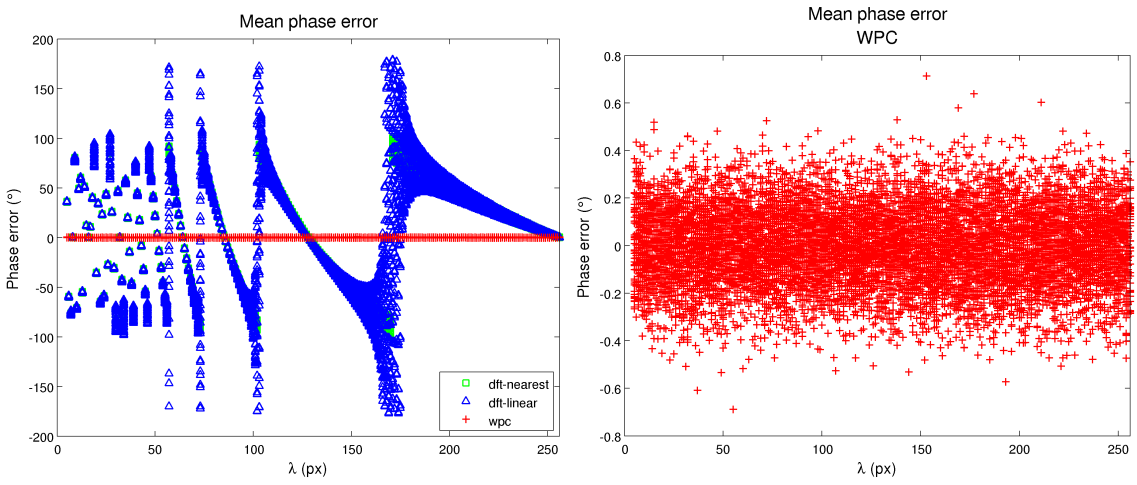
where x_i are the N computed mean parameter estimates and x_{T_i} are the known ground truth values. The standard error is

$$\sigma_E = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(E_i - \bar{E})^2}, \tag{5.22}$$

where $\bar{E}$ is the mean of all computed errors [PTVF07, Ch. 14]. The RMS error and standard error provide an estimate of accuracy and error in the accuracy estimate, respectively.

Nearest neighbor and linear interpolation are used when estimating phase from the DFT when the target frequency falls between DFT bins. Magnitude interpolation is straight forward, as one can interpolate the magnitude only of the DFT components. Phase interpolation, however, should be performed on the complex values of the DFT to handle averaging angles appropriately. It is possible that the interpolated complex value will have very small magnitude, for example if the interpolation requires averaging a phase of 0° and 180°. If the magnitude of the interpolated complex values is too small (*e.g.* less than −100 dB), the angle estimate is not reliable and should be discarded. Note that the magnitude of interpolated complex values is not the same as the interpolated magnitude of the DFT.

Figure 5.6 plots the phase estimation error for each estimation method using a rectangular window. WPC performs well throughout the range of tested frequencies, $u = [5 \ldots N]$. DFT, however, only performs well at frequencies that are periodic in the window domain, $\{u \mid u = \frac{k}{N}, k = [1 \ldots N-1]\}$.



(a) (b)

Figure 5.6: Phase estimation for a target wavelength using the DFT and WPC methods on a signal containing one frequency. a) The mean phase error for all methods using a rectangular window. DFT with nearest neighbor and linear interpolation are nearly identical. b) The mean phase error from (a), for the WPC method only. Note the change in y axis scale.

Table 5.1: Comparison of phase estimation using DFT and WPC on 1D single frequency signals with 10% Gaussian-distributed noise. Nearest neighbor and linear interpolation are used to estimate frequency components that fall between DFT bins, as indicated.

Window	**Method**	**RMSE** θ (°)	$\sigma_E\theta$ (°)
Rectangular	DFT-nearest	50.5966	0.4994
	DFT-linear	52.3382	0.5171
	WPC	0.1610	0.0016
Hamming	DFT-nearest	50.6685	0.5001
	DFT-linear	50.5222	0.4987
	WPC	0.1898	0.0019
Hanning $\alpha = 1.0$	DFT-nearest	50.4066	0.4966
	DFT-linear	49.2724	0.4855
	WPC	0.5764	0.0057
Kaiser $\alpha = 1.5$	DFT-nearest	50.4774	0.4960
	DFT-linear	51.5416	0.5066
	WPC	0.9864	0.0098
Gaussian $\alpha = 2.5$	DFT-nearest	50.5029	0.4985
	DFT-linear	49.1729	0.4854
	WPC	1.7394	0.0173

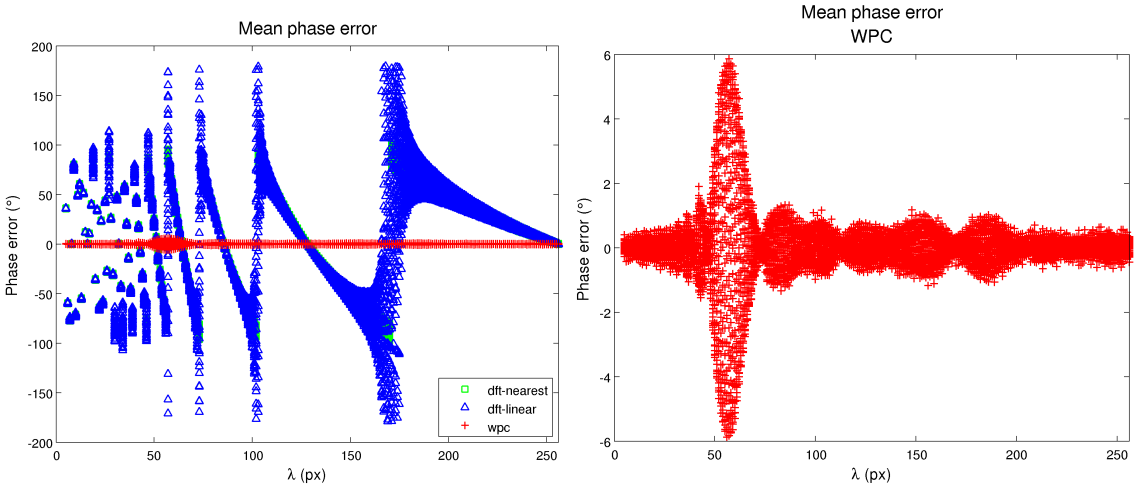
Table 5.1 records the RMS error and standard error for phase estimation using DFT and WPC with several popular windowing functions. Parameters are provided for some windowing functions, using the terms provided by Harris [Har78]. These parameters adjust the width and fall off of the functions' Fourier transforms. In each case, the best of several parameter trials are reported. None of the DFT methods reliably estimates phase parameters over the entire frequency range, regardless of windowing function. WPC, however, performs considerably better, with the best result obtained with the rectangular windowing function (no windowing).

This evaluation highlights one of the shortcomings of using the DFT for phase estimation using a small set of frequencies: there are only a few frequencies for which the estimation will be accurate! Chapter 6 discusses the characteristics of well-designed patterns for structured illumination microscopy, but to foreshadow a little, it is desirable to have a number of long wavelength components in the pattern, as this enables tracking over longer ranges and with greater degrees of defocus. WPC enables the use of many more possible wavelengths for which phase estimation will be accurate.

The situation changes, however, in the presence of other frequency components. Figure 5.7 shows the phase estimation error for a signal of the form

$$f(x) = A\cos(2\pi ux + \theta) + 0.1A\cos(2\pi u_c x). \tag{5.23}$$

That is, the single frequency signal of Equation 5.20 with a corrupting signal added with 10% amplitude. In this case, the corrupting signal's wavelength is $\lambda = 56.5$, and therefore does not align with the target frequency in any test case. Figure 5.7 illustrates that the ability of WPC to estimate phase at any frequency is degraded by the presence of this corrupting signal, and is significantly more degraded at frequencies near the corrupting signal's frequency. The maximum error in this situation is approximately 6°, or 1.7%.



(a) (b)

Figure 5.7: Phase estimation for a target frequency using the DFT and WPC methods on a signal containing two frequencies. a) The mean phase error for all methods using a rectangular window. DFT with nearest neighbor and linear interpolation are nearly identical. b) The mean phase error from (a), for the WPC method only. Note the change in y axis scale.

Table 5.2: Comparison of phase estimation using DFT and WPC on 1D signals containing two frequencies with 10% Gaussian-distributed noise. Nearest neighbor and linear interpolation are used to estimate frequency components that fall between DFT bins, as indicated.

Window	**Method**	**RMSE** θ (°)	$\sigma_E\theta$ (°)
Rectangular	DFT-nearest	50.6088	0.4996
	DFT-linear	52.4114	0.5177
	WPC	0.9384	0.0093
Hamming	DFT-nearest	50.6824	0.5003
	DFT-linear	50.6913	0.5000
	WPC	1.0677	0.0106
Hanning $\alpha = 1.0$	DFT-nearest	50.7132	0.5006
	DFT-linear	51.1447	0.5047
	WPC	1.1182	0.0111
Kaiser 1.5 $\alpha = 1.5$	DFT-nearest	50.5523	0.4990
	DFT-linear	51.6387	0.5098
	WPC	1.2807	0.0128
Gaussian 2.5 $\alpha = 2.5$	DFT-nearest	50.5149	0.4986
	DFT-linear	49.4103	0.4874
	WPC	2.0495	0.0204

Table 5.2 records the RMS error and standard error for phase estimation of the corrupted signals using DFT and WPC and the windowing functions from Table 5.1. The results are familiar—windowing does not improve phase estimates of single frequency components. While the accuracy of WPC is not as good in the presence of corrupting signals, it performs better than DFT. Whether this is sufficient for tracking patterns in structured illumination microscopy depends to a large extent on the specimen being observed. This issue is discussed in greater detail in Section 5.4.

Having made a case for the utility of WPC, I now turn to evaluate the lateral tracking techniques applied to structured illumination microscopy.

5.3 Evaluation Using Simulated Images

A combination of simulation and real experiments are discussed to evaluate the two lateral microscope stage tracking approaches described in Section 5.1.2, weighted phase correlation (WPC) and model-based spatial correlation (MSC). The experimental data uses a square grid micropattern printed on a cover slip, as described in Section 5.4. Simulated images of this pattern undergoing translation are used to evaluate the tracking methods using known ground truth displacements. The simulated pattern is based on a square transmission electron microscope (TEM) grid composed of a set of orthogonal $6\,\mu$m bars spaced $25\,\mu$m apart. This arrangement results in a set of regular square regions with a $19\,\mu$m hole in the center. The simulation is designed to mimic the way the pattern used in real experiments is imaged in the microscope, considering the following parameters:

- The physical size of the grid pattern ($25\,\mu$m)
- The objective lens used for observation (40X, 0.65NA)
- The optical transmission of the pattern

- The orientation of the pattern
- The axial position of the pattern
- The size of the CCD sensor elements ($9 \times 9\,\mu$m)
- The shot noise from image acquisition (Poisson noise with 100 photoelectron per count ADC gain)

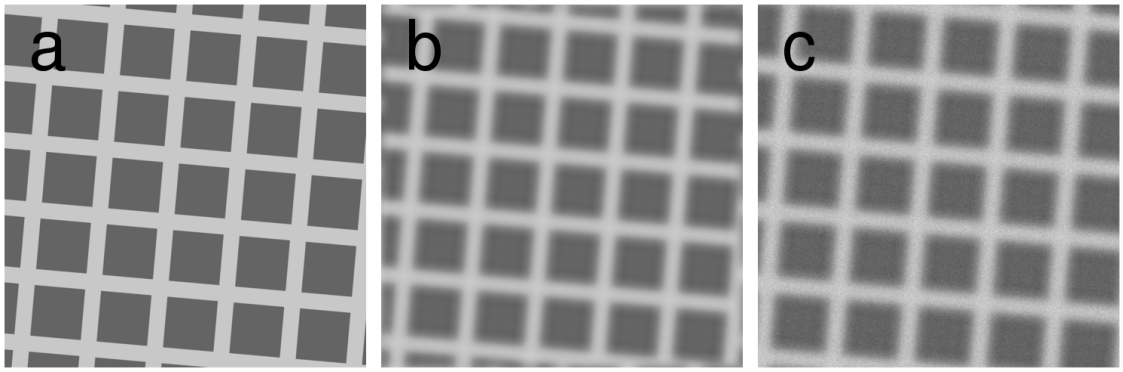


Figure 5.8: Steps used to simulate imaging of a structured pattern with transmission $\alpha = 0.5$. a) Simulated image predicted by geometrical optics. b) Image convolved with analytical PSF for $dz = 5\,\mu$m. c) Resampled image with Poisson-distributed noise.

The first step in simulating an image of the grid pattern is to compute the image formed on a virtual microscope's CCD predicted by geometrical optics. The pattern is given a transmission α and is illuminated by a uniform light source of intensity L. An image of the pattern is generated by evaluating an analytical function of the grid with a specific orientation θ and translation $[dx, dy]^T$ with 4×4 oversampling for each CCD element. An example of the image predicted by geometrical optics is shown in Figure 5.8a.

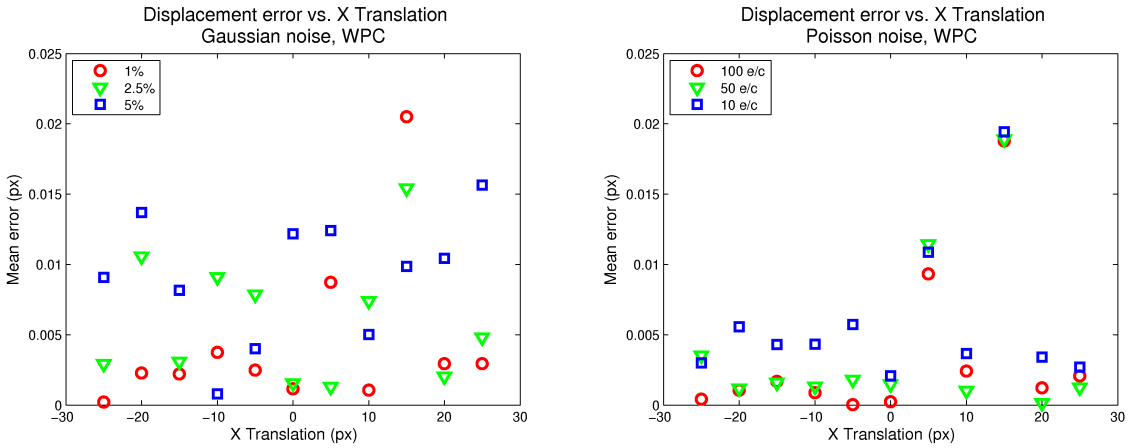
This supersampled image is convolved with the point spread function predicted by Fourier optics for the grid displaced from the microscope's focal plane by a distance dz, as described in Section 2.1.2. An example of the image predicted by Fourier optics is shown in Figure 5.8b.

The Fourier optics image is down sampled to obtain the ideal intensities read by each CCD pixel. Noise is added to each pixel to model shot noise using either a Gaussian or Poisson noise distribution. The Poisson noise model simulates the shot noise seen in a real camera—the number of photons arriving at a sensor site is governed by a Poisson random process, as explained in Chapter 2 [HK94]. In image simulation, a parameter modeling analog-to-digital converter (ADC) gain (photoelectrons per count) converts ideal intensity values to photoelectron counts at each pixel. A Poisson-random number with variance equal to the electron count is generated at each pixel and converted back to an intensity value. With high ADC gains (high photoelectron/count ratios), however, the Poisson distribution can be closely approximated by a Gaussian distribution [Pit93]. For the Gaussian noise model, Gaussian-distributed noise is added to each pixel with a standard deviation that is a percentage of the brightest image value (equal to the lamp intensity). The noisy image is finally quantized to whole-integer intensity values to produce the final simulated image of the pattern, as seen in Figure 5.8c.

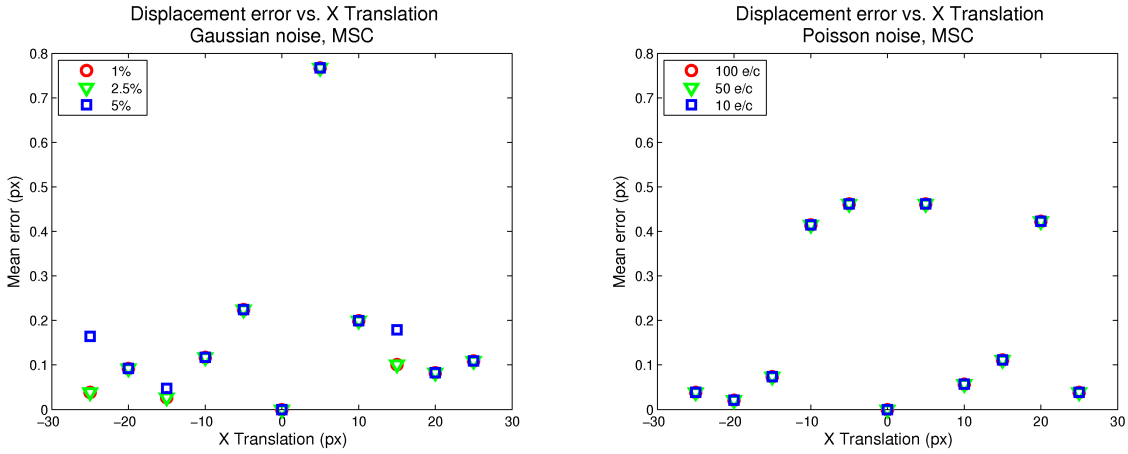
This process enables evaluating the tracking algorithm for varying pattern transmissions, rotations, translations, focus positions, lens parameters, and noise models. The following analysis again makes use of the approach outlined in [TV98, A1]. I investigate varying each of the system parameters to evaluate the effect on the tracking system. For each test case, 5 images are generated with noise to simulate repeated measures. For each simulated image, the rotation of the pattern and the displacement from a reference image is measured using WPC and model-based spatial correlation (MSC). The difference between the mean parameter estimate and the ground truth (the mean error) indicates how accurately each tracking algorithm handles a particular test scenario. The standard deviation of measurements indicates the algorithm's robustness to noise. Where appropriate, the RMS error and standard error is also computed over all varied parameters.

5.3.1 Shot Noise

The first set of test cases evaluates the effect of camera shot noise on the tracking algorithm. The simulated grid is given a transmission of $\alpha = 0.5$, placed unrotated and in focus, and translated between -25 and 25 pixels along the x-axis. The grid image is corrupted with either Gaussian-distributed noise (with standard deviation a percentage of the maximum intensity value) or Poisson-distributed noise (with a simulated ADC gain).



(a)



(b)

Figure 5.9: Displacement error for tracking a simulated in focus grid pattern with transmission $\alpha = 0.5$ corrupted with different noise models. Left: Gaussian noise with standard deviation a percentage of maximum image intensity. Right: Poisson noise with different ADC gains (electrons / count). a) Error using WPC tracking. b) Error using MSC tracking. Note the change in y axis scale.

Figure 5.9 shows the mean displacement error for different noise models with the two tracking methods. The tracking accuracy for WPC is consistently high, independent of noise model, with higher levels of noise disrupting the accuracy more. The WPC method attains far more accuracy than the MSC method in this scenario. The MSC method is nearly invariant to noise levels in this range. These results also indicate that tracking error does not depend on translation distance.

Table 5.3: RMS displacement error and standard error (σ_E) for different noise models using the two lateral tracking methods. Results are for tracking an in focus square grid pattern with transmission $\alpha = 0.5$. σ_s denotes the standard deviation of the shot noise.

		WPC		**MSC**	
Model	σ_s (counts)	**RMSE D** (px)	σ_E**D** (px)	**RMSE D** (px)	σ_E**D** (px)
Gaussian $\sigma = 1\%$	2	0.0071	0.0017	0.2581	0.0641
Gaussian $\sigma = 2.5\%$	5	0.0074	0.0014	0.2581	0.0641
Gaussian $\sigma = 5\%$	10	0.0101	0.0013	0.2666	0.0621
Poisson g = 100 e/c	1.41	0.0064	0.0017	0.2700	0.0603
Poisson g = 50 e/c	2	0.0068	0.0018	0.2700	0.0603
Poisson g = 10 e/c	4.47	0.0076	0.0015	0.2699	0.0603

Table 5.3 shows the estimated displacement RMS error and standard error for all translations under the different noise models. This table reports the shot noise standard deviation, σ_s, for each scenario; for WPC, similar shot noise levels produce comparable tracking errors using both the Gaussian and Poisson models. This indicates that the distributions are interchangeable for modeling shot noise, especially when high ADC gains are used. A Poisson noise model with a gain of 100 electrons/count is used in all subsequent evaluations, to approximately match the noise level measured for the Pulnix camera calibrated in Section 2.3.

The RMS displacement error remains bounded by 0.010 pixel for all noise models using WPC, which corresponds to an error of 2.3 nm for the simulated microscope's configuration. Some caution should be used in interpreting this value, as this error

represents the scenario of imaging an in focus pattern with precisely known spacing and orientation and no corrupting specimen.

The MSC method fair far worse in this comparison, maintaining a nearly-constant error of 0.25 pixel, or 56.3 nm. To foreshadow a little, Chapter 6 demonstrates that this level of error is due to the pattern and optimization strategy, and MSC can be improved with different patterns.

The remainder of the system parameters fall into two categories: those that do not degrade the pattern signal (rotation and translation) and those that do (transmission and focus). I will evaluate parameters that do not degrade the signal first.

5.3.2 Rotation

The pattern and camera sensor can be placed at any arbitrary orientation. The tracking algorithm relies on being able to both accurately determine the orientation of the pattern and track the pattern at any orientation. These tasks are evaluated independently, with two methods used for each. Orientation is evaluated with the frequency-based estimation described in Section 5.1.1 and the model-based refinement associated with MSC tracking. The model-based method depends on initial estimates of both the pattern position and orientation while the frequency-based method does not. In this evaluation, the estimate provided by the frequency-based method is used to initialize the model-based method to reflect how the two methods are used in practice. The position estimate is provided by the known ground truth position of the pattern.

Figure 5.10 shows the mean orientation error obtained by the two methods for pattern rotations within $\theta = [0 \ldots 90]^\circ$. Orientations outside this range repeat for this square pattern. These data show that using the frequency-based estimation, the orientation error is bounded within $\pm 0.4^\circ$ of the true value. The offset of a pixel at the edge of the 256 pixel^2 test images whose orientation has been miscalculated by this amount is $128 \tan(0.4) \approx 0.89$ pixel ($0.20\,\mu$m). Over all estimates, the RMS error us-

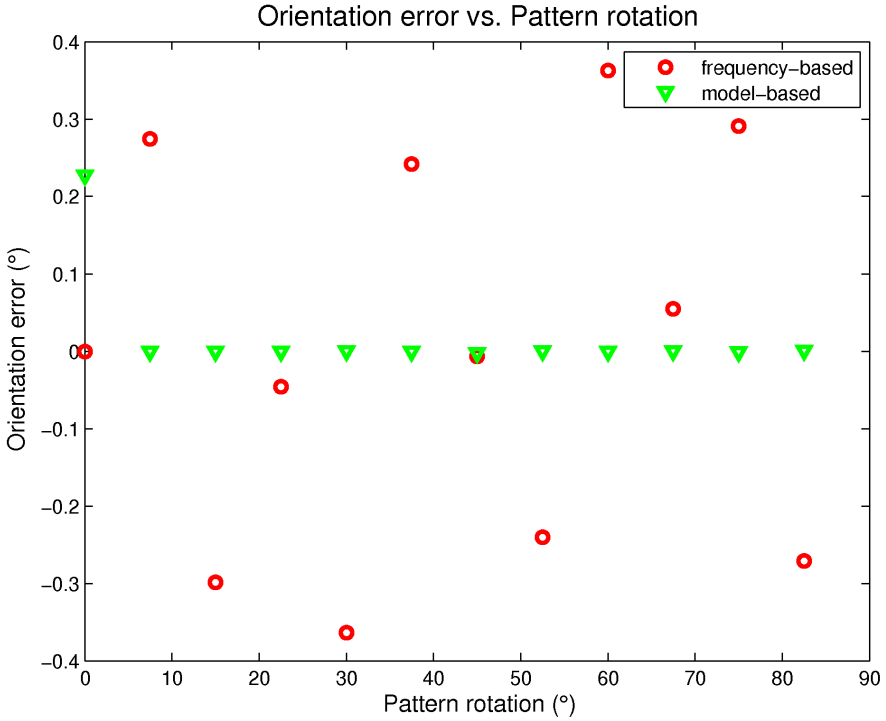
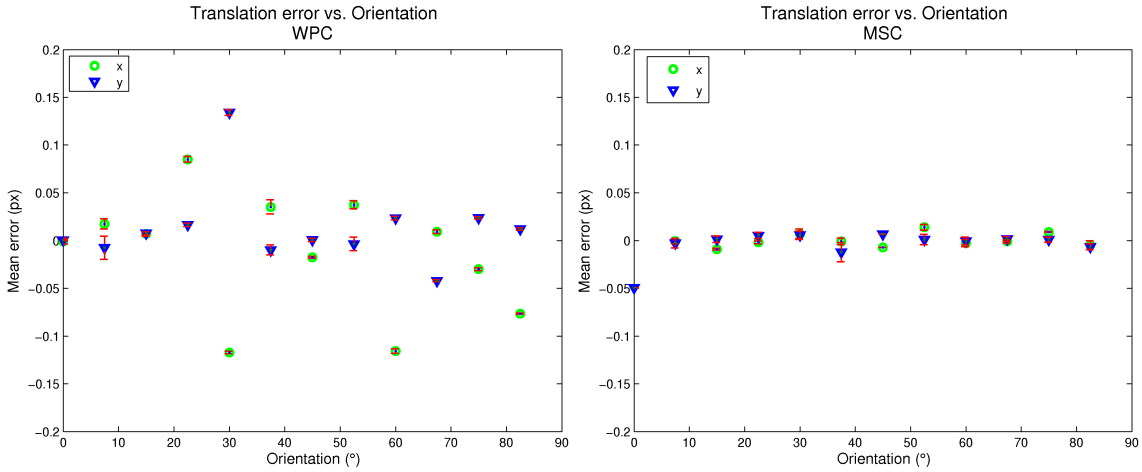


Figure 5.10: Orientation estimation error in response to different pattern rotations ($\theta \in [0 \ldots 90]^\circ$) for an in focus square grid pattern of transmission $\alpha = 0.5$, corrupted with Poisson-distributed noise (gain = 100 e/c). The initial orientation for the model-based estimation is provided by the frequency-based estimation. Error bars are too small to be resolved on this graph.

ing frequency-based estimation is 0.24° with standard error 0.073°. The model-based estimation provides nearly a factor of four improvement, with an RMS error of 0.066° with standard error 0.019°. This presents a strong argument for refining the orientation estimate after obtaining a position estimate.



(a) (b)

Figure 5.11: Translation error obtained for tracking a grid pattern displaced $d\mathbf{x} = [10, 0]^T$ pixel with different rotations. Vertical error bars indicate the standard deviation for each measurement taken over 5 samples. a) WPC translation error in x and y. b) MSC translation error in x and y.

Figure 5.11 shows the mean translation errors obtained using the two tracking methods where the pattern orientations are provided from ground truth values and the pattern translation is $d\mathbf{x} = [10, 0]^T$ pixel. Once again, MSC outperforms WPC for most scenarios, with exceptions at 0 and 45° (the x translation error for MSC at 45° is off the chart at 0.45 pixel). The large errors with MSC are again attributable to the particular pattern being tracked. Table 5.4 summarizes these results (ignoring these gross errors) along with tracking errors for small pattern displacements, discussed next.

5.3.3 Translation

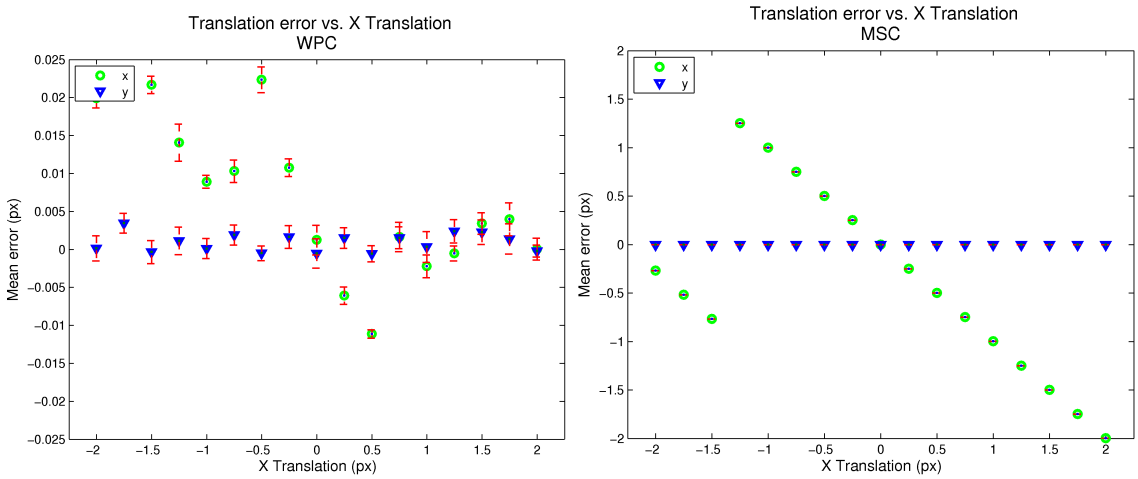
The previous test case demonstrated that the tracking method maintains accuracy at arbitrary orientations—even though the method is also dependent on estimating those

Table 5.4: RMS error and standard error in translation estimation using WPC and MSC for different pattern orientations and small translations. Gross errors have been discounted from the MSC orientation estimates (justification in text). All measurements in pixels.

	WPC		**MSC**	
Test Case	**RMSE x**	**RMSE y**	**RMSE x**	**RMSE y**
Orientation	0.0566	0.0330	0.0023	0.0115
Translation (large)	0.0064	0.0006	0.2695	0.0156
Translation (small)	0.0121	0.0015	1.0036	0.0000

orientations precisely. In an evaluation of the range of pattern translations, then, it suffices to investigate the translations in one direction for an axis-aligned pattern. The maximum frame-to-frame displacement along a pattern axis is limited to half the wavelength of the pattern's fundamental frequency along that axis. A greater displacement will alias to a displacement in the opposite direction, so larger displacements cannot be recovered accurately without additional knowledge. The error for large displacements of the pattern was already investigated during the discussion of shot noise in Section 5.3.5. The RMS translation errors for the Poisson-distributed noise with ADC gain of 100 photoelectrons / count are reported in Table 5.4.

Because the WPC tracking algorithm is based on phase comparisons and MSC is based on matching simulated images to an observed image, there is nothing that locks either algorithm to measuring whole-pixel displacements. Figure 5.12 shows the translation error in each direction for $\frac{1}{4}^{\text{th}}$ pixel displacements along the x-axis in the range $dx = [-2 \ldots 2]$ pixel. Table 5.4 records the RMS errors for these test cases. For this scenario, WPC vastly outperforms MSC, maintaining a worst-case error of 0.025 pixel (5.6 nm). The optimization of MSC, on the other hand, appears to be trapped in a local minimum, always reporting zero translation, which gives rise to the obvious structured error in Figure 5.12b. Once again, this is a problem that the square grid pattern poses to the optimization strategy; this will be discussed further in Chapter 6.



(a) (b)

Figure 5.12: Translation estimation error using WPC and MSC for small translations. Vertical error bars indicate the standard deviation for each measurement taken over 5 samples.

5.3.4 Transmission

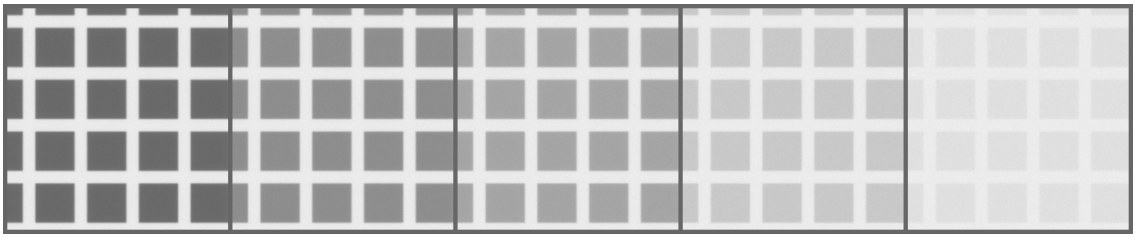


Figure 5.13: Simulated square grid pattern with varying transmissions, from left to right, $\alpha = \{0.45, 0.60, 0.70, 0.85, 0.95\}$.

The remaining test parameters degrade the quality of the signal used for tracking. The pattern transmission and lamp intensity together determine how much information is available in an image sequence for phase-based tracking. Optical transmission is the percentage of light that passes through a semitransparent object. The grid pattern acts as a transmission mask, segmenting the image plane into bright and dark regions. Figure 5.13 shows simulated images of square grid patterns with varying transmissions and constant lamp intensity.

Both the pattern transmission and distance to the focal plane affects the SNR of the

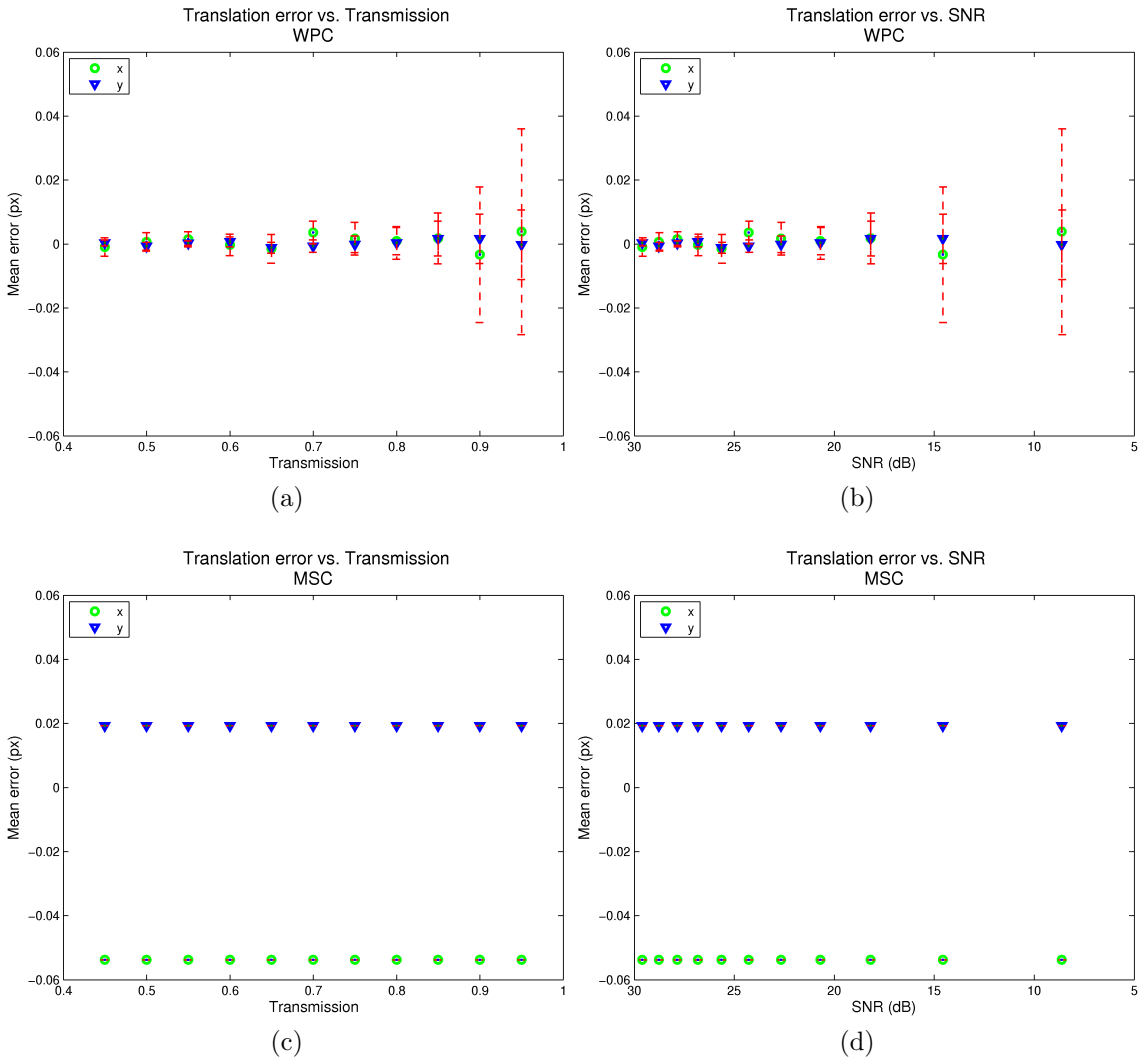


Figure 5.14: Translation estimation error for square grid pattern with varying optical transmission. Vertical error bars indicate the standard deviation for each measurement taken over 5 samples. Translation error for WPC as a function of (a) transmission and (b) SNR. Mean translation error for MSC as a function of (c) transmission and (d) SNR.

information available to track. For this reason, the SNR is provided for comparisons across these signal-degrading parameters. The value reported is the mean of SNRs at the frequencies used for tracking computed from simulated images containing the pattern only and noise only.

Figure 5.14 shows the mean displacement error and standard deviation measured for in focus patterns of varying transmissions with a displacement of $d\mathbf{x} = [10, 0]^T$. For WPC, the tracking precision begins to degrade for transmissions greater than $\alpha = 0.8$, or SNRs below 20 dB. However, because the mean measurements remain close to the ground truth values, the tracking accuracy can presumably be increased with repeated measures at the same pattern position. The pattern transmission has no effect on the MSC method in this scenario; and the error for this method is once again highly structured. At large SNRs, the error in the MSC method is significantly greater than that of the WPC method.

5.3.5 Focus

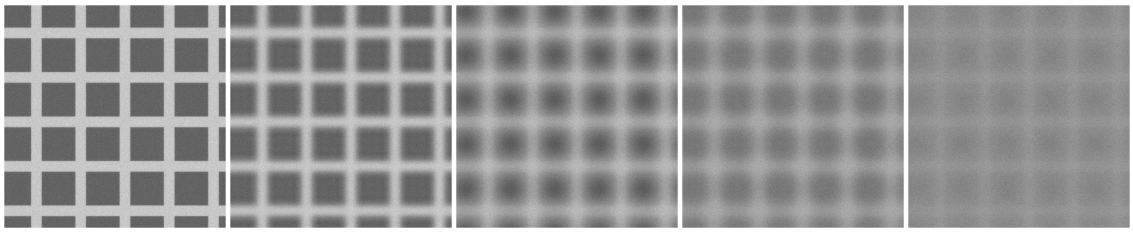


Figure 5.15: Simulated square pattern with transmission $\alpha = 0.5$ at different depths. From left to right, $dz = \{0, 5, 10, 15, 20\}$ μm.

The test scenarios presented thus far have demonstrated the accuracy of the WPC-based tracking method for an in focus pattern. In structured illumination microscopy, however, the pattern lies on a different plane from the specimen, and therefore is out of focus. The microscope PSF acts as a non-ideal low pass filter. As the pattern moves further out of focus, higher frequency components are attenuated to a greater degree, the pattern is smoothed, and the SNR decreases. Because higher frequency

components are attenuated more than lower frequency components, the signal from high frequency components used in tracking will fall off faster than the signal from low frequencies. This may limit the usefulness of higher harmonics in the square wave pattern. Figure 5.15 shows the square pattern at different axial distances from the focal plane of the microscope.

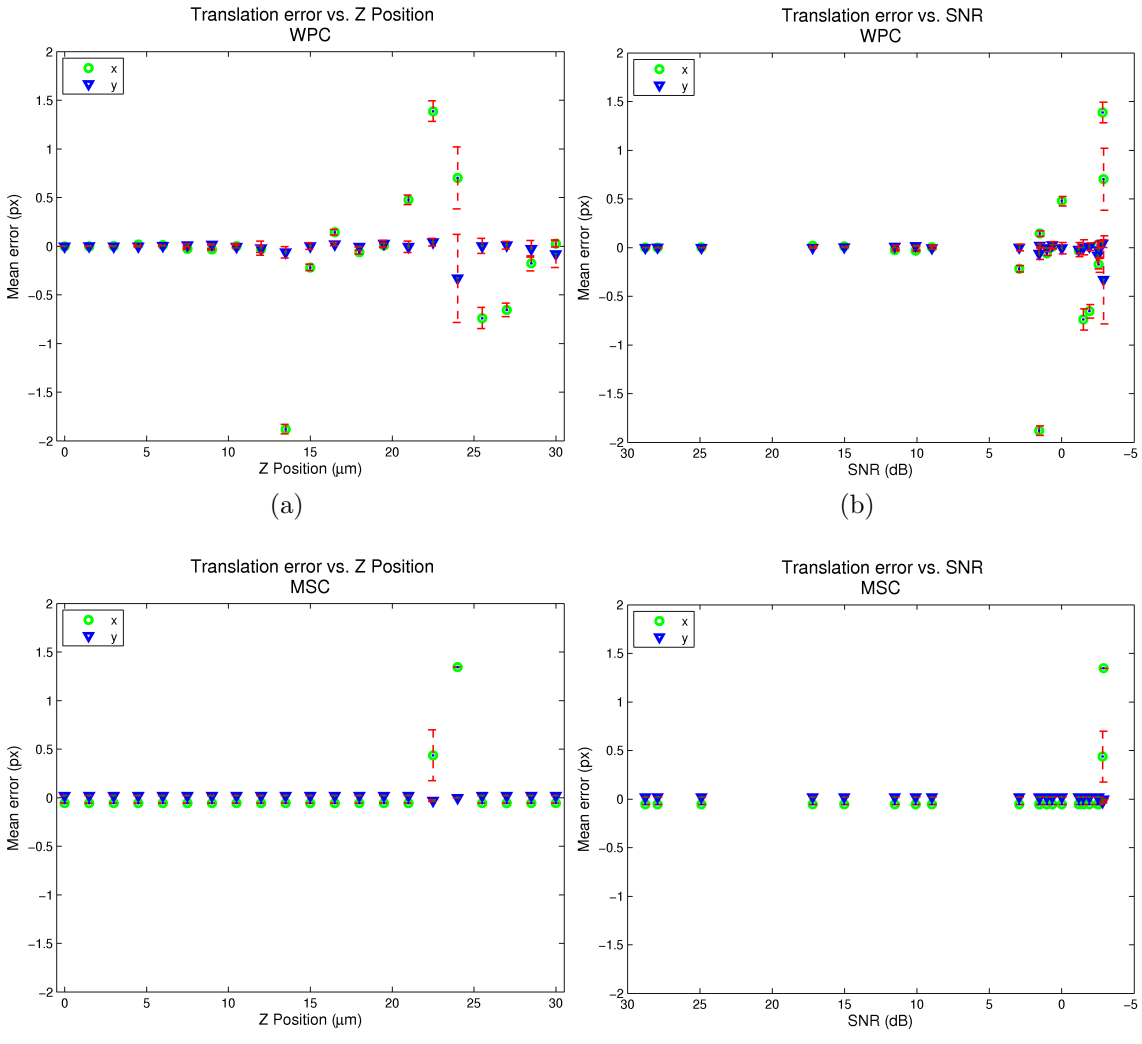
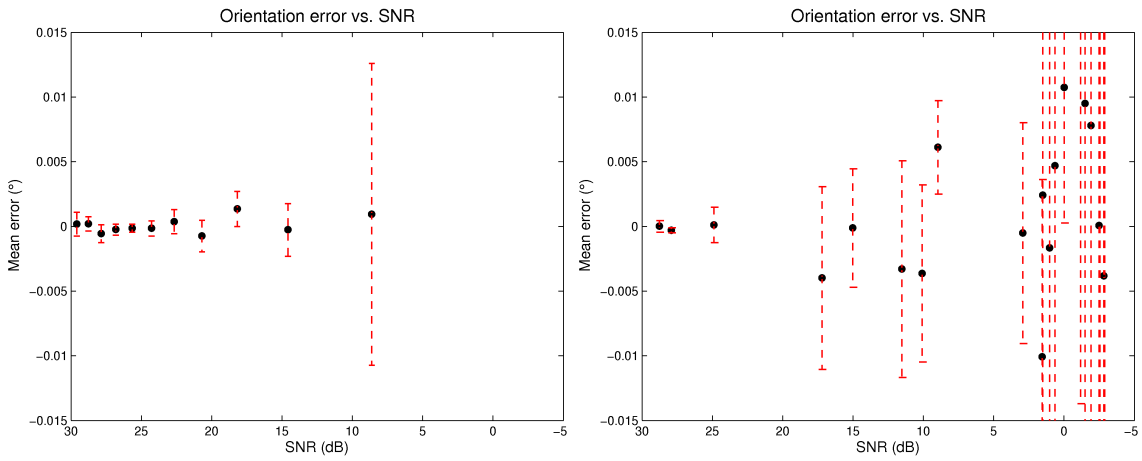


Figure 5.16: Translation estimation error for square grid pattern with varying distance to the focal plane. Vertical error bars indicate the standard deviation for each measurement taken over 5 samples. Translation error for WPC as a function of (a) z position and (b) SNR. Mean translation error for MSC as a function of (c) z position and (d) SNR.

Figure 5.16 shows the effect of defocus on translation estimation for images simulated

with a 40X, 0.65NA objective lens. The displacement of the square grid pattern from the focal plane ranges within $dz = [0 \ldots 30]\,\mu$m (the analytical PSF is symmetric about $dz = 0$). The pattern is displaced laterally by $d\mathbf{x} = [10, 0]^T$. These data show that the accuracy and precision of the displacement estimates fall off as distance from the focal plane increases. Tracking with WPC using four harmonics begins to deteriorate significantly above $dz = 20\,\mu$m. Note, however, that monotonically increasing z position does not correspond to monotonically decreasing SNR. (See Section 6.1 for more information on this.) For instance, the error for x translation at $14\,\mu$m is based on a SNR of about 2 dB. These results demonstrate the value of reporting the mean SNR at the frequencies used for pattern tracking. The WPC translation estimate degrades below 5 dB. The translation estimates from MSC show fewer large errors, with consistent performance down to 0 dB SNR. At high SNR, MSC again under performs compared to WPC.



(a) Transmission (b) Focus

Figure 5.17: Orientation error using frequency-based estimation for varying transmission and focus. Vertical error bars indicate the standard deviation for each measurement taken over 5 samples. a) Orientation error vs. SNR for varying transmission. b) Orientation error vs. SNR for varying z position.

As a final comparison, Figure 5.17 shows the effect of SNR on frequency-based orientation estimation. In all cases, the pattern rotation is 0°. Though the focus test case covers a greater range of SNRs, the general trend is the same in both situations. The

mean orientation estimate stays close to the ground truth value—within 0.015 pixel. This suggests that orientation estimates will be improved with a large number of samples, as is the case when tracking over a large number of video frames. The fact that the pattern orientation stays constant over the course of an experiment helps mitigate the effect of errors in individual orientation estimates.

5.4 Evaluation Using Real Images

The test cases presented in Section 5.3 provide upper bounds on tracking accuracy under the condition that no specimen is present. Real microscopy images were collected to demonstrate the technique in practical application. The greatest limiting factor in experimental evaluation was manufacturing the semitransparent micropatterns. In this section, I present a set of experiments using real microscopy data from a proof-of-concept system. I compare the results obtained with the real system to a simulation that closely matches the experiment. I argue that the deficiencies seen in the real system can be overcome by improving the micropattern manufacturing process.

The micropattern was created using 1000 mesh Gilder transmission electron microscope (TEM) grids from Ted Pella, Inc [Pel09]. As mentioned previously, these grids are composed of repeating 25 μm square regions with 19 μm holes and 6 μm bars. To create a micropattern, the grids are taped to #0 (approximately 110 μm thick) glass cover slip and metal is deposited onto the glass using a thermal evaporator. The grids are then removed, leaving a metal layer on the cover slip that is a negative image of the grid. Varying the thickness of the deposited metal controls the optical transmission of the micropattern. The metal deposition consists of 2 nm Cr and varying thicknesses of Au ranging between 10 and 30 nm.

This method of pattern manufacture yields highly variable results. Because the grids are fixed to the cover slip along the outside edges, some grids warp during metal

deposition, leaving unclear images of the grid. This leads to significant variation in optical transmission, even for metal deposits of the same thickness. These shortcomings, however, are purely a consequence of this particular procedure. Photolithography using precise masks would greatly improve the quality and repeatability of the pattern manufacture process. The system as described serves only as a proof of concept for the tracking technique, not a prescription for manufacturing.

5.4.1 Measuring Optical Transmission

The best micropatterns were selected by visually assessing the clarity of the pattern images and by measuring the pattern's optical transmission—the percentage of light transmitted through the metal layer. Knowing the optical transmission of a micropattern helps predict what sort of tracking accuracy should be obtainable, as discussed in Section 5.3.

The micropattern segments the image plane into two regions: a *bright* region, R_b where all light is transmitted, and a *dark* background region, R_d, where some constant fraction of light is transmitted. For a constant transmission, α, the intensity levels measured in bright and dark regions are related by

$$\mu_{R_d} = \alpha \mu_{R_b}. \tag{5.24}$$

The transmission constant α can therefore be measured by segmenting a single image of the micropattern into bright and dark regions, and finding the ratio of the light intensity measured in each region.

As discussed in Section 2.2, the expected value of a pixel measured by a CCD is given by Equation 2.34

$$\mu(x, y) = k(x, y)E(x, y)A + \mu_t(x, y)A, \tag{5.25}$$

where k is a per-pixel gain from fixed-pattern and flat-field non-uniformities, E is the photoelectron count measured at a pixel, A is the ADC gain, and μ_t is the expected dark current value. After flat-field and dark-current correction, the set of pixels in one region provides an estimate of the intensity level in the region, $E_R A$, that would be measured by an ideal CCD,

$$\mu_R = \frac{1}{|R|} \sum_{(x,y)\in R} I_c(x,y) \approx E_R A, \tag{5.26}$$

where $|R|$ is the number of elements in the region R, and I_c is a flat-field and dark-current corrected image.

Using a single image to measure transmission, however, assumes that the image sensor has a linear response to intensity. Indeed, CCDs usually have a very linear response, except at low intensities [HK94]. To account for the nonlinearity at low intensity levels, Equation 5.24 should be modified with a constant offset

$$\mu_{R_d} = \alpha \mu_{R_b} + \beta. \tag{5.27}$$

This modification assumes the camera has an *affine* response within the middle intensity ranges. A series of images of a micropattern taken at different illumination intensities can verify that a camera has an affine response and can also estimate the optical transmission of the mask. This is completed by fitting the mean intensity values from the bright and dark regions in each image using Equation 5.27.

Micropattern image segmentation is required to complete the transmission measurement. This segmentation divides the image plane into disjoint bright and dark regions. There should be no overlap between regions, but there may be pixels that do not belong to either region. Because the intensity values of the pixels in each region have a normal distribution, *K-means clustering* provides the basis for an effective segmentation method [HW79]. K-means clustering, with $k = 2$, applied to pixel intensities in an image of the pattern divides the pixels into two clusters around two centroids with intensity

values $\{I_A, I_B\}$. Each pixel belongs to the cluster whose centroid value is closest to its intensity value. The clustering is optimal in the sense that the sum of squared intensity differences between the centroid value and all cluster pixel intensity values is minimal over the two clusters.

This initial clustering enables computing the expected values μ_{R_b} and μ_{R_d} and the standard deviation of intensities in the clusters, σ_{R_b} and σ_{R_d}. Given these, values, a segmentation threshold for each region is provided by

$$\begin{aligned} &\{(x,y) \in R_b \quad | \quad I_c(x,y) \geq \mu_{R_b} - \rho\sigma_{R_b}\} \\ &\{(x,y) \in R_d \quad | \quad I_c(x,y) \leq \mu_{R_d} + \rho\sigma_{R_d}\}, \end{aligned} \tag{5.28}$$

where ρ is an empirically-determined constant that indicates how broad a region of intensities to include in each segmentation. Each segmented region is subjected to a morphological *close* operation—which grows and then contracts region boundaries—to eliminate isolated pixels.

Figure 5.18a shows a sample of images of a micropattern taken with different lamp intensities. Figure 5.18b shows the segmentation of the image plane based on the first of these images. Figure 5.18c plots Equation 5.27 for all images in the series. Linear least-squares regression confirms that the camera sensor has an affine response over the middle of the intensity range, with a correlation coefficient, $r^2 = 0.998$. This particular micropattern has a transmission coefficient $\alpha = 0.580$, which is among the lowest transmission coefficients obtained using the TEM grid manufacturing process. The offset $\beta = -0.937$ indicates that this image sensor is nearly linear.

5.4.2 Matching Experiment to Simulation

Two experiments are presented here that evaluate the two tracking methods using microscopy images of the square grid patterns. The first experiment evaluates tracking on

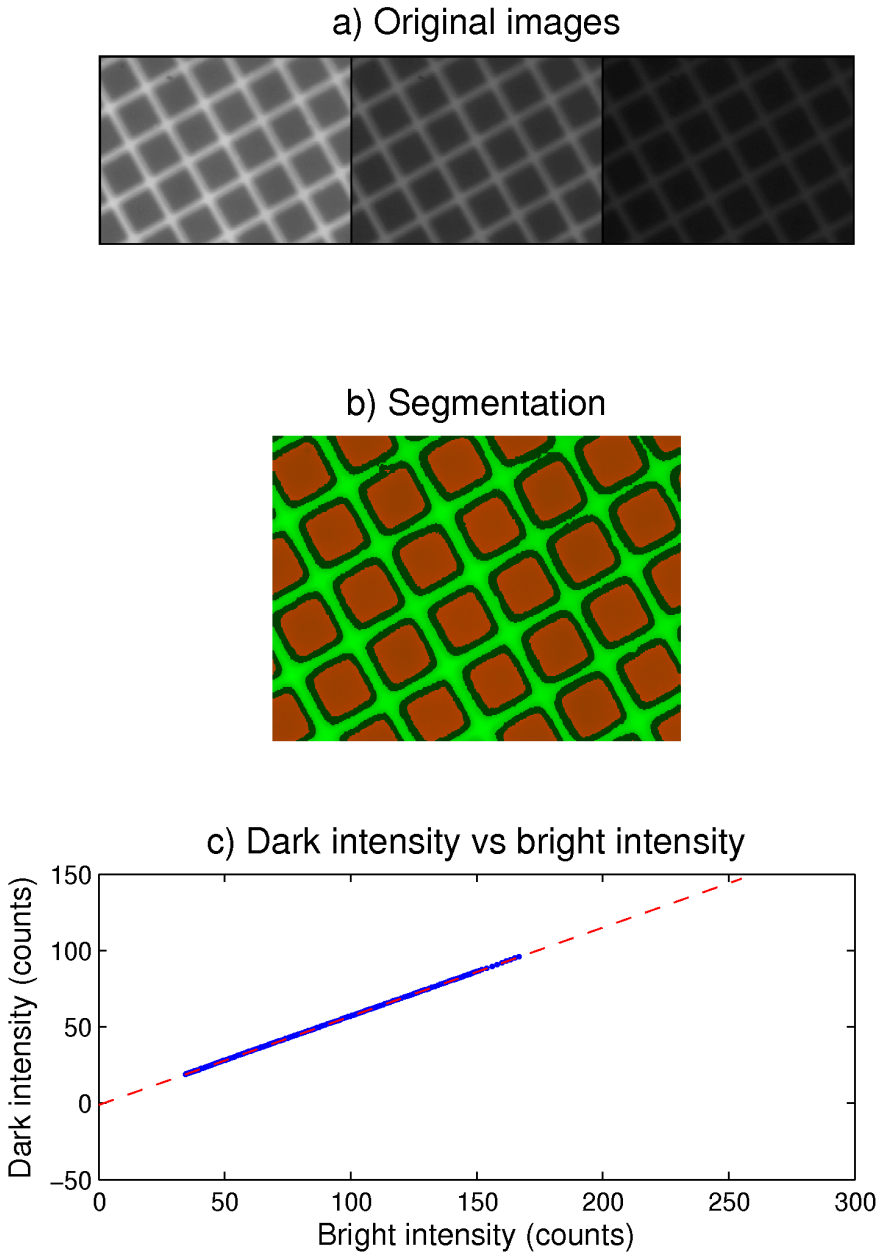
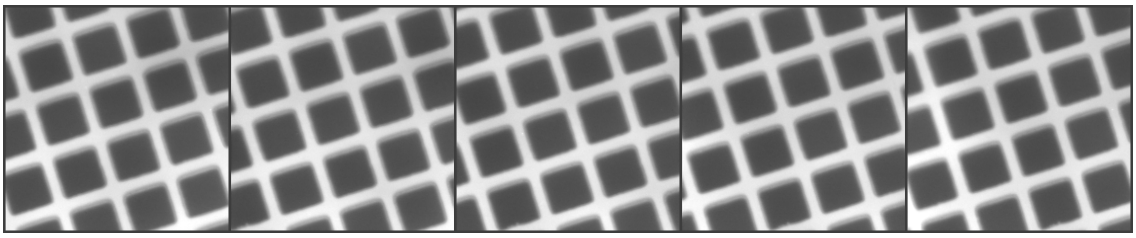
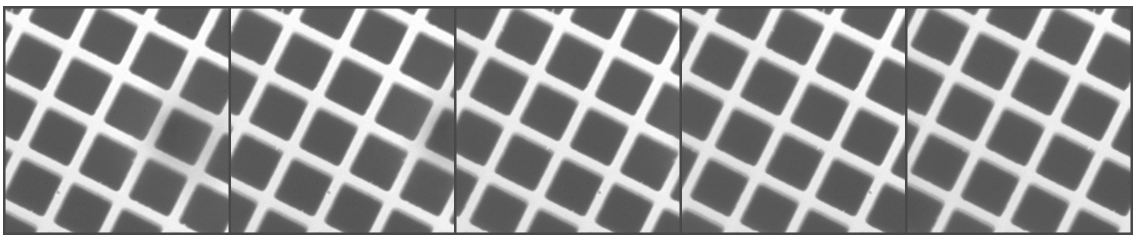


Figure 5.18: Measuring optical transmission from images of a micropattern. a) A sequence of micropattern images obtained at different illumination levels. b) Segmentation of the image plane, pixels in R_d are displayed in red (medium gray), pixels in R_b are displayed in green (bright gray), and pixels belonging to neither region are displayed in black. c) Mean dark intensity *vs.* mean bright intensity obtained from 200 images at different intensity levels. Dotted line shows least-squares regression for Equation 5.27, $\alpha = 0.581$, $\beta = -0.937$.



(a)



(b)

Figure 5.19: Microscopy images of a square grid micropattern. a) Images of a micropattern being translated by a MCL nano-positioning stage. b) Images of a micropattern cropped from a single high-resolution image used to simulate stage motion.

the micropattern grid absent a specimen. To capture a series of images of the micropattern moving with known displacement, a prepared slide is placed on a Mad City Labs (MCL) Nano-LPS100 nano-positioning stage with software control via a MCL Nano-Drive controller. The imaging system consists of an inverted Nikon Eclipse TE2000-E microscope with a 40X 0.7NA objective lens and a Pulnix TM-6710CL camera, which captures monochrome, 8-bit, 648×484 pixel^2 images at 120 fps. The microscope is focused on the micropattern which has a measured optical transmission of $\alpha = 0.46$. The stage is moved in $10\,\mu$m increments along one lateral axis, with the Pulnix camera acquiring 5 images at each location. Each image is flat-field and dark-current corrected after capture as described in 2.3. The MCL stage has a $100\,\mu$m range of motion along each axis (x, y, and z), but the total travel for this experiment is restricted to $80\,\mu$m to avoid intermittent misalignment issues reported at the far ends of the stage range. Phase-based and spatial correlation tracking is performed on 440×440 pixel^2 images to avoid introducing any bias from having more samples along one image axis. Figure 5.19a

shows a selection of the images acquired during this experiment.

Another set of images are used to simulate stage motion to closely match the experimental images. A high resolution image of the micropattern is formed with a Leica DM5000 B microscope with a 40X, 0.85NA objective lens and a 1.2X tube lens for a 48X total magnification. Images are acquired with a SPOT Flex FX1520 camera, configured to capture monochrome, 12-bit, 1024×1024 pixel^2 images. The high resolution image was scaled to match the intensity range and transmission of the experimentally-acquired images. Cropped regions of this image were selected to simulate translating the stage in $10\,\mu$m increments over $80\,\mu$m. Each image was corrupted with Poisson-distributed noise with ADC gain of 100 photoelectrons per count (as measured for the Pulnix camera in Section 2.3) to create 5 simulated images at each location. Figure 5.21b shows a selection of simulated images.

Table 5.5: RMS and standard error for tracking a micropattern using WPC and MSC. Images are of a $25\,\mu$m square grid pattern in focus.

	Method	**RMSE x**		**RMSE y**		σ_E**x**	σ_E**y**
		μm	pixel	μm	pixel	pixel	pixel
Experiment	WPC	0.8184	3.6373	0.3851	1.7116	1.3316	0.7258
	MSC	0.2193	0.9747	0.3900	1.7333	0.1173	0.3969
Simulation	WPC	0.0708	0.3147	0.2162	0.9609	0.2387	0.4596
	MSC	0.1343	0.5969	0.2837	1.2609	0.2218	0.4564

Table 5.5 records the RMS tracking errors obtained on the experimental and simulated images using the WPC and MSC methods. Both methods perform significantly worse on these data than on the simulated micropattern images presented in Section 5.3. Note that the bars in the micropattern of Figure 5.19a are much narrower than the bars in simulated micropattern images, for example Figure 5.8. This is because metal diffuses around the TEM grid bars during the deposition process, narrowing the region that should remain completely transparent. This amount of diffusion is different for every micropattern manufactured with this process. This disparity will have a larger

effect on the MSC method, as the grid in the modeled images will completely cover the grid in the observed images for a wider range of model parameters.

The second experiment investigates micropattern tracking in the presence of a specimen. The micropattern-prepared glass cover slips are mounted on tissue sections prepared by biologists at the University of North Carolina at Chapel Hill (UNC). The cover slips are placed pattern-side down, and secured with mounting media, which leaves a small gap between the specimen and the micropattern. Stained octopus muscle sections serve as specimens for this study. Figure 5.20 shows the octopus muscle and micropattern at 10X magnification.

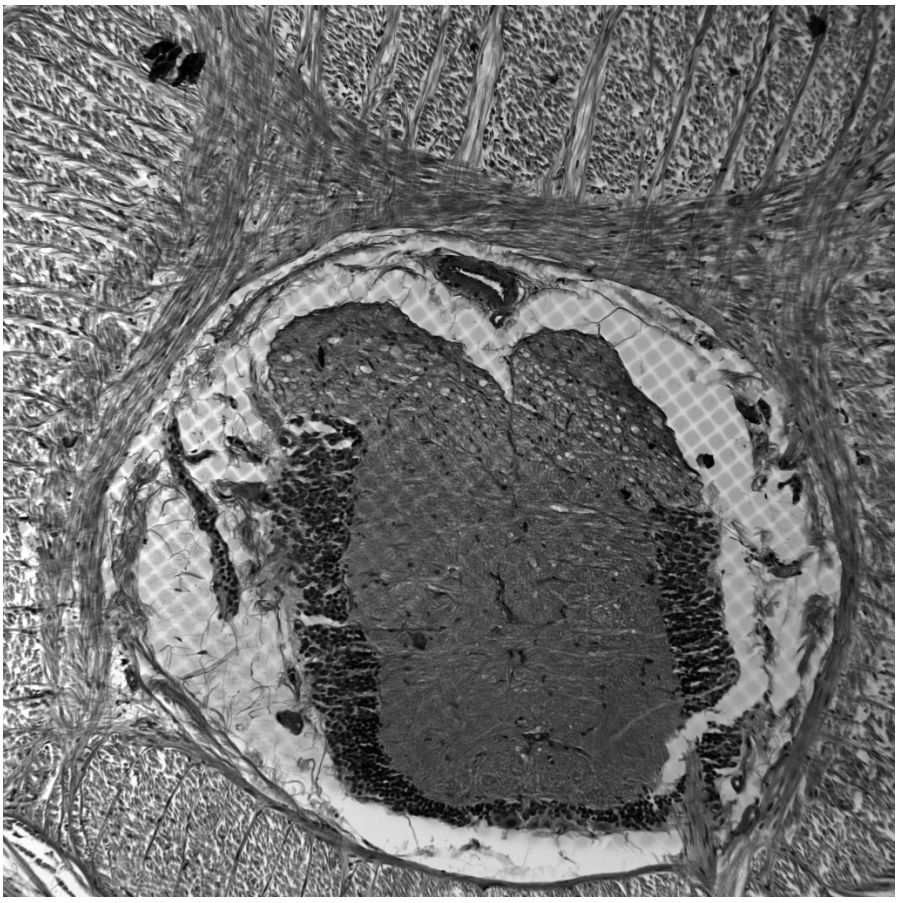
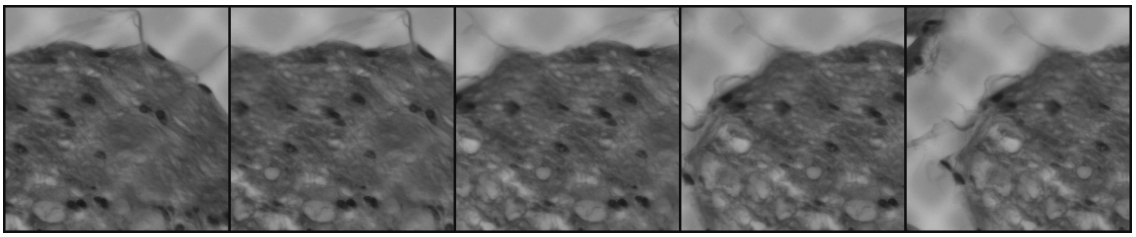


Figure 5.20: Metal deposit grid micropattern mounted on a section of octopus muscle tissue viewed at 10X magnification. The micropattern is clearly visible in the gaps near the center of the section. The edge of the micropattern is near the bottom of the image. The micropattern lies approximately 8 μm above the tissue. Octopus section from Bill Kier.

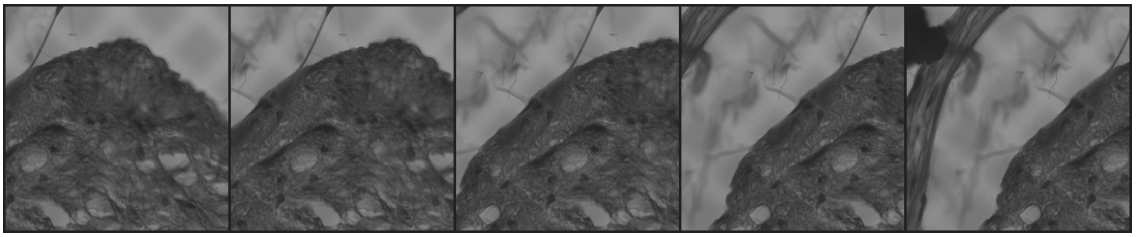
Image capture of experimental data follows a similar process to that used in the grid-only experiment above. A prepared slide is placed on a MCL nano-positioning stage and imaged with an inverted Zeiss Axiocam microscope with a 40X 0.7NA objective lens and a Pulnix TM-6710CL camera. The microscope is focused on the specimen tissue—for this specimen, the micropattern is approximately $8\,\mu$m above the specimen. The measured optical transmission of the micropattern is $\alpha = 0.837$. The stage is moved in $10\,\mu$m increments along the x axis a total of $80\,\mu$m, with the Pulnix camera acquiring 5 images at each location. Each image is flat-field and dark-current corrected after acquisition. Figure 5.21a shows a selection of the images acquired during this experiment.

A high resolution image of a non-micropatterned octopus muscle section is used to simulate images that closely match the experimental images. The image is formed with a Leica DM5000 B microscope with a 40X, 0.85 NA objective lens and a 1.2X tube lens for a 48X total magnification. 12-bit, 1024×1024 pixel^2 images are acquired with a SPOT Flex FX1520 camera. A micropattern image is simulated as described in Section 5.3 with a 40X, 0.65 NA objective. The micropattern is given an optical transmission of $\alpha = 0.837$ and is displaced $8\,\mu$m from the focal plane. The high resolution image was scaled to match the intensity range of the experimentally-acquired images, and multiplied by the simulated micropattern image. Cropped regions of the modulated image were selected to simulate translating the stage in $10\,\mu$m increments over $80\,\mu$m (measured by the simulated pattern). Each image was corrupted with Poisson-distributed noise with ADC gain of 100 photoelectrons per count to create 5 simulated images at each location. Figure 5.21b shows a selection of simulated images.

Both sets of images are tracked using MSC and WPC. All image frames are registered to the first frame in the image sequence. Micropattern spacing parameters in the experimental images are optimized for the fundamental frequency and four harmonics from images of the micropattern alone. These four frequencies are used to perform the



(a) Experiment



(b) Simulation

Figure 5.21: Experimental and simulated images of a micropattern mounted over octopus muscle tissue. a) Images acquired using a real micropattern, 40X magnification. b) Simulated images created to match the experimentally-acquired images.

phase correlation.

Table 5.6: RMS and standard error for tracking a micropattern and specimen using WPC, MSC, and image registration. Tracked images are of octopus tissue and a 25 μm square grid micropattern, approximately 8 μm above the tissue.

	Method	**RMSE x**		**RMSE y**		σ_E**x**	σ_E**y**
		μm	pixel	μm	pixel	pixel	pixel
Experiment	WPC	1.6544	7.3529	0.8989	3.9951	2.2582	1.5751
	MSC	1.0442	4.6409	0.8901	3.9560	1.4844	1.2947
	Registration	0.3280	1.4578	0.2948	1.3102	0.9511	0.4996
Simulation	WPC	0.5832	2.5920	1.3307	5.9142	1.6147	2.5533
	MSC	0.4543	2.0191	1.8349	8.1551	1.3560	3.3093
	Registration	0.0204	0.0907	0.0191	0.0849	0.0564	0.0622

The RMS tracking error and standard error for the different tracking scenarios are enumerated in Table 5.6. Both methods perform worse on the octopus data set than the grid-only data set, with the error in WPC doubling and the error in MSC increasing about six fold. This data does, however, provide a very important insight—the error characteristics obtained in simulation match very well with the experimental data.

Considering the x and y axis errors together, both WPC and MSC demonstrate approximately a 1.6 μm (7.1 pixel) RMS error in *both* the experimental data and the simulation data. This indicates that simulations that include the corruption from a specimen provide a reasonable substitute for experimental data—improving the tracking on simulated images should provide a similar improvement in experimental data. Simulations are used in Chapter 6 to explore the limits of tracking with optimal patterns.

5.4.3 Comparison to Image Registration

For comparison purposes, a multiple resolution image registration is performed on the experimental and simulation octopus sequences. The multiple resolution registration initializes an alignment estimate between image pairs with coarse representations of each image. This estimate is refined using finer scale image representations, eventually minimize the NCC metric on the full-scale images. Image registration relies on all image data to obtain tracking estimates, not just the pattern. This approach provides approximately 300 nm RMS error for the experimental data, but this falls to just 20 nm (0.09 pixel) for the simulated data. This indicates the experimental data may have additional characteristics that negatively impact tracking—*e.g.* radial distortion or errors in the stage positions reported by the MCL stage.

5.5 Microscopy Mosaicking

I close this chapter by presenting an application of WPC-based micropattern tracking. The micropattern is introduced to the optical train of the bright-field microscope to provide stage position information for non-motorized stages. Stage position information can be used to construct mosaic images from multiple frames of a microscopy video.

A complete tracking application would enable a microscopist to examine a specimen while a camera continuously acquires images. The stage position would be tracked, and

the acquired images would be assembled and presented to the microscopist with proper spatial context. The assembled images would enable the microscopist to examine regions larger than a single field of view at high magnification. This would provide a natural view of specimen structure, and may better enable a microscopist to understand the relationships between different regions of a specimen.

When moving a stage micrometer by hand, it is difficult to restrict motion to less than half the pattern wavelength in each direction (for the TEM grid micropattern, 12.5 μm) for every frame of an image sequence. In these cases, WPC tracking is subject to an ambiguity of whole pattern distances. Image correlation can easily resolve this ambiguity for fixed specimens—image registration restricted to several whole-pattern displacements along each pattern axis provides a disambiguation of image pair alignments. WPC or MSC on the full-resolution image provides fine alignment.

Figure 5.22 shows an image mosaic assembled from 20 microscopy images using this approach. Images are formed on a Leica DM5000-B microscope with a 40X, 0.85NA objective lens and a 1.2X tube lens. A SPOT Flex FX1520 camera acquires monochrome, 14-bit, 1024×1024 pixel images at 4 fps with 5 μs exposures. The stage is moved continuously by hand during image acquisition. During tracking, normalized cross-correlation on images reduced to 256×256 pixel displaced by $[-3 \ldots 3]$ micropattern cycles disambiguates among possible translations reported by pattern tracking. The microscopy mosaic image is constructed by merging all images into a common reference frame, overwriting data with each new frame (no blending is performed). The resulting mosaic is visually coherent, with some image seams visible upon close inspection. Improving the micropattern manufacturing process would improve the tracking accuracy, as discussed in Chapter 6, and would thereby improve the quality of microscopy mosaics assembled from WPC stage tracking information.

This example demonstrates that the WPC tracking technique provides reasonably reliable information in some situations. One would not usually use structured illumi-

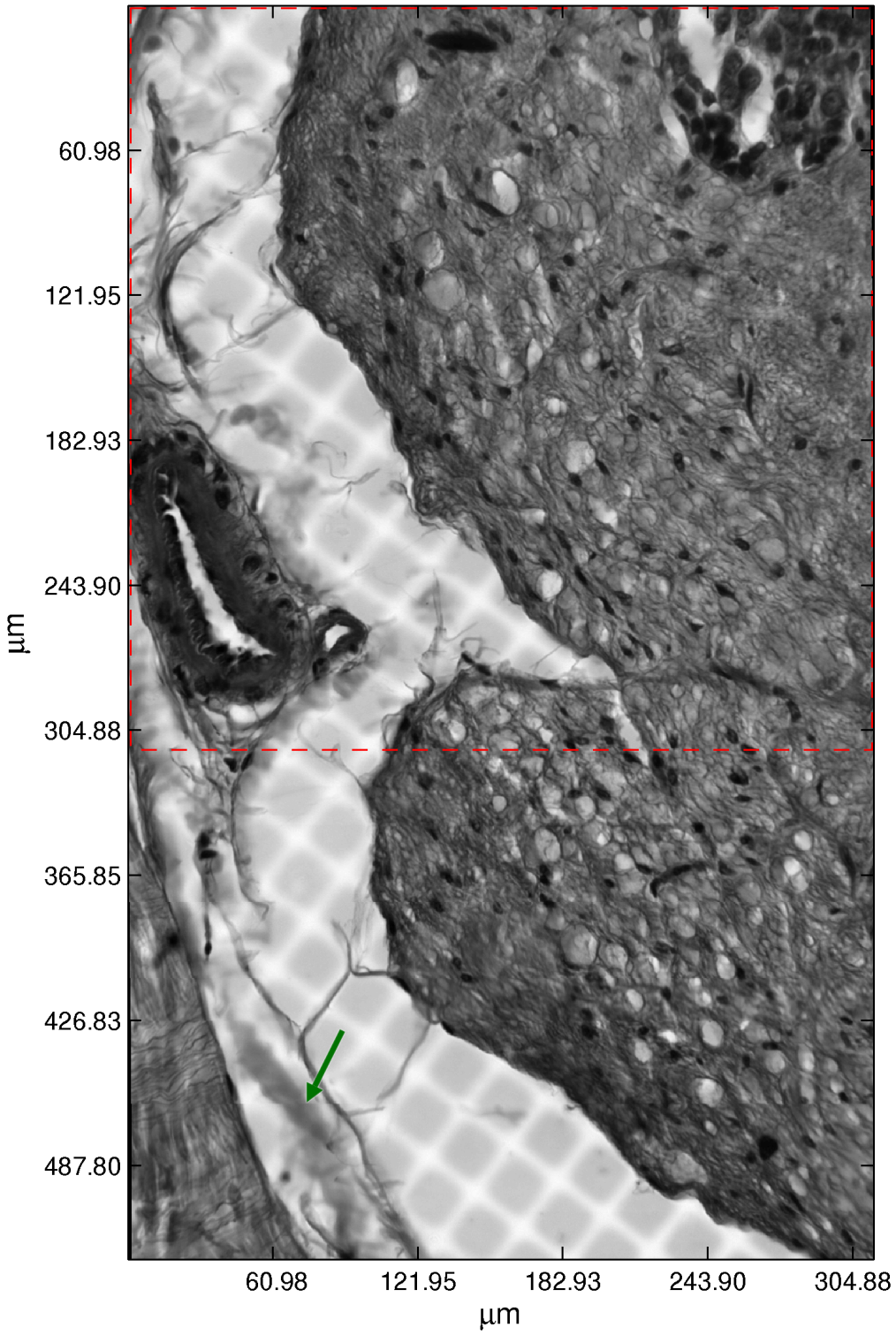


Figure 5.22: A microscope mosaic image assembled from 20 images of octopus muscle tissue. The stage is moved by hand during image acquisition, and the micropattern is tracked using WPC and coarse image registration. Dotted lines show the size of a single image frame. Arrow points to a visible seam.

nation to create mosaics of the type seen in Figure 5.22. In fact, image registration on these images obtains a better alignment with fewer visible seams, and image blending techniques would improve the visual coherence of the mosaics even more [HPS08]. However, WPC additionally handles sparse specimens, moving specimens, and z position tracking.

5.6 Discussion

In summary, this chapter has introduced a structured illumination approach for stage tracking in bright-field microscopy. A semitransparent pattern layer is introduced to the microscope's light path, fixed to the specimen's frame of reference. The pattern transforms the stage tracking problem into a multiple layer image analysis problem. A search for magnitude peaks at pattern frequencies in the frequency domain yields an estimate of pattern orientation that does not rely on knowing the pattern position. Two methods are proposed to track the pattern. Model-based spatial correlation (MSC) determines pattern translation using an image formation model to simulate multiple images of the pattern, and optimizing model parameters to find the best match to an observed image. Weighted phase correlation (WPC) computes a phase estimate for individual target frequencies in an image by first reweighting the image according to the target frequency's wavelength, and then computing a single component of the discrete Fourier transform (DFT). Once the pattern's position has been estimated, a model-based rotation optimization can provide an improved estimate of the pattern orientation.

The accuracy of the orientation and tracking methods were examined in detail. The frequency-based orientation method provides estimates with a 0.25° RMS error for an in focus square grid pattern. The model-based orientation method improves this estimate to 0.066° RMS error once the position of the pattern is determined.

A different story arises for pattern tracking. The WPC technique demonstrates very

accurate tracking for axis-aligned square grid patterns with no specimen present. Lateral tracking accuracy remains around 0.010 pixel (2.3 nm for the simulated microscope) RMS error in this case, falling to 0.057 pixel for arbitrary orientations. This accuracy is maintained down to approximately 18 dB SNR due to pattern transmission and below 10 dB due to defocus effects.

The WPC method relies on a phase estimate from individual pattern frequencies. Point sampling in the frequency domain requires infinite support in the spatial domain, and WPC attempts to obtain this by redistributing and reweighting the signal to analyze the target frequency. In the presence of uncorrelated noise, the WPC method maintains subpixel tracking accuracy. In the presence of correlated noise, however, the phase estimate becomes unreliable, and tracking accuracy diminishes. Nevertheless, there are some practical imaging scenarios in which WPC may excel. For example, multimode microscopes capture fluorescence and bright-field images in quick succession [SSW+03]. For observations of fluorescent specimens that do not show up well in bright-field, one could use the bright-field channel to obtain tracking information from structured illumination. Because fluorescence microscopy uses epi-illumination (both the excitation and emission light travel through the objective lens), placing the micropattern below the specimen would cause no interference with the fluorescent imaging.

In comparison, MSC struggles to provide 0.25 pixel RMS error in the absence of rotation, though it seems to improve remarkably for many non-axis-aligned patterns. Uniform accuracy is maintained down to below 10 dB SNR due to pattern transmission and to 0 dB due to defocus effects.

The error obtained with both methods increases significantly on real microscopy images of the a micropattern grid. Tracking degrades even more in the presence of a specimen. However, the errors for experimental and simulated images with a specimen present agree quite well. This provides motivation to continue exploring pattern layer tracking in simulation.

For tracking contiguous, fixed specimens, one should use an image registration technique; in the presence of sufficient specimen contrast, image registration fares far better than either WPC or MSC. However, these method provide something that image registration cannot—they enable tracking with a fixed frame of reference in the presence of sparse specimen information and for moving specimens. Structured illumination provides the additional ability to extend tracking into 3D. Throughout this chapter I have claimed that some of MSC's accuracy issues are due to the pattern used for evaluation—so what is a better pattern? These issues are investigated in greater detail in Chapter 6.

Chapter 6

Pattern Design for Structured Illumination Microscopy

In Chapter 5, I developed weighted phase correlation (WPC) and model-based spatial correlation (MSC) as techniques that provide lateral tracking of a semitransparent pattern layer in structured illumination microscopy. The pattern used in this evaluation was obtained from a TEM grid which produced dark square "holes" separated by thin bright "bars". This pattern was chosen because the grids were readily available to purchase from a microscopy supply company and it was relatively simple for a materials scientist to deposit a metal image onto glass cover slips. This provided a working system with which to demonstrate the concept of structured illumination microscopy. The TEM grid pattern, however, is not the optimal pattern for structured illumination microscopy.

This chapter investigates pattern design for structured illumination microscopy, with the end goal being an understanding of how choices made in pattern design affect the ability to recover tracking information from the pattern layer. Specifically, I discuss how the choice of pattern influences the ability to track the stage independently from the specimen, determine focus position, and disambiguate among possible stage positions.

The following design goals are generally applicable to stage tracking in a broad range of microscopy experiments:

1. The pattern should enable tracking laterally and axially (in X, Y, and Z) indepen-

dent of specimen motion.

2. The pattern should enable tracking over large frame-to-frame displacements.

3. The pattern should enable determining the absolute stage position.

4. The pattern should be easily removable from the acquired images through image processing.

Note that there is an important distinction between goals 2 and 3. The former goal refers to tracking the stage relative to some fixed reference position. The latter goal refers to determining absolute stage position from a single image—a slide-based absolute positioning system.

Pattern design necessarily involves some trade-offs. Both WPC and MSC depend on finding the displacement of a regular pattern layer present in an image. The resolution of these methods increases as that pattern matching is based on higher frequency components, as seen in Equation 5.9: $E_{px} = \frac{E_\phi}{360f}$. But, to enable tracking over large frame-to-frame displacements the pattern should have a large wavelength, favoring low frequencies. Another consideration that favors low frequencies is that as the pattern moves farther out of focus, the high frequency components are attenuated more and therefore reach the noise floor first.

In order to track the stage independent of the specimen the pattern signal should dominate the specimen signal. This requires high-contrast, low transparency patterns. Patterns that are more transparent, however, enable removing the pattern from the specimen images with less information loss.

As I will show, to enable axial (focus) tracking, the pattern should contain multiple frequency components. The more frequency components that are present, the more terms can be used to determine focus. Choosing which frequencies to include, however, requires considering the frequencies present in the specimen. If the specimen has high

magnitude at many frequencies, it may be difficult to find an appropriate set of frequencies to use as a pattern. Additionally, the microscope point-spread function (PSF) may attenuate a large proportion of high frequencies too much to be useful.

The remainder of this chapter is organized as follows. Section 6.1 discusses pattern selection and design, including how to maximize pattern tracking accuracy in the presence of an independently moving specimen. Section 6.2 discusses how to expand the frequency analysis of micropatterns to enable axial tracking of the microscope stage. Section 6.3 evaluates the approaches discussed here using simulated structured illumination microscopy images. Section 6.4 summarizes results and discusses strengths and limitations of the methods presented here.

6.1 Pattern Design

While it would be ideal to be able to design a single pattern that could provide optimal tracking in all scenarios, the reality is that tracking performance is largely dependent on the imaging application. The size of the pattern in images depends on the magnification provided by the microscope and the size of the camera's image sensor. Furthermore, some specimens have strong magnitudes at particular frequencies which interfere with the signal provided by the pattern. In the following sections, I discuss how to select a pattern given an objective lens, camera sensor, specimen, and expected axial distance between the pattern and specimen.

6.1.1 Coordinate System

One can specify wavelengths and frequencies in object or image coordinates, depending on what parameters are being considered. The relationship between a wavelength in object coordinates (measured in μm, for example), L_o, and the corresponding wavelength

in image coordinates (measured in pixels), L_i, is

$$L_i = \frac{ML_o}{s}. \tag{6.1}$$

A spatial frequency in object coordinates (measured in cycles per μm, for example), f_o, is

$$f_o = \frac{1}{L_o} = \frac{M}{sL_i} = \frac{Mf_i}{s}, \tag{6.2}$$

and a spatial frequency in image coordinates (measured in cycles per pixel), f_i, is

$$f_i = \frac{1}{L_i} = \frac{s}{ML_o} = \frac{sf_o}{M}. \tag{6.3}$$

6.1.2 Pattern Spacing

The imaging application influences the selection of the pattern's physical size. The lower limit in pattern size is determined by three factors: the numerical aperture (NA) of the objective lens, the spacing of the sensor elements in the CCD used to record images, and the shot noise. The Abbe limit relates the objective NA to the smallest spacing in a resolvable pattern from Equation 2.24:

$$d = \frac{\lambda}{2\mathrm{NA}}. \tag{6.4}$$

The Wittaker-Shannon sampling theorem states that the Nyquist frequency, f_N, is the maximum frequency represented by a sampling rate f_s [FvDFH97, Ch. 14]:

$$f_N = \frac{f_s}{2}. \tag{6.5}$$

In other words, accurately reconstructing a signal containing a specific maximum frequency requires sampling the signal at twice that frequency.

The size of the smallest pattern wavelength that can be resolved in an image taken with a CCD is determined by twice the distance between sensor elements, $\frac{2s}{M}$, where M is the magnification of the objective lens and s is the size of a CCD sensor element.

The noise in acquired images can be reduced using a Gaussian filter with standard deviation equal to the standard deviation of the noise expected in the image, σ_n [TV98, Ch. 3]. Applying a low-pass filter to structured illumination images diminishes the signal from the patterns, especially those with short wavelengths. In the spatial domain, a Gaussian filter kernel with standard deviation σ has 98% of its energy contained in a width of 5σ (or radius 2.5σ). So a good guideline is to choose pattern wavelengths that exceed at least twice the noise level to avoid excessive smoothing of the pattern signal during filtering.

Taken together, the Abbe limit, Nyquist rate, and noise level provide a lower bound for an observable pattern wavelength:

$$L_{min} = \max\left(\frac{\lambda}{2\mathrm{NA}}, \frac{2s}{M}, \frac{2s\sigma_n}{M}\right). \tag{6.6}$$

The upper limit to pattern spacing depends on the field of view. Higher magnification applications require smaller pattern structures to include more than one pattern instance within a single field of view. For patterns with wavelengths greater than one field of view, the WPC method does not have enough information to determine how to weight each component in its correlation computation (see Equation 5.17). Though the mathematics still holds, the accuracy of the phase estimates is severely diminished, as shown below. Additionally, the structure of the specimen places constraints on pattern spacing. Takita *et al.* note that natural images have a predominance of low frequency information [TAS+03], so there may be significant interference between pattern and specimen signals in this region. Although not as firm as the lower bound, the upper bound for a usable pattern size is on the order of the field of view along one axis of a

microscope image:

$$L_{max} = \frac{sN}{M}, \tag{6.7}$$

where N is the number of pixels along an image axis.

6.1.3 Choosing Optimal Pattern Frequencies

Assuming that the pattern and specimen are moving independently, the pattern provides the *only* information in an image for tracking with structured illumination microscopy. To provide the most accurate tracking, then, one needs to maximize the information present in the pattern (the signal) with respect to the other factors in the image (the noise). In this sense, noise includes both camera noise and the specimen, as both of these elements interfere with pattern tracking.

Chapter 2 discusses the role the microscope point-spread function (PSF) plays in image formation. The image that a microscope forms of a planar object is the convolution of the image predicted by geometrical optics with a slice of the 3D PSF determined by where the object is placed in front of the objective lens. Equivalently, in the frequency domain, the magnitudes of the frequencies in the object are attenuated by the Fourier transform of the microscope's PSF. From Equation 2.21,

$$\mathcal{I}_i(f_X, f_Y) = \mathfrak{H}(f_X, f_Y)\mathcal{I}_g(f_X, f_Y), \tag{6.8}$$

where $\mathcal{I}(f_X, f_Y)$ is the Fourier transform of $I(x, y)$ and $\mathfrak{H}$ is the Fourier transform of the normalized PSF, also known as the optical transfer function (OTF).

Analysis of the microscope OTF enables choosing the optimal pattern frequencies for a tracking application. The OTF is the frequency domain representation of the normalized point-spread function (PSF)—it determines by how much pattern frequencies are attenuated. So, maximizing the SNR for a pattern involves choosing the pattern that, when multiplied by the OTF, has maximum magnitude at the pattern frequencies. The

selection of possible patterns and OTFs from all possibilities is driven by the constraints of the experimental parameters. The critical experimental parameters include objective lens characteristics, image sensor size, the expected axial distance between specimen and pattern, and the composition of the specimen itself.

Figure 6.1 shows a frequency-focus slice of magnitudes in the OTF of a simulated 40X, 0.65NA objective lens. Here focal depths range within $z = [0 \dots 25]\,\mu$m and frequencies range within $f = [0...\frac{1}{2}]$ cycles per pixel. For this lens and sensor pair, the Abbe frequency limit is above the Nyquist frequency. This figure illustrates why one cannot select a set of pattern frequencies that provide accurate tracking capabilities when the pattern is in focus, and expect tracking to perform uniformly throughout the focus range. Different frequencies will be attenuated by different amounts at each focal depth, and this attenuation is not monotonic across frequency or depth.

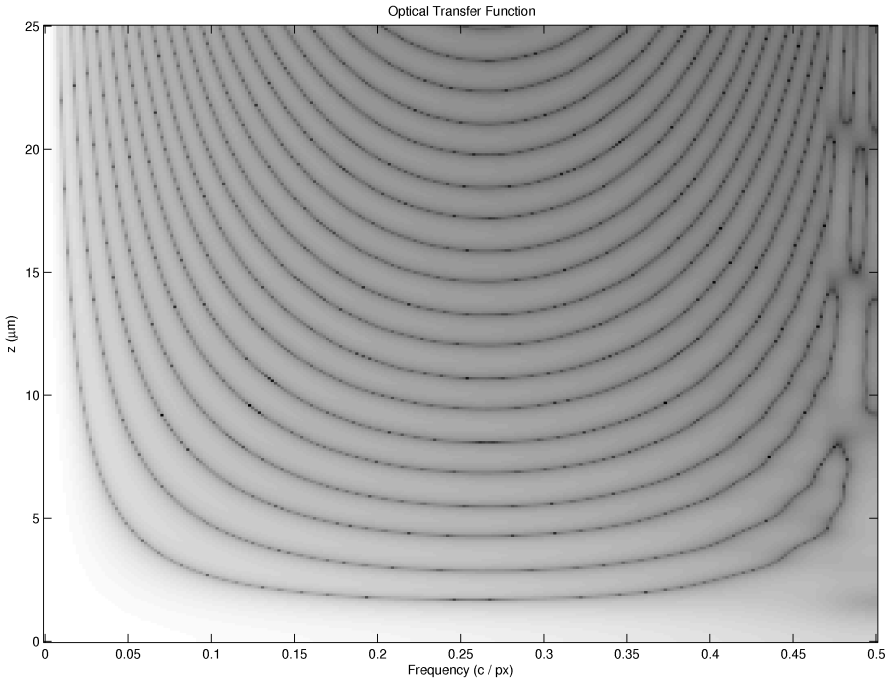
Because the OTF is radially symmetric for a focal slice, in the following analysis, I restrict the pattern design to 1D laterally. When designing a 2D pattern one would select the same frequencies determined by 1D analysis, possibly distributing different frequencies along each axis.

Let $P(x)$ be a pattern defined by a set of frequencies $f := \{f_i \mid i = 0, 1, \dots n\}$. Let $\mathcal{P}(f)$ be the Fourier transform of $P(x)$. The magnitude at frequency f_i is given by $|\mathcal{P}(f_i)|$. Let $H(x; z)$ be the normalized PSF of an objective lens at a focal plane z. Here, the semicolon specifies that the focus axis is treated differently from the lateral axis on which the pattern resides. Specifically, $\mathcal{H}(f; z)$ is the Fourier transform of a slice of the PSF, one slice of the OTF; the magnitude at frequency f_i is given by $|\mathcal{H}(f_i; z)|$.

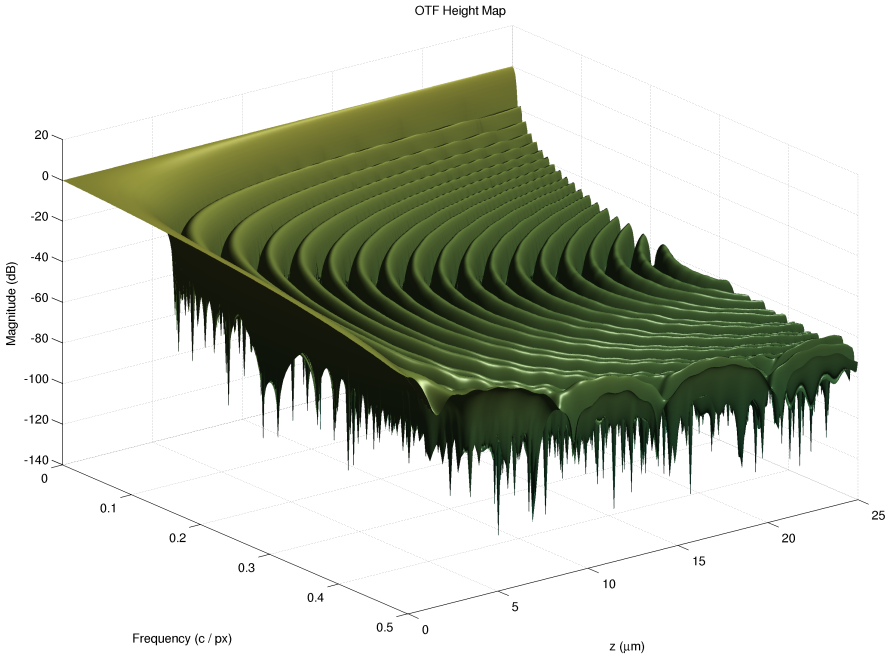
The total magnitude at all frequencies in a pattern at a distance, z, from the focal plane is given by

$$S(P; z) = \sum_i |\mathcal{P}(f_i)||\mathcal{H}(f_i; z)|. \tag{6.9}$$

Equation 6.9 provides a scoring function for how strongly a particular pattern with frequencies f_i is represented at a particular focal depth, and this indicates how well this



(a)



(b)

Figure 6.1: The frequency-focus plane of the optical transfer function (OTF) for a 40X, 0.65NA objective lens from 0 to 25 μm. a) One slice of the OTF displayed as an image; magnitudes are mapped with a logarithmic scale to make all magnitudes visible; bright values indicate high magnitudes. b) The OTF as a height map with same scaling, to better show relative values.

pattern will trackable compared to other patterns with different frequencies.

There is usually a known range of distances from the focal plane to the pattern in a given experiment. This distance could be fixed, or it could stay centered around a value with some variation, or it could span a whole range with equal probability. For this reason it is sensible to introduce a weighting function $W(z)$ that weights the importance of focal distances in an experiment. This can be thought of as a probability density function for axial stage positions.

Combining the pattern scoring function for individual focal depths and the weighting function for the importance of focal depths, a metric for how well a particular pattern is suited to tracking in an experiment is given by:

$$S(P) = \sum_z W(z) S(P; z) = \sum_z W(z) \sum_i |\mathcal{P}(f_i)||\mathcal{H}(f_i; z)|. \tag{6.10}$$

Note that the weighting function in Equation 6.10 is multiplied by all frequencies in the pattern, and the result is summed across all focal depths. Rearranging terms therefore provides a function that specifies the score for *all* frequencies in the OTF, irrespective of the magnitude in a particular pattern:

$$S(f) = \sum_z W(z) |\mathcal{H}(f; z)|. \tag{6.11}$$

The frequency selection thus far has not taken into account the noise present in the system—the camera noise and interference from the specimen. Camera noise has constant magnitude at all frequencies, so one needs to ensure that any frequency selected for the pattern does not get attenuated by the OTF to near or below the noise floor. Equation 6.11 treats all frequencies equally—any frequency that is heavily attenuated by the OTF will receive a low score—and so provides the set of best frequencies to choose. Section 6.1.4 will discuss determining whether the pattern is trackable.

Spectral analysis of the specimen to be observed provides a way to consider the track-

ing interference caused by its presence in images. Because the specimen can have any arbitrary orientation, the magnitudes at frequencies at all orientations should be considered. Given a specimen image, $I(x, y)$, and its Fourier transform in polar coordinates, $\mathcal{I}(f, \theta)$, the radially-summed magnitude provides a measure of specimen interference:

$$Q(f) = \sum_{\theta} |\mathcal{I}(f, \theta)|. \tag{6.12}$$

To adjust this to the same scale as the axial weighting function, it is sensible to express the magnitudes on a logarithmic scale, and shift the scores to fall within $Q(f) \in [0 \ldots 1]$. This weighting term can now apply an additional constraint to Equation 6.11 that is the inverse of the specimen magnitude at all frequencies, $Q'(f) = \max(Q(f)) - Q(f)$:

$$S(f) = Q'(f) \sum_{z} W(z)|\mathcal{H}(f; z)|. \tag{6.13}$$

The frequency weighting function can also be truncated to enforce the minimum and maximum pattern frequencies from Equations 6.6 and 6.7.

Now that a frequency scoring criterion has been established, the final task is to use it to select which frequencies to include in the pattern. Choosing the n frequencies with maximum score maximizes Equation 6.13, but these frequencies are likely to be bunched around whichever frequency had the maximum score. Section 6.2 discusses how to determine focus, with the observation that focus accuracy increases when the frequencies used are spread out. For this reason it makes sense to apply some additional constraints to the frequency selection. For example, one could select the frequencies that have local maxima in the scoring function; this distributes the pattern frequencies throughout frequency space, but uses the best one in each area. Or, one could select the frequency with maximum score and some of its harmonics; this ensures the pattern is periodic over a small window.

In summary, to determine the optimal frequencies to use in a pattern:

1. Select an objective lens and compute its OTF.

2. Construct a focus weighting function, $W(z)$, that specifies the importance of different axial distances from the pattern.

3. Construct a frequency weighting function, $Q(f)$, that specifies the importance of different frequencies, using the parameters of the imaging system and the Fourier transform of an image of the specimen.

4. Multiply the OTF by the focus and specimen weighting functions.

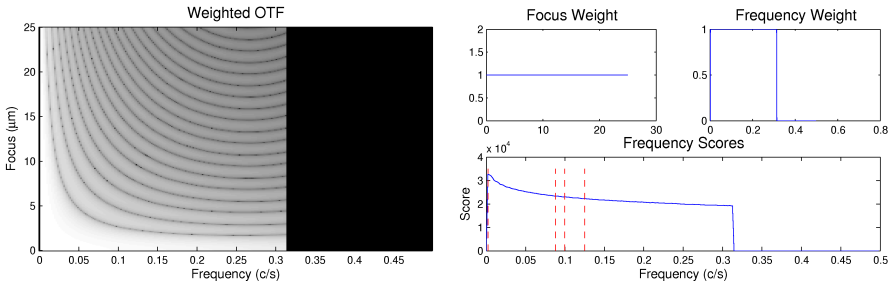
5. Sum the weighted OTF along the focus axis to determine the frequencies that will provide the best tracking signal.

6. Select the n frequencies that maximize this signal.

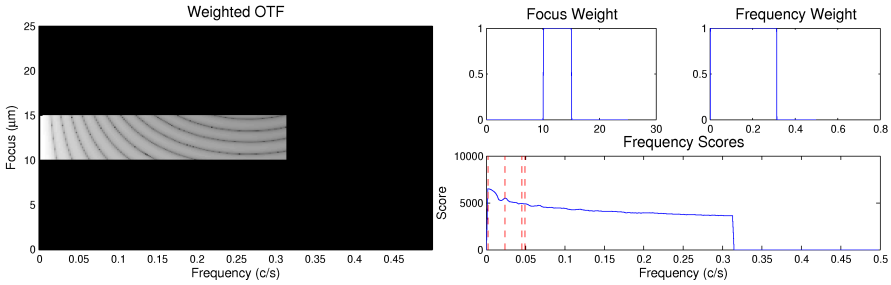
Figure 6.2 provides some concrete examples of weighting functions applied to the OTF of Figure 6.1. In Figure 6.2a, the only constraints applied are the minimum and maximum frequency thresholds. The high frequency cutoff occurs because of the noise filtering requirement, where $\sigma_n = 1.6$, as determined for a Pulnix camera in Section 2.3. For this scenario, the predominant feature of the frequency scoring function is due to the attenuation of higher frequencies by the OTF. Low frequency patterns will provide the best tracking signal. The vertical dashed lines in the frequency score plot indicate the four local maxima with highest frequency, but in this case there is little to distinguish these points from any others.

In Figure 6.2b, the focus range has been restricted to fall within $dz = [10 \ldots 15]\,\mu$m of the pattern. The frequency score again shows a preference for low frequency patterns, with local maxima at 0.025 and 0.05 cycles per sample that would be good options for patterns that have frequency spread throughout frequency space.

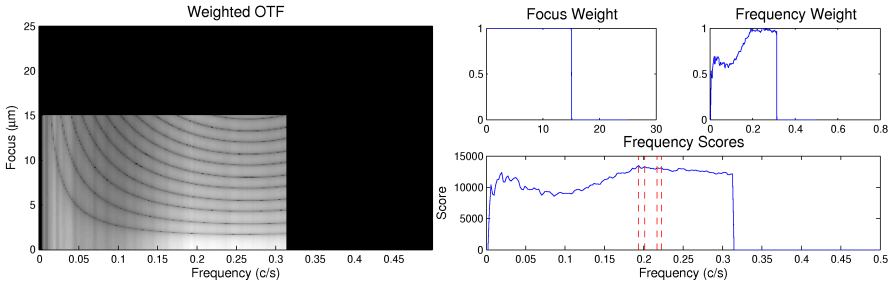
In Figure 6.2c, the focus range is restricted within $dz = [0 \ldots 15]\,\mu$m and the frequency weighting has been constrained by the inverse spectrum of a frog brain tissue



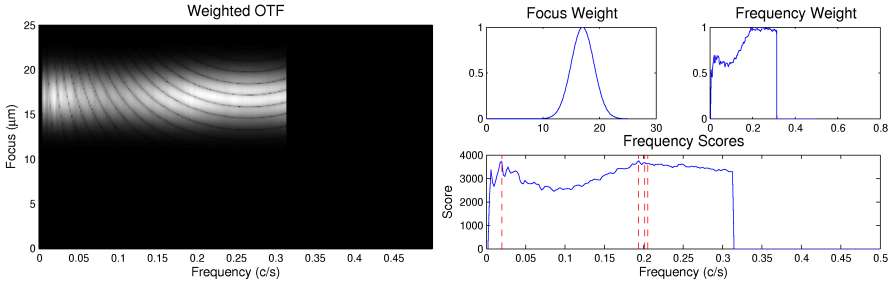
(a)



(b)



(c)



(d)

Figure 6.2: Optimal frequency selection for different experimental conditions. In all cases, the OTF is computed for a 40X 0.65NA objective lens. Vertical dashed lines in the Frequency Score plots represent the local maxima with best score. a) Constraints imposed only by maximum and minimum pattern frequency for optical system. b) Constraints for imaging pattern at depths with $[10 \ldots 15\,\mu\text{m}]$. c) Constraints for imaging a frog brain tissue within $[0 \ldots 15]\,\mu$m of the pattern. d) Constraints for imaging frog brain tissue where the mean distance to the pattern is $17\,\mu$m with standard deviation $2\,\mu$m.

specimen. Because the tissue has a predominance of low frequencies, the optimal frequencies to include in the pattern are pushed out to around 0.2 cycles per sample.

In Figure 6.2d, the frequency weighting is the same as above, but the focus weighting has been provided by a normal distribution of stage positions, with $dz \sim N(17\,\mu\text{m}, 2\,\mu\text{m})$. In this case, a mixture of low frequencies and high frequencies would provide the best tracking pattern.

6.1.4 The Tracking Noise Floor

Once the pattern frequencies have been chosen, the question remains of how well the pattern will be trackable in an experiment. An answer to this question is provided by analyzing the signal-to-noise ratio (SNR) of the system, where once again the signal is provided by the pattern and the noise is provided by camera noise and the specimen.

The SNR is the ratio of information in the signal to interference in the noise. The information present in a signal is given by the magnitude at the frequencies in the signal. The relationship between a signal amplitude in the spatial domain and its magnitude in the frequency domain is:

$$M = \frac{AN}{2}, \tag{6.14}$$

where N is the number of samples in the signal.

The maximum amplitude in a signal is determined by the dynamic range required to represent it. For digital images, dynamic range is measured by bit depth, where the dynamic range of an image, $I(x, y)$, is

$$b = \log_2 \left(\frac{\max(I(x,y))}{\min(I(x,y))} \right). \tag{6.15}$$

At maximum amplitude, a signal contains both the highest and lowest values allowed by the dynamic range. The maximum amplitude for a signal represented by b bits is

then:

$$A_{max} = \frac{2^b}{2} = 2^{b-1}, \tag{6.16}$$

and the maximum magnitude at any frequency is

$$M_{max} = \frac{A_{max} N}{2} = 2^{b-2} N. \tag{6.17}$$

Given a pattern with dynamic range b_p bits, a specimen with dynamic range b_s bits, and noise of standard deviation σ_n, the signal to noise ratio for saturated pattern and specimen is:

$$\mathrm{SNR}(f, z) = \frac{2^{b_p-2} N \mathcal{H}(f, z)}{2^{b_s-2} N + \bar{M}_n}, \tag{6.18}$$

where $\bar{M}_n$ is the mean magnitude of the noise, provided by $\bar{M}_n = \sigma_n \sqrt{N}$. This is a worst-case estimate of SNR, as it assumes that the specimen has maximum dynamic range at all frequencies. It is preferable instead to compute the SNR only at the frequencies in the pattern.

When the SNR falls below 1 (or 0 dB), the pattern signal is said to be below the noise floor. In practice, the signal from a pattern is required to remain some amount above the noise floor. This technique is evaluated in Section 6.3.3.

6.1.5 Increasing the Signal-to-Noise Ratio

Increasing the SNR requires either increasing the pattern signal or decreasing the specimen signal. The dynamic range required to record the bright and dark regions in a specimen is fixed—that is, a specimen that covers 4 bits of dynamic range to record detail in its brightest and darkest regions on an 8-bit camera will also cover 4 bits of dynamic range on a 12-bit camera. For a specific specimen, an increase in pattern SNR can therefore be obtained only by increasing the pattern's dynamic range. The pattern's optical transmission is limited to the range $\alpha = [0 \ldots 1]$, and every halving of

the minimum optical transmission increases the pattern dynamic range by one bit and consequently decreases the dynamic range remaining for representing the specimen by one bit. Increasing the camera's dynamic range and decreasing the optical transmission of the pattern provides a practical method to increase tracking SNR without affecting the dynamic range used to record specimen information.

The other factor that affects the pattern's SNR is the distance of the pattern from the focal plane. At larger distances from the focal plane, more light diffracts from the bright regions of the pattern into the dark regions, lowering the pattern's effective dynamic range. Note that the optical transmission of the pattern remains the same, but the ratio of light in the brightest and darkest regions of the pattern diminishes. Therefore another practical method to increase the pattern SNR is to decrease the distance between the specimen and pattern.

6.2 Determining Focus

The pattern used in structured illumination microscopy can provide axial as well as lateral tracking. The method relies on analyzing the defocus of the pattern in observed images. In the focus model of Chapter 2, the PSF is symmetric about the optimal focus plane ($z = 0$) and is radially symmetric about the optical axis. The microscope stage is moved axially (in z) to align a plane of interest in the specimen with the focal plane. As the stage moves, different parts of the specimen are in focus at the image plane while other parts move out of focus. As discussed in Section 6.1, the frequencies in the pattern are attenuated by the OTF dependent on the distance from the focal plane.

Knowledge of the micropattern and microscope PSF leads to a method for axial tracking of the microscope stage. The focus model predicts the attenuation of micropattern components expected at different focal distances. The model predictions can be compared to experimental images to obtain an estimate of the micropattern distance

from the focal plane—this estimate provides axial tracking for the microscope stage. Because observations are made with the specimen in focus, knowing on which side of the specimen the micropattern is placed resolves the ambiguity from the axial symmetry of the PSF.

Let $f_i = (r_i, \theta_i)$ be a set of frequencies known to be present in a pattern, specified in polar coordinates where r_i is the distance from the center of frequency space and θ_i is the orientation at which the component appears. The magnitudes observed in an image $I(x, y)$ are given by $|\mathcal{I}(r_i, \theta_i)|$ with the transform to polar coordinates. The magnitudes predicted by a model pattern image $P(x, y; dz)$ for a distance from the focal plane dz are given by $|\mathcal{P}(r_i, \theta_i; dz)|$. Given an observed image and a model of the microscope and pattern, focus estimation seeks to minimize an objective function, for example

$$E(dz) = \sqrt{\sum_i (|\mathcal{I}(r_i, \theta_i)| - |\mathcal{P}(r_i, \theta_i; dz)|)^2}, \tag{6.19}$$

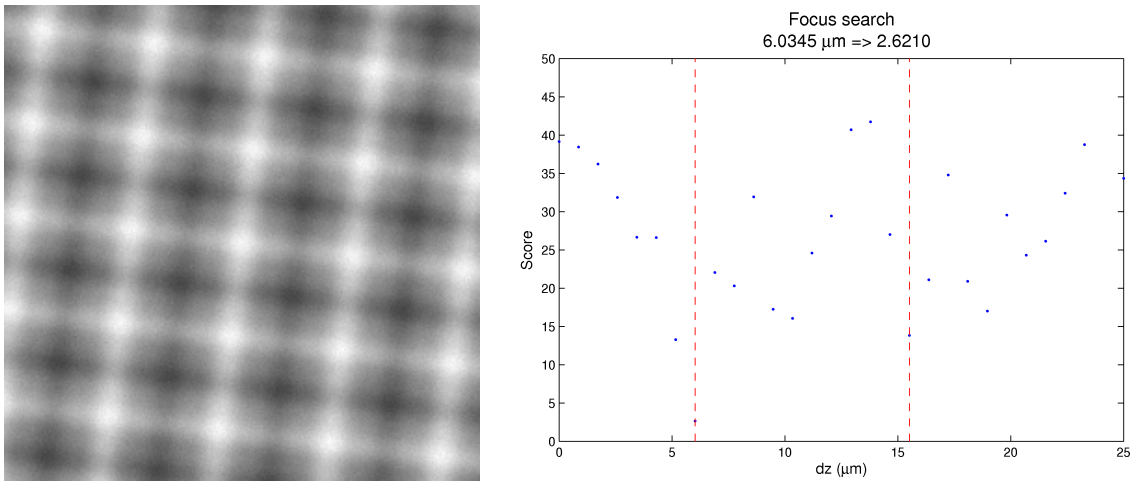
where the distance between observed and modeled magnitudes is given by the ℓ^2-norm.

The focus estimate can be applied to any number of frequencies present in the pattern, but that number may be limited depending on pattern design. Due to the non-monotonic nature of the OTF (see Figure 6.1), the focus estimation objective function of Equation 6.19 is characterized by many local minima over practical focus ranges. Arriving at the global minimum and a reliable estimate for z position relies on either a search at a fine sampling of the objective function or restricting the search range based on experiment parameters.

The former option involves a practical balance between sampling resolution and performance, as evaluating the PSF requires significantly more processing for larger axial displacements. This performance hit can be mitigated by first conducting a coarse search through focus, applying optimization around all local minima, and selecting the best one. This is the approach adopted here, where local minima within 10% of the

smallest value in the range of focus scores are considered.

The latter option is also practical in many imaging situations. One can obtain an initial estimate of the distance between the specimen and the micropattern using a stage micrometer. Once an initial focus range is established, one can restrict the search space by assuming that focus will change little from frame to frame. That is, the focus search can be conducted within a bounded range around the prior focus estimate. With a suitably-constrained search range, optimization of the focus objective function can proceed with a 1D minimization technique, such as golden section search [PTVF07, Ch. 10].



(a) (b)

Figure 6.3: Estimating focus by fitting a pattern model to observed data. a) The pattern is composed of sinusoidal waves with four frequencies along each axis, has a minimum transmission of 0.25 and is placed 6 μm from the focal plane. b) The focus objective function has local minima at approximately 6 and 16 μm, marked by vertical dashed lines.

As an example, Figure 6.3 demonstrates focus estimation for a simulated sinusoidal pattern using four frequencies along each pattern axis. The pattern has transmission $\alpha = 0.25$ and is placed 6 μm from the focal plane. Images are simulated for a 40X, 0.65NA objective lens, 9×9 μm CCD image sensor, an ADC gain of 100 photoelectrons per pixel and Poisson-distributed noise. Figure 6.3b shows the objective function of

Equation 6.19 evaluated for displacements of $dz = [0 \ldots 25]\,\mu$m with the observed image in Figure 6.3a. With this sampling, the minimum occurs at $6\,\mu$m, but another local minimum occurs at $16\,\mu$m. So, accurate z tracking depends not only on there being sufficient pattern SNR, but also sufficient ability to discriminate among different regions of the focus estimation function. In this case, optimization in the regions near the local minima resolves a focus estimate of $6.01\,\mu$m.

The focus measure depends on knowing the orientation of the pattern in the image. A method for estimating pattern orientation was discussed in Section 5.1.1. The focus measure does not depend on the lateral position of the pattern—from the Fourier shift theorem, a translation in the spatial domain affects only the phase in frequency domain. The WPC tracking method does not depend on focus, but MSC does. These considerations suggest the natural order for practically obtaining stage tracking information:

1. Estimate the orientation of the pattern in the frequency domain from the maximum magnitudes at each pattern frequency.

2. Estimate the axial (Z) position of the pattern by comparing the magnitudes of the pattern frequencies to magnitudes predicted by the focus model.

3. Estimate the lateral (X,Y) position of the pattern using either WPC or MSC.

4. Refine the orientation estimate using MSC.

5. Repeat steps 2-4 until convergence.

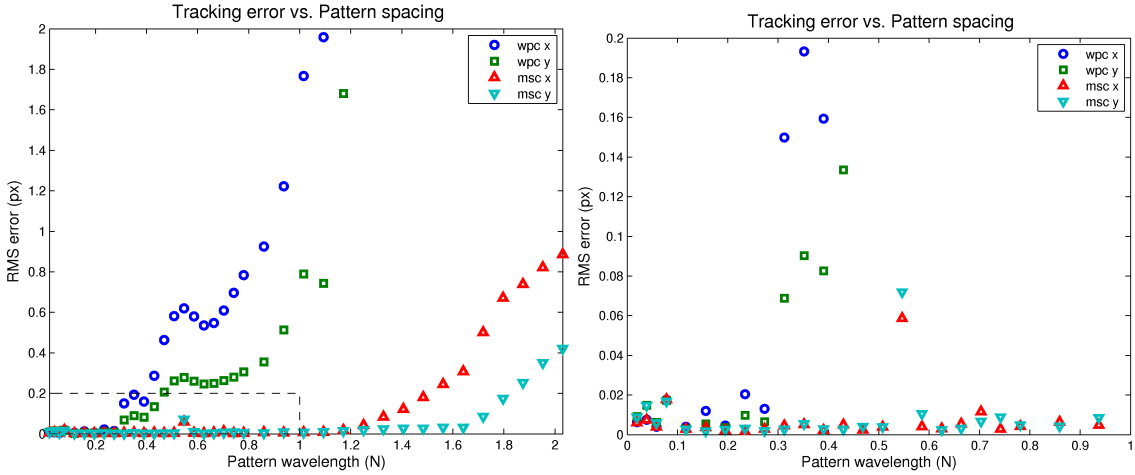
6.3 Evaluation Using Simulated Images

Tracking evaluation can proceed in a similar manner to the evaluation method described in the previous chapter. Evaluating a test scenario begins with choosing appropriate system parameters, including the pattern design, objective lens, camera parameters, and specimen. Test images are then simulated for a variety of stage translations with noise

added. The stage position is estimated for each translation by averaging the translation obtained by the tracking algorithm for a number of different noisy images. Given the errors for each translation the root mean squared (RMS) error is computed for the test scenario.

6.3.1 Pattern Spacing

To evaluate the effect of pattern spacing on tracking accuracy, tracking is performed on simulated images of single frequency sinusoidal patterns. The wavelength of the patterns ranges within $L_n = [\frac{10}{256} \ldots \frac{512}{256}]$. For each wavelength, the pattern is translated along the x axis between $dx = [-\frac{L_n}{2} \ldots \frac{L_n}{2}]$ at 10 even spacings. For each image, the pattern displacement is measured with WPC and MSC in reference to an untranslated pattern image. The pattern is given a transmission of $\alpha = 0.25$, placed $dz = 5\,\mu$m from the focal plane, and imaged with a 40X, 0.65NA objective lens to form 8 bit, 256×256 pixel^2 images. Each image is corrupted by Poisson-distributed noise modeled by 100 photoelectron/pixel ADC gain.



(a) (b)

Figure 6.4: RMS tracking error as a function of pattern spacing. Units along the x axis are in fractions of the field of view. a) RMS tracking error for wavelengths up to two times the field of view. b) Same data as in (a), zoomed in to the region marked by the dashed lines.

Figure 6.4a shows the RMS tracking error in x and y for each tracking method as a function of pattern wavelength. It is immediately apparent that WPC tracking deteriorates significantly for wavelengths longer than $\frac{1}{4}^{th}$ the field of view. This result correlates well with the results from the 1D analysis of Section 5.2, which showed that there is a fixed phase error in WPC.

MSC maintains consistent tracking accuracy for wavelengths up to 1.2 times the field of view. Figure 6.4b shows a subset of the data in Figure 6.4a. This view shows that the RMS error for MSC remains below 0.02 pixel in most cases.

6.3.2 Tracking Without Specimen

Section 6.1.3 showed that optimal structured illumination patterns containing low frequencies are optimal in the absence of a specimen. The tracking test cases from Chapter 5 were rerun for an optimized pattern consisting of orthogonal sinusoidal waves.

Table 6.1: Comparison of the RMS error for tracking square grid and sinusoidal patterns. Bold values indicated the best accuracy obtained for a particular test case considering both x and y accuracy.

		WPC		**MSC**	
Pattern	**Test Case**	**RMSE x** (px)	**RMSE y** (px)	**RMSE x** (px)	**RMSE y** (px)
Grid	Orientation	0.0566	0.0330	0.0023	0.0115
	Translation (large)	0.0064	0.0006	0.2695	0.0156
	Translation (small)	0.0121	0.0015	1.0036	0.0000
Optimal Frequencies	Orientation	4.085	1.8157	**0.0035**	**0.0049**
	Translation (large)	0.0944	0.1033	**0.0025**	**0.0040**
	Translation (small)	0.0040	0.0053	**0.0043**	**0.0034**

Table 6.1 compares the RMS errors obtained for tracking the grid pattern from Chapter 5 and optimized sinusoidal patterns using WPC and MSC. The test cases refer to the orientation and translation evaluations performed Section 5.3, and the data for tracking the grid pattern comes from this section, repeated here for ease of comparison.

In this evaluation, all simulated patterns have transmissions of $\alpha = 0.5$, are imaged in focus with a 40X, 0.65NA objective lens to obtain 8-bit, 256×256 pixel images. The fundamental plus three higher harmonic frequencies are used for tracking the square grid pattern; the sinusoidal pattern contains the four frequencies with maximum score determined in Section 6.1.3 along each axis. The bold values in this table represent the pattern and tracker combination that obtained the best accuracy for a particular test case, evaluated by smallest RMS magnitude, $E = \sqrt{\mathrm{RMS}_x^2 + \mathrm{RMS}_y^2}$. In all cases the optimized sinusoidal patterns tracked by MSC showed the least error, remaining below 0.005 pixel (1.1 nm). Note that WPC performs poorly in this case because the pattern contains only long wavelengths, which Section 6.3.1 showed to be error-prone for this tracking method.

The objective function for the sinusoidal patterns used in MSC is well-suited to gradient-based optimization, where the direction in which to search for a minimum is obtained from an estimate of the objective function's gradient. Gradient-based methods are known to converge faster than Nelder-Mead simplex-based optimization when applied to the right problem [PTVF07, Ch. 10]. Applied to the square grid pattern, gradient-based optimization converges more slowly and to a less accurate solution compared to simplex-based search. Applied to the sinusoidal pattern, gradient-based optimization converges more quickly and to a more accurate solution than simplex-based search. Therefore, switching from the grid pattern to patterns composed of relatively few sinusoids enables using a faster optimization strategy that obtains more accurate results.

6.3.3 Tracking Moving Specimens

Chapter 5 demonstrated that tracking a pattern layer using MSC and WPC was possible in the presence of a specimen that moved in concert with the pattern. This situation is common in histology studies, but in many such cases specimen-based image registration

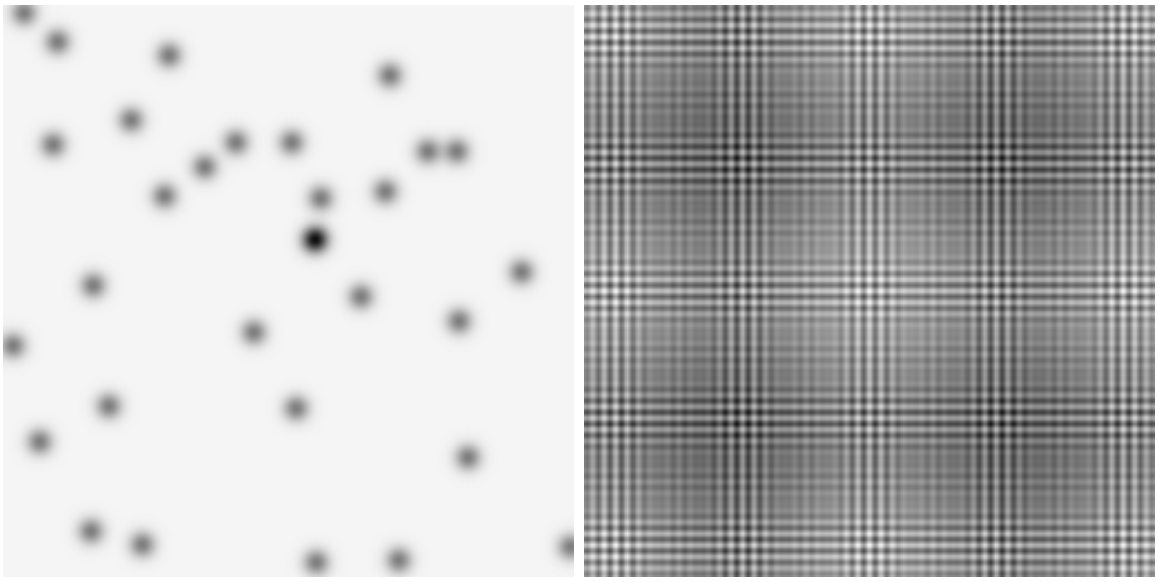
techniques can also provide the tracking data. An obvious exception is when there is not enough specimen information to provide reliable image registration, for example in gaps within a specimen or between sections of tissue.

The full power of structured illumination microscopy includes the ability to track a pattern layer independent of the specimen motion. This expands stage tracking capability from observations of stationary specimens to experiments involving live, moving specimens.

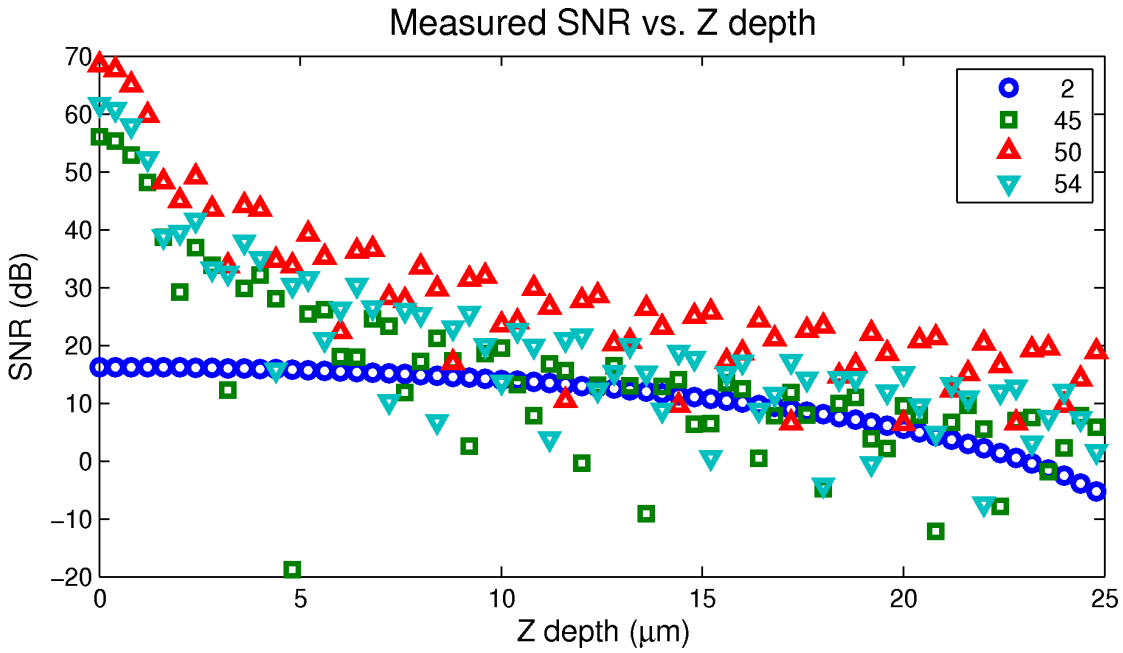
Sparse Specimens

Figure 6.5a shows a simulated image of a specimen with sparsely-distributed contrast. This image represents what one might see when looking at a cluster of opaque beads, for example in a bead diffusion experiment, or at a colony of bacteria. Optimal pattern selection on this specimen image, as discussed in Section 6.1.3, reveals that the best frequencies to use to track a pattern in the region $dz > 5\,\mu$m from the plane appear around 0.18 cycles per pixel. Because patterns at this frequency do not allow for much stage displacement between image frames, the low frequency with the highest score was also selected to be included in this pattern. Figure 6.5b shows the pattern generated with these optimal frequencies. Figure 6.5c shows the SNR for these frequencies as a function of pattern z depth, computed at the pattern frequencies in the presence of the specimen and Poisson-distributed noise with an ADC gain of 100 photoelectrons per count. The legend on this graph displays pattern frequencies expressed as cycles per image.

Tracking the pattern in the presence of a moving specimen was simulated by moving the pattern and keeping the specimen fixed. To maximize the pattern SNR further from the focal plane, the pattern was given a minimum transmission of $\alpha = 0$. The pattern was simulated at different focal depths while keeping the specimen in focus. At each focal depth, the pattern was displaced along one axis within $dx = [-\frac{L_{max}}{2} \ldots \frac{L_{max}}{2}]$,



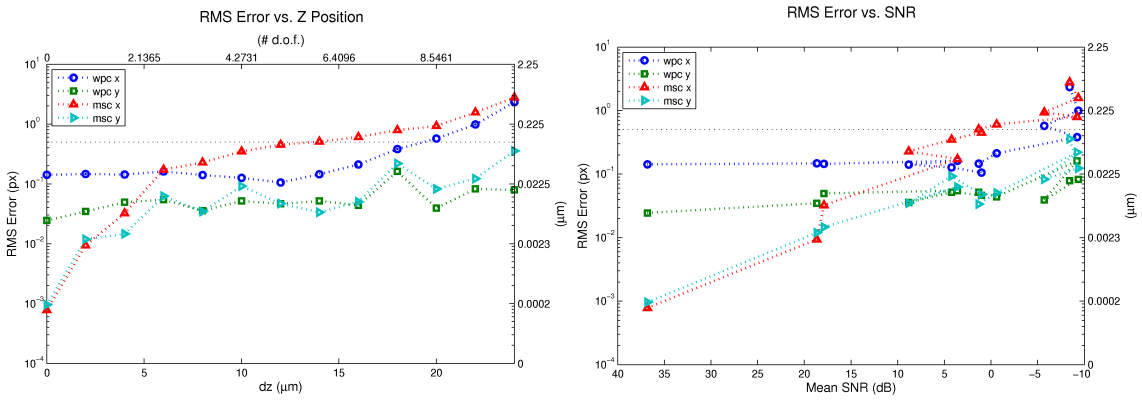
(a) (b)



(c)

Figure 6.5: Predicted SNR for tracking a pattern in the presence of a sparse specimen as a function of pattern distance from the focal plane. a) A simulated specimen with sparsely distributed contrast. b) The pattern composed of three optimal frequencies and one low frequency. c) The SNR estimated for tracking the pattern in (b) in the presence of the specimen in (a) as a function of pattern distance from the focal plane. The legend indicates the frequencies evaluated, shown in cycles per image.

where L_{max} is the maximum wavelength in the pattern.



(a) (b)

Figure 6.6: RMS tracking error in the presence of a specimen with sparse contrast. a) RMS tracking error as a function of the distance between the pattern and specimen. X axes are labeled in μm (bottom) and multiples of a depth of field (top) for the simulated lens. Y axes are labeled in pixel (left) and μm (right). The dotted horizontal line marks 0.5 pixel. b) RMS tracking error as a function of mean SNR. Y axes are labeled in pixel (left) and μm (right).

Figure 6.6a shows RMS tracking errors for tracking the optimized pattern at different distances from the in focus specimen. For this data set, MSC obtains tracking errors below 0.03 pixel for z positions less than 5 μm, outperforming WPC in this range. The tracking error for MSC remains below 0.5 pixel for z positions less than 15 μm. WPC outperforms MSC above 5 μm, maintaining an error below 0.5 pixel to almost 20 μm from the focal plane. This represents a distance of 8.5 times the depth of field for the simulated lens.

Figure 6.6b shows the same RMS tracking errors as a function of mean pattern SNR, considering the frequencies in Figure 6.5. From this plot, the tracking methods maintain an accuracy of better than 0.5 pixel down to the noise floor (0 dB) for WPC and down to 5 dB for MSC.

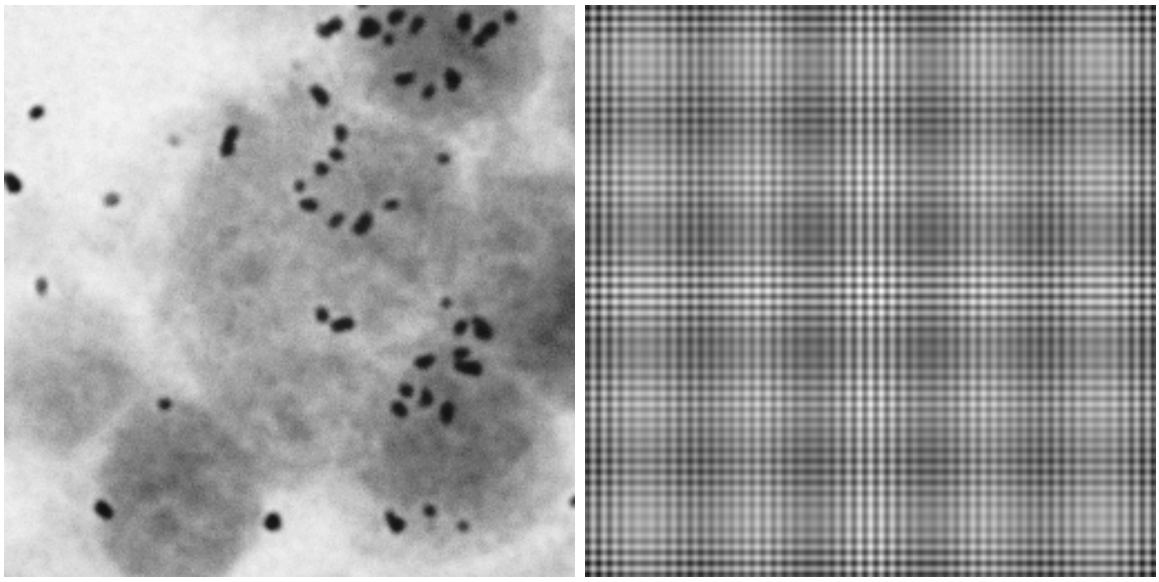
Frog Brain Tissue

To simulate tracking in the presence of specimens with densely-distributed contrast, test images are create using single real microscopy images that remain stationary multiplied by simulated pattern images. The microscopy images come from the Burmeister lab at the University of North Carolina at Chapel Hill (UNC) Department of Biology. Mangiamele and Burmeister use radioactive markers to localize the expression of immediate early genes that signal neural activity in túngara frog (*Physalaemus pustulosus*) brains [MB08]. Figure 6.7 shows a section of frog brain tissue; the gray regions are neurons and the dark spots are silver grains from photographic emulsion used to localize the radioactive markers. Although the brain tissue image remains stationary during these tests, moving the pattern simulates the effect of the specimen and pattern layers moving independently.

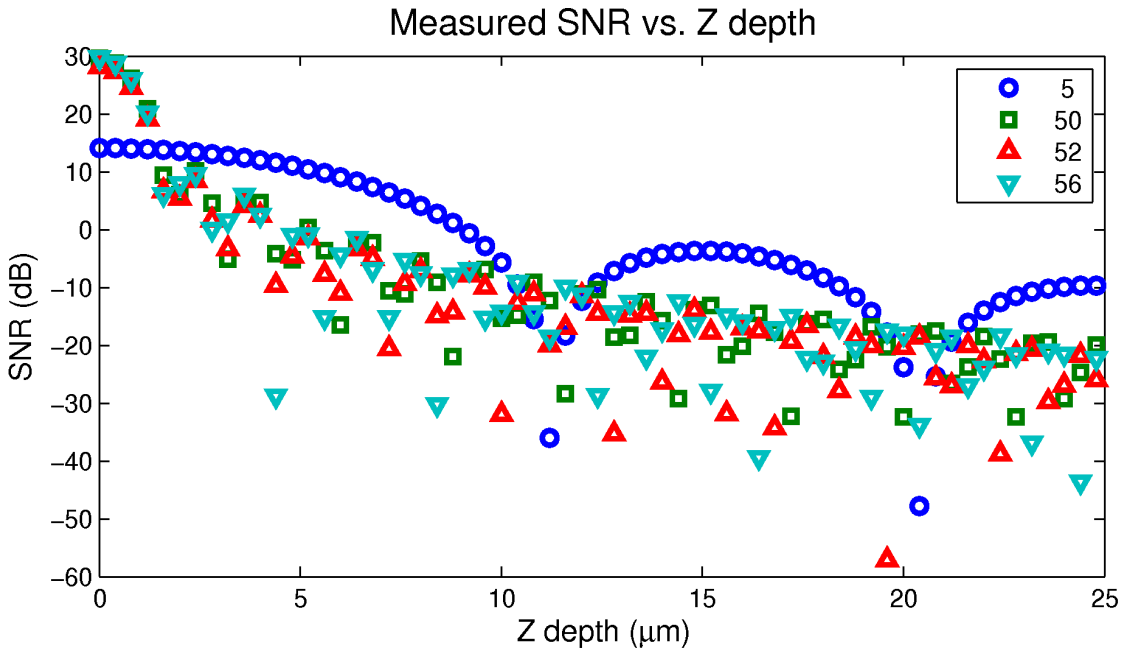
Figure 6.8a shows RMS tracking errors for tracking the optimized pattern at different distances from the in focus specimen. For this data set, MSC obtains tracking errors below 0.5 pixel for z positions less than $4\,\mu$m, outperforming WPC. This provides subpixel tracking for patterns that remain within one depth of field of the specimen. The tracking error for both MSC and WPC becomes greater than 1 pixel for axial distances above $5\,\mu$m.

Figure 6.8b shows the same RMS tracking errors as a function of mean pattern SNR, considering the frequencies in Figure 6.7. From this plot, MSC maintains an accuracy of better than 0.5 pixel ($0.1123\,\mu$m) almost down to 10 dB. Note that this tracking accuracy is well below the Abbe resolution limit of $0.423\,\mu$m for this lens. The tracking deterioration for axial displacements above $5\,\mu$m is well predicted by Figure 6.7c.

This example further demonstrates why the tracking in Chapter 5 suffered for the square grid patterns in the presence of specimens. The square grid patterns are composed of a fundamental frequency and harmonics at decreasing magnitudes. At low frequencies the specimen provides the greatest interference, diminishing the tracking ac-

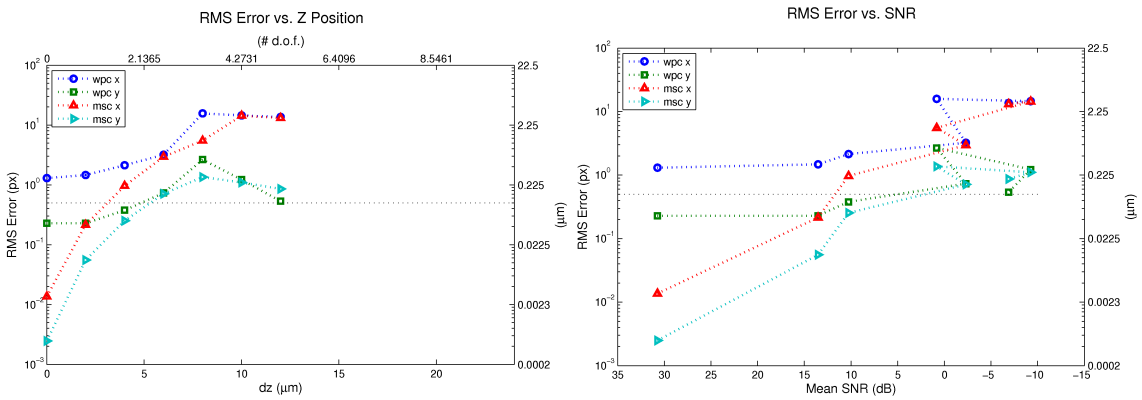


(a) (b)



(c)

Figure 6.7: Predicted SNR for tracking a pattern in the presence of frog brain tissue as a function of pattern distance from the focal plane. a) An image of túngara frog brain tissue, section from Lisa Mangiemele. b) The pattern composed of four optimal frequencies, including one low frequency. c) The SNR estimated for tracking the pattern in (b) in the presence of the specimen in (a) as a function of pattern distance from the focal plane. The legend indicates the frequencies evaluated, shown in cycles per image.



(a) (b)

Figure 6.8: RMS tracking error in the presence of frog brain tissue, providing dense specimen contrast. a) RMS tracking error as a function of the distance between the pattern and specimen. X axes are labeled in μm (bottom) and multiples of a depth of field (top) for the simulated lens. Y axes are labeled in pixel (left) and μm (right). The dotted horizontal line marks 0.5 pixel. b) RMS tracking error as a function of mean SNR. Y axes are labeled in pixel (left) and μm (right).

curacy for frequencies in the pattern with the largest magnitude. At higher frequencies, the magnitudes in the pattern are too low to rise above the camera noise.

6.3.4 Estimating Focus

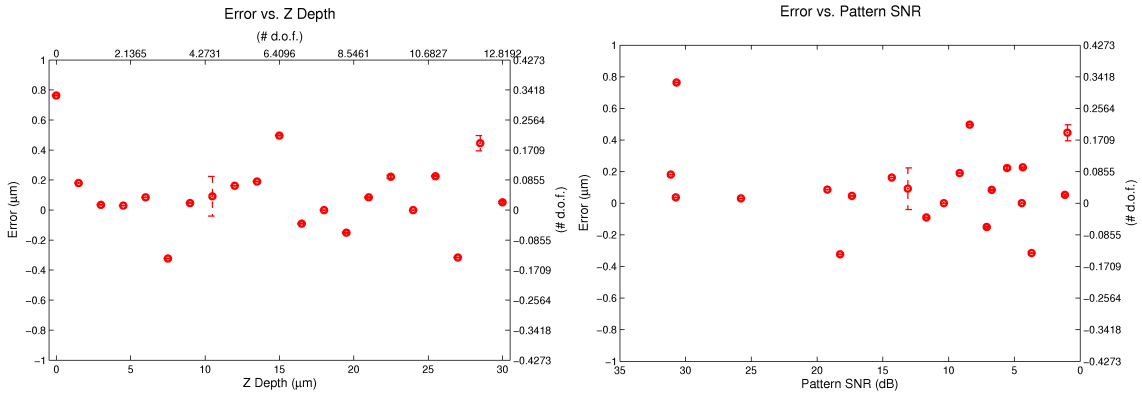
Focus estimation using the approach described in Section 6.2 was evaluated for simulated experiments both with and without a specimen. The axial tracking range is limited by the depth of field of the objective lens—as the pattern moves farther outside the depth of field for the lens, the signal diminishes to a level that cannot be tracked reliably. The depth of field for a microscope system, given by Equation 2.25, is

$$d = \frac{\lambda n}{\mathrm{NA}^2} + \frac{ne}{M\mathrm{NA}}, \tag{6.20}$$

where λ is the wavelength of light used in the observation, n is the index of refraction of the objective immersion medium, and e is the resolvable distance of the image sensor—three times the physical size of the sensor elements. The simulation discussed here

assumes illumination with 550 nm green light, the use of a dry objective ($n = 1.0$), and a CCD sensor element size of $9 \times 9\,\mu$m.

A pattern with minimum transmission $\alpha = 0$ was simulated at different depths from the focal plane using a 40X, 0.65NA objective lens, and Poisson-distributed shot noise modeled with 100 photoelectron per count ADC gain. This optical system has a depth of field of $2.34\,\mu$m. Axial distances to the pattern were within the range $dz = [0 \ldots 30]\,\mu$m and the initial search range is restricted to $dz = [-3 \ldots 3]\,\mu$m, with the assumption that axial position changes slowly from frame to frame. For each focus position, five images are generated with different noise, and the mean estimated z position is recorded.

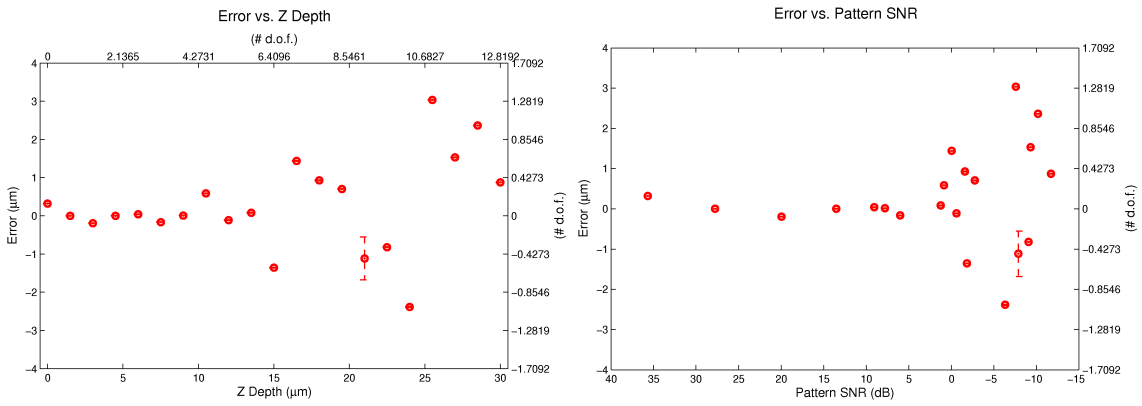


(a) (b)

Figure 6.9: Focus estimation error for a sinusoidal pattern. Four different frequencies along each axis are used to determine z position. a) Mean z position error as a function of axial distance to the focal plane. b) Mean z position error as a function of mean pattern SNR measured at the pattern frequencies.

Figure 6.9a shows mean focus estimation errors for simulated images of a sinusoidal pattern with four different frequency components along each pattern axis. The pattern frequencies were selected using the same criteria used for lateral tracking outlined in Section 6.1.3. In the absence of a specimen, low frequencies are preferred. Except when the pattern is completely in focus, the focus estimation error remains below $0.5\,\mu$m to at least 12.5 times the depth of field. The RMS focus error is $0.266\,\mu$m with a standard error of $0.244\,\mu$m. Figure 6.9b shows the effect of mean pattern SNR on the axial

tracking, demonstrating that tracking persists down to the noise floor.



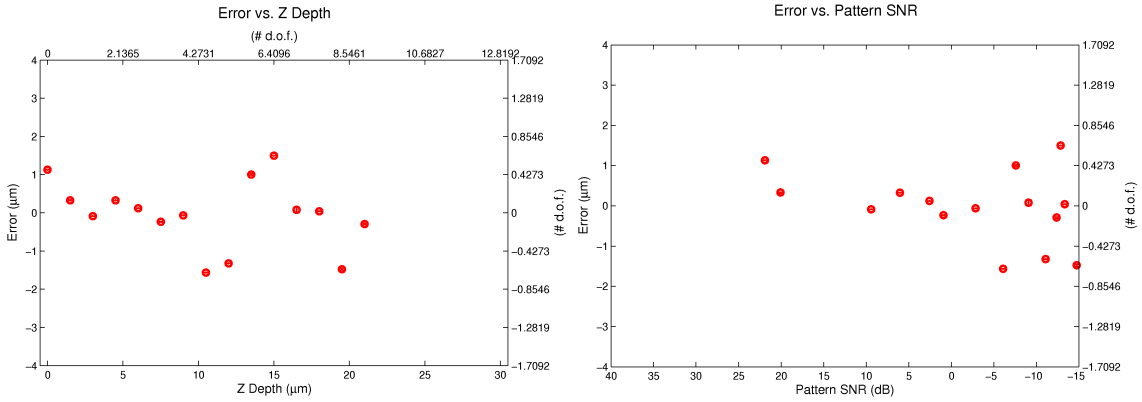
(a) (b)

Figure 6.10: Focus estimation error for an optimal sinusoidal pattern with a frog brain section present, imaged with a 40X, 0.65NA objective lens. Mean focus error *vs.* (a) z position and (b) *vs.* SNR.

Figure 6.10 shows the z tracking error for a simulated sinusoidal pattern multiplied by an image of an in focus túngara frog brain section. The frequencies for this pattern were selected using the optimal lateral tracking criteria, with the additional constraint that a wide range of frequencies are chosen—this amounted to selecting frequencies with local maximum scores that were not among the absolute maximum scores. The tracking search range is bounded within $[-3 \ldots 3]\,\mu$m of the ground truth value, applying the assumption that z position changes slowly from frame to frame. The best tracking accuracy occurs when the pattern is within $dz = [0 \ldots 15]\,\mu$m from the focal plane, maintaining less than $0.5\,\mu$m error in all cases except at $10\,\mu$m, where there is a temporary drop in the SNR. The range extends at least 6 times the depth of field for the lens. In this range, the mean pattern SNR spans $[35 \ldots 5]$ dB. At lower SNRs the tracking accuracy deteriorates significantly.

In contrast, lateral (x and y) tracking for the sinusoidal pattern in the presence of the frog tissue has a RMS error of $0.113\,\mu$m for SNRs above 10 dB. The theoretical axial resolution of a bright-field microscope is half the lateral resolution, so some degree of diminished accuracy is to be expected [RT94, Ch. 3]. The reduced lateral tracking

accuracy corresponds well with the theoretical reduced axial resolution for the bright-field microscope.



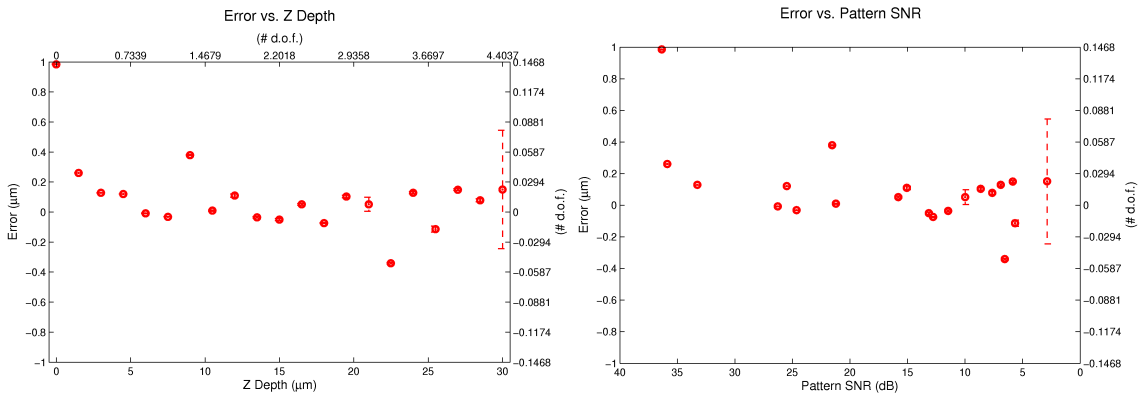
(a) (b)

Figure 6.11: Focus estimation error for a non-optimal sinusoidal pattern with a frog brain section present, imaged with a 40X, 0.65NA objective lens. Mean focus error *vs.* (a) z position and (b) *vs.* SNR.

As a comparison, Figure 6.11 shows the z tracking error for a non-optimized sinusoidal pattern in the presence of the frog brain section. The ranges in these graphs are selected to match those in Figure 6.10. Note that z distances only up to $20\,\mu$m are evaluated in Figure 6.11. Frequencies for this pattern were chosen arbitrarily, and the impact on tracking error is apparent—tracking error is less than $1\,\mu$m in the range $dz = [2 \ldots 8\,\mu\text{m}]$. The pattern SNR drops below the noise floor when the pattern is greater than $10\,\mu$m from the focal plane.

Increasing the z tracking range using structured illumination microscopy requires increasing the depth of field of the objective lens. Lower magnification lenses usually have smaller NAs, and consequently a significantly larger depth of field, as there is a dual effect on the depth of field Equation 2.25. Intuitively, at lower magnifications and NAs, the PSF projects onto a smaller region of the image sensor, so out of focus elements are not smoothed as much over the image plane.

Figure 6.12 shows the z estimation error for simulation experiments conducted with a 20X, 0.40NA objective lens, which has a depth of field of $6.81\,\mu$m. The axial distance to



(a) (b)

Figure 6.12: Focus estimation error for sinusoidal patterns with a frog brain section present, imaged with a 20X, 0.40NA objective lens. Focus error vs. (a) z position and (b) vs. SNR.

the pattern ranges within $dz = [0 \ldots 30]\,\mu$m and the initial focus search is constrained to within $[-3 \ldots 3]\,\mu$m of the true focus position. The objective's increased depth of field enables obtaining z tracking estimates over a larger range. The focus estimate is inaccurate when the pattern is placed close to the specimen because the pattern appearance changes little in this region, and fitting the focus objective function is therefore particularly prone to error. The tracking estimate becomes reliable when the pattern is about $1\,\mu$m from the specimen. The estimate has less than $0.5\,\mu$m error within the range $dz = [1 \ldots 30]\,\mu$m, and the mean pattern SNR ranges from $[35 \ldots 3]$ dB over this range.

6.4 Discussion

This chapter has discussed practical design issues to consider when selecting patterns for structured illumination microscopy. Modeling the microscope OTF is crucial in determining optimal pattern design. The model used here has some known limitations—it is formed from an idealized symmetric PSF. A more complete model would take into account spherical aberrations of the objective lens (rays passing through different parts

of the objective lens focus at slightly different points) and the effect of the cover slip on the OTF. However, the results presented here demonstrate how to apply constraints from experimental parameters to design structured illumination patterns for microscopy tracking.

Pattern selection is further constrained by the requirement that the pattern be manufacturable with a high degree of precision to minimize tracking error. Larger scale patterns can be printed with photographic reduction to create the gradients required for sinusoidal patterns. Photolithography provides a method to produce smaller structures, usually characterized by sharp edges. Current research in microlithography extends the ability to print fine structures with gradients accurately with a reconfigurable, maskless approach [KBPH07].

Using a frequency selection method that maximizes the pattern SNR under experimental constraints, I have demonstrated lateral and axial tracking at or slightly above the noise floor. Different pattern designs and tracking strategies work better under different scenarios. With the simulated 40X, 0.65NA objective lens, which has an Abbe resolution of $0.423\,\mu$m and depth of field of $2.34\,\mu$m:

No specimen present: The pattern should contain low frequencies. Lateral tracking with the MSC technique obtains an RMS error of 1.1 nm (4.9×10^{-3} pixel, 2.6×10^{-3} Abbe units) when the pattern is in focus. Axial tracking is accurate to within $0.50\,\mu$m (2.2 pixel, 1.2 Abbe units) when the pattern is less than $15\,\mu$m from the focal plane.

Observing a specimen with sparse contrast: The pattern should have one low frequency to enable large displacements between frames, and several high frequencies that are above the dominant specimen frequencies. If the pattern will remain less than $5\,\mu$m from the focal plane, lateral tracking with MSC obtains RMS errors better than $0.023\,\mu$m (0.10 pixel, 0.053 Abbe units). From 5 to $15\,\mu$m from the focal plane, WPC obtains RMS errors better than $0.046\,\mu$m (0.20 pixel, 0.11 Abbe

units).

Observing a specimen with dense contrast: The pattern should have one low frequency to enable large displacements between frames, and several high frequencies that are above the dominant specimen frequencies. Lateral tracking is accurate to within $0.11\,\mu$m (0.50 pixel, 0.27 Abbe units) if the pattern remains less than $5\,\mu$m from the focal plane. Above this, the SNR drops too low to guarantee tracking accuracy. For axial tracking, the pattern should have a wide spread of frequencies, selected from local maxima in the SNR optimization. Tracking is accurate to within $0.5\,\mu$m for axial distances of less than $15\,\mu$m, 6.4 times the depth of field.

Chapter 7

Conclusion and Future Work

The following are the contributions made by the research discussed in this dissertation.

I presented a novel technique for recovering a model of stationary objects in multiple layer images. Median gradient estimation of log-intensity images followed by Fourier transform-based integration constructs a model of the transmission through stationary objects in the bright-field microscope light path. This model provides a specimen-free field correction that separates the stationary components from moving components in the images. The median gradient technique provides accurate stationary layer recovery when moving semitransparent objects cover each image location in no more than half the video frames.

Removing the stationary layer from microscopy videos of cilia-driven mucus motion visually enhances the motion. Additionally, the repaired video is more conducive to optical flow computations because there is no longer competing stationary information in the background layer. The stationary occlusion removal technique provides advantages over flat-field calibration because it creates a specimen-specific nonuniform illumination correction and it can be applied to existing videos without calibration data.

I presented a novel application of structured illumination that provides 3D stage tracking in bright-field microscopy. Structured illumination microscopy introduces a semitransparent pattern layer to the optical path of the bright-field microscope. This

image layer transforms the stage tracking problem into a multiple layer image analysis problem.

Magnitude comparisons at the frequencies present in the pattern layer provide estimates of the pattern orientation with RMS error of 0.25° for an in focus square grid pattern. Model-based optimization after pattern tracking improves orientation estimates to 0.066°.

Analysis of the magnitudes present at pattern frequencies also provides an estimate of pattern distance from the focal plane, enabling z position tracking for the microscope stage. Axial tracking with mean error less than 0.5 μm is obtained for distances less than 15 μm in the presences of sparse or dense specimen contrast. This corresponds to a tracking range of up to 6 times the depth of field for the objective lens.

I presented two methods for tracking the semitransparent structured illumination pattern layers. The phase correlation-based approach could be made fast enough for online tracking and provides accuracies to within 0.01 pixel for an unoccluded light path and to within 0.2 pixel for sparsely occluding, moving specimens up to 15 μm out of focus.

The model-based spatial correlation tracking approach obtained accuracies of 0.01 pixel when tracking in focus patterns in the presence of moving specimens with sparse and densely-distributed contrast. This method maintained accuracies of 0.5 pixel when tracking specimens with densely-distributed contrast up to 5 μm out of focus.

Furthermore, I presented a framework for designing optimal structured illumination patterns tuned to specific requirements of microscopy experiments. This framework determines the frequencies that will maximize the pattern SNR at different depths and in the presence of specimens. The computed pattern SNRs enable predicting how well a pattern will be trackable at different axial positions.

The structured illumination microscopy techniques presented here constitute an image analysis system that provides 3D stage tracking for microscopes that do not have

electronically-driven stages. The system cost involves a camera used for recording images, a computer for analysis, and microscope slides marked with micropatterns. The requirements for the tracking camera are relatively modest. The Pulnix camera used in many of the experiments described throughout this dissertation captures 648×484 pixel^2 monochrome images at 120 fps, and costs approximately $\$3,000$ United States Dollars (USD) (Tim O'Brien, personal correspondence). This provides a reasonable upper bound to the tracking camera cost—most of the analysis in this dissertation was performed on 256×256 pixel^2 images, so a lower resolution camera would also suffice.

In contrast, the MCL nanopositioning stage used to provide ground-truth positions in the microscopy evaluations presented here provides a travel range of $100\,\mu$m along three axes and costs approximately $\$40,000$ USD. A recently purchased Ludl positioning stage with 120 mm travel range and 50 nm accuracy costs approximately $\$15,000$ USD (Ricky Spero, personal correspondence). So, the structured illumination microscopy system presented here would provide cost-effective 3D stage tracking for some types of microscopy experiments.

7.1 Future Work

There are many opportunities to continue this research in interesting directions. The following suggests a few of these.

Stationary Occlusion Removal for Structured Illumination Microscopy

A few modifications to the stationary occlusion removal techniques presented in Chapter 4 would adapt this technique to removing the semitransparent micropattern from structured illumination microscopy images. This would be desirable if the final goal of the observation is an unmodulated view of the specimen and not just tracking informa-

tion. When observing moving specimens, the pattern layer acts as a stationary partial occluder, which can be removed using the stationary occlusion removal approach with no modifications.

In structured illumination microscopy of fixed specimens, the pattern and specimen move together. Even in this case, knowledge of the appearance of the pattern layer enables finding an appropriate transmission mask that removes the effect of the pattern. There are two approaches, depending on whether the pattern is repetitive with a small wavelength or unique over a large area. In the former case, the transmission mask can be determined from the many views of the pattern gathered over time. Registering all views so that the pattern in each image appears in the same location while the specimen being observed appears in different locations creates a situation in which stationary occlusion removal would compute the pattern layer, given a sufficient number of frames. In the latter case, the transmission mask must be computed based on knowledge of the pattern design and the defocus of the optical system.

Multiple Focus System

One of the limits of structured illumination microscopy as described here is that the signal from the pattern degrades as it moves farther from focus, and tracking becomes hopeless when the signal drops below the noise floor. A practical solution to this problem may be to use two cameras focused at different planes, one to provide tracking information and the other to provide specimen observations. Most research microscopes split the light in the optical train between the image sensor, which records digital images, and the ocular lens, in which the researcher observes the specimen and searches for regions of interest. The optical path length along these two directions is usually adjusted to set the same focal plane for both of these views. However, one could replace the ocular lens with another camera (a tracking camera), and adjust its position up or down the optical path to focus on the pattern layer. This adjustment would be made at the tracking

camera, so as not to affect the image of the specimen acquired by the specimen camera. The alignment of the microscope would not provide Köhler illumination for the tracking camera, but flat-field calibration would provide some improvement in that regard.

"Global Positioning" Patterns

In Chapters 5 and 6, I described tracking methods that determine position relative to an initial starting position. All of the patterns considered repeated over a relatively small wavelength, up to the size of a field of view. It may be possible to design patterns that provide absolute position information from a single view. For example, sinusoids with relatively prime wavelengths repeat only at intervals of the product of the two wavelengths. The position within that range can be obtained by comparing the phases of the two sinusoids. Extending to 2D would require a coordinate system established by several patterns with relatively prime wavelengths. At a minimum, three sinusoids would be required, arranged to provide a triangular region within which each set of phases would be unique. A pair of sinusoids along each axis would provide a Cartesian coordinate system. Using a few relatively prime wavelengths would disambiguate a larger range of motions. For example, along an axis with sinusoidal wavelengths of $L_1 = 58$ pixel and $L_2 = 75$ pixel, motion would be unambiguous up to $\frac{58*75}{2} = 2175$ pixel in each direction, equivalent to several fields of view for many image sensors. Expanding the pattern to contain more wavelengths could extend this disambiguation to span an entire slide or a whole collection of slides. This latter case could be useful in histology where sections of a single tissue sample may be spread over a large number of slides.

Semitransparent Augmented Reality Markers

Human-scale augmented reality systems often use patterns printed on markers to provide an image feature that can be recognized in images. The augmented reality system determines the geometry of the marker in the scene, and projects an arbitrary digital

model onto the marker, appropriately scaled and oriented. In some cases, the marker is still visible in the augmented reality image. Using semitransparent pattern markers would enable removing the marker completely from the scene, enabling the projection of semitransparent virtual objects that maintain visual consistency in the scene. Additionally, if the shadow of the semitransparent marker could be recognized in the video, an appropriately scaled and oriented shadow of the virtual object could be projected into the scene.

Removing Atmospheric Effects

Many videos, especially of outdoor sports events and nature videos, are corrupted by atmospheric effects, such as falling rain and snow, and water droplets splashed onto camera lenses. Falling snow and rain behaves similarly to a moving, semitransparent layer in the video, and it should therefore be possible to extend some of the concepts in stationary occlusion removal to digitally enhance such videos. This is an active area of research—recent research has used spatio-temporal correlation analysis to detect streaks in videos characteristic of rain and snow [GN04, BKN07, GN07]. Water droplets on lenses constitute an image layer that refracts light, redirecting the intensities in an image from one location to another. In some cases, an inverted image of the entire scene appears within a water droplet. The occlusion removal technique has been shown to provide some enhancement of video corrupted by water droplets, but a more complete solution would involve modeling the refocusing of light through the water droplets.

Appendix A

ImageTracker: Motion Analysis Software

One component of my research has included designing image analysis software for microscopists. My software package, *ImageTracker*, contains analysis algorithms developed by myself (such as stationary occlusion removal [ET07]) and others (including the local-global optical flow algorithm of Bruhn, Weickert, and Schnörr [BWS05]). ImageTracker is free software distributed by Computer Integrated Systems for Microscopy and Manipulation (CISMM).

This appendix includes a description of the ImageTracker software as well as instructions for building the application on Windows and Linux systems.

A.1 Introduction

ImageTracker is an application designed to track and measure motion in image sequences. Videos often contain motion of several types, and analysis of these motions can be difficult. For example, videos of cells dividing show the cell moving within solution, kinetochores moving back and forth across the cell equatorial plane, and microtubule structures migrating from the kinetochores to the poles. ImageTracker enables analysis of this type of motion through a multiscale approach. The video is first stabilized with respect to global motion and flow analysis of smaller scale structures follows.

ImageTracker was written to analyze microscopy videos. Several challenges common to video microscopy may impede the accuracy of analysis, depending on the nature of the video being analyzed. Specifically, accuracy may suffer on videos that have low signal-to-

noise ratio (SNR), low spatial resolution resolution, sparse temporal resolution (objects move too much between fames), and short object life (features appear and disappear frequently). These challenges are not unique to the algorithms in ImageTracker—in fact, addressing these challenges is the subject of much image analysis research.

The text of this guide refers to the Windows version of ImageTracker, but the screen shots are from the Linux version. The difference between ImageTracker versions on these two platforms is purely aesthetic.

A.2 Installing and Running

Remove previous versions of ImageTracker from the computer before installing updates. To uninstall ImageTracker, select Start → All Programs → NSRG → ImageTracker → Uninstall ImageTracker from the Windows Start menu.

Download and execute the installer from the Computer Integrated Systems for Microscopy and Manipulation (CISMM) download page, http://cismm.org/downloads/, and accept all default parameters during installation. This will create a folder on your hard drive, C:\NSRG\bin\ImageTracker by default, and will add to your programs menu a folder NSRG\ImageTracker.

Run ImageTracker by selecting Start → All Programs → NSRG → ImageTracker → ImageTracker. The main ImageTracker window is shown in Figure A.1. The window consists of a menu bar on the top, a data control panel on the left, and an image panel on the right. ImageTracker will launch new dialog windows for different purposes. Most of these windows may be moved around the screen, resized, or temporarily hidden.

Program information will be written to a logging file at various points during execution. This information may be useful for making sure the application is still running or troubleshooting the application if something goes wrong. You can view the log messages by selecting View→Logger from the ImageTracker menu. Each log message has an asso-

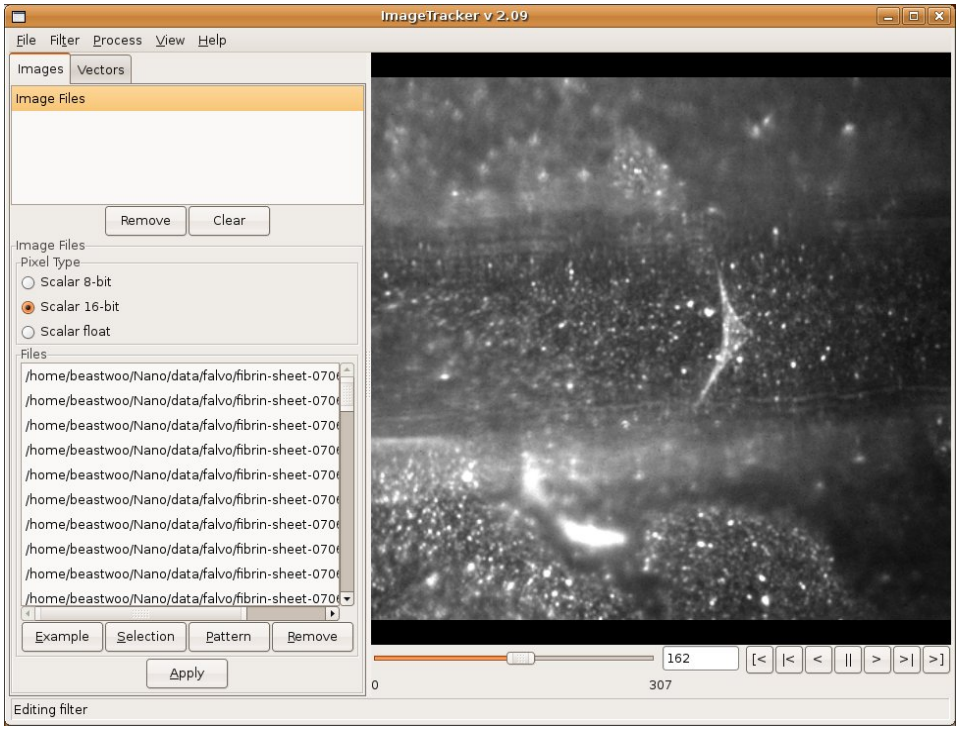


Figure A.1: The main window for the ImageTracker application. Microscopy image from Daniel Millard.

ciated logging level on the following scale of increasing severity: Verbose, Debug, Info, Warn, Error. From the logging window you can limit what messages get logged—all messages at or above the selected logging level will be written. Info is the recommended logging level; lower levels can produce enough messages to noticeably slow down the application. The log messages displayed in the logger dialog do not update automatically. Press the Refresh button to refresh the logger messages.

Occasionally messages from libraries ImageTracker depends on may be printed to other locations. Some of the Insight Toolkit (ITK) classes used in image registration print messages to the console (standard out) if ImageTracker is launched from a command line (this is not common on Windows). Visualization Toolkit (VTK) classes sometimes print messages to a separate logging window; this usually indicates something has gone very wrong with the visualization system.

When ImageTracker is performing an image processing task, a progress bar will

appear displaying a graphical depiction of about how much of the task has completed as well as an estimate of the time remaining to completion. A Cancel button on the progress dialog enables a user to request that a process terminate. The task may not terminate immediately after pressing the Cancel button, however. ImageTracker will wait until the next convenient point to stop the process. A few processing tasks may ignore requests to stop all together—this may be because at the next convenient stopping point the task is nearly finished anyway.

A.3 Working with Images

ImageTracker operates on stacks of individual image files—a set of images constitutes a video. ImageTracker can read many types of images, including TIFF, PNG, JPG, MHA, and VTK. Currently, ImageTracker does not support Metamorph stacks, video files (e.g. AVI, WMV), or raw camera files. Software such as ImageJ, supplied by the National Institutes of Health (NIH), can be used to convert images from many file formats into a format ImageTracker can read.

A.3.1 Loading Images

ImageTracker can have one active image stack and one active vector stack loaded at a time. To load an image stack, select the Images tab from the data control panel and select the Image Files item from the filter list at the top. An Image Files control panel appears below the filter list, as seen in Figure A.2. Below the filter list is a list of file names that will be loaded into ImageTracker. Modify this file list by using one of the buttons at the bottom of the list.

Example Select a single image, and let ImageTracker find files in the same directory with the same naming convention. This is the easiest and most common way to specify a series of files.

Figure A.2: The Image Files control panel specifies which image files are loaded in ImageTracker.

Selection Select multiple files to add to the file list. The file names will be sorted alphabetically as they are added. A maximum of around 1000 file names can be selected at one time.

Pattern Specify a file series naming convention pattern by selecting a root directory, file prefix, number format, file extension, and numeric range. Press the Enter key as you modify each field in the pattern specification dialog to update the example name displayed below. (This option is seldom used, but is appropriate for loading a subset of a large sequence of files.)

Remove Remove any selected file names from the list.

When the list of file names is complete, click the Apply button to load the named files into ImageTracker. The first image in the list will be displayed in the image panel. The Apply button should be pressed whenever any changes are made to the Image File

control panel to have those changes reflected in ImageTracker.

A.3.2 Image Controls

In the main ImageTracker window, images loaded in the Image Files panel are displayed in the main image display panel on the right, as shown in Figure A.1. Controls below the image display panel adjust the index of the frame that is displayed. Move the slider to scan through the loaded images. To jump to a specific frame, enter the frame index in the text box and press Enter. A set of buttons to the right of the slider enables playing, rewinding, pausing, single frame advance and rewind, and skipping to either end of the image sequence.

Images and vector visualizations are displayed simultaneously. The frame index range is based on the larger of the number of loaded images and loaded vector images. If the two sequences are of different sizes, the visualization from the shorter sequence will no longer update once the frame index exceeds the length of that sequence.

The display properties of the image display panel can be adjusted by selecting View → Window/Contrast from the ImageTracker menu. This dialog controls only the display of images and does not change the actual intensity values. The Maximum and Minimum sliders change the intensity values displayed in the image panel. The Range combo box controls the range of values displayed on the intensity sliders. The Auto button automatically adjusts the maximum and minimum display values to the maximum and minimum intensity values in the currently displayed image. You may need to adjust the display contrast when switching between different types of images (*e.g.* changing from 8-bit to 16-bit images).

To change the magnification of the current image, right click on the image, and drag the mouse up or down. To move the current image, center click on the image and drag the mouse in the desired direction. To rotate an image, left click on the image and drag the mouse.

Image size and intensity statistics for the currently displayed image can be shown by selecting View → Image Info from the menu. This will display the Logger window with the most recent message containing the image information. Note that a log level of Info or lower needs to be selected in order for the image information to be displayed.

A.3.3 Loading Vector Images

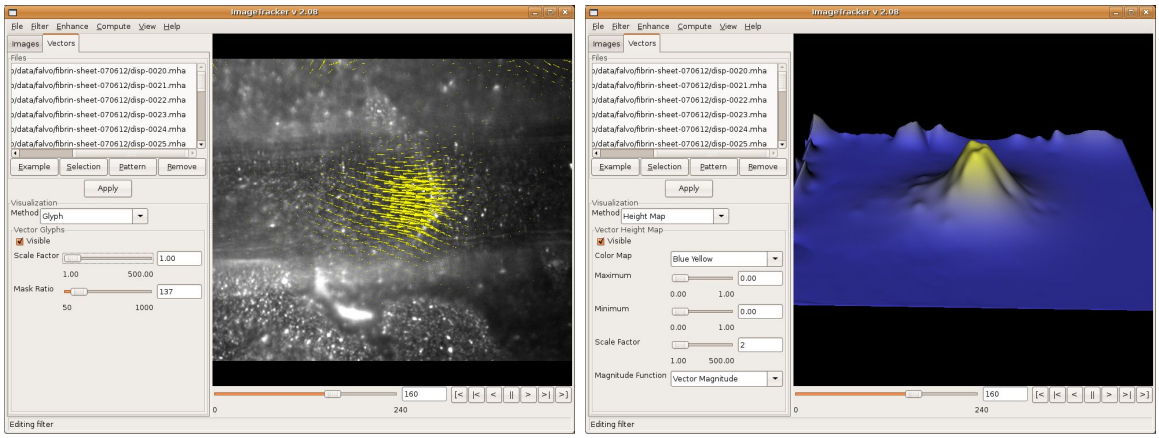
A *vector image* is an image in which every pixel contains vector information instead of intensity information. For example, the output of an optical flow computation is a vector image where vectors explain the motion of intensities between images.

Vector data is loaded in a similar way to image data. Click on the Vectors tab of the data control panel to bring up the Vector Files control panel. A list of vector image files can be modified in the same way as the list of image files on the Images tab. The Visualization control panel below the vector file list provides options for how to display the vector image data.

Vector images can be displayed with two different visualizations: Glyph and Height Map. The Method combo box in the vector visualization control panel selects between these two options. Controls below this combo box enable modifying aspects of the visualization.

If the controls for the current visualization are not visible (this sometimes happens when the vector images are automatically loaded as the result of a processing task), press the Apply button below the vector files list. This will update the visualization control panel.

The Glyph visualization displays regularly-spaced vector glyphs across a vector field, as seen in Figure A.3a. Each glyph is scaled according to the vector magnitude and oriented according to the vector direction. The vector visualization control enables rescaling glyphssetting the scaling to 1.0 will match the glyph lengths to the underlying vector magnitudes. Drawing a glyph at every pixel location would clutter the visualization; the



(a) (b)

Figure A.3: Vector file visualization in ImageTracker. a) Vector glyph visualization and (b) vector height map visualization.

Mask Ratio slider adjusts the spacing between adjacent glyphs. Some spacings produce distracting patterns of glyph placement. Adjusting the Mask Ratio up or down by a few values will usually correct this.

Height map visualization maps the magnitude of the vector image to the height of a surface, as seen in Figure A.3b. It may be helpful to rotate the image plane to better view the surface when using this visualization. The Color Map combo box controls the colormap used to paint the surface, where color is a redundant encoding of vector magnitude. The Scale Factor slider controls the scaling of the height map. The Magnitude Function combo box selects between height mapping the vector magnitude or a single vector component.

A.3.4 Saving Visualizations

To save a visualization displayed by ImageTracker, select File → Save View Images. This will bring up a dialog that enables selecting an output directory, file naming convention, and the range of frames to save. The images saved will be exactly what is displayed in the ImageTracker image panel. To save the result of applying a set of filters, see Section A.4.

It is important to not let any other window obscure the ImageTracker image display panel when saving visualizations. Doing so will corrupt the saved images, as the underlying image capture procedure copies data from the screen buffer. The frame index slider provides progress information and a message will be displayed when ImageTracker has finished saving the visualizations.

Due to an error in the VTK class that handles image capture, if a set of visualizations is saved, the image display panel is resized, and then another set of visualizations is saved, ImageTracker will crash.

A.4 Filters

ImageTracker has two main image manipulation modes: filtering and processing. Filtering applies an operation on images as they are displayed in the ImageTracker image panel, and results are visible instantly. (This is similar to filtering in other imaging applications such as ImageJ or Gimp.) Processing typically involves more complex computations that create a new set of output images. The output images must be loaded into ImageTracker to see the results of a processing task, though this is often handled automatically by ImageTracker.

Filtering is often used to prepare an image set for processing. The list at the top of the Images tab in the control panel contains the names of all filters that are being actively applied. The Image Files filter is a special filter that is always present and can be used to load image files as shown above. Other filters can be added to the list by selecting them from the Filters menu. Filters are chained and applied in the order that they appear in this list. Selecting any filter in the filter list displays that filter's configuration in the panel below. Changes to filter parameters are visible instantly in the image window.

The Remove button at the bottom of the filter list removes the currently selected

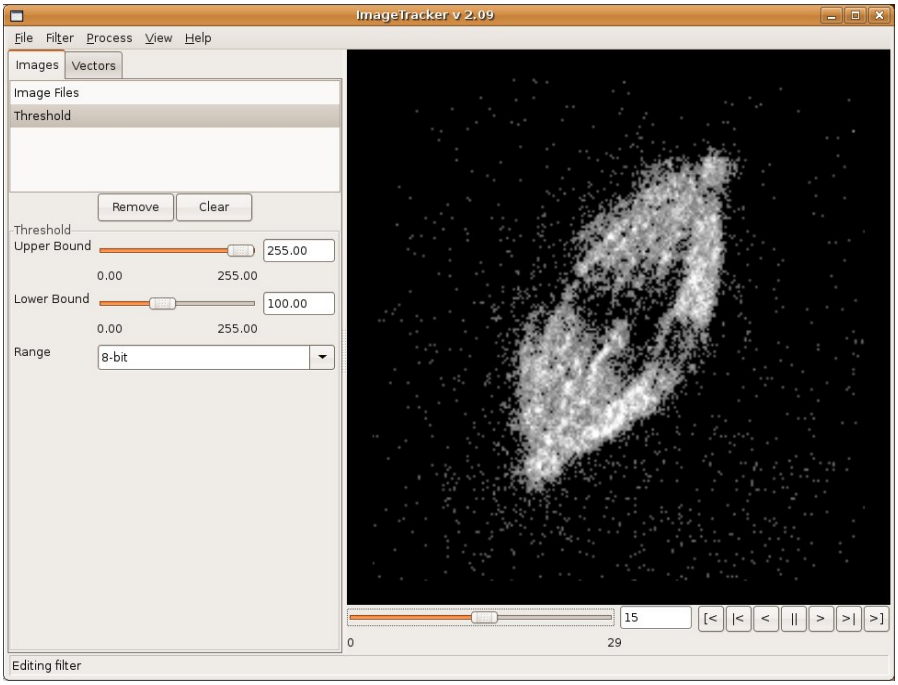


Figure A.4: Result of applying the threshold filter to a fluorescence microscopy image of a mitotic spindle. The threshold filter limits the range of intensities present in an image. Microscopy image from Lisa Cameron.

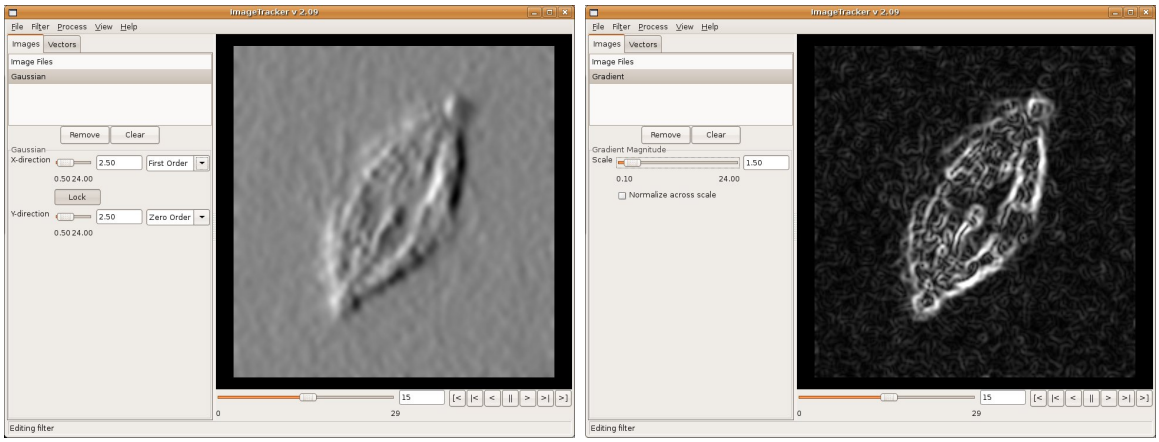
filter from the filter list. The Image Files filter is a special filter that cannot be removed (with no image files loaded there is no image data to filter). The Clear button resets the filter list to its initial state with no filters or image files loaded.

A.4.1 Threshold

The threshold filter sets any pixels from an input image that lie outside an intensity range to zero (black). An example of applying a threshold is shown in Figure A.4. The threshold control panel has the following controls.

Upper and Lower Bound Sliders that adjust the range of pixel intensities that pass through the threshold filter.

Range Controls the range of pixel intensities displayed on the threshold bound sliders. There are presets for 8-bit and 16-bit images, and an option for setting the slider range to match the intensities of the current input image.



(a) (b)

Figure A.5: a) Result of applying the Gaussian filter to a fluorescence microscopy image of a mitotic spindle. A first order Gaussian is applied along the x-axis, while a zeroth order Gaussian is used on the y-axis. b) Result of applying the gradient magnitude filter to the same image.

A.4.2 Gaussian

The Gaussian filter convolves an input image with a Gaussian kernel. Convolution with a zero-order Gaussian smooths (blurs) the image. The scale (standard deviation) of the Gaussian filter determines the degree of smoothing applied. Convolution with a first-order Gaussian along one dimension computes the derivative of the image along that dimension at the scale of the Gaussian. An example of applying a Gaussian filter is shown in Figure A.5a. The Gaussian filter control panel has the following controls.

X- and Y-direction Scale Sliders that select the Gaussian kernel scale along the x and y dimension of the input image. The scale determines the standard deviation of the Gaussian kernel in pixel units.

Lock When the Lock button is depressed, the X- and Y-direction Scale sliders will move together. Otherwise, the scales can be adjusted independently.

X- and Y-direction Order Controls the order of the Gaussian filter applied along each dimension. Zeroth, first, and second order Gaussians can be selected.

A.4.3 Gradient Magnitude

The Gradient Magnitude filter is often used to find edges in an image. The gradient (first derivative) of the input image is computed in each direction with a Gaussian kernel at a spatial scale (see Section A.4.2), yielding I_x and I_y. The gradient magnitude is then given by: $G = \sqrt{I_x^2 + I_y^2}$. Edges, regions with large gradients, are emphasized by this operation. An example of applying a Gradient Magnitude filter is shown in Figure A.5b. The Gradient Magnitude control panel has the following controls.

Scale The spatial scale of the Gaussian kernels used in gradient computation. Larger values find the edges of larger image features.

Normalize Across Scale When this option is checked, the brightness of the filtered image will not fade at larger spatial scales. When this option is disabled, use the Window/Contrast control to rescale the contrast if the image becomes dim.

A.4.4 Logarithm

The Logarithm filter takes the logarithm (base 10) of an image. Zero values in the input image are mapped to zero values in the output image. This filter has no options.

A.4.5 Flat-field

As described in detail in Section 2.3, constant image flat-fielding and subtraction are techniques commonly used in microscopy to correct for imaging artifacts. Flat-fielding divides input images by a constant image to correct for uneven illumination or occluding objects in bright-field imaging. An appropriate flat-field image would be an image of a blank slide or "bright" calibration (or the mean of several such images) or the transmission map obtained from the Remove Occlusions process (discussed in Section A.5.1). Figure A.6 shows an example of applying a flat-field filter to an image.

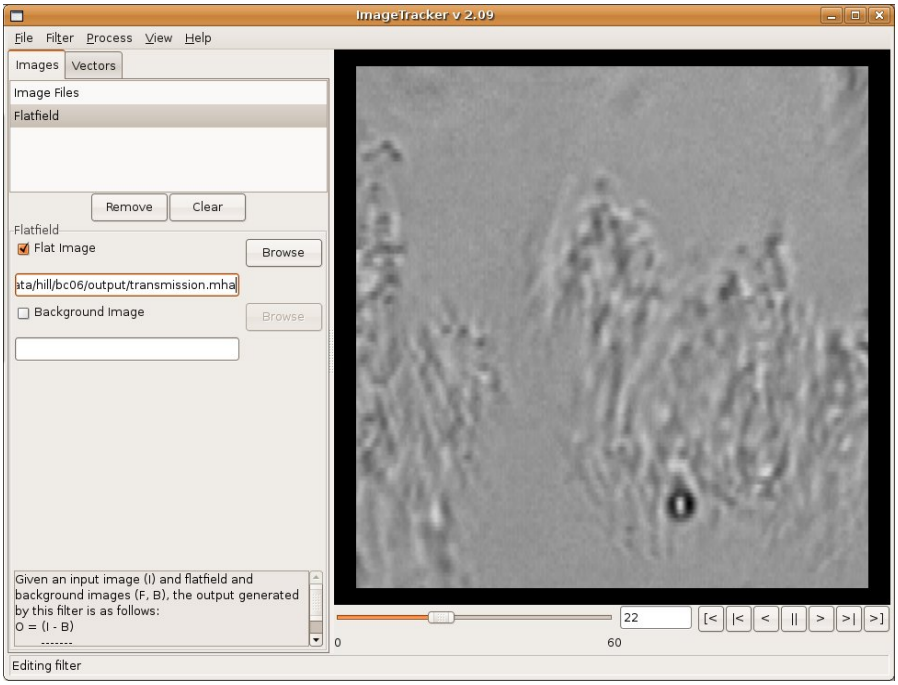


Figure A.6: Result of applying a flat-field filter to a bright-field image of cilia beating on a cell culture. The flat-fielding has removed the non-moving cell bodies from the image.

Background subtraction subtracts a constant image to correct for sensor noise or background signal. An appropriate background image would be a "dark" calibration image taken with the lamp off (or the mean of several such images). The Flat-field control panel enables using either or both of these techniques simultaneously. When both background, B, and flat-field, F, images are supplied, the output image, O, is given by: $O = \frac{I - B}{F}$. That is, the background image is subtracted from the input image before flat-fielding is applied.

The following controls compose the Flat-field control panel.

Flat Image Specifies whether flat-fielding should be applied. The control prompts for the flat-field image when checked for the first time. To select a different flat-field image, use the Browse button.

Background Image Specifies whether background subtraction should be applied. The control prompts for the background image when checked for the first time. To

select a different background image, use the Browse button.

A.4.6 Saving Filter Images

A set of filtered images can be saved by selecting File → Save Filter Images from the menu. This will save the output from the filter pipeline that is displayed in the image filter list, *not* the visualization displayed in the image panel. The Save Filter Images dialog includes controls for specifying the output directory and file naming convention to use. There are also several output options:

Pixel Type Specifies the data type of pixels in the output images. Use 16-bit for images with intensities in the normal camera output range (floating point values will be truncated). Use floating point for computed images that have small fractional values.

Rescale Intensities When this box is checked, the output image intensities will be rescaled to match the full range allowed by the output pixel type.

ImageTracker can save image files in any format supported by ITK. This includes TIFF, PNG, JPG, MHA, and VTK among many others. Some file types (*e.g.* TIFF and PNG) only allow integer pixel type data, and saving floating point pixel data will likely produce an error. When saving floating point images, select a file extension that has floating point support (*e.g.* MHA and VTK).

A.5 Processes

Processes in ImageTracker are generally more computationally intense than filters, and may operate on more than one image at a time. For example, optical flow computation finds a motion field that explains the intensity changes between images in a sequence. The optical flow process results in a set of output vector images saved to disk. Once

the process launches, ImageTracker performs the flow computation on all image pairs in the sequence. When the process is complete, ImageTracker usually loads the result, depending on the process.

Whenever ImageTracker needs to save image data, a dialog box will prompt for output. The output options are always the same.

Directory The directory in which to save all output files.

Prefix The prefix to apply to saved images.

Number format A C `fprintf()`-style number format to append to each image in the output sequence, *e.g.* %04d, which means a zero-padded 4 digit integer.

Extension The file extension to apply; ImageTracker will sometimes choose this for you.

File index range The beginning and ending index for the file sequence; ImageTracker will usually choose this for you.

ImageTracker saves images in 16-bit gray scale format. 16-bit images may appear completely dark in image viewers that only handle 8-bit images. ImageTracker, of course, is a 16-bit image viewer, and can be used to view the images it outputs.

A.5.1 Occlusion Removal

As discussed in Chapter 4, occlusion removal may be able to remove fixed elements from bright-field microscopy videos in which the features of interest are moving. This method depends on the specimen moving and on the fixed occlusions transmitting a constant portion of the microscope's illumination. The result of this process is a set of corrected images in addition to the transmission map that holds the computed constant transmission factor at every location on the image plane. ImageTracker can open this

type of image (it has floating point pixels), which can be useful to see the structure of occluding regions. In the transmission map, bright values indicate areas of the image plane that constantly transmitted a larger percentage of light than darker areas.

The transmission map computed from a short image sequence can be used as a flat-field image applied to a longer image sequence captured with the same imaging parameters.

Select Process→Remove Occlusions to show the Remove Partial Occlusions dialog. The following parameters are options in this dialog.

Metric Specifies the computation that will be used to estimate the constant amount of light absorbed at each image location. The Mean metric is faster but the Median metric performs better if there are slow moving objects in the specimen.

Max Transmission Adjusts the overall brightness of the final images. Select 100 to only increase the brightness of the image everywhere. Select 50 to maintain the average intensity between the input and output images. Note that some intensity values of the output images may become clipped to the maximum allowable output level. If this happens—the resulting images look washed out or saturated—decrease the value of this setting.

Fourier Padding This value is seldom changed. This setting corresponds to a windowing applied to images when using the Median transmission metric. The default value of 0.5 will use a window of half the image size on each boundary when computing Fourier transforms.

Output The common process output options are available.

Transmission Map This is an image file in which to store the transmission map. Use an image format that supports floating point, such as MHA or VTK.

Open output when finished The output images are automatically opened when the process finishes if this box is checked.

A.5.2 Stabilize

Stabilization (registration) aligns image sequences with respect to global translation and rotation. To align a pair of images, the transform that best aligns intensities in the two images is found. The effect of registering all images in an image sequence is that strong image features remain stationary from frame to frame. Note that using a threshold filter (Section A.4.1) as a preprocess can often significantly improve the registration by segmenting the objects of interest from the background.

Multiple resolution registration aligns images iteratively using a series of coarse to fine image resolution scales [LK81, ISNC05]. This method provides several advantages over registration at a single scale: it is more robust to large initial displacements, it enables targeting the scale of features to consider for registration, and it typically performs faster. The method relies on the user specifying a series of image resolutions. Registration is performed in multiple passes, starting at the most coarse resolution. The result of each registration is propagated to the next finer resolution. Lower resolution images are generated by smoothing and rescaling a higher resolution image to half its width and height.

Launch the Multi-Resolution Registration dialog by selecting Process→Stabilize from the ImageTracker menu. The alignment dialog consists of the following options.

Maximum Smoothing Specifies the upper (coarsest) bound of the multiresolution registration. Adjust this setting to smooth all but the coarsest image features that should be aligned.

Minimum Smoothing Specifies the lower (finest) bound of the multiresolution registration. Adjust this setting to smooth noise and any small scale features that

should not be considered for alignment.

Optimization The parameters in this section are seldom changed; they tweak the inner-workings of the multiresolution registration algorithm.

Iterations The maximum number of alignments to try at each resolution scale.

Initial Maximum Step Length The largest translation (in pixel units) to try at each alignment iteration at the coarsest scale.

Initial Minimum Step Length The smallest translation (in pixel units) to try at each alignment iteration at the coarsest scale; for motion below this, the algorithm continues to the next finer resolution.

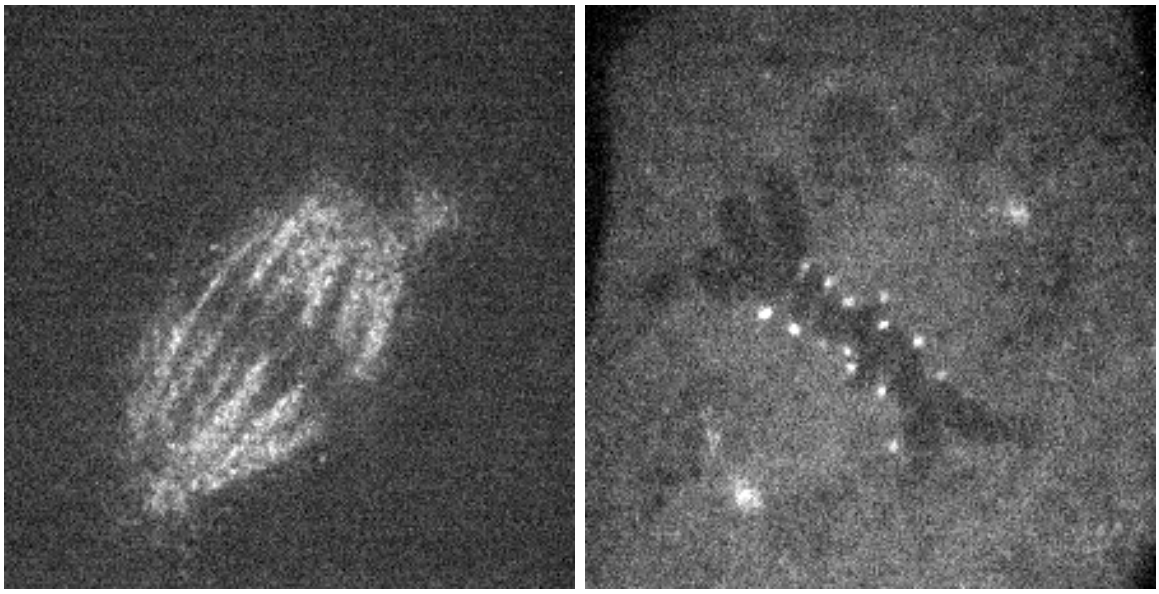
Inter-level Step Scale The factor by which to divide the Step Lengths at each finer scale.

Output The common output options are available.

Transform File A text file in which to store the image alignment transforms. This file is useful for apply the same alignment to a different set of images (*e.g.* for multichannel imaging). See Section A.5.3 for details.

Open output when finished The output images are automatically opened when the process finishes if this box is checked.

ImageTracker adds black pixels to transformed images where there is no original image data. Because ImageTracker also rescales images to maximize contrast when displaying an image sequence initially, it may appear that image intensities have shifted in transformed images if the input series did not include pixels with zero intensity.



(a) (b)

Figure A.7: A pair of images from a multichannel fluorescence labeling of a mitotic spindle. The (a) spindle and (b) kinetochores are labeled with different fluorophores, and separate images are captured serially. Images courtesy of Dr. Lisa Cameron.

A.5.3 Apply Transform

Sometimes more than one image sequence is captured during an experiment. In some situations it may be easier to compute registration transforms from one series of images and apply them to another. For example, Figure A.7 shows separate fluorescent labeling of a mitotic spindle and kinetochores. The spindle image channel contains enough information to obtain a good alignment while the kinetochore channel does not. Select Process→Apply Transform to launch the Apply Transform dialog. This process has the following options.

Transform File Specify the text file that contains the transform information from the previous image registration. This should be output from a Stabilize process.

Output The common output options are available.

Open output when finished The output images are automatically opened when the

process finishes if this box is checked.

A.5.4 CLG Optical Flow

ImageTracker provides several methods to determine the image motion from an image sequence. Optical flow is an estimate of real world motion projected onto the image plane. In the case of 2D motion where the image plane is parallel to the motion plane, optical flow is a good estimate of real world motion. This scenario is common in microscopy. Flow computation results in a vector image in which each pixel represents the computed velocity in the image's x and y directions. Vector images can be loaded and viewed from ImageTracker's Vectors tab, as described in Section A.3.3.

The Combined Local Global (CLG) method of optical flow computation is described by Bruhn *et al.* in [BWS05]. The implementation provided in ImageTracker is the multiresolution, simple temporal difference variant described in that paper. The algorithm iteratively estimates optical flow by minimizing an energy functional that includes a local intensity constancy term (due to Lucas and Kanade, [LK81]) and a global smoothness term (due to Horn and Schunck, [HS81]).

To compute optical flow, select Process→CLG Optical Flow from the menu. When the CLG Optical Flow dialog is active, the image displayed in the image panel will be filtered with a Harris feature detector [SMB00]. This image is useful for determining the Derivative and Integration scale parameters. The goal is to adjust these scale parameters such that the features of interest appear as distinct blurry blobs. The CLG process has the following options.

Derivative Scale The spatial scale used to compute image derivatives; adjust this to expected noise level present in the images.

Integration Scale The spatial scale used when integrating local regions of the image. Adjust this to the size of the predominant features that are being tracked.

Number of Levels The number of resolution levels to use in the computation. More levels are required if there are larger motions between frames.

Iterations Specifies the number of iterations to complete when minimizing the error functional at each resolution level; more iterations will presumably provide a lower error residual but will take longer.

Regularization Adjusts the constant scaling of the smoothness term in the energy functional. A larger number will result in smoother flow fields, a smaller number will allow more turbulent flow fields. I have empirically found that a value close to the mean intensity of an image sequence is often a good place to start for this term. For many 8-bit images, 80-180 is a reasonable range.

Output The common output options are available.

Open output when finished The output images are automatically opened when the process finishes if this box is checked. Vector image visualizations are controlled on the Vectors tab.

Click the Run button to start the optical flow computation. ImageTracker's implementation of the algorithm is notoriously slow to run; the progress bar will provide an estimate of how long the process will take after computing the flow for the first image pair. When analyzing long image sequences, it is advisable to try the flow computation on a shorter image sequence to find an appropriate set of parameters and then analyze the full image sequence with the same settings.

A.5.5 Horn and Schunck Optical Flow

This process is seldom used because the CLG Optical Flow implementation described in Section A.5.4 is superior in several ways. The Horn and Schunck optical flow computation is a classic motion estimation algorithm that uses only the global term from the

CLG computation [HS81, BWS05]. This process remains for comparison purposes only. The following parameters control this process.

Derivative Scale The spatial scale used to compute image derivatives; adjust this to the noise level present in the images.

Iterations Specifies the number of iterations to complete when minimizing the error functional for each image pair; more iterations will presumably provide a lower error residual but will take longer.

Regularization Adjusts the constant scaling of the smoothness term in the energy functional. A larger number will result in smoother flow fields, a smaller number will allow more turbulent flow fields. I have empirically found that a value close to the mean intensity of an image sequence is often a good place to start for this term. For many 8-bit images, 80-180 is a reasonable range.

Output The common output options are available.

Open output when finished The output images are automatically opened when the process finishes if this box is checked. Vector image visualizations are controlled on the Vectors tab.

A.5.6 Integrate Flow

The optical flow methods described in Sections A.5.4 and A.5.5 compute displacement from one frame to the next—this is the estimated image velocity in the time between frames. Integrating a series of velocity fields over time yields the net displacement from the first frame of a sequence. ImageTracker includes a 4th-order Runge-Kutta flow integration [PTVF07, Ch. 17]. Access flow integration through Process→Integrate Flow. The following parameters control this process.

Step size Specifies the step size (in frames) of the integration. The integration process saves results at the nearest whole step size, to match the spacing of input flow data. So, with a step size of 0.25 frames, displacement images will be saved every 4 steps.

Output The common output options are available.

Open output when finished The output images are automatically opened when the process finishes if this box is checked. Vector image visualizations are controlled on the Vectors tab.

A.6 Building from Source

ImageTracker leverages a number of open source software packages. To build ImageTracker from source, one first needs to acquire or build these packages.

A.6.1 Windows

The following instructions create a build of ImageTracker on a Windows XP machine using Microsoft Visual Studio 2003 .NET.

CMake

CMake is a cross-platform, compiler-independent build system. It is available from Kitware at www.cmake.org. Download and install the latest stable release for Windows.

fftw

fftw is a fast Fourier transform library. Download the Windows DLLs from www.fftw.org. Then, follow the instructions included with the fftw package to create "lib import libraries."

ITK

ITK is a powerful image processing and analysis library available from Kitware at www.itk.org. Download the latest source code release. Follow the instructions for using CMake to configure ITK for Visual Studio with the following important exceptions.

- In CMake, click on the "Show Advanced Values" check box to display advanced ITK configuration parameters.
- Turn "BUILD_EXAMPLES" and "BUILD_TESTING" OFF.
- Turn "USE_FFTWF" and "USE_FFTWD" ON.
- Press "Configure". If you receive errors about not being able to find fftw components, fix the flagged CMake parameters by pointing to the appropriate files on your system.
- Press "Configure" and "OK" to build the Visual Studio project file.

In Visual Studio, build Debug (if you plan to do any development) and Release versions of the generated ITK solution.

VTK

VTK is a data visualization library available from Kitware at www.vtk.org. Download, configure, and build the latest source code release using CMake and Visual Studio. Alternatively, obtain the version from the VTK source control repositories. Build both a Debug (if you plan to do any development) and Release version.

wxWidgets

wxWidgets is a platform-independent graphical user interface (GUI) development library. Download and install MSW version 2.6.3 from www.wxwidgets.org. Image-

Tracker uses a feature of wxWidgets that is disabled by default, which can be enabled by following these steps.

- Open wx.dsw and convert to a .NET solution.
- In the Solution Explorer, locate base→Setup Headers→`setup.h`. Confusingly, there are two headers called `setup.h`. The second file in the list is the one to edit (it says "configuration for the library" at the top).
- Find `#define wxUSE_STD_IOSTREAM 0` and change the 0 to 1 to enable standard C++ streams.
- Build a Release version of the .NET project.

ImageTracker

The ImageTracker source is available by request or from the CISMM source control repository (requires a valid login from UNC Department of Computer Science). The `CVSROOT` is `/afs/cs.unc.edu/proj/stm/src/` and the project name is `ImageTracker`.

- Create an ImageTracker solution file using CMake. An out-of-source build is suggested to keep the code and binaries separate.
- The CMakeLists.txt file should configure everything, but it may be necessary to point to the compiled ITK, VTK, wxWidgets, or fftw libraries.
- Generate the Visual Studio solution file, and build the ImageTracker project.
- After ImageTracker builds, copy the fftw DLLs to the ImageTracker executable directory.

A.7 Linux

Instructions for building ImageTracker on Linux is distribution dependent. The following instructions build ImageTracker on a new Ubuntu 7.10 (Gutsy) install, and should be easily adapted for more recent Ubuntu releases or any other Linux distribution.

Prerequisites

Install the following packages, *e.g.* with the command

```
$ sudo aptitude install package
```

Following each package is a list of dependencies that aptitude (or another repository manager) will likely find automatically.

build-essential dpkg-dev g++ g++-4.1 libc6-dev libstdc++6-4.1-dev linux-libc-dev patch

libncurses5-dev

fftw3-dev fftw3

libgl1-mesa-dev libx11-dev libxau-dev libxdmcp-dev mesa-common-dev x11proto-core-dev x11proto-input-dev x11proto-kb-dev xtrans-dev

libglu1-mesa-dev

libxt-dev libice-dev libsm-dev

libwxgtk2.8-dev libwxbase2.8-dev

libgtk2.0-dev libatk1.0-dev libcairo2-dev libexpat1-dev libfontconfig1-dev libfreetype6-dev libglib2.0-dev libpango1.0-dev libpng12-dev

libxcomposite-dev libxcursor-dev libxdamage-dev libxext-dev

libxfixes-dev libxft-dev libxi-dev libxinerama-dev libxrandr-dev

libxrender-dev x11proto-composite-dev x11proto-damage-dev

x11proto-fixes-dev x11proto-randr-dev x11proto-render-dev

x11proto-xext-dev x11proto-xinerama-dev zlib1g-dev

python-wxglade libwxbase2.8-0 libwxgtk2.8-0 python-wxgtk2.8 python-wxversion

CMake

The CMake package for Ubuntu sometimes lags behind the latest release, so it is suggested to build from source. Download the source from www.cmake.org. Enter the following commands to build and install CMake.

```
$ tar -xvf cmake-2.4.7.tar.gz
$ cd cmake-2.4.7/
$ ./bootstrap
$ make
$ sudo make install
$ cmake --version
cmake version 2.4-patch 7
```

ITK

Download the latest source code release from www.itk.org. Do an out of source build.

```
$ tar xvf InsightToolkit-3.4.0.tar.gz
$ ccmake (path to itk source)/InsightToolkit-3.4.0
BUILD_EXAMPLES OFF
BUILD_TESTING OFF
USE_FFTWD ON
USE_FFTWF ON
Generate
$ make
```

VTK

Download the latest source code from www.vtk.org. Do an out of source build.

```
$ tar xvf vtk-5.0.3.tar.gz
$ ccmake (path to vtk source)/VTK
$ make
```

ImageTracker

Checkout ImageTracker from the CISMM source control. Do an out of source build. The following generates KDevelop project files, which is optional.

```
$ ccmake -G KDevelop3 ../../source/ImageTracker
CMAKE_BUILD_TYPE Debug (optional)
ITK_DIR path to itk
VTK_DIR path to vtk
BUILD_IT_TESTS OFF
$ make ImageTracker
```

Bibliography

[ABMD92] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE TIP*, 1:205–220, 1992. 61

[BA96] Michael J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Comput. Vis. Image Underst.*, 63(1):75–104, 1996. 42

[BB95] S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM Computing Surveys*, 27(3):433–466, 1995. 39, 40, 76

[BB08] Brian Button and Richard C. Boucher. Role of mechanical stress in regulating airway surface hydration and mucus clearance rates. *Respiratory Physiology & Neurobiology*, 163(1-3):189 – 201, 2008. Respiratory Biomechanics. 4

[BKN07] Peter Barnum, Takeo Kanade, and Srinivasa Narasimhan. Spatio-temporal frequency analysis for removing rain and snow from videos. In *Proceedings of the First International Workshop on Photometric Analysis For Computer Vision - PACV 2007*, page 8 p., Rio de Janeiro Brésil, 2007. INRIA. ISBN 2-7261-1297 8. 165

[Bov05] Al Bovik, editor. *Handbook of image and video processing*. Academic Press, 2nd edition, 2005. 61

[BW75] Max Born and Emil Wolf. *Principles of Optics*. Pergamon Press, 5th edition, 1975. 17, 45

[BWS05] Andres Bruhn, Joachim Weickert, and Christoph Schnorr. Lucas/kanade meets horn/schunck: combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005. 39, 66, 166, 185, 187

[CAC+02] Yung-Yu Chuang, Aseem Agarwala, Brian Curless, David H. Salesin, and Richard Szeliski. Video matting of complex scenes. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 243–248, New York, NY, USA, 2002. ACM Press. 41

[Car08] Peter M. Carlton. Three-dimensional structured illumination microscopy and its application to chromosome structure. *Chromosome Research*, 16(3):351–365, May 2008. 45

[CCSS01] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. *IEEE CVPR*, 02:264, 2001. 41

[CT65] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965. 80

[CWG01] Michael K. Cheezum, William F. Walker, and William H. Guilford. Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophysical Journal*, 81(4):2378–2388, October 2001. 17

[ET07] Brian Eastwood and Russell M. Taylor. Occlusion removal in video microscopy. *Computer Analysis of Images and Patterns, Springer Berlin*, 4673:125–132, 2007. 2, 47, 166

[FC88] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE PAMI*, 10(4):439–451, 1988. 57

[FHC03] M. H. Freeman, C. C. Hull, and W. N. Charman. *Optics*. Butterworth-Heinemann, New York, 11th edition, 2003. 14

[FKS00] Jan T. Frohn, Helmut F. Knapp, and Andreas Stemmer. True optical resolution beyond the rayleigh limit achieved by standing wave illumination. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13):7232–7236, June 2000. 45

[FP03] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2003. 26, 27

[FvDFH97] James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley Publishing Company, Inc, 1997. 23, 26, 83, 130

[FZB02] H. Foroosh, J.B. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *Image Processing, IEEE Transactions on*, 11(3):188–200, March 2002. 39, 80

[Gas78] Jack D. Gaskill. *Linear Systems, Fourier Transforms, and Optics*. Wiley series in pure and applied optics. John Wiley and Sons, 1978. 18, 19

[GAS00] Mats G. L. Gustafsson, David A. Agard, and John W. Sedat. Doubling the lateral resolution of wide-field fluorescence microscopy using structured illumination. volume 3919, pages 141–150. SPIE, 2000. 45

[GCB+04] Wei Geng, P. Cosman, C.C. Berry, Zhaoyang Feng, and W.R. Schafer. Automatic tracking, feature extraction and classification of c. elegans phenotypes. *Biomedical Engineering, IEEE Transactions on*, 51(10):1811–1820, Oct. 2004. 6

[GHCH05] R. Garnett, T. Huegerich, C. Chui, and Wenjie He. A universal noise removal algorithm with an impulse detector. *Image Processing, IEEE Transactions on*, 14(11):1747–1754, Nov. 2005. 53

[GLT96] Hansjoerg Gaertner, Peter Lehle, and Hans J. Tiziani. New highly efficient binary codes for structured light methods. volume 2599, pages 4–13. SPIE, 1996. 44

[GN04] Kshitiz Garg and Shree K. Nayar. Detection and removal of rain from videos. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:528–535, 2004. 165

[GN07] Kshitiz Garg and Shree K. Nayar. Vision and rain. *International Journal of Computer Vision*, 75(1):3–27, October 2007. 165

[Goo68] Joseph W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, Inc., 1st edition, 1968. 17, 45

[GSZ08] Kun Gai, Zhenwei Shi, and Changshui Zhang. Blindly separating mixtures of multiple layers with spatial shifts. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008. 43

[Har78] Frederic J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978. 57, 84, 86, 95

[HB06] Rainer Heintzmann and Pier A. Benedetti. High-resolution image reconstruction in fluorescence microscopy with patterned excitation. *Applied Optics*, 45(20):5037–5045, 2006. 45

[HG08] Huy Tho Ho and Roland Goecke. Optical flow estimation using fourier mellin transform. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008. 40

[HK94] G. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE PAMI*, 16(3):267–276, 1994. 26, 27, 30, 49, 58, 99, 113

[Hog03] W.S. Hoge. A subspace identification extension to the phase correlation method [mri application]. *Medical Imaging, IEEE Transactions on*, 22(2):277–280, Feb 2003. 39

[HPBH04] D. B. Hill, M. J. Plaza, K. Bonin, and G. Holzwarth. Fast vesicle transport in PC12 neurites: velocities and forces. *European Biophysics Journal*, 33(7):623–632, 2004. 4, 49, 66

[HPS08] W. Y. Hsu, W. F. Paul Poon, and Y. N. Sun. Automatic seamless mosaicing of micrsocope images: enhancing appearance with color degradation compensation and wavelet-based encoding. *Journal of Microscopy*, 231(3):408–418, 2008. 124

[HS81] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. 39, 185, 187

[HS06] Katsunori Hoshi and Ryuzo Shingai. Computer-driven automatic identification of locomotion states in caenorhabditis elegans. *Journal of Neuroscience Methods*, 157(2):355–363, 2006. 6

[HW79] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979. 113

[IRP94] Michal Irani, Benny Rousso, and Shmuel Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, February 1994. 42

[IS97] Shinya Inoué and Kenneth R. Spring. *Video Microscopy : The Fundamentals*. Springer, 2nd edition, 1997. 49, 50

[ISNC05] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, http://www.itk.org/ItkSoftwareGuide.pdf:PDF, 2nd edition, 2005. 40, 56, 182

[JWTT04] Jiaya Jia, Tai-Pang Wu, Yu-Wing Tai, and Chi-Keung Tang. Video repairing: Inference of foreground and background under severe occlusion. *IEEE CVPR*, 01:364–371, 2004. 50

[KBPH07] Mélanie V. Kessels, Marwa El Bouz, Robin Pagan, and Kevin Heggarty. Versatile stepper based maskless microlithography using a liquid crystal display for direct write of binary and multilevel microstructures. *Journal of Micro/Nanolithography, MEMS and MOEMS*, 6(3):033002, 2007. 158

[KS07] William M. Kier and Michael P. Stella. The arrangement and function of octopus arm musculature and connective tissue. *Journal of Morphology*, 268:831 – 843, 2007. 71

[KWZK95] Sing Bing Kang, J.A. Webb, C.L. Zitnick, and T. Kanade. A multibaseline stereo system with active illumination and real-time image acquisition. pages 88–93, Jun 1995. 44

[Lew95] J. P. Lewis. Fast normalized cross-correlation. *Vision Interface*, 1995. 77

[LK81] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision (darpa). In *Proceedings of the 1981 DARPA Image Understanding Workshop*, pages 121–130, April 1981. 38, 182, 185

[LNA+06] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. Light field microscopy. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 924–934, New York, NY, USA, 2006. ACM Press. 45

[MB08] LA Mangiamele and SS Burmeister. Acoustically evoked immediate early gene expression in the pallium of female tngara frogs. *Brain, Behavior and Evolution*, 72:239–250, 2008. 151

[McA08] Justin S. McAlister. Evolutionary responses to environmental heterogeneity in central american echinoid larvae: Plastic versus constant phenotypes. *Evolution*, 62(6):1358–1372, 2008. 5

[MMG05] A Mansouri, FS Marzani, and P Gouton. Development of a protocol for ccd calibration: Application to a multispectral imaging system. *INTERNATIONAL JOURNAL OF ROBOTICS & AUTOMATION*, 20(2, Sp. Iss. SI):94–100, 2005. 30

[Mur01] Douglas B. Murphy. *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley-Liss, Inc., 2001. 8, 10, 15

[NM65] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965. 78

[NN94] M. Noguchi and S.K. Nayar. Microscopic shape from focus using active illumination. In *International Conference on Pattern Recognition*, volume 1, pages A:147–152, October 1994. 44, 70

[NWN95] Shree K. Nayar, Masahiro Watanabe, and Minori Noguchi. Real-time focus range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1186–1198, 1995. 44

[Paw06] James B. Pawley. *Handbook of Biological Confocal Microscopy*. Springer, 3rd edition, 2006. ISBN: 978-0-387-25921-5. 20

[PD84] Thomas Porter and Tom Duff. Compositing digital images. In *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 253–259, New York, NY, USA, 1984. ACM Press. 40

[Pel09] Ted Pella. Gilder TEM grids. Commercial internet web site, available at http://www.tedpella.com/grids_html/gilder.htm., March 2009. 111

[Pit93] Jim Pitman. *Probability*, chapter 3, pages 140–258. Springer, 1993. 99

[PTVF07] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 3rd edition, 2007. 55, 79, 93, 143, 147, 187

[QAA94] N. Qian, R. A. Andersen, and E. H. Adelson. Transparent motion perception as detection of unbalanced motion signals. i. psychophysics. *Journal of Neuroscience*, 14:7357–7366, 1994. 37

[RBY+99] Ramesh Raskar, Michael S. Brown, Ruigang Yang, Wei-Chao Chen, Greg Welch, Herman Towles, Brent Seales, and Henry Fuchs. Multi-projector displays using camera-based registration. In *IEEE Visualization: Proceedings of the conference on Visualization '99*, pages 161–168, Los Alamitos, CA, USA, 1999. IEEE Computer Society Press. 70

[RC96] B.S. Reddy and B.N. Chatterji. An fft-based technique for translation, rotation, and scale-invariant image registration. *Image Processing, IEEE Transactions on*, 5(8):1266–1271, Aug 1996. 39

[RHM+03] J. Ryu, B. K. P. Horn, M. S. Mermelstein, S. Hong, and D. M. Freeman. Application of structured illumination in nano-scale vision. volume 2, pages 17–17, June 2003. 45

[RT94] Theodore George Rochow and Paul Arthur Tucker. *Introduction to Microscopy by Means of Light, Electrons, X Rays, or Acoustics*. Plenum Press, 2nd edition, 1994. 155

[RT00] M.A. Ruzon and C. Tomasi. Alpha estimation in natural images. *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 1:18–25, June 2000. 41

[SAA00] R. Szeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 1:246–253 vol.1, 2000. 43

[SB96] Alvy Ray Smith and James F. Blinn. Blue screen matting. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 259–268, New York, NY, USA, 1996. ACM. 41

[SI05] Bernard Sarel and Michal Irani. Separating transparent layers of repetitive dynamic behavior. *IEEE International Conference on Computer Vision*, 1:26–32, 2005. 43

[SLKS05] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. Symmetric stereo matching for occlusion handling. *IEEE Computer Vision and Pattern Recognition*, 2:399–406, 2005. 50

[SM90] Masahiko Shizawa and Kenji Mase. Simultaneous multiple optical flow estimation. *ICPR*, 1:274–278, 1990. 42, 50, 54

[SMB00] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000. 38, 185

[SN03] Yoav Y. Schechner and Shree K. Nayar. Generalized mosaicing: High dynamic range in a wide field of view. *International Journal of Computer Vision*, 53(3):245–267, 2003. 49

[SP96] A. Said and W.A. Pearlman. An image multiresolution representation for lossless and lossy compression. *IEEE Transactions on Image Processing*, 5:1303–1310, 1996. 61

[SSW+03] ED Salmon, SL Shaw, JC Waters, CM Waterman-Storer, PS Maddox, E Yeh, and K. Bloom. A high-resolution multimode digital microscope system. *Methods in Cell Biology*, 56:185–215, 2003. 6, 125

[ST94] Jianbo Shi and Carlo Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, June 1994. 38

[TAS+03] Kenji Takita, Takafumi Aoki, Yoshifumi Sasaki, Tatsuo Higuchi, and Koji Kobayashi. High-accuracy subpixel image registration based on phase-only correlation. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E86-A(8):1925–1934, 2003. 40, 80, 131

[TRK01] Yanghai Tsin, Visvanathan Ramesh, and Takeo Kanade. Statistical calibration of CCD imaging process. *IEEE ICCV*, 01:480, 2001. 27, 30, 49

[TRL+08] C. Tauber, P.F.G. Rodriguez, V. Loriette, N. Chenouard, B. Dubertret, and J.-C. Olivo-Marin. Fast and automatic reconstruction of structured illumination microscopy images with multiscale products. pages 1–4, Dec. 2008. 45

[TT07] George Tsibidis and Nektarios Tavernarakis. Nemo: a computational tool for analyzing nematode locomotion. *BMC Neuroscience*, 8(1):86, 2007. 6

[TV98] Emanuele Trucco and Alessandro Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1st edition, 1998. 30, 38, 92, 99, 131

[VO90] P. Vuylsteke and A. Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(2):148–164, Feb 1990. 44, 70

[WN97] Masahiro Watanabe and Shree K. Nayar. Telecentric optics for focus analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1360–1365, 1997. 16

[WN98] Masahiro Watanabe and Shree K. Nayar. Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27(3):203–225, May 1998. 16

[YWT08] Sai-Kit Yeung, Tai-Pang Wu, and Chi-Keung Tang. Extracting smooth and transparent layers from a single image. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7, June 2008. 41