Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________________ ____________

Blair J. Rossetti Date

Bioimage Informatics in the Big Data Era: Algorithms for High-Dimensional Spectral, Volumetric, and Temporal Image Processing

By

Blair J. Rossetti
Doctor of Philosophy

Computer Science and Informatics

James Nagy, Ph.D.
Advisor

Lee Cooper, Ph.D.
Committee Member

Ashish Sharma, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Bioimage Informatics in the Big Data Era: Algorithms for High-Dimensional Spectral, Volumetric, and Temporal Image Processing

By

Blair J. Rossetti
B.A., Wheaton College, MA, 2009
M.S., Emory University, GA, 2018

Advisor: James Nagy, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Computer Science and Informatics
2019

Abstract

Bioimage Informatics in the Big Data Era: Algorithms for High-Dimensional Spectral, Volumetric, and Temporal Image Processing

By Blair J. Rossetti

Big data is revealing new challenges in the area of bioimage informatics—the study of systems and methods for handling, processing, and visualizing biological images. Biological events are inherently multidimensional, and they occur at different timescales, frequencies, and resolutions. Yet, for many years, the limitations of digital imaging technologies have prevented researchers from gaining a holistic view of their problem. The goal of modern bioimage informatics is to leverage knowledge in big data, computer vision, and machine learning to efficiently and effectively process high-dimensional data sets. Although many open-source algorithms are available for processing bioimages, advanced numerical methods are still needed for handling big spectral, volumetric, and temporal data. In this work, we discuss novel contributions to the areas of spectral unmixing, volumetric reconstruction, and video-based tracking. Specifically, we present (1) a robust algorithm for unmixing large numbers of fluorescent labels from contaminated spectral micrographs; (2) an offline/online multiresolution workflow for the reconstruction of three-dimensional subvolumes of interest from serial gigapixel whole slide images; and (3) an automated analysis pipeline and graphical interface to aid *in situ* tracking of insect movement that uses low-cost, human-readable tags.

Bioimage Informatics in the Big Data Era: Algorithms for High-Dimensional Spectral, Volumetric, and Temporal Image Processing

By

Blair J. Rossetti
B.A., Wheaton College, MA, 2009
M.S., Emory University, GA, 2018

Advisor: James Nagy, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Computer Science and Informatics
2019

Acknowledgments

This work represents the realization of a goal to which I have aspired for most of my life, and it belongs as much to me as to those around me. I have been fortunate in my life and career to have been mentored, guided, and supported by many truly remarkable people. In these brief words, I express only a fraction of the appreciation I hold for my thesis committee, mentors, colleagues, family, and friends.

Advisers

I am profoundly grateful to have had the opportunity to work with two dedicated and inspiring advisers, Jim Nagy and Jun Kong. Over the past four years, Jim and Jun have encouraged me to set high goals, push the boundaries of my knowledge, and live outside of my comfort zone. They provided me with the freedom to explore new areas and the support to overcome roadblocks along the way.

Committee Members

I would like to extend my sincere thanks to Ashish Sharma and Lee Cooper for helping me navigate my graduate career. On countless occasions, their advice has saved me from pitfalls, improved the quality of my research, and ensured that I stay on target.

Mentors, Colleagues, and Friends

This dissertation is a milestone on a path that began for me almost fifteen years ago, and I am deeply indebted to the mentors who have helped me along the way. I would like to give a special thanks to Jeff Bowen, Bob Morris, Louie Kerr, and Gary Borisy for giving me the resources, knowledge, and opportunities to pursue my research passions. I must also convey my great appreciation for the thoughtful advice and insight given to me by Lars Ruthotto and Yuanzhe Xi.

I would be remiss if I did not acknowledge the people who afforded me compassion and companionship on this journey. I sincerely thank my colleagues in the Scientific

Computing Group, BMI cohort, and fellow computer science students.

Family

This dissertation is dedicated to my wife, Rachel; my parents, Chuck and Diana; and my siblings, Bianca and Ross. Without your unwavering love and support, this work would not exist. You have helped me realize when I am strong enough to push on, and you have shown me the wisdom in knowing when to ask for help.

I am fortunate, and I am forever grateful.

Funding

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1444932. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

> We will begin by asking if the prevailing notion of what computer systems can achieve in biological image-processing may not be overly modest. This question seems to us fundamental, for such an underestimate might produce not only a limitation of objectives, but also a restriction of the class of problems considered amenable to computer processing.
>
> "The Analysis, Synthesis, and Description of Biological Images"
>
> — Lipkin, Watt, & Kirsch, (1966)

In a time where seemingly all information is recorded, stored, and collated, it is perhaps a bit ironic that we cannot pinpoint the origin of the phrase *big data*. To the best of our knowledge, however, the first mention of the phrase in computer science was the 1997 seminar entitled "Big Data and the Next Wave of InfraS-tress"[1] by Silicon Graphics scientist John Mashey [50, 139, 148]. In his presentation, Mashey described big data as being a looming source of what he called *infrastress*[2], or "the bad effects of faster change in computer susbsystems and usage…than in underlying infrastructure" [149]. Infrastress is what arises when hardware speed and capacity

---

[1] A later version of Mashey's slide deck can be found at `http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf` [149].

[2] Infrastress is a portmanteau of the words *infrastructure* and *stress*.

outpace the software on which they rely. For every bump in hardware performance, computer scientists and software engineers have been left scrambling to keep up. This imbalance leads to bottlenecks, artificial limits, instability, and complicated workarounds. Mashey believed that infrastress would only last a couple of years until the existing imbalance was corrected [149]. However, the sense of infrastress has only increased over the past twenty years and the phrase *big data* is now in common parlance [29, 30, 31, 103]. This phenomena raises a few pressing questions. First, *what does big data mean today and does it differ from what Mashey described in the late-nineties*? Second, *how is big data impacting modern science*? Third, *what algorithms and methods can be used to rectify the current hardware/software imbalance*? The goal of this work is not to answer these questions outright, as such answers may not even exist, but to explore how these questions relate to a subfield of computer science called bioimage informatics[3]. Let's begin by addressing our first question and settling on a definition for big data.

1.1 Big Data and the Big Data Era

One of the first attempts to characterize big data came from META Group[4] analyst Doug Laney. In his 2001 report discussing trends in e-commerce, Laney outlined what is now referred to as the three Vs of big data [118]. He suggested that any solution for managing the large influx of e-commerce data must take into account its *volume* (i.e., storage requirements), *velocity* (i.e., rate of generation/analysis), and *variety* (i.e., mixture of data types). The three Vs have since been generalized to describe the three ways data can be *big*[5]. For example, we find large *volume* three-dimensional

---

[3] Bioimage informatics is not to be confused with imaging informatics which deals with the processing of radiography data.

[4] META Group was acquired by Gartner in 2005.

[5] The three Vs model has expanded to include value [30], veracity [48], variability [65], and visualization [202]; however, these additional attributes describe desired traits rather than fundamental properties of data.

images in microscopy [133], high *velocity* streaming videos in ethology [76], and a wide *variety* of multimodal satellite data in remote sensing [31]. Each of these data sets represents a different big data problem.

The three Vs are useful in classifying the types of big data, but *big* is still a relative term. For Mashey, *big* meant any data set that could saturate a 4 GB memory cards, 40 GB storage drive, or 1 GHz processor [149]. Ten years earlier, Stytz, Frieder & Frieder had an even smaller interpretation of *big*:

> The large amount of data processed, up to 35 MB (megabytes) produced per patient by a single modality study, hinders rapid image formation. The challenge posed to the computer scientist lies in the development of techniques for rapid, accurate manipulation of large quantities of data to produce images that are useful to a physician. [214]

These definitions of *big* are almost laughable by modern standards. For comparison, 35 MB is equivalent to five seconds of 4K video at 30 fps, and the same quantity of information can be transferred in 280 ms over a gigabit network connection. This change in what we consider *big data* is a reflection of the 58 % year-over-year growth in general computing capacity [91].

In an effort to localize the phrase to the last twenty years, *big data* is now commonly defined as any collection of information that cannot be analyzed using traditional methods in a reasonable amount of time. However, this definition is also a moving target since *traditional methods* are constantly evolving. Freeman & Glass articulated this same imbalance between data size and existing methods in their 1969 publication on image digitization:

> Currently, there is much interest in the use of digital computers for the processing of pictorial data. Usually the applications are problems that involve such large amounts of pictorial data that manual processing is either not feasible or not economical. [62]

From this passage, we must conclude that big data, as it has been defined, is not a new phenomenon[6]. Why, then, did Chen, Mao & Liu claim in 2014 that the "era of big data has come beyond all doubt" [30]?

There is something undoubtedly new about the challenges of modern data processing, yet our attempts to define *big data* as an entity have proven futile. Scientific research will continue to push the boundaries of existing hardware, and our data will always be *big*. It appears best, then, not to think of big data as a collection of information but, instead, as the field of research that aims to find scalable solution to the storage, manipulation, and analysis data sets. Through this lens, the *Big Data Era* marks the point where the field of big data has begun to mature across a variety of domains.

1.2 Bioimage Informatics

As with most fields, big data is revealing new challenges in the area of *bioimage informatics*—the study of systems and methods for handling, processing, and visualizing biological images. Biological event are inherently multidimensional, and they occur at different timescales, frequencies, and resolutions. Yet, for many years, the limitations of digital imaging technologies have prevented researchers from gaining a holistic view of their problem. Advances in optical sensing and the rise of computational imaging are providing new avenues for measuring complex biological processes [16, 26, 231]. The goal of modern bioimage informatics is to leverage knowledge in big data, computer vision, and machine learning to efficiently and effectively process high-dimensional data sets. Although many open-source algorithms are available for processing bioimages [46, 153, 192], advanced numerical methods are needed for handling big *spectral*, *volumetric*, and *temporal* data [9]. We will briefly explore applications within these three areas.

---

[6] A similar observation led Lugmayr et al. [143] to develop the Cognitive Big Data framework.

Bioimages are considered to have five possible dimensions: space (x, y, z), color (λ), and time (t). Although most photographs have three color channels for red, green, and blue light, bioimages can contain ten or hundreds of colors. Spectral imagers are optical sensors that measure the intensity of light across the visible and near-visible wavelengths (typically 400–800 nm). When coupled with a microscope, spectral imaging systems allow scientists to record the emission profiles of fluorophore-labeled specimen across an entire scene. Overlapping emission profiles are difficult to visually inspect, and they require the use of spectral unmixing algorithms to interpret. *Spectral unmixing* is the process of separating a mixture of light into its component profiles or signatures. Unmixing algorithms have shown great promise for applications where samples contain many fluorescent labels; however, existing methods perform poorly on images contaminated by autofluorescent molecules and stray sources of light. Therefore, robust spectral unmixing methods are required to analyze high-dimensional spectral micrographs under noisy, real-world conditions.

Two-dimensional (x,y) images represent either a projection or a cross-section of a three-dimensional scene. Loss of depth (z) information makes it difficult to fully understand the nature of three-dimensional biological processes. Gigapixel imaging of serial histology samples presents a new modality for the study of three-dimensional disease morphologies. However, serial gigapixel imaging has also created new problems for the *registration* and *reconstruction* of large volumes. A single color gigapixel (i.e., 10^9 pixels) image can occupy nearly a gigabyte of memory. With tens or hundreds of serial gigapixel images, the memory requirements quickly become intractable. Gigapixel image registration is an out-of-core problem that precludes the direct application of many conventional registration techniques. Clinicians require fast access to subvolumes of interest from within gigapixel histology data sets in order to evaluate and study disease hallmarks.

Biological processes occur over a variety of timescales. A neuronal action potential is measured in milliseconds; whereas, the pupation of a butterfly may take several weeks. Temporal (t) imaging is essential for recording the dynamics of biological events. For instance, measurements of animal movement are critical for understanding behavioral and ecological properties such as the spread of disease within communities. Video-based animal tracking is a low-cost mechanism for elucidating the roles that specific individuals play in disease emergence (e.g., by identifying a subset of superspreaders). Monitoring animals often necessitates long-duration video recordings across multiple sites. Since these experiments produce thousands of hours of video data, automated methods are required to condense the tracking data to a manageable size for review and analysis.

1.3 Contributions of Work

In this work, we discuss novel contributions to the areas of *spectral unmixing*, *volumetric reconstruction*, and *video-based tracking*. These contributions focus on different applications of big spectral, volumetric, and temporal data:

Spectral Data: We present a robust algorithm for unmixing large numbers of fluorescent labels from contaminated spectral micrographs. As part of this algorithm, we also describe a new method for measuring fluorophore reference spectra [186].

Volumetric Data: We present an offline/online multiresolution workflow for the reconstruction of three-dimensional subvolumes of interest from serial gigapixel whole slide images [185].

Temporal Data: We present an automated analysis pipeline and graphical interface to aid *in situ* tracking of insect movement that uses human-readable tags and low-cost video stations [184].

1.4 Outline

This dissertation is divided into three primary chapters corresponding to spectral data (Chapter 2), volumetric data (Chapter 3), and temporal data (Chapter 4). Each chapter begins with a brief historical overview that aims to provide context to the scale and structure of each data type. We then discuss what sensors and optical systems are required to capture each type of bioimage. This background sets the stage for a discussion of applications in each data area and a review of relevant algorithms. Finally, each chapter presents a novel contribution to the area of bioimage informatics. The details of each method are described and followed by an evaluation of their performance and discussion of any limitations.

Chapter 2

Spectral Data

In his 1672 letter to the Royal Society, Isaac Newton presented his “New Theory about Light and Colours” [164]. At the time, it was known that one could produce a rainbow of colors by passing white light through a triangular prism. However, the physical principles behind this “Phenomena of Colours” was highly contested. Newton dismissed the theory put forth in René Descartes’ 1644 *Principia philosophiae* suggesting that prisms *create* the observed colors by modifying white light [49]. Instead, Newton took the opposite position by claiming that “light itself is a heterogeneous mixture of differently refrangible rays” [164]. In other words, colors *combine* to form white light. To prove this, Newton described his *Experimentum Crucis* in which he used a prism to separate white light into its component colors and a lens to recombine the colors back into white light (see Figure 2.1). The critical insight gained from this experiment was the notion that color is a property of light itself and not a property of the prism or the lens as Descartes’ theory suggested.

Today we know that light consists of waves of particular frequencies, and these waves travel at different speeds in different materials. When light passes from one material to another, its speed changes but its frequency stays the same. As a result, the light bends, or refracts, by some angle. The relationship between the incident

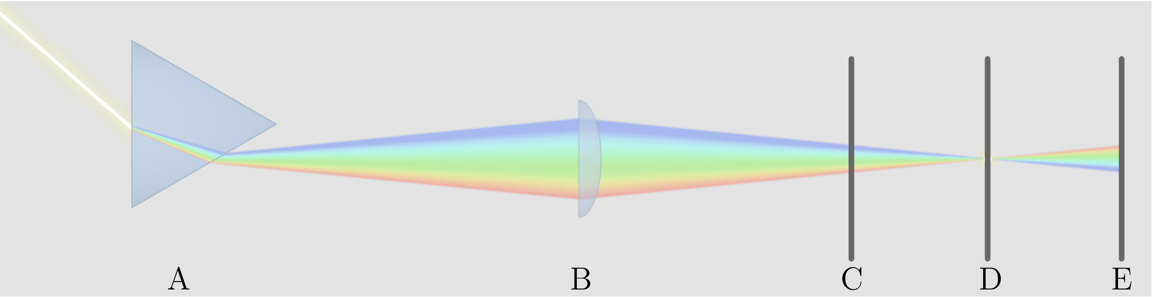


Figure 2.1: Illustration of Newton's *Experimentum Crucis* from his 1672 letter "New Theory about Light and Colours." Incident white light is dispersed by a prism (A). The diverging component colors are collected and focused by a lens (B). Light cast on a piece of white paper placed at plane C would show a rainbow with colors from blue (top) to red (bottom). Moving the paper to plane D would reveal that the colors converge to form white light. Finally, moving the paper to plane E would show an inverted rainbow with colors from red (top) to blue (bottom).

angle, denoted θ_1, and the refraction angle, denoted θ_2, is given by Snell's law:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}, \tag{2.1}$$

where n_1 and n_2 are the refractive indices of the two materials. The *refractive index* for a given material can be expressed by

$$n_m = \frac{\lambda_0}{\lambda_m}, \tag{2.2}$$

where λ_0 is the wavelength of light in a vacuum and λ_m is the wavelength of light in material m. As we can see, the refraction of light is dependent on the wavelength of the incident light. This dependence between wavelength and refraction is the reason why short-wavelength blue light has a larger refraction angle than long-wavelength red light. Newton's experiments on the dispersion of light led to a critical shift in thinking about light as a combination of spectral colors.

In this chapter, we examine the spectral characteristics of light, the information they provide, and how we can design algorithms to leverage this information. We begin by examining the human perception of color and how this process is mimicked in digital cameras. This discussion is followed by an exploration of how the spectral components of light are measured and the applications of these methods in fluorescence microscopy. The remainder of the chapter is spent on algorithms for processing high-dimensional spectral images. We outline the existing spectral unmixing methods and reveal some of the challenges in analyzing real-world data. We identify broken assumptions and derive new algorithms for the processing of contaminated spectral micrographs. Finally, we compare these new methods to the current standards in spectral image analysis.

2.1 Spectral Imaging

Newton's 1672 *Experimentum Crucis* provided proof of the heterogeneity of light; however, his thinking on how light and colors are detected by the human eye were fundamentally flawed. In his 1665 notes entitled "Of Colours", Newton wrote[1]:

> Light seldom striks upon the parts of grosse bodys (as may bee seen in its passing through them), its reflection & refraction is made by the diversity of æthers, & therefore its effect on the Retina can only bee to make this vibrate which motion then must bee either carried in the optick nerve to the sensorium or produce other motions that are carried thither. [163]

In this quote, Newton suggests that light cannot pass through the curved "optick nerve" because light particles travel in a straight lines. Therefore, he believed that impacts from light particles caused vibrational waves that traveled into the brain. While we still do not have a complete understanding of how the brain perceives color, we now know the fundamental biological components that allow humans to detect light of different wavelengths.

---

[1] The Newton Project has published an online edition of Newton's writings from 1642–1727 which can be accessed at `http://www.newtonproject.ox.ac.uk/`.

The human eye contains three types of photoreceptors—rods, cones, and intrinsically photosensitive retinal ganglion cells. Our perception of color, however, is largely dictated by the three[2] subtypes of cone cells found in the retina. Short, medium, and long cone cells are so named for the portions of the electromagnetic spectrum that they can detect. Short cones, or *S-cones*, are most sensitive to wavelengths around 442 nm; whereas, *M-cones* and *L-cones* are sensitive to wavelengths around 543 nm and 570 nm, respectively [211]. By detecting different *combinations* of short-wavelength blue light, medium-wavelength green light, and long-wavelength red light, our cone cells allow us to distinguish millions of different spectral profiles.

Digital cameras have been designed to function in much the same way as our cone cells. Camera sensors are made up of blue, green, and red photodetectors that match the spectral sensitivity of our S-, M-, and L-cones, respectively (see Figure 2.2). *Spectral sensitivity* is a wavelength-dependent measure of light detection efficiency. Let's denote the spectral sensitivity of the blue, green, and red photodetectors as $S_B(\lambda)$, $S_G(\lambda)$, and $S_R(\lambda)$, where λ represents the wavelength of light. For a given image pixel, a camera measures the intensity of blue (B), green (G), and red (R) light according to the detector *response functions*

$$\begin{aligned} B &= \int L(\lambda) S_B(\lambda) d\lambda, \\ G &= \int L(\lambda) S_G(\lambda) d\lambda, \\ R &= \int L(\lambda) S_R(\lambda) d\lambda, \end{aligned} \tag{2.3}$$

where $L(\lambda)$ is the incident light spectrum. When displayed as an additive overlay (e.g., as with an electronic display), these components form an RGB triplet that replicates the original color perceived by our eye–brain combination.

For each location in an observed scene, digital cameras integrate a continuous

---

[2] A 2010 study by Jordan et al. [100] suggests that roughly 12% of women are tetrachromatic (i.e., they have four subtypes of cone cells).

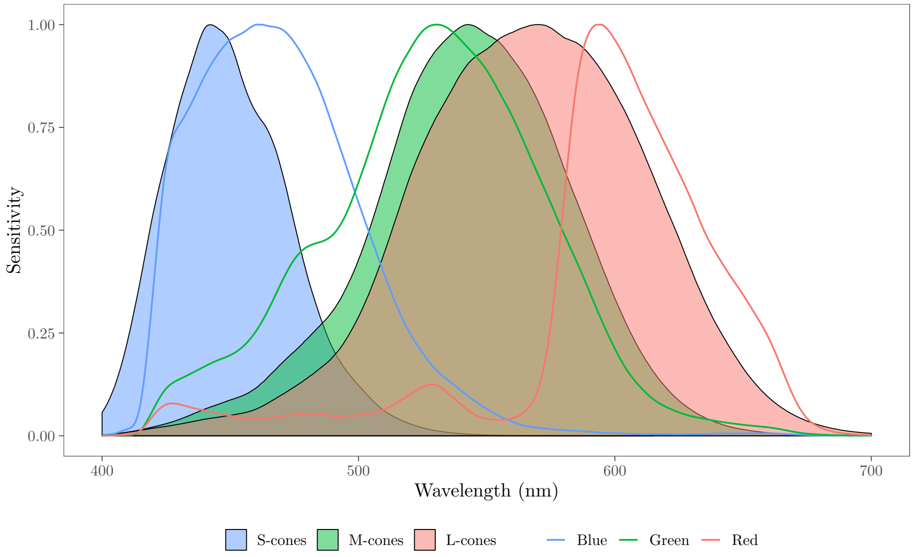


Figure 2.2: Spectral sensitivity of S-, M-, and L-cone cells (shaded regions) compared to the blue, green, and red photodetectors of a Nikon D5100 (colored lines) [44, 211].

spectrum of light according to the RGB response functions. Together, these response functions provide a coarse discretization of $L(\lambda)$ (see Figure 2.3 E). This rough spectral approximation limits our ability to differentiate sources of light with similar spectra. In other words, there exist multiple spectral profiles that produce the same RGB color. Fortunately, there is no inherent reason why camera sensors need only record three bands of spectral wavelengths. Instead of measuring B, G, and R by Eq. 2.3, we can integrate the incident light over an array of M detectors with non-overlapping spectral sensitivities (see Figure 2.3). The M spectral response functions measure the intensity, denoted I_m, at each contiguous spectral band and can be written as

$$I_m = \int L(\lambda) S_m(\lambda) d\lambda, \text{ for } m = 1, \ldots, M, \tag{2.4}$$

where $S_m(\lambda)$ is the spectral sensitivity of the m^{th} detector. For fixed bounds, the range of wavelengths detected within each band decreases as M increases. Therefore, more spectral bands results in a better approximation of the true light spectrum $L(\lambda)$. This idea is the crux of a technique called *spectral imaging*.

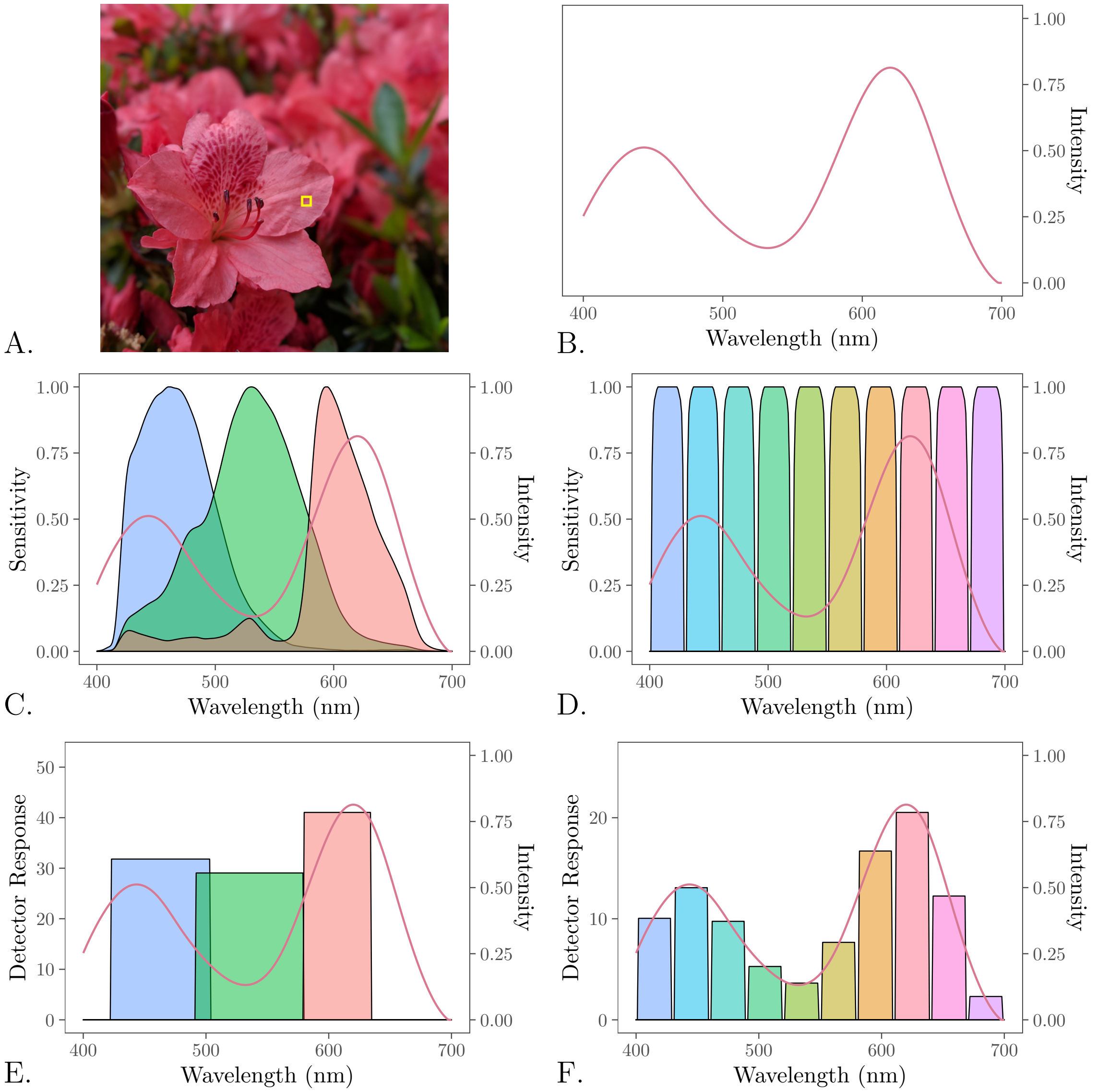


Figure 2.3: Illustration comparing the RGB response functions to ten spectral response functions. A scene of flowers (A) contains reflected light with different spectra profiles. The yellow bounded region produces the continuous spectrum, denoted $L_{flower}(\lambda)$, shown in B. $L_{flower}(\lambda)$ is overlaid on the subsequent plots for reference. The plots in panels C and D show the spectral sensitivities of the blue, green and red detectors (C) and the spectral array detectors (D). The results of computing the RGB response functions in Eq. 2.3 are shown as overlapping bars in panel E. The results of computing the spectral response functions in Eq. 2.4 are shown as non-overlapping bars in (F). The widths of the bars in panels E and F are equal to the full width at half maximum of the corresponding spectral sensitivities. The RGB response functions produce a much coarser approximation of $L_{flower}(\lambda)$ as compared to the spectral response functions.

Spectral imaging is a technique for measuring the spectrum of light at each pixel location in a scene. The resulting spectral image is represented as a third-order tensor, or three-dimensional matrix, containing a single intensity value for every (x, y, λ) pixel (see Figure 2.4 A). Since spectral imaging systems are restricted to using one- or two-dimensional photodetectors[3], special methods are required for capturing spatial-spectral data. There are four main categories of spectral imaging systems: wavelength scanning, spatial scanning, spatiospectral scanning, and snapshot (see Figure 2.4 B-E). *Wavelength scanning* systems use a two-dimensional image sensor to measure the light intensities of an entire scene one spectral band at a time. Similarly, *spatial scanning* systems use either one- or two-dimensional sensors to measure all spectral bands for one pixel or row of pixels at a time. *Spatiospectral scanning* mixes these two methods by acquiring a different spectral band for every row of pixels [81]. Scanning methods are able to achieve high spatial and spectral resolution, but do so at the expense of time. For time-sensitive applications, *snapshot* methods sacrifice spatial and spectral resolution in order to simultaneously measure the spectral data for the entire scene in one image. This is often accomplished by projecting each spectral band onto different regions of a two-dimensional sensor. For more details on spectral image acquisition, we refer the reader to reviews by Gao & Smith [66], Garini et al. [68], Li et al. [131], Lu & Fei [142], Shaw & Burke [200].

The *spectral resolution*, or width of each spectral band, dictates how well a spectral image can approximate $L(\lambda)$; whereas, the *detection range* determines which portion of the electromagnetic spectrum is observed. For a fixed detection range, higher spectral resolution means that more spectral bands are measured. Systems that collect an order of 10^0 to 10^1 spectral bands are generally referred to as *multispectral* imagers. *Hyperspectral* imagers, on the other hand, capture an order of 10^1 to 10^2 spectral bands. As a result, spectral data can range anywhere from a couple of megabytes to

---

[3] Three-dimensional photodetectors, such as multi-well arrays [54], have not yet been introduced into commercial spectral imaging systems.

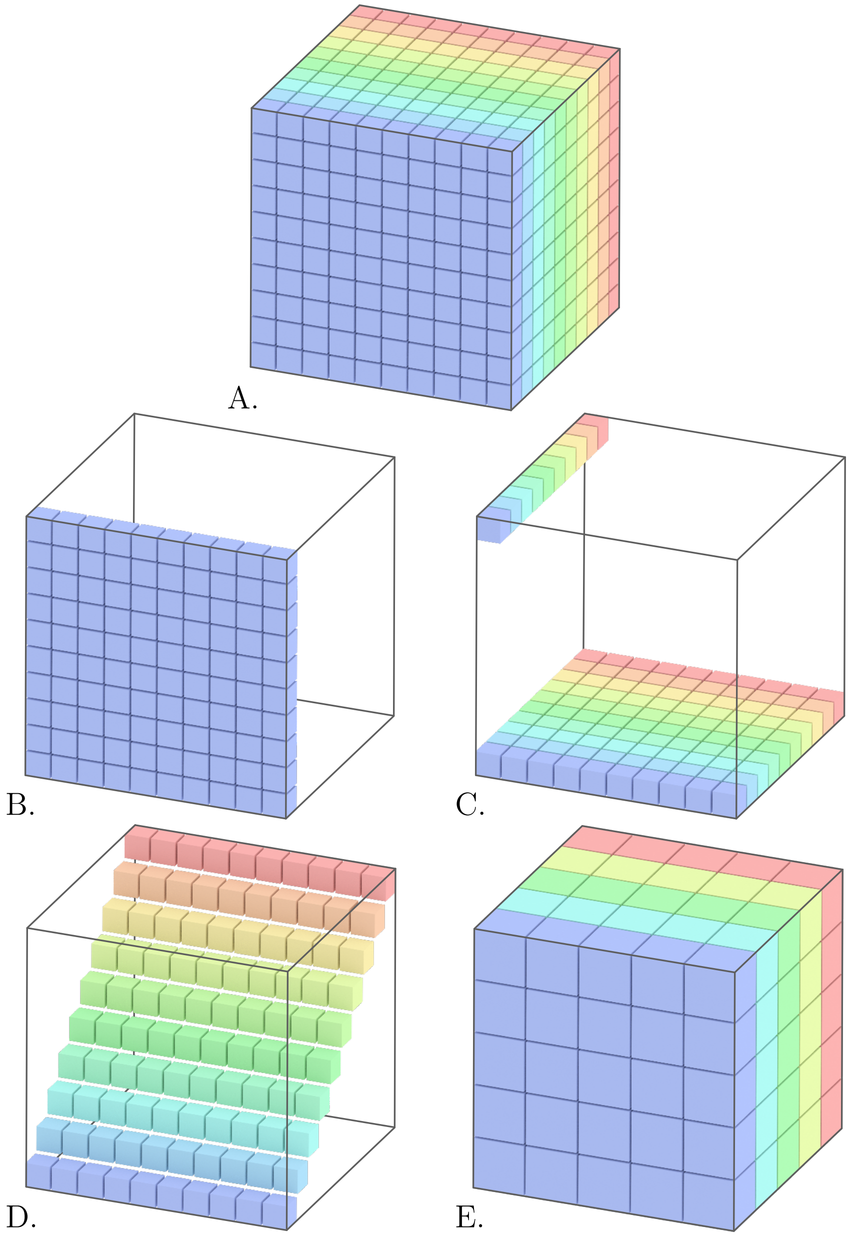


Figure 2.4: Illustration of a spectral image and the four methods of spectral imaging. A spectral image (A) is represented as a series of two-dimensional spatial images where each image corresponds to a different band of wavelengths. There are four categories of spectral imaging systems: wavelength scanning (B), spatial scanning (C), spatiospectral scanning (D), and snapshot imaging (E). Each depiction shows which dimension of data is recorded at each time step.

many gigabytes per image. The decision to use multispectral or hyperspectral imaging depends on the required spectral resolution and the amount of available light. For low-light situations, hyperspectral imaging is often not possible because there are not enough photons of each wavelength to produce an adequate signal-to-noise ratio image.

Multispectral and hyperspectral imaging have a wide range of applications in biology [68, 252], biomedicine [142], remote sensing [70], food safety [135], forensics [178], art conservation [7], and more [108]. In each application area, spectral imaging allows for the detection of otherwise invisible features. In this work, we focus on a specific type of spectral imaging called *spectral microscopy* and discuss how it can be used to explore micron-scale biological objects. More specifically, we derive algorithms for the analysis of spectral micrographs under real-world conditions.

2.1.1 Applications in Microscopy

Fluorescence microscopy is an imaging modality for measuring the light emitted by fluorescent molecules, called fluorophores, at spatial scales near or equal to the diffraction limit of light[4]. There are two main types of fluorophores—labels and sensors. Fluorescent *labels* are molecules that have fixed spectral signatures[5] that are used to differentiate and localize distinct objects or regions of interest. In *labeling* experiments, fluorophores with distinct emission spectra are chemically bound to different biological molecules. Fluorescent *sensors*, on the other hand, are molecules who's spectra shift in response to changing micro-environments [84, 111], molecular interactions [223], biomechanical forces [42], and more [208]. Fluorescent sensors can either be bound to target molecules or be free-floating. Both fluorescent labels and sensors

---

[4] There are now many super-resolution microscopy techniques available that use optical and computational methods to exceed the diffraction limit; however, these technique carry a large number of constraints that make sample preparation and imaging challenging [196].

[5] The emission spectrum of any fluorescent molecule will be altered by its environment; however, fluorescent labels are designed to be reasonably stable in most environments.

are commercially available with a wide range of spectral properties [115, 152].

Conventional fluorescence microscopes work similarly to digital cameras in that they both rely on filters to separate light. Unlike a digital camera, however, the goal of a fluorescence microscope is not to replicate the color vision but to completely separate the light of each unique emission spectrum. Which fluorophores and, more importantly, how many fluorescent labels/sensors can be used in an experiment is dictated by the spectral characteristics of the filter being used (i.e., what wavelengths of light they let pass). There are a number of tools available to help optimize the choice of fluorescent labels/sensors based on the configuration of a given microscope (e.g., SPEKcheck by Phillips et al. [172] and FPbase by Lambert [115]). These tools attempt to identify the set of fluorescent labels that minimize spectral *cross talk*, a problem where the filter used for one fluorescent label does not adequately exclude the fluorescence emission of other labels (see Waters [236] and references therein). Yet, even with such optimization, most fluorescence microscopes can only accommodate up to four fluorescent labels or two fluorescent sensors before cross talk becomes unavoidable. To make matters worse, real-world biological images are contaminated with background fluorescence from external sources of light and autofluorescence from naturally fluorescing organic molecules.

Spectral microscopy has become the method of choice when needing to avoid cross talk, mitigate autofluorescence, and simultaneously visualize many fluorescent labels or sensors [85, 99, 128]. Instead of relying on conventional filters, spectral microscopes use specialized optics such as prisms, diffraction gratings, continuous or variable filters, and interferometers to approximate the spectrum of emitted light for every pixel (see reviews by Garini & Tauber [67] and Li et al. [131][6]). Most commercial systems have a spectral resolution of 1–15 nm/band, a detection range between 380–800 nm, and collect tens (i.e., multispectral) or hundreds (i.e., hyperspectral) of spectral bands

---

[6] For an even broader treatment of spectral imaging, we refer the reader to the "Spectral Imaging" special issue of *Cytometry Part A* [126].

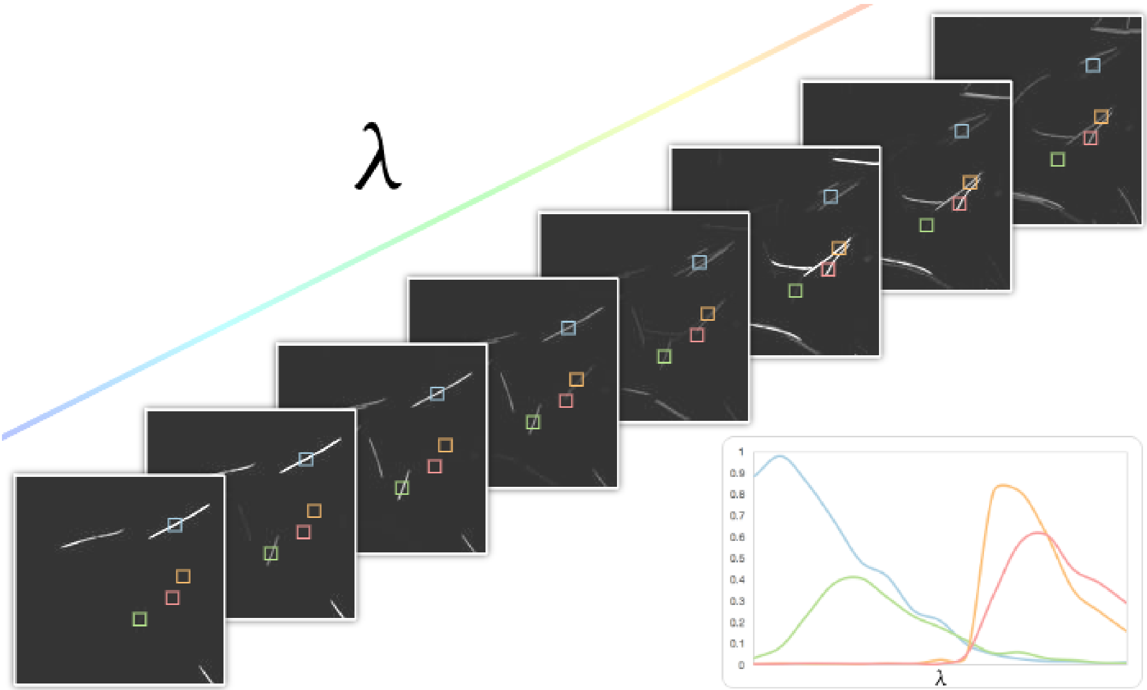


Figure 2.5: Spectral micrograph of objects labeled with one of four different fluorophores. Each two-dimensional image indicates to the amount of light recorded at a particular band of wavelengths. Plotting the intensity at the center of each colored bounding box produces the corresponding spectral profiles plotted at the bottom right. Each fluorescent label exhibits a different spectrum.

[35, 66]. Regardless of the acquisition mode, spectral microscopes provide access to the underlying spectral information so that the spectra of fluorescent labels can be compared and the spectral shift of fluorescent sensors can be measured (see Figure 2.5).

Spectral microscopy has been used in a range of label and sensor applications including karyotyping [198], Förster resonance energy transfer [216], Brainbow imaging [137], and live-cell imaging [227]. Perhaps the most prominent use of spectral microscopy is the simultaneous visualization of multiplexed fluorescent labels. Valm et al. [228, 229] showed that spectral microscopy, along with computational methods, can distinguish over one hundred different biological targets in a single field of view. In these experiments, binary combinations of fluorophores were used as barcodes to uniquely identifying different populations of bacteria. More recently, Mark Welch

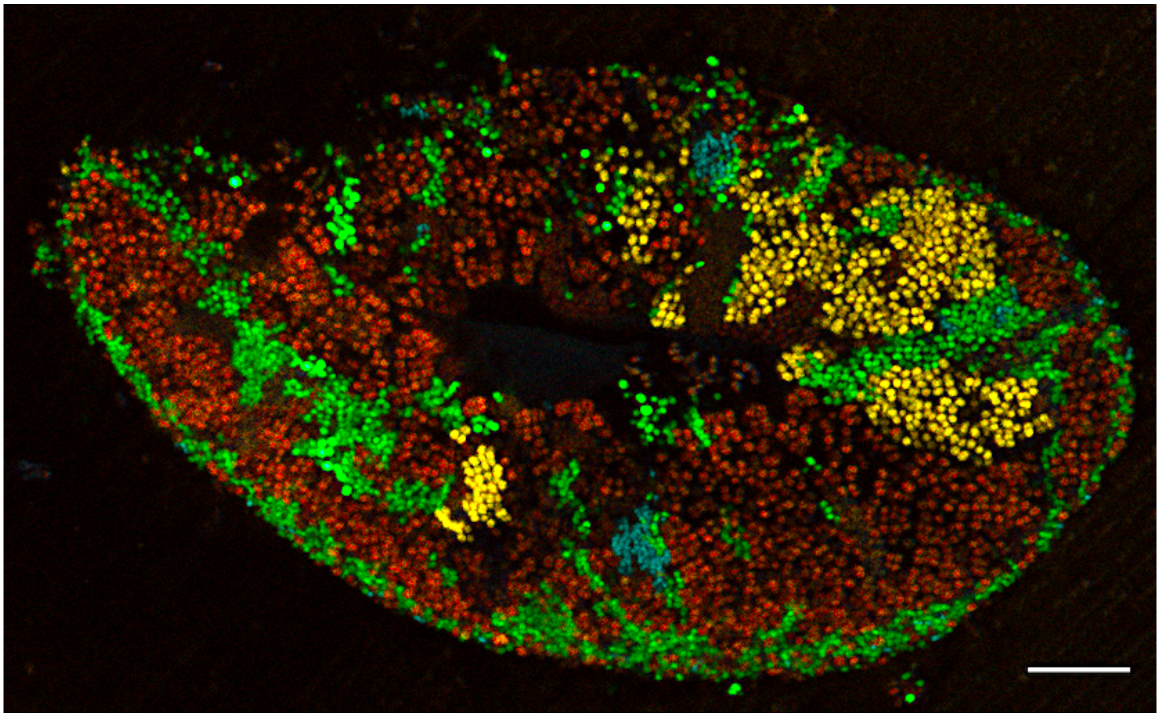


Figure 2.6: Processed spectral micrograph of a bacterial biofilm where each color identifies a different type of bacteria: blue is *Actinomyces*, green is *Streptococcus*, yellow is *Neisseriaceae*, and red is *Corynebacterium*. The scale bar represents 10 µm.

et al. [147] showed that spectral microscopy could be used to image complex bacterial biofilms. In these experiments, the fluorescent labels were bound to different taxonomic groups of bacteria using a method called fluorescence *in situ* hybridization (FISH). As the name suggests, FISH works by hybridizing fluorescent oligonucleotide probes to target cells containing complementary 16S ribosomal RNA sequences (see Bouvier & Del Giorgio [21] and references therein). Under optimal conditions, spectral microscopy allows for the visualization of dynamic microbial communities (see Figure 2.6). However, real-world spectral images are plagued by several problems: background fluorescence, autofluorescence, and spatially overlapping fluorophores. As a result, the spectrum at a given pixel may represent a mixture of light from other sources. We rely on a class of algorithms called spectral unmixing in order the extract the components of mixed spectra.

2.2 Spectral Mixing and Unmixing

In this section, we examine the properties of spectral mixing in well-behaved images (i.e., data without contamination from sources such as background fluorescence and autofluorescence). When a pixel location contains only one fluorescent label, the measured emission spectrum is a scaled version of that fluorophore's spectral signature (i.e., the canonical spectrum of the fluorophore under the given environmental conditions). If two or more fluorescent labels exist at the same pixel location, then the measured spectrum represents a combination of their spectral signatures. In this scenario, we refer to each unique component of the mixture as an endmember[7]. Exactly how the spectral signatures of endmembers combine to form an observed spectrum is described mathematical by a *spectral mixing model.*

2.2.1 Linear Mixing Model (LMM)

Let's assume that we have observed the spectrum at a pixel location using the spectral response functions defined in Eq. 2.4. We can denote the observed M-dimensional nonnegative spectrum as $\mathbf{y} \in \mathbb{R}^M_{\geq 0}$. Recall that M is the number of contiguous spectral bands. We can similarly denote the spectra of N endmembers as $\mathbf{s}_n \in \mathbb{R}^M_{\geq 0}$ for $n = 1, \ldots, N$. Exactly how these N endmember spectra mix to generate the observed spectrum $\mathbf{y}$ is determined by a spectral mixing model. In nearly all cases of spectral microscopy analysis, the mixing of light is assumed to follow a *linear mixing model* (LMM) [252]. The LMM states that

$$\mathbf{y} = \sum_{n=1}^{N} \mathbf{s}_n w_n + \mathbf{e}, \tag{2.5}$$

---

[7] Endmember is a term borrowed from mineralogy that refers to the purest form of an element that exists in a mixture.

where $w_n \in \mathbb{R}_{\geq 0}$ is the nonnegative weight for the n^{th} endmember, and $\mathbf{e} \in \mathbb{R}^M$ is a noise vector (typically assumed to follow a Gaussian or Poisson distribution). Since $M > N$, this is an overdetermined system. The LMM is commonly described as a linear combination of endmembers; however, it is important to note that the weights are nonnegative. Instead, we can more accurately describe the LMM as a conical combination of endmembers. The nonnegativity of the weights reflects the fact that a negative amount of an endmember does not have a physical meaning in the context of our problem[8]. In exchange for the simplicity of the LMM, we make the fairly large assumption that any nonlinear affects are negligible. Nonlinear mixing can arise as a result of light scattering, constructive or destructive interference, Förster resonance energy transfer, or contamination from background fluorescence.

Let's now consider an entire spectral image with P pixels, denoted $\mathcal{Y} \in \mathbb{R}_{\geq 0}^{P_y \times P_x \times M}$, where $P_y P_x = P$. To simplify our problem, we will start by defining an unfold$(\cdot)$ operation that takes a three-dimensional spectral image and flattens it into a two-dimensional matrix:

$$\text{unfold}(\mathcal{Y}) = \mathbf{Y}, \tag{2.6}$$

where $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{M \times P}$ is our spectral image matrix. Effectively, we have vectorized the spatial data for every spectral band in $\mathcal{Y}$. For completeness, we define fold$(\cdot)$ as the operation that reforms the spatial dimension and takes unfold$(\mathcal{Y})$ back to a tensor: $\text{fold}(\text{unfold}(\mathcal{Y})) = \mathcal{Y}$. Given an observed spectral image matrix, we can write the LMM in matrix notation as

$$\mathbf{Y} = \mathbf{S}\mathbf{W} + \mathbf{E}, \tag{2.7}$$

where $\mathbf{S} \in \mathbb{R}_{\geq 0}^{M \times N}$ is the endmember matrix with columns being endmember spectra, $\mathbf{W} \in \mathbb{R}_{\geq 0}^{N \times P}$ is the weight matrix with columns being the endmember weights for a given pixel, and $\mathbf{E} \in \mathbb{R}_{\geq 0}^{M \times P}$ is the noise matrix. Since the weights are nonnegative,

---

[8] In general, destructive interference of light waves is unlikely to break the assumption of nonnegativity.

$\mathbf{W}$ can be interpreted as the abundance of each endmember at every pixel. By folding $\mathbf{W}$ into the third-order tensor $\mathcal{W} \in \mathbb{R}_{\geq 0}^{P_y \times P_x \times N}$ (i.e., fold($\mathbf{W}$) = $\mathcal{W}$), we can conveniently visualize the endmember weights as a series of spatial abundance maps. In the literature, $\mathcal{W}$ is referred to as the *unmixed image* because it separately shows the contributions that each endmember makes to the mixed spectra in $\mathcal{Y}$. Since $\mathbf{W}$ is seldom, if ever, known in advance, the unmixed image must be determined indirectly from the observed spectral image in a process called *spectral unmixing.*

2.2.2 Linear Unmixing

Spectral unmixing is an inverse problem that aims to decompose a mixed spectrum into a set of component endmembers and their associated weights. When spectral mixing is assumed to follow the LMM, then the inverse problem is called linear unmixing. *Linear unmixing*, like most inverse problems, is ill-posed. According to the definition given by Jacques Hadamard [82], a problem is well-posed if: (1) a solution exists, (2) the solution is unique, and (3) the solution is stable. Should any of these conditions be violated, the problem is considered ill-posed. As we will see, the presence of noise means that linear unmixing is unlikely to have an exact solution, and there is often no guarantee that an approximate solution will be unique.

A variety of linear unmixing methods have been used by the remote sensing community for the analysis of hyperspectral geospatial data, and there exists a rich literature on advanced hyperspectral unmixing algorithms [17, 52, 90, 105, 106, 145]. Although some methods from remote sensing have been directly applied to spectral micrographs [86, 142], there are several key differences between the problem conditions that make unmixing algorithms for geospatial data unsuitable for spectral microscopy. As compared to remote sensing data,

- spectral micrographs typically measure fluorescence rather than reflectance;

- spectral micrographs are photon-limited and contain fewer spectral bands (i.e., lower spectral resolution);
- spectral micrographs may be illuminated by multiple sources;
- spectral micrographs may contain deadbands where illumination optics obscure the detection of some spectral wavelengths;
- and spectral micrographs are often heavily contaminated by background fluorescence and autofluorescent organic molecules.

As a result of these differences, there is an ongoing need to develop spectral unmixing methods specifically tailored towards spectral microscopy applications [9]. Existing methods for spectral microscopy unmixing can generally be classified as either a nonnegative matrix factorization or nonnegative least squares problem. In the remainder of this section, we describe these two problems and their prominent variants.

Nonnegative Matrix Factorization (NMF)

The LMM suggests that an observed spectral image can be decomposed into the product of two nonnegative matrices. Therefore, it seems appropriate to interpret linear unmixing as a nonnegative matrix factorization (NMF) problem. Despite already being a popular analysis method for remote sensing [169], NMF was launched into prominence as a result of a paper by Lee & Seung [123]. Unlike other matrix decomposition methods such as principal component analysis (PCA) or independent component analysis (ICA), NMF has the advantage of producing a set of basis vectors that are identifiable and interpretable as endmembers [63]. NMF is typically formulated as a constrained nonlinear optimization problem in which we aim to minimize the distance between a nonnegative matrix $\mathbf{Y}$ and the product of its nonnegative factors $\mathbf{S}$ and $\mathbf{W}$:

$$\min_{\mathbf{S},\mathbf{W}} D(\mathbf{Y}, \mathbf{SW}), \text{ subject to } \mathbf{S} \succeq \mathbf{0}, \mathbf{W} \succeq \mathbf{0}, \tag{2.8}$$

where D is some distance metric related to the assumed noise model, $\succeq$ is element-wise $\geq$, and $\mathbf{0}$ is the appropriately-sized zero matrix. Since neither $\mathbf{S}$ nor $\mathbf{W}$ are known *a priori*, NMF is a type of blind source separation. While it is not necessary to know either of $\mathbf{S}$ or $\mathbf{W}$, it is critical to have an accurate estimate of their structure in order to make the problem more well-determined. For example, a particularly useful piece of information is the number of nonnegative basis vectors (i.e., endmembers).

A common criticism against NMF is that there is no unique solution to Eq. 2.8. For any pair of matrices $\mathbf{S}^*$ and $\mathbf{W}^*$ that satisfy $\mathbf{Y} = \mathbf{S}^*\mathbf{W}^*$, it is always possible to define a nonsingular matrix $\mathbf{C}$ such that $\mathbf{Y} = (\mathbf{S}^*\mathbf{C})(\mathbf{C}^{-1}\mathbf{W}^*)$. In the trivial case, $\mathbf{C}$ can be a scaling or permutation matrix. Nevertheless, uniqueness results have been shown for special cases of NMF where the data is separable or sufficiently spread [51, 72, 95, 120]. For many practical applications, NMF has been shown to provide unique solutions up to some scaling and permutation [73].

The method of minimizing Eq. 2.8 turns out to be nontrivial since the objective function is nonconvex in $\mathbf{S}$ and $\mathbf{W}$. Methods for solving NMF problems are derived from the observation that $D(\mathbf{Y}, \mathbf{SW})$ is convex in $\mathbf{S}$ when $\mathbf{W}$ is fixed and convex in $\mathbf{W}$ when $\mathbf{S}$ is fixed. This holds due to the symmetry of the problem; in particular, note that

$$D(\mathbf{Y}, \mathbf{SW}) = D(\mathbf{Y}^T, \mathbf{W}^T\mathbf{S}^T). \tag{2.9}$$

The convex subproblems of NMF can be alternatingly optimized using a two block coordinate descent method (see Algorithm 1). The main variation between NMF algorithms lies in how these subproblems are solved or approximated. For a thorough treatment of NMF algorithms, we refer the reader to the work by Cichocki et al. [32].

Algorithm 1: Block Coordinate Descent Framework for NMF

```
Input  : spectral data $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{M \times P}$ and rank $N$
Output: optimal solutions $\mathbf{S}^* \in \mathbb{R}_{\geq 0}^{M \times N}$ and $\mathbf{W}^* \in \mathbb{R}_{\geq 0}^{N \times P}$
$\mathbf{W}^* \leftarrow$ initial guess;
$\mathbf{S}^* \leftarrow$ initial guess;
repeat
    // approximate solutions to steps 4 & 5 are often sufficient
    $\mathbf{W}^* \leftarrow \arg\min_{\mathbf{W}^* \succeq \mathbf{0}} D(\mathbf{Y}, \mathbf{S}^*\mathbf{W}^*)$;
    $\mathbf{S}^* \leftarrow \arg\min_{\mathbf{S}^* \succeq \mathbf{0}} D(\mathbf{Y}^T, \mathbf{W}^{*T}\mathbf{S}^{*T})$;
until a stopping condition is met;
```

The distance metric used in Eq. 2.8 depends on which noise distribution we assume for $\mathbf{E}$. There are three sources of noise in most microscopy systems: shot noise, thermal noise, and read noise [236]. *Shot noise* and *thermal noise* are derived from the intrinsic uncertainty of measuring discrete units such as photons or electrons. *Read noise* refers to the electronic interference introduced when converting the detected photons into an electrical signal. Whereas shot noise and thermal noise follow a Poisson distribution, read noise is generally assumed to be Gaussian distributed. If an image is shot-noise-limited, we aim to minimize Poisson distributed noise. This problem is equivalent to minimizing the negative log-likelihood, or Kullback-Leibler (KL) divergence [124]. The generalized KL divergence is written as

$$D_{\mathrm{KL}}(\mathbf{Y}, \mathbf{SW}) = \sum_{m,p} \Big(y_{mp} \log \frac{y_{mp}}{\sum_n s_{mn} w_{np}} - y_{mp} + \sum_n s_{mn} w_{np} \Big). \tag{2.10}$$

A more common assumption is that Gaussian distributed noise dominates $\mathbf{E}$. To minimize Gaussian noise, we equivalently minimize the squared Euclidean distance defined as

$$D_{\mathrm{E}}(\mathbf{Y}, \mathbf{SW}) = \frac{1}{2} \lVert \mathbf{Y} - \mathbf{SW} \rVert_F^2 = \frac{1}{2} \sum_{m,p} \Big(y_{mp} - \sum_n s_{mn} w_{np} \Big)^2, \tag{2.11}$$

where $\lVert \cdot \rVert_F$ is the Frobenius norm. Both D_{KL} and D_{E} have been used extensively for spectral unmixing thanks to the work by Lee & Seung [124]. However, Cichocki et al. [32, chap. 2] have outlined a variety of distance metrics to accommodate different noise profiles.

The nonconvexity of NMF typically requires that we add some type of regularization to our objective function to guide us to a sensible solution. Given that most spectral microscopy images contain large regions of dark background pixels, the most obvious option is to impose sparsity on the weights $\mathbf{W}$. As pointed out by Gillis & Glineur [74], the first-order Karush-Kuhn-Tucker (KKT) conditions for Eq. 2.8 show that NMF already provides some level of sparsity in its solutions. Consider the following generalized KKT conditions for NMF:

$$\mathbf{S} \succeq \mathbf{0}, \qquad \mathbf{W} \succeq \mathbf{0}, \tag{2.12}$$

$$\nabla_{\mathbf{S}} D \succeq \mathbf{0}, \qquad \nabla_{\mathbf{W}} D \succeq \mathbf{0}, \tag{2.13}$$

$$\mathbf{S} \odot \nabla_{\mathbf{S}} D = \mathbf{0}, \qquad \mathbf{W} \odot \nabla_{\mathbf{W}} D = \mathbf{0}, \tag{2.14}$$

where $\odot$ is the element-wise product. From $\mathbf{W} \odot \nabla_{\mathbf{W}} D = \mathbf{0}$, we can see that any stationary point requires that either of the elements w_{np} or $\frac{\partial D}{\partial w_{np}}$ be equal to zero. The same argument can be made to explain the implicit sparsity of $\mathbf{S}$. Nevertheless, many applications of spectral unmixing benefit from additional penalties against dense solutions.

Perhaps the most popular regularized NMF method for spectral microscopy unmixing is PoissonNMF developed by Neher et al. [162]. As one of the only open-source unmixing tools available in ImageJ [9], PoissonNMF has become a common method for spectral microscopy unmixing. As its name suggests, PoissonNMF is intended for use with images dominated by Poisson distributed noise. The PoissonNMF algorithm is a sparsity-regularized form of the KL-divergence that can be written as

$$\min_{\mathbf{S},\mathbf{W}} \underbrace{\sum_{m,p}\sum_{n} s_{mn}w_{np} - y_{mp}\log(\sum_{n} s_{mn}w_{np})}_{\text{Negative Log-Likelihood}} + \underbrace{\gamma\sum_{p}\frac{\sum_n w_{np}}{\sqrt{\sum_n w_{np}^2}}}_{\ell^1/\ell^2 \text{ Sparsity}}, \text{ s.t. } \mathbf{S}\succeq\mathbf{0}, \mathbf{W}\succeq\mathbf{0}, \tag{2.15}$$

where the γ parameter controls the level of sparsity. Although it is not described in Eq. 2.15, the PoissonNMF ImageJ plugin can operate in both a blind and semi-blind manner where select endmembers can be fixed during the minimization. As reported by Neher et al. [162], a limitation of PoissonNMF is that it yields unsatisfactory results when unmixing more than four endmembers. Similar semi-blind NMF approaches using the Gaussian noise model for read-noise-limited data have been suggested by Huang et al. [96] and Tong et al. [220]. However, these methods are also limited to unmixing three to five endmembers. Megjhani et al. [154] proposed a powerful morphologically constrained spectral unmixing (MCSU) algorithm using an NMF-like dictionary learning method. MCSU builds a dictionary of morphological motifs unique to each fluorescent label, and shows impressive results for up to eight fluorescent labels. Unfortunately, MCSU requires that the reference images share the same morphologies found in the test images and that the morphologies differ between fluorophores.

Nonnegative Least Squares (NLS)

For spectral unmixing by NMF, we are only required to provide an estimate of the number of endmembers N (i.e., the rank of $\mathbf{S}$). In some spectral microscopy applications, however, it may be possible to estimate each of the endmember spectra. This greatly simplifies our unmixing problem because we now only need to solve one of the two convex NMF subproblems. Furthermore, if we assumed that the noise follows a Gaussian distribution, then spectral unmixing becomes a nonnegative least squares

(NLS) problem defined as

$$\mathbf{W}_{\text{NLS}} = \arg\min_{\mathbf{W}} \frac{1}{2}\left\|\mathbf{Y} - \mathbf{S}\mathbf{W}\right\|_F^2, \text{ subject to } \mathbf{W} \succeq \mathbf{0}. \tag{2.16}$$

When $\mathbf{S}$ has full column rank, NLS is strictly convex and has a unique solution. In the context of spectral microscopy, requiring $\mathbf{S}$ to have full column rank means that each endmember spectrum must be linearly independent[9]. Most unmixing algorithms that come packaged with commercial microscopes solve the NLS problem using the active-set algorithm by Lawson & Hanson [122].

The convexity of NLS makes it a convenient method for solving spectral unmixing problems; however, it carries the strong assumption that $\mathbf{S}$ in known. Evaluating the validity of this assumption is critical to ensuring a good unmixing solution. Fortunately, there are certain cases where $\mathbf{S}$ can be approximated to a high degree. Since biological samples are labeled with a known set of fluorophores, it is often possible to generate a set of reference samples. A *reference sample* and its corresponding spectral *reference image* is a separate control sample that is labeled with only one of the fluorophores from the set (see Figure 2.7). We can denote the corresponding spectral reference images as $\mathcal{R}_n \in \mathbb{R}_{\geq 0}^{Q_y \times Q_x \times M}$ and the spectral reference matrices as $\text{unfold}(\mathcal{R}_n) = \mathbf{R}_n \in \mathbb{R}_{\geq 0}^{M \times Q}$ for $n = 1, \ldots, N$, where $Q_y Q_x = Q$ is the number of pixels.

Spectral microscopy practitioners estimate endmembers from reference images using a variety of methods based on the arithmetic mean. In general, endmembers are determined by the average spectral signature over all foreground pixels or some user-defined region of interest within the reference image. Since the illumination conditions can vary between reference and test images, it is common to normalize the

---

[9] Linear independence is different than statistical independence since we do not require the endmember spectra to be orthogonal.

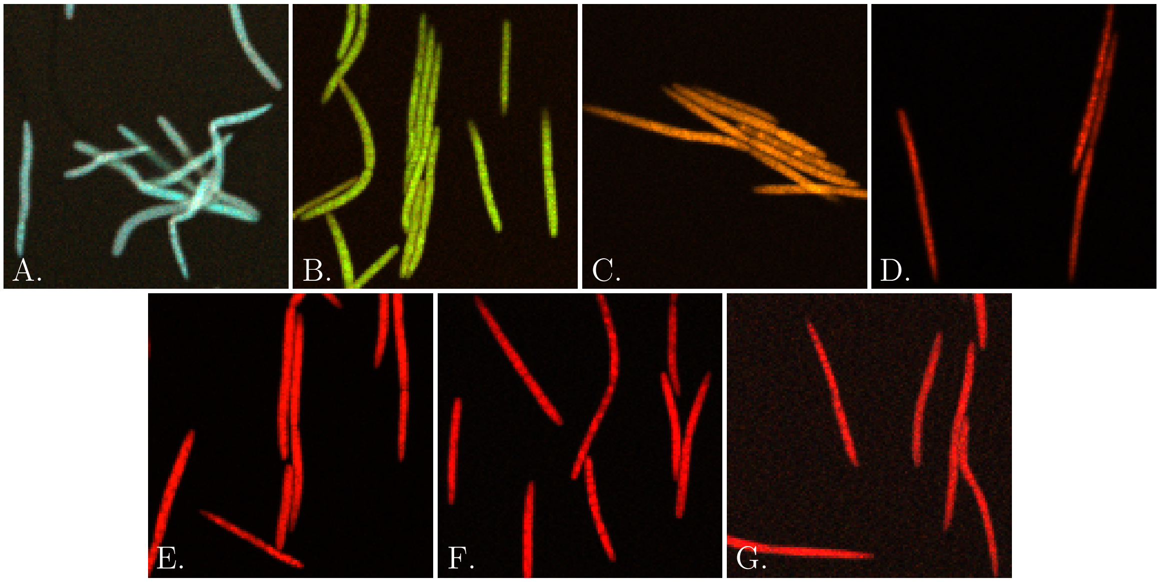


Figure 2.7: Example of spectral reference images of fluorescent labeled bacteria. Each spectral image has been color-coded to mimic its appearance to the human eye. Each image corresponds to a different fluorescent label: DY-415 (A), DY-490 (B), ATTO 520 (C), ATTO 550 (D), Texas Red-X (E), ATTO 620 (F), and ATTO 655 (G). Notice that the human eye is unable to distinguish between the red fluorescent labels (D-G).

mean endmembers by their ℓ^∞- or ℓ^1-norm. We can write this in matrix form as

$$\begin{aligned} \bar{\mathbf{s}}_n &= \frac{1}{\|\mathbf{h}_n\|_1}\mathbf{R}_n\mathbf{h}_n \\ \mathbf{s}_n &= \frac{1}{\|\bar{\mathbf{s}}_n\|_\infty}\bar{\mathbf{s}}_n \text{ for } n = 1, \ldots, N, \end{aligned} \tag{2.17}$$

where $\mathbf{h}_n \in \mathbb{R}^Q$ is a binary vector with 1 indicating the foreground and 0 indicating the background of the n^{th} reference image. Foreground/background thresholding can be performed using any of a number of different thresholding algorithms.

When $\mathbf{Y}$ is known to only contain endmembers corresponding to measurable fluorescent labels, NLS will provide sufficiently accurate unmixing results. However, it is not always possible to know which endmembers exist in an image. In particular, there are many different types of autofluorescent molecules, and knowing which autofluorescence endmembers are present is often impractical. In addition, a poorly prepared reference sample or an improper estimation method will lead to undesired unmixing results [252]. As such, NLS typically lacks the flexibility to adequately handle many of the unmixing problems that arise in realistic applications.

2.3 Spectral Mixing and Unmixing of Contaminated Micrographs

Under ideal imaging conditions, both NLS and NMF methods can produce decent solutions to the spectral unmixing problem. However, real-world spectral images are almost never well-behaved. Contamination can arise from multiple sources including nonuniform illumination, stray and scattered light, nonlinear interactions, and autofluorescent molecules [99, 236, 252]. The LMM lumps all of these sources of contamination into the noise term $\mathbf{E}$. As a result, spectral unmixing methods based on the LMM will distribute the unaccounted intensity across many endmember weights—a problem called *proportion indeterminacy* [250]

In order to perform spectral unmixing on real-world images, we need to examine the assumptions made by the LMM and observe where NMF and NLS fail. With this understanding, we can build a new unmixing method that is suitable for contaminated spectral imagines. In this section, we formally define an *affine mixing model* (AMM) that generalizes the LMM by including a term to absorb any background fluorescence. From this model, we derive an affine NMF method for estimating endmembers from reference images. We then propose a semi-blind sparse affine spectral unmixing (SSASU) method for images contaminated with both autofluorescence and background fluorescence.

2.3.1 Affine Mixing Model (AMM)

According to the LMM, all observed spectra are the result of a conical combination of light from different endmembers and some (hopefully) small amount of noise. Yet, all fluorescence images contain some amount of global background fluorescence that originates from a variety of sources in the sample and optical path [236]. This background fluorescence represents a type of intensity offset that is present in all observed spectra. By neglecting this source of light, the LMM forces the noise term, $\mathbf{E}$, to account for the offset intensity. This becomes a problem in NMF and NLS when $\mathbf{E}$ is assumed to follow a Gaussian distribution centered at zero. As a result, regularization terms intended to enforce sparsity will be ineffective under these conditions and the solution for the weight matrix $\mathbf{W}$ will be dense—a form of proportion indeterminacy. Laurberg & Hansen [121] identified a similar problem and proposed a generalized NMF method that included a term to account for the offset component found in a variety of data types. Following this idea, we can formulate a new mixing model for spectral images containing background fluorescence. The affine mixing model (AMM) is defined as

$$\mathbf{Y} = \mathbf{S}\mathbf{W} + \mathbf{b}\mathbf{1}^T + \mathbf{E}, \tag{2.18}$$

where $\mathbf{b} \in \mathbb{R}^M_{\geq 0}$ is the nonnegative background spectrum and $\mathbf{1}$ is a vector of ones. The offset component of the observed spectral imaging is given by $\mathbf{b}\mathbf{1}^T$. Conveniently, the AMM becomes equivalent to the LMM when no background fluorescence exists (i.e., $\mathbf{b} = \mathbf{0}$). It is important to note that the AMM assumes that the background spectrum is ubiquitous across the entire image. This assumption is most likely to fail when illumination optics are misaligned; otherwise, the offset term provides a mechanism to enforce sparsity on the weight matrix.

The main goal of spectral unmixing is to determine the abundance of each fluorescent label; however, fluorescent labels are only a subset of the endmembers found in a contaminated spectral image. Instead, we can think of the endmember matrix, $\mathbf{S}$, as being composed of two different endmember types—fluorophore endmembers and autofluorescence endmembers. As we discussed above, it is often possible to estimate the endmember spectra of fluorescent labels using reference images. Autofluorescence endmember spectra, on the other hand, are notoriously more difficult. While it is possible to use an unlabeled sample as a reference for autofluorescence, there is no guarantee that the same autofluorescence molecules will exist in every spectral image. Therefore, it makes sense to slightly modify Eq. 2.18 by partitioning $\mathbf{S}$ into the K fluorophore endmembers that remain constant across spectral images and the L autofluorescence endmembers that are unique to each observed image. If we denote the fluorophore endmember matrix as $\bar{\mathbf{S}} \in \mathbb{R}_{\geq 0}^{M \times K}$ and the autofluorescence endmember matrix as $\breve{\mathbf{S}} \in \mathbb{R}_{\geq 0}^{M \times L}$, then we can write the AMM as

$$\mathbf{Y} = \left[\bar{\mathbf{S}}, \breve{\mathbf{S}}\right] \mathbf{W} + \mathbf{b}\mathbf{1}^T + \mathbf{E}. \tag{2.19}$$

Since $\mathbf{S} = \left[\bar{\mathbf{S}}, \breve{\mathbf{S}}\right]$, Eq. 2.18 and Eq. 2.19 are equivalent. However, this more explicit formulation allows us to separately treat endmembers from different sources.

2.3.2 Semi-blind Sparse Affine Spectral Unmixing (SSASU)

NMF- and NLS-based spectral unmixing methods presents different trade-offs that must be considered before a particular method is chosen for analysis. NMF provides flexibility in that endmember spectra need not be known *a priori*. This flexibility, however, means that we must solve a difficult nonconvex problem. NLS is strictly convex with a unique and easy to determine solution; however, inaccurate or missing endmember spectra can lead to erroneous solutions. As a compromise, we present a hybrid semi-blind source separation approach based on the AMM called semi-blind sparse affine spectral unmixing (SSASU). By assuming a Gaussian noise model, we can write SSASU as

$$\min_{\breve{\mathbf{S}},\mathbf{W},\mathbf{b}} \frac{1}{2}\left\|\mathbf{Y} - [\bar{\mathbf{S}}, \breve{\mathbf{S}}]\mathbf{W} - \mathbf{b}\mathbf{1}^T\right\|_F^2 + \gamma\left\|\mathbf{W}\right\|_1, \qquad \text{subject to } \breve{\mathbf{S}} \succeq \mathbf{0}, \mathbf{W} \succeq \mathbf{0}, \mathbf{b} \succeq \mathbf{0}, \tag{2.20}$$

where γ is a parameter for controlling the level of sparsity and $\|\cdot\|_1$ is the sum over all matrix elements. A related affine model (incorrectly called a linear model) was proposed by Woolfe et al. [239] to address autofluorescence in spectral images; however, this model did not include a penalty for dense solutions.

In this problem, semi-blind refers to the fact that Eq. 2.20 minimizes over $\breve{\mathbf{S}}$ and not $\bar{\mathbf{S}}$. The fluorophore endmember spectra are fixed in SSASU because we assume that they can be accurately estimated from reference images. Since the autofluorescence endmembers vary from one spectral image to another, $\breve{\mathbf{S}}$ is learned directly from the data. SSASU is based on the AMM and includes a background term to account for an offset component which allows for the effective enforcement of sparsity. Since the objective function is nonconvex, the sparsity penalty is also useful in helping Eq. 2.20 converge to a desired solution.

We solve SSASU using a three block coordinate descent framework, similar to Algorithm 1, where each iteration separately updates $\breve{\mathbf{S}}$, $\mathbf{W}$, and $\mathbf{b}$. We follow the multiplicative update scheme suggested by Lee & Seung [124] in which each update can be viewed as a type of gradient descent. Because the typical gradient descent update is additive and can introduce negative elements, we aim to derive updates of the form

$$\begin{aligned} \breve{\mathbf{S}} &\leftarrow \breve{\mathbf{S}} \odot \frac{\nabla^-_{\breve{\mathbf{S}}} D_{\text{SSASU}}}{\nabla^+_{\breve{\mathbf{S}}} D_{\text{SSASU}}}, \\ \mathbf{W} &\leftarrow \mathbf{W} \odot \frac{\nabla^-_{\mathbf{W}} D_{\text{SSASU}}}{\nabla^+_{\mathbf{W}} D_{\text{SSASU}}}, \\ \mathbf{b} &\leftarrow \mathbf{b} \odot \frac{\nabla^-_{\mathbf{b}} D_{\text{SSASU}}}{\nabla^+_{\mathbf{b}} D_{\text{SSASU}}}, \end{aligned} \tag{2.21}$$

where ∇^- and ∇^+ represent the negative and positive terms of the gradient, respectively, and the division is performed element-wise.

To derive the updates for $\breve{\mathbf{S}}$, $\mathbf{W}$, and $\mathbf{b}$, we first rewrite the SSASU objective function from Eq. 2.20 as

$$D_{\text{SSASU}} = \frac{1}{2}\operatorname{Tr}\left(\left(\mathbf{Y} - \mathbf{SW} - \mathbf{b}\mathbf{1}_P^T\right)\left(\mathbf{Y} - \mathbf{SW} - \mathbf{b}\mathbf{1}_P^T\right)^T\right) + \gamma\mathbf{1}_N^T\mathbf{W}\mathbf{1}_P, \tag{2.22}$$

where $\operatorname{Tr}(\cdot)$ is a matrix trace, $\mathbf{1}_P$ is a P-dimensional vector of ones, and $\mathbf{1}_N$ is an N-dimensional vector of ones. Note that we have substituted $\mathbf{S}$ for $\left[\bar{\mathbf{S}}, \breve{\mathbf{S}}\right]$ in Eq. 2.22 to simplify the notation. If we expand the terms inside the trace in Eq. 2.22, we get

$$\begin{aligned} D_{\text{SSASU}} =& \frac{1}{2}\operatorname{Tr}\Big(\mathbf{YY}^T - \mathbf{YW}^T\mathbf{S}^T - \mathbf{Y}\mathbf{1}_P\mathbf{b}^T - \mathbf{SWY}^T + \mathbf{SWW}^T\mathbf{S}^T & (2.23)\\ & + \mathbf{SW}\mathbf{1}_p\mathbf{b}^T - \mathbf{b}\mathbf{1}_P^T\mathbf{Y}^T + \mathbf{b}\mathbf{1}_P^T\mathbf{W}^T\mathbf{S}^T + \mathbf{b}\mathbf{1}_P^T\mathbf{1}_P\mathbf{b}^T\Big) + \gamma\mathbf{1}_N^T\mathbf{W}\mathbf{1}_P. & (2.24) \end{aligned}$$

In this form, it becomes much easier to determine the partial derivatives of D_{SSASU} with respect to $\mathbf{S}$, $\mathbf{W}$, and $\mathbf{b}$. Using the properties of matrix traces we get

$$\nabla_{\mathbf{S}} D_{\text{SSASU}} = \mathbf{SWW}^T + \mathbf{b}\mathbf{1}_P^T\mathbf{W}^T - \mathbf{YW}^T \tag{2.25}$$

$$\nabla_{\mathbf{W}} D_{\text{SSASU}} = \mathbf{S}^T\mathbf{SW} + \mathbf{S}^T\mathbf{b}\mathbf{1}_P^T + \gamma\mathbf{1}_N\mathbf{1}_P^T - \mathbf{S}^T\mathbf{Y} \tag{2.26}$$

$$\nabla_{\mathbf{b}} D_{\text{SSASU}} = \mathbf{SW}\mathbf{1}_P + \mathbf{b}\mathbf{1}_P^T\mathbf{1}_P - \mathbf{Y}\mathbf{1}_P. \tag{2.27}$$

From these partial derivatives, we can define the gradient updates using the heuristics in Eq. 2.21. By separating the negative and positive terms of the partial derivatives, we obtain the update rules

$$\begin{aligned}\mathbf{S} &\leftarrow \mathbf{S} \odot \frac{\mathbf{Y}\mathbf{W}^T}{\left(\mathbf{S}\mathbf{W} + \mathbf{b}\mathbf{1}_P^T\right)\mathbf{W}^T}, \\ \mathbf{W} &\leftarrow \mathbf{W} \odot \frac{\mathbf{S}^T\mathbf{Y}}{\mathbf{S}^T\left(\mathbf{S}\mathbf{W} + \mathbf{b}\mathbf{1}_P^T\right) + \gamma}, \\ \mathbf{b} &\leftarrow \mathbf{b} \odot \frac{\mathbf{Y}\mathbf{1}_P}{\left(\mathbf{S}\mathbf{W} + \mathbf{b}\mathbf{1}_P^T\right)\mathbf{1}_P}.\end{aligned} \tag{2.28}$$

Finally, we need to restrict the update rule for $\mathbf{S}$ to the autofluorescence endmember matrix $\breve{\mathbf{S}}$. Let's first partition the weight matrix into a fluorophore weight matrix, denoted $\bar{\mathbf{W}} \in \mathbb{R}_{\geq 0}^{K \times P}$, and an autofluorescence weight matrix, denoted $\breve{\mathbf{W}} \in \mathbb{R}_{\geq 0}^{L \times P}$, such that $\mathbf{W}^T = \left[\bar{\mathbf{W}}^T, \breve{\mathbf{W}}^T\right]^T$. By restricting the update rule for $\mathbf{S}$ to the autofluorescence components, we obtain the autofluorescence endmember update rule

$$\breve{\mathbf{S}} \leftarrow \breve{\mathbf{S}} \odot \frac{\mathbf{Y}\breve{\mathbf{W}}^T}{\left(\breve{\mathbf{S}}\breve{\mathbf{W}} + \mathbf{b}\mathbf{1}_P^T\right)\breve{\mathbf{W}}^T}. \tag{2.29}$$

To prevent against scaling problems, each update of $\breve{\mathbf{S}}$ is typically followed by a normalization of the columns. Combining the update rules from Eq. 2.28 and Eq. 2.29 along with the normalization procedure, we have the complete iterative scheme for SSASU (see Algorithm 2).

Endmember estimation

Unlike standard NMF, SSASU assumes that the endmember spectra for the fluorescent labels are known *a priori.* As we have seen with NLS, estimating the K fluorophore endmember spectra is possible from a set of reference images $\mathbf{R}_k \in \mathbb{R}_{\geq 0}^{M \times Q}$ for $k = 1, \ldots, K$. However, the arithmetic mean method defined in Eq. 2.17 assumes the same zero-centered Gaussian noise as the LMM. Fluorophore endmember spectra estimated by the mean of the foreground pixels will generally contain an offset component. Even slight inaccuracies in endmember estimates are known to dramatically

Algorithm 2: SSASU

```
Input  : spectral data $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{M \times P}$, fluorophore endmember spectra $\bar{\mathbf{S}} \in \mathbb{R}_{\geq 0}^{M \times K}$,
         autofluorescence endmember rank $L$, and sparsity parameter $\gamma$
Output: optimal solutions $\breve{\mathbf{S}}^* \succeq \mathbf{0}$, $\mathbf{W}^* \succeq \mathbf{0}$, and $\mathbf{b}^* \succeq \mathbf{0}$
$\mathbf{W}^* \leftarrow$ initial guess;
$\breve{\mathbf{S}}^* \leftarrow$ initial guess;
$\mathbf{b}^* \leftarrow$ initial guess;
repeat
    $\hat{\mathbf{Y}} \leftarrow [\bar{\mathbf{S}}, \breve{\mathbf{S}}^*]\mathbf{W}^* + \mathbf{b}^*\mathbf{1}^T$;
    $\mathbf{W}^* \leftarrow \mathbf{W}^* \odot \Big(\big([\bar{\mathbf{S}}, \breve{\mathbf{S}}^*]^T\mathbf{Y}\big) \oslash \big([\bar{\mathbf{S}}, \breve{\mathbf{S}}^*]^T\hat{\mathbf{Y}} + \gamma\big)\Big)$;
    $\breve{\mathbf{S}}^* \leftarrow \breve{\mathbf{S}}^* \odot \Big(\big(\mathbf{Y}\breve{\mathbf{W}}^{*T}\big) \oslash \big(\hat{\mathbf{Y}}\breve{\mathbf{W}}^*\big)\Big)$;
    foreach $\breve{\mathbf{s}}_i \in \breve{\mathbf{S}}^*$ do
        $\breve{\mathbf{s}}_i \leftarrow \breve{\mathbf{s}}_i / \|\breve{\mathbf{s}}_i\|_\infty$;
    end
    $\mathbf{b}^* \leftarrow \mathbf{b}^* \odot \Big(\big(\mathbf{Y}\mathbf{1}\big) \oslash \big(\hat{\mathbf{Y}}\mathbf{1}\big)\Big)$;
until a stopping condition is met;
// $\odot$ and $\oslash$ represent element-wise multiplication and division,
   respectively
```

effect the determination of the endmember weights by inducing proportion indeterminacy [250]. We address this problem by proposing an affine NMF method for estimating endmembers.

To estimate endmembers from reference images, we first define an affine model similar to Eq. 2.18 for reference images as

$$\mathbf{R}_k = \mathbf{s}_k \ddot{\mathbf{w}}_k^T + \ddot{\mathbf{b}}_k \mathbf{1}^T + \mathbf{E}_k, \tag{2.30}$$

where $\ddot{\mathbf{w}}_k \in \mathbb{R}_{\geq 0}^{Q}$ and $\ddot{\mathbf{b}}_k \in \mathbb{R}_{\geq 0}^{M}$ are the weights and background spectrum for the n^{th} reference image. Since the reference image contains only one fluorophore, the weight matrix gives the intensity scaling of $\mathbf{s}_k$ for each pixel. The inverse problem of estimating each endmember spectra is equivalent to a blind source separation problem solved by

$$\begin{aligned} &\min_{\mathbf{s}_k,\ddot{\mathbf{w}}_k,\ddot{\mathbf{b}}_k} \frac{1}{2}\left\|\mathbf{R}_k - \mathbf{s}_k\ddot{\mathbf{w}}_k^T - \ddot{\mathbf{b}}_k\mathbf{1}^T\right\|_F^2, \\ &\text{subject to } \mathbf{s}_k \succeq \mathbf{0}, \ddot{\mathbf{w}}_k \succeq \mathbf{0}, \ddot{\mathbf{b}}_k \succeq \mathbf{0}. \end{aligned} \tag{2.31}$$

Since $\mathbf{s}_n$, $\ddot{\mathbf{w}}_n$, and $\ddot{\mathbf{b}}_n$ are all unknown, the objective function in Eq. 2.31 is nonconvex. Again, we use a block coordinate descent method to find a solution. The update rules for the three variables are

$$\begin{aligned} \mathbf{s}_k &\leftarrow \mathbf{s}_k \odot \frac{\mathbf{R}_k\ddot{\mathbf{w}}_k}{\left(\mathbf{s}_k\ddot{\mathbf{w}}_k^T + \ddot{\mathbf{b}}_k\mathbf{1}^T\right)\ddot{\mathbf{w}}_k}, \\ \ddot{\mathbf{w}}_k &\leftarrow \ddot{\mathbf{w}}_k \odot \frac{\mathbf{R}_k^T\mathbf{s}}{\left(\ddot{\mathbf{w}}_k\mathbf{s}_k^T + \mathbf{1}\ddot{\mathbf{b}}_k^T\right)\mathbf{s}_k}, \\ \ddot{\mathbf{b}}_k &\leftarrow \ddot{\mathbf{b}}_k \odot \frac{\mathbf{R}_k\mathbf{1}}{\left(\mathbf{s}_k\ddot{\mathbf{w}}_k^T + \ddot{\mathbf{b}}_k\mathbf{1}^T\right)\mathbf{1}}, \end{aligned} \tag{2.32}$$

for $k = 1, \ldots, K$. An advantage of these update rules compared to the mean estimation method is that they operate over the entire reference image $\mathbf{R}_k$. Therefore, we are not required to threshold foreground from background or define a region of interest. This is important to emphasize because the estimated spectrum can change dramatically depending on how the thresholding or regions are defined.

2.4 Comparison of Methods

We assessed the robustness of endmember estimation by affine NMF and arithmetic mean across seven different reference images: DY-415, DY-490, ATTO 520, ATTO 550, Texas Red-X, ATTO 620, and ATTO 655. Each endmember estimate was compared against a ground truth endmember by measuring the spectral angle.

We evaluated our SSASU method by unmixing a set of ten real-world spectral images each of which were labeled with seven fluorophores (see Table 2.1) and contaminated by background and autofluorescence (see Appendix A for details on sample preparation and imaging). For comparison, we performed the same evaluation with

Table 2.1: Oligonucleotide probes and their taxonomic targets

Probe ID	Taxon	Target	Fluorophore
Smit651-DY415-2	Species	*Streptococcus mitis*	DY-415
Ssal372-DY490-2	Species	*Streptococcus salivarius*	DY-490
Prv392-AT520-2	Genus	*Prevotella*	ATTO 520
Vei488-AT550-1	Genus	*Veillonella*	ATTO 550
Act118-TRX-1	Genus	*Actinomyces*	Texas Red-X
Nei1030-AT620-2	Family	*Neisseriaceae*	ATTO 620
Rot491-AT655-2	Genus	*Rothia*	ATTO 655

the two most commonly used unmixing methods—PoissonNMF and NLS. The success of each method was measured by the Relative Reconstruction Error (RRE) and Proportion Indeterminacy (PI).

2.4.1 Endmember Estimation

Since poorly estimated endmembers can degrade the overall performance of unmixing, it remains important to evaluate the accuracy of the estimates. Yet, defining a ground truth set of endmembers is difficult because all images will contain some level of background fluorescence. Instead, we compare the endmember estimates to fluorometer data reported in the literature [152]. While the fluorometer data is measured under different optical and environmental conditions, these data still provide a useful baseline to check estimates made from reference images. For the following comparison, endmembers were estimated by the mean method using Eq. 2.17 and by the affine NMF method using Eq. 2.31. The binary mask required by the mean method was computed using the the Triangle algorithm because of its robustness to different illumination conditions [248].

While neither the mean nor the affine NMF estimation methods perfectly matched the fluorometer data due to differing environmental conditions and the spectral accuracy of the microscope [35], Figure 2.8 shows that affine NMF estimated endmembers are less contaminated by background fluorescence than the mean estimated

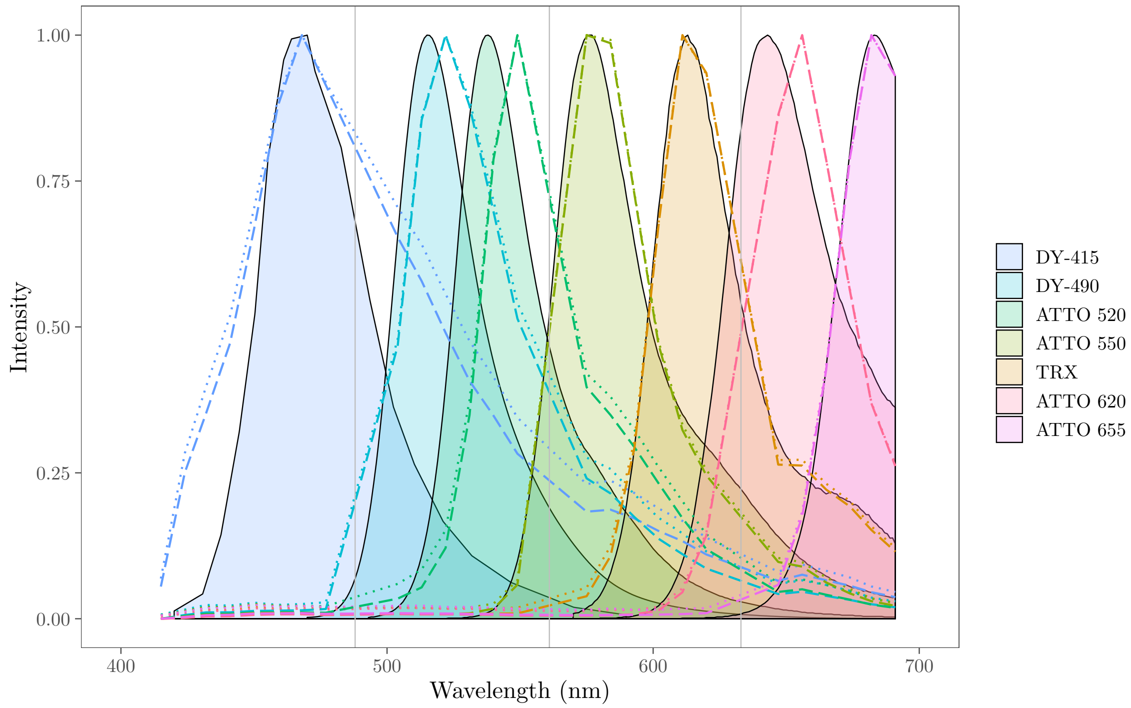


Figure 2.8: Comparison of seven mean and affine NMF estimated endmember spectra with fluorometer measurements. The shaded regions represent the fluorometer data, the dotted lines represent the mean estimates, and the dashed lines represent the affine NMF estimates. The gray vertical lines show the wavelength where dichroic mirrors blocked the measurement of emitted light (i.e., locations of missing spectral data).

endmembers. As expected, fluorophores with higher-energy emission spectra are effected more by background generated from mounting media and other sources. We evaluated this quantitatively by calculating the spectral angle between each mean and affine NMF estimated endmember and its corresponding fluorometer-measured spectrum. The spectral angle, which is related to cosine similarity, is calculated as $\theta(\mathbf{s}, \hat{\mathbf{s}}) = \arccos\left(\frac{\mathbf{s}\cdot\hat{\mathbf{s}}}{\|\mathbf{s}\|_2\|\hat{\mathbf{s}}\|_2}\right)$, where $\mathbf{s}$ is the fluorometer spectrum and $\hat{\mathbf{s}}$ is the estimated spectrum. As reported in Table 2.2, affine NMF estimated endmembers are as good or better at approximating the true endmember spectra than estimation by arithmetic mean.

Table 2.2: Spectral angle between estimated endmembers and their corresponding fluorometer spectra

Method	DY-415	DY-490	ATTO 520	ATTO 550	TRX	ATTO 620	ATTO 655
Mean	0.484	0.324	0.455	0.088	0.065	0.203	0.053
Affine NMF	0.421	0.293	0.455	0.087	0.052	0.203	0.024

2.4.2 Spectral Unmixing

The unmixing performance of SSASU, NLS, and PoissonNMF was evaluated on the basis of two relative error metrics that ranged from zero (better) to one (worse). The ability to reconstruct the observed signal was determined by the relative reconstruction error (RRE), and was defined by

$$RRE(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F}{\|\mathbf{Y}\|_F}, \tag{2.33}$$

where $\mathbf{Y}$ was the observed image and $\hat{\mathbf{Y}}$ was the reconstructed image. It is worth noting that the RRE provides a measure of underfitting/overfitting and does not indicate the quality of a solution. For example, an algorithm can produce a physically meaningless solution to the unmixing problem with a near-zero RRE by fitting the noise in addition to the signal. Therefore, we use the RRE in conjunction with a metric that evaluates the ability of the unmixing algorithm to separate autofluorescence from the fluorophore endmembers.

Since the fluorophores used to label the test images were known to bind to distinct bacteria, each pixel in the image contains at most one fluorophore (with the rare exception of areas where different bacteria overlap). In this case, a poor solution for an observed spectrum will produce positive weights for more than one endmember. This is a type of overfitting known as proportion indeterminacy (PI). We measured PI by checking the non-orthogonality of the weight matrix. Since each pixel contains only one endmember, the property $\mathbf{W}^*\mathbf{W}^{*T} = \mathbf{D}$ should hold, where $\mathbf{D} = diag(\|\mathbf{w}_1^*\|_2^2, \ldots, \|\mathbf{w}_N^*\|_2^2)$ and $\mathbf{w}_n^*$ is the n^{th} column of $\mathbf{W}^{*T}$. From this property, we define a measure of PI as

$$PI(\mathbf{W}^*) = \frac{\|\mathbf{D} - \mathbf{W}^*\mathbf{W}^{*T}\|_F}{\|\mathbf{D}\|_F}. \tag{2.34}$$

Together, RRE and PI indicate the fit and quality of each unmixing solution.

SSASU requires two parameters to be determined prior to unmixing: the sparsity parameter, γ, and the rank of the autofluorescence endmember matrix, $L = rank(\breve{\mathbf{S}})$. While these parameters can be tuned on a per-image basis, we kept the parameters constant across all test images to illustrate the robustness of the algorithm. The parameters were empirically chosen to be $\gamma = 0.009$ and $L = 1$. PoissonNMF also requires the user to set a sparsity parameter γ_P and the autofluorescence endmember rank K_P. These parameters were set empirically to $\gamma_P = 6$ and $L_P = 1$ for all test images. The fluorophore endmembers, $\mathbf{S}$, used for SSASU, NLS, and PoissonNMF were set to the values estimated by the affine NMF method. The autofluorescence endmember for NLS was estimated by affine NMF from a no-probe control sample (i.e., an unlabeled sample taken from the dorsum of the tongue).

SSASU, NLS, and PoissonNMF were each able to effectively reconstruct the test images with RREs below 0.14 (see bottom of Figure 2.9). Across all test images, NLS had the highest RRE. This was expected because NLS has the fewest free variables and is therefore more limited in its ability to fit the observed data. Although PoissonNMF had the lowest RRE in most cases, both PoissonNMF and SSASU performed similarly and neither exhibited signs of overfitting.

Despite having similar RRE across all test images, the quality of the solutions varied dramatically between the different methods. SSASU outperformed both NLS and PoissonNMF in all but one test case at reducing proportion indeterminacy (see top of Figure 2.9). For eight of the ten test images, PoissonNMF was least able to cleanly separate the fluorophores. When comparing the worst PI for each method,

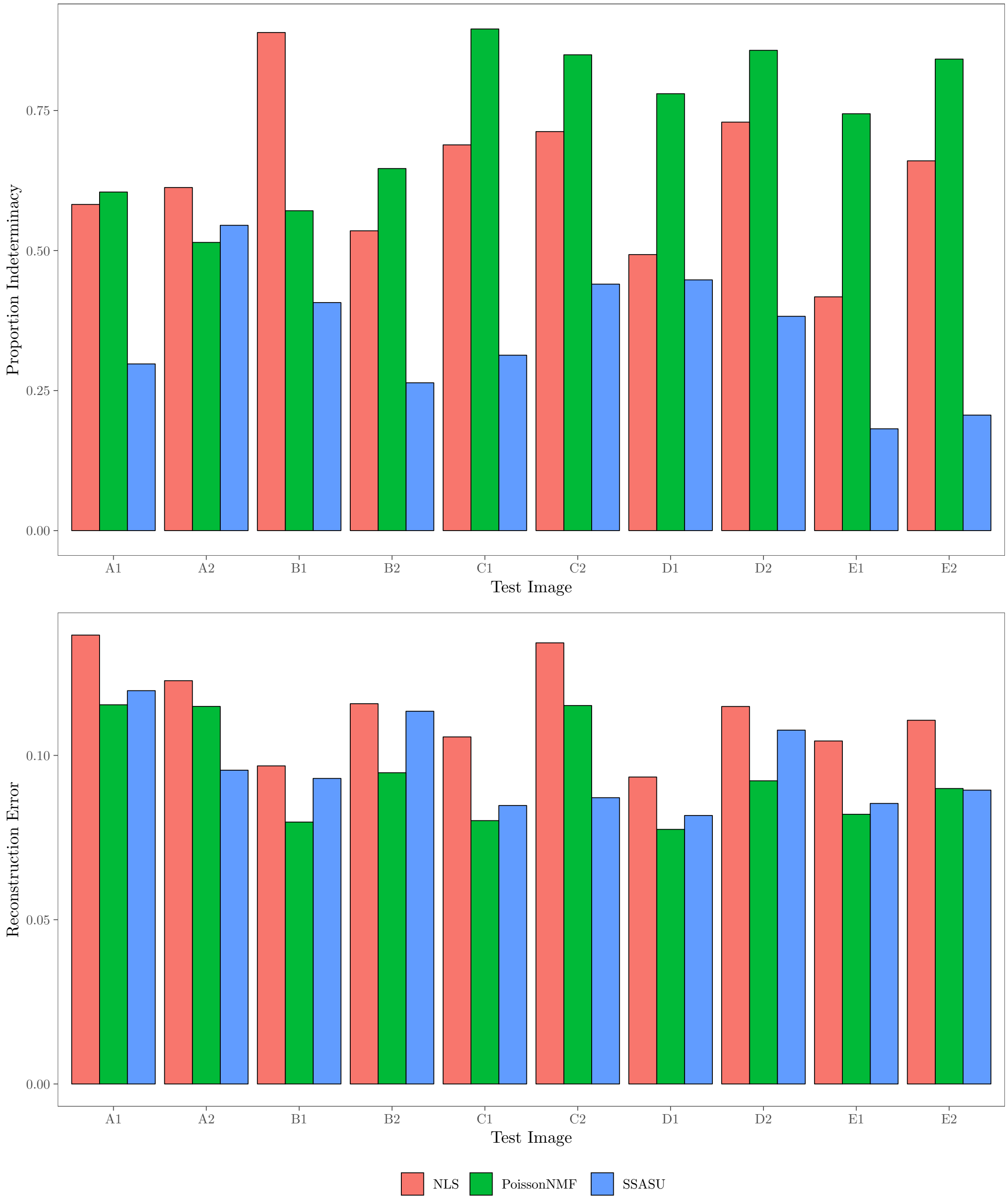


Figure 2.9: Comparison of unmixing performance for SSASU, NLS, and PoissonNMF across ten test images taken from five samples. The proportion indeterminacy (top) measures the non-orthogonality of the weight matrices and illustrates how well each method separates the fluorophore endmembers in the presence of autofluorescence. The relative reconstruction error (bottom) evaluates each method's ability to reconstruct the observed spectra image.

SSASU showed a clear improvement over both NLS and PoissonNMF at 0.55, 0.89, and 0.90, respectively.

The performance illustrated by these metrics can be observed qualitatively in the unmixing results for test image E2 (see Figure 2.10). In the results for NLS (Figure 2.10 A-H), autofluorescence has contaminated nearly all of the unmixed channels. This same autofluorescence light was efficiently captured in the autofluorescence channel of the SSASU unmixed image (Figure 2.10 I). We also note that *Prevotella* (ATTO 520) was not present in image E2, yet the ATTO 520 channel of the NLS unmixed image contained a significant amount of light (Figure 2.10 D). The composite view of all unmixed fluorophore channels (i.e., all channels excluding autofluorescence) clearly illustrates the ability of SSASU to efficiently separate both the autofluorescence and the fluorophores (Figure 2.10 Q & R).

Samples can contain many different types of autofluorescent molecules, and exactly which autofluorescence endmembers exist in an image is difficult to ascertain in advance. For applications using NLS, it is common to estimate the autofluorescence endmember from a no-probe control. Figure 2.11 shows how the autofluorescence endmember estimated from the no-probe control image compared to those learned by SSASU. It is clear that the no-probe control endmember poorly characterized the autofluorescence encountered in the test images. The ability to learn the autofluorescence spectrum directly from an image allows SSASU to adapt to highly varied samples. This flexibility lets SSASU fit the observed data with a sparse set of endmember weight, thereby reducing proportion indeterminacy as compared to NLS.

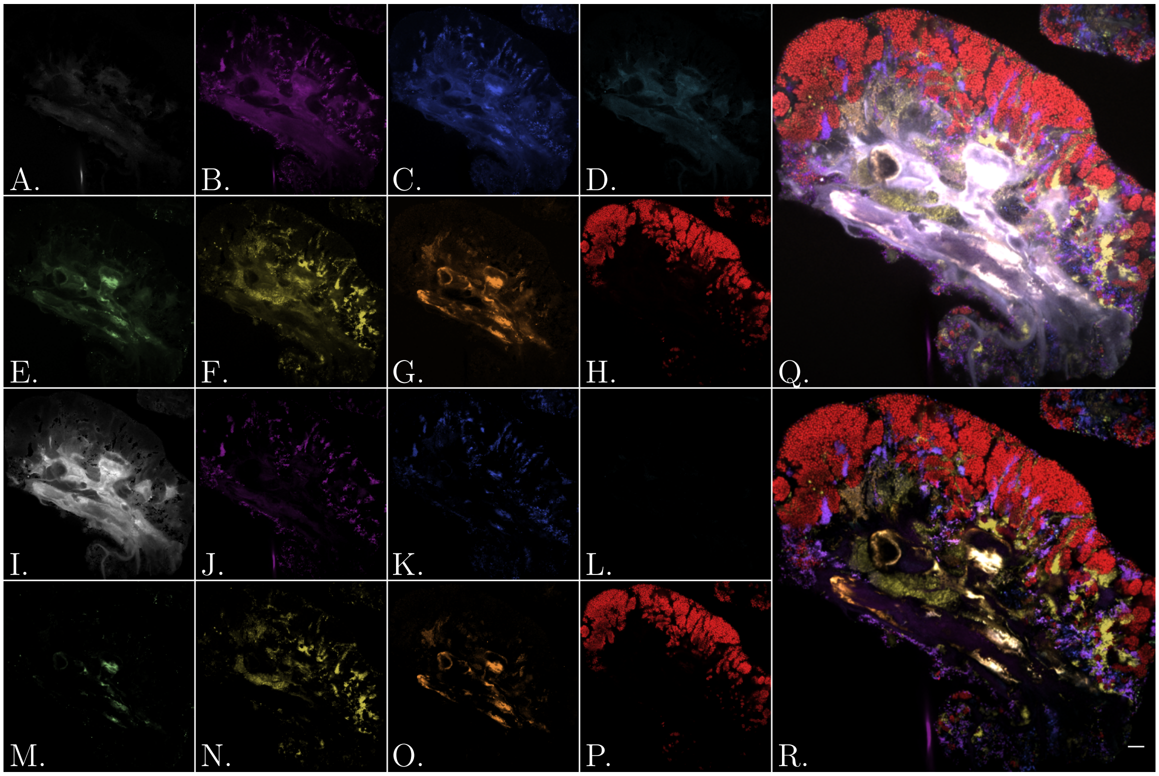


Figure 2.10: Montage of unmixed images for NLS (top half) and SSASU (bottom half). Panels A-P show the unmixed channels for autofluorescence (A, I); *S. mitis*/DY-415 (B, J); *S. salivarius*/DY-490 (C, K); *Prevotella*/ATTO 520 (D, L); *Veillonella*/ATTO 550 (E, M); *Actinomyces*/Texas Red-X (F, N); *Neisseriaceae*/ATTO 620 (G, O); and *Rothia*/ATTO 655 (H, P). A larger composite view of the non-autofluorescence unmixed channels is shown for NLS in panel Q and for SSASU in panel R. The scale bar in panel R indicates 10 μm.

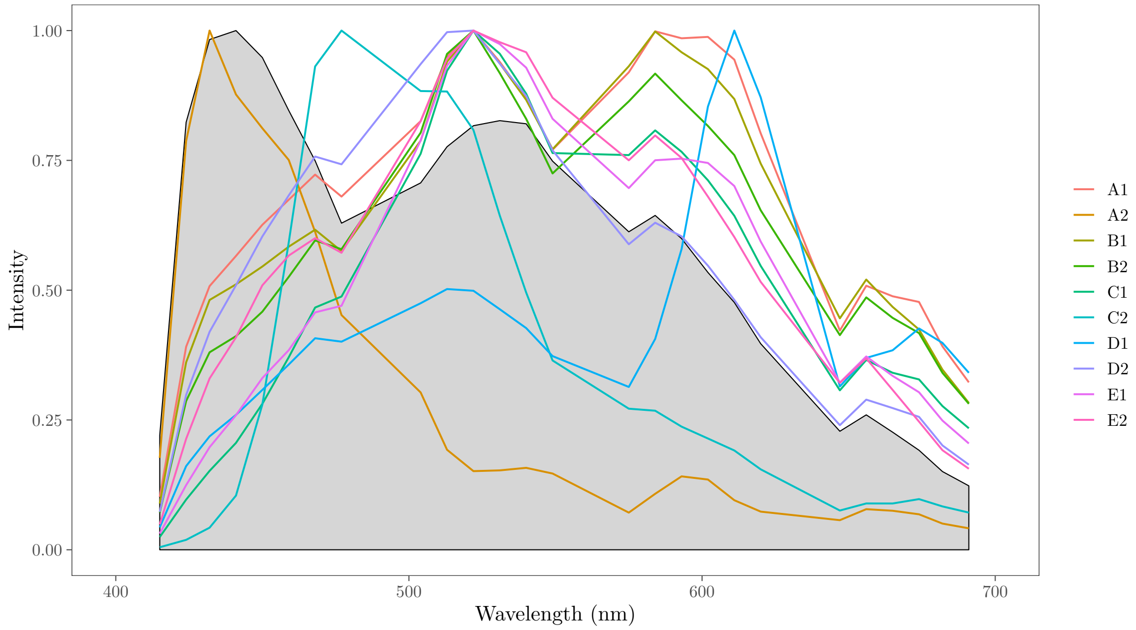


Figure 2.11: Comparison of the autofluorescence endmember estimated from the no-probe control reference image (gray region) to the autofluorescence endmembers learned by SSASU.

2.5 Discussion

Spectral microscopy and unmixing make it possible to visualize biological samples labeled with a large set of fluorophores. However, the choice of unmixing algorithm is important for achieving the desired results. In this chapter, we proposed and evaluated a semi-blind sparse affine spectral unmixing (SSASU) algorithm aimed at separating fluorescence endmembers in the presence of autofluorescence and background fluorescence. In all but one test case, SSASU was able to outperform both NLS and PoissonNMF in mitigating autofluorescence. While our method is more flexible than NLS, we note that, like other NMF methods, SSASU is more computationally expensive and does not guarantee a unique solution. Therefore, we recommend SSASU for situations where spectral micrographs are contaminated with one or many sources of autofluorescence.

In addition, we described an affine NMF method for estimating endmembers from reference images. We showed that affine NMF estimation was as good or better than the mean estimation method across all test cases. In addition, affine NMF does not depend on a thresholding of foreground and background. This makes affine NMF more robust to images with uneven illumination profiles.

There are several clear extensions of this work. First, formulating a version of SSASU for tensors would allow for the unmixing of spectral images that use sequential excitation. Second, allowing minor adjustments to fluorophore endmember spectra on an image-by-image basis would allow the algorithm to accommodate for endmember variability (i.e., changes in the endmember spectra as a result of the micro-environment).

Chapter 3

Volumetric Data

By the end of the nineteenth century, industrialized printing had made scientific textbooks and the illuminating illustrations they contained accessible to the masses. Across the emerging disciplines, scientific illustrations quickly became the standard for detailing experiments, summarizing results, and relaying abstract concepts. Philosopher Bruno Latour highlighted the enduring importance of two-dimensional images in his 1986 essay "Visualization and Cognition: Thinking with Eyes and Hands" when he wrote:

> Scientists start seeing something once they stop looking at nature and look exclusively and obsessively at prints and flat inscriptions. In the debates around perception, what is always forgotten is this simple drift from watching confusing three-dimensional objects, to inspecting two-dimensional images which have been made less confusing. [119]

Latour posits that the prevalence of scientific illustration stems, in part, from the mobility of flat media. Figures on paper can be easily shared, stored, and reproduced; whereas, three-dimensional objects are much more cumbersome. In the digital age, however, virtual reality headsets, 3D printers, and three-dimensional rendering software have greatly simplified the sharing of three-dimensional data. These physical and virtual models have become indispensable tools for gaining a multilevel understanding of specimen. Yet, even during the rise of flat drawings, the importance of three-dimensional models in providing a complete and tangible representation of a subject was recognized by scientists such as embryologist and inventor Wilhelm His Sr.

Embryology, the study of early development, flourished during the last half of the nineteenth century as scientists raced to define a theory of embryogenesis. At the time, scientists would carefully dissect embryos at different developmental stages and illustrate their findings. To facilitate this process, His developed one of the first microtomes in 1870. The *microtome* replaced hand dissection by providing a mechanism for generating repeatable serial sections through a specimen. These serial sections revealed important internal structures that would help explain embryonic development; however, many scientists found it difficult to relate the two-dimensional sections to the embryo's three-dimensional form. Controversial scientist Ernst Haeckel lamented that "the next generation of 'scientific zoologists' will know only cross-sections and stained tissues, but neither whole animals nor their mode of life" [83, 93]. To avoid the problem described by Haeckel, His argued that the only way to simultaneously understand the microscopic internal and macroscopic external features of an embryo was to physically reconstruct wax models of specimen from projections of serial sections. In His' 1870 "Description of a Microtome", he wrote:

> Gaining plastic views through the synthetic combination of sectional images is undoubtedly a long and laborious detour, but it cannot be avoided whenever the object is too fine to reveal its relief to us directly. Anyone who takes the trouble to give body to their views in a malleable material, such as in wax or in clay, will soon discover how important the uninterruptedness of the sections is for such reconstructions of plastic views and

> sectional images.[1] [92, 93]

What His proclaimed in the late nineteenth century has now been shown experimentally. Humans are particularly bad at mentally reconstructing three-dimensional objects from two-dimensional cross-sections. This is even true when the object's structure is known in advanced [64, 240].

Just as an engineer might learn how an unknown system works by taking it apart and putting it back together, His felt that a complete understanding of an embryo could be gained by precisely breaking it down with a microtome and carefully reconstructing the serial sections in a malleable material. This type of slice-based volumetric modeling has become the basis of many areas of modern research. Volumetric data made from serial slices are found across biology [27], medicine [59], geology [107], material science [249], and manufacturing [136].

In this chapter, we will explore slice-based methods for acquiring volumetric data, the limitations and problems these three-dimensional methods present, and algorithms used to reconstruct three–dimensional models from consecutive volume slices. We will start by examining different types of three-dimensional imaging systems and their use in microscopy. Each method has its own set of limitations that impede and restrict our ability to reconstruct the original three-dimensional form. We then outline existing methods for assembling volumetric data from serial images. Finally, we propose and evaluate a new method aimed at reconstructing large-scale histology samples.

---

[1] Original text: "Die Gewinnung plastischer Anschauungen durch synthetische Combination von Durchschnittsbildern ist unstreitig ein weiter und mühsamer Umweg, aber er ist nicht zu umgehen überall da, wo die Objecte zu fein sind, um uns ihr Relief unmittelbar zu enthüllen. Wie wichtig aber für solche Reconstructionen plastischer Anschauungen und Durchschnittsbilder die Lückenlosigkeit der Schnitte seien, das wird jeder bald erfahren, der sich die Mhe nimmt, seinen Anschauungen in einem bildsamen Material in Wachs oder in Thon Körper zu geben" [92].

3.1 Volumetric Imaging

There are many different types of three-dimensional imaging methods, such as spectral (x, y, λ), video (x, y, t), and multimodal (x, y, m) to name a few; however, the term *three-dimensional imaging* is more commonly used in reference to methods that capture information about the x, y, and z spatial dimensions. Of the many imaging modalities, three-dimensional imaging is arguably the most familiar and comprehensible to humans. We, like many predatory animals, rely on binocular vision[2] for stereopsis—the perception of depth. Our brains infer the depth of an object by comparing the views provided by our left and right eyes. In general, objects with a greater horizontal shift, or disparity, between the two views are closer than objects with less horizontal shift.

Binocular vision is one example of a class of three-dimensional imaging called *range imaging*. Some other examples of range imaging techniques include: time-of-flight, structured illumination, lidar, structure from motion, lightfield, and photogrammetry [195]. Each of these range imaging methods create a depth map or point cloud in which every point in a scene is associated with a distance relative to the observer. Range imaging has a variety of uses in medicine, quality control, forensics, game development, art, and cultural heritage [57, 195]. Although range imaging provides three-dimensional information, the resulting depth maps only relates to the surface features of the objects in the scene. These method are unable to measure information about interior details or occluded regions (see Figure 3.1).

Depth maps and point clouds are useful in describing external three-dimensional structures, but they do not allow for inspection of important internal features. To gain a holistic understanding of an entire three-dimensional volume, we must turn to a different class of imaging called *volumetric imaging*. Instead of producing depth maps,

---

[2] Binocular vision, as opposed to monocular vision, refers to the observation of a scene by two sources.

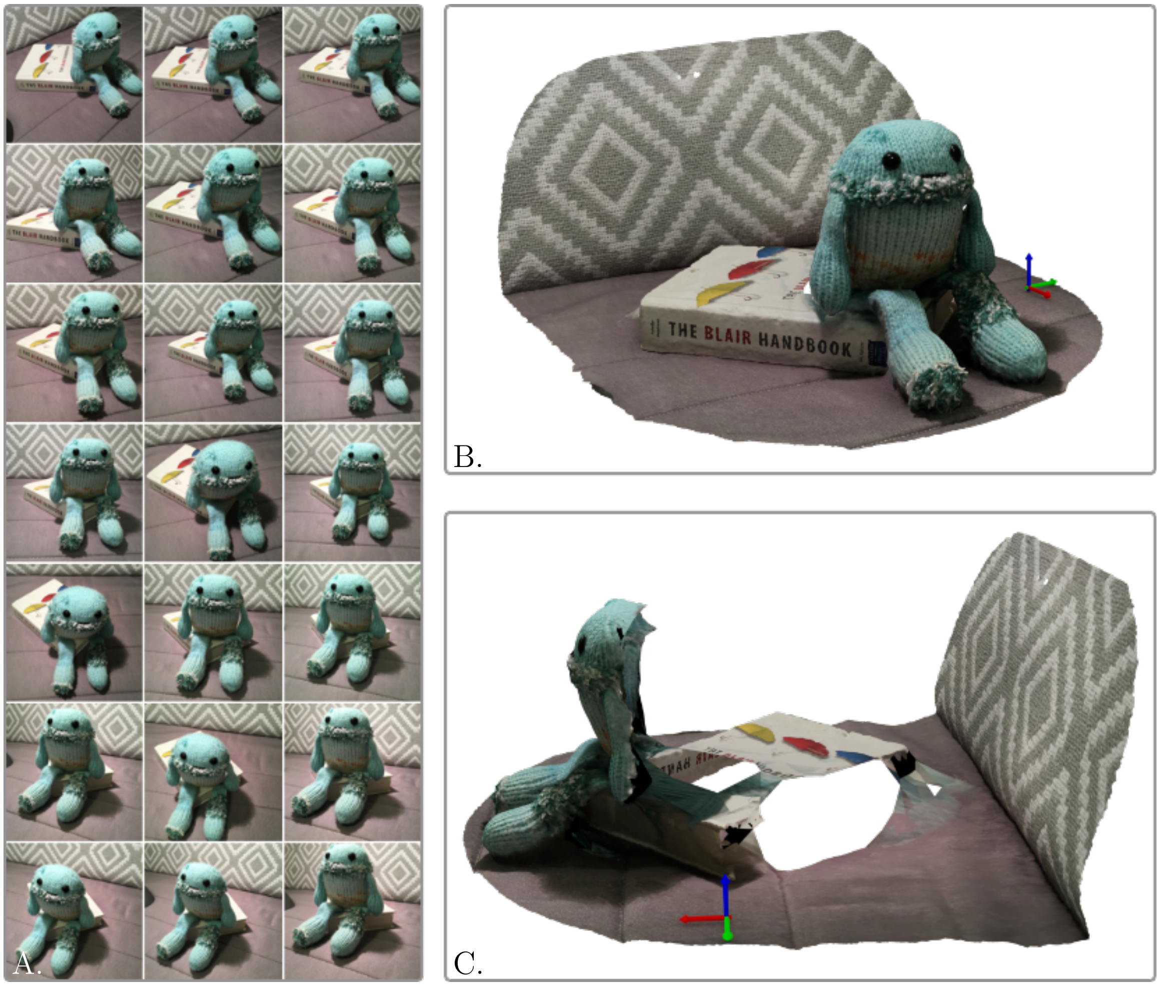


Figure 3.1: Three-dimensional photogrammetry range imaging. A series of two-dimensional images is acquired from different angles around the object of interest (A). A three-dimensional point cloud is generated by determining the spatial correspondence between matching feature points within the images. The color data from the two-dimensional images are then mapped onto the surfaces defined by the point cloud (B). Since range imaging only measures the light reflecting off an object, three-dimensional information cannot be determined for any occluded regions in the camera views. This missing data translates into holes and empty regions in the resulting model (C). The x, y, and z axes are indicated in panels B and C by the red, green, and blue arrows, respectively.

volumetric imaging measures intensity information for every voxel[3] within a volume. If the object of interest is partially transparent to the portion of the electromagentic spectrum used for imaging, then volumetric imaging can be performed *in toto*[4]. Examples of *in toto* volumetric imaging methods include: confocal microscopy, light-sheet microscopy, magnetic resonance imaging (MRI), positron emission tomography (PET), and other computed tomography modalities [12, 104, 215]. These nondestructive volumetric methods work by imaging a series of contiguous two-dimensional virtual slices through the specimen. In the case of confocal microscopy, a specialized aperture blocks any light originating from outside the plane of focus. As such, a three-dimensional volume can be generated by moving the focal plane through the z dimension and acquiring a two-dimensional (x, y) image at each step (see Figure 3.2). These methods have the advantage of imaging the entire three-dimensional volume in-place; however, they require the specimen to be translucent.

In toto volumetric imaging is ineffective when the object is opaque to the electromagnetic wavelengths used for imaging. In these situations, we must return to the methods pioneered by Wilhelm His Sr. and the anatomists of the nineteenth century. His proposed that the entire volume be examined by: (1) sectioning a specimen into sequential thin slices using a microtome, (2) transferring the form of each section onto a workable medium, and (3) reconstructing the three-dimensional volume from the series of media. As described by Born [20], wax was the medium of choice for late nineteenth- and early twentieth-century anatomists. Each serial section was magnified and projected onto a thin wax plate, and the wax was trimmed to match the contours and internal details of each section. Subsequently, the set of trimmed wax plates were fused together to form a complete three-dimensional model. Today, *in sectio*[5] imaging remains an important component of modern research where it is

---

[3] Voxel refers to an element in a volume (i.e., *vo*·lume *el*·ement), and it is the three-dimensional equivalent of a pixel (i.e., *pi*·cture *el*·ement).

[4] *In toto* means as a whole.

[5] *In sectio* means in sections.

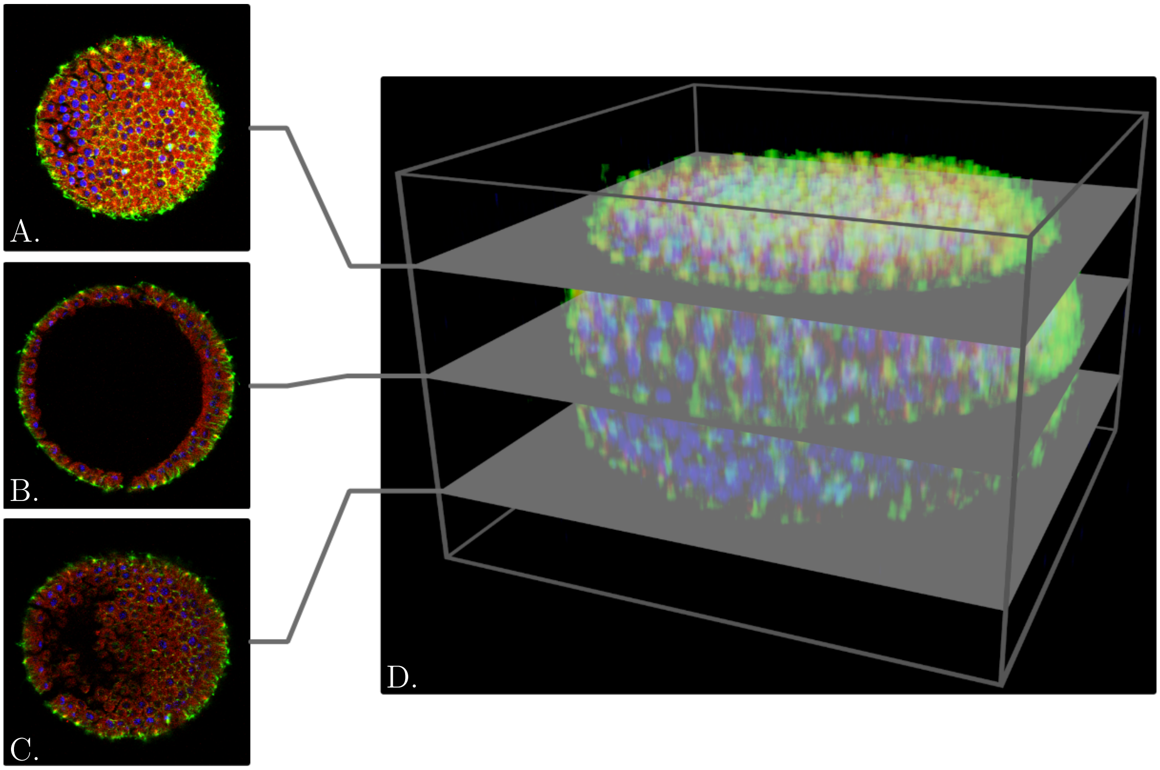


Figure 3.2: Three-dimensional confocal microscopy volumetric imaging. A confocal microscope is used to measure a series of two-dimensional optical sections through a transparent embryo containing fluorescent labels (A-C). The serial sections are reconstructed and rendered as a three-dimensional model of the original embryo (D). Each voxel in the volumetric image (D) contains intensity information related to its specific (x, y, z) location.

most commonly used in medicine and histology [217]. Although we now use digital images instead of wax, the principles remain the same. Serial sections are cut on microtomes, digitally imaged, and virtually reconstructed to recreate the original three-dimensional structure.

As an destructive method, serial sectioning and reconstruction has more points of failure than *in toto* volumetric imaging. Specifically, *in sectio* imaging suffers from artifacts at both the sectioning and imaging steps (see Pichat et al. [173] and references therein). In order to safely section a specimen, it must first be fixed and embedded in a rigid medium (e.g., paraffin wax or optimal cutting temperature compound). Processing tissue for fixation and embedding can cause structures in the tissue to deform or degrade [217]. During cutting, the tissue sections are compressed by the knife and can be torn or fractured. Folds and wrinkles can form when placing the sections onto glass slides, and de-embedding can introduce contaminates such as dust and dirt. Vignetting, out-of-focus regions, and uneven illumination [125] are a few examples of artifacts that can appear during imaging. As part of the reconstruction process, these artifacts must be taken into account and corrected.

In his 1951 paper "Section Compression Photographically Rectified," Osborne Heard noted:

> [I]t appears that section deformation is inherent and inescapable. Therefore when the exact form of the section is critically important, as for instance in the precise registration required in modelling embryos from serial sections, deformation must either be accepted as a necessary evil, or ways and means must be found to correct it. [88]

Heard went on to describe the first optical method for correcting compression artifacts induced by sectioning. The method used a pair of mirrors to apply a transformation to the projected image of a section. By aligning the projections of each pair of sections, it was possible to remove compression artifacts and sequentially reconstruct the entire three-dimensional volume. What is critical to note is that a successful three-dimensional reconstruction relied on the accurate registration of each sequential pair of sections. This workflow remains the basis of modern reconstruction methods for sets of digital images.

3.2 Registration

Image registration is the process of aligning a pair of images to a common coordinate system. The image pair may be multiple views of the same scene (e.g., photogrammetry [213]), the same view at multiple time points (e.g., image-guided surgery [5]), the same view using multiple sensor types (e.g., multimodal image fusion [53]), or multiple views of a related scene (e.g., reconstruction of serial sections [173]). In more formal terms, the objective of image registration is to find a transformation φ that maps a target image $\mathbf{T} \in \mathbb{R}^{Q_y \times Q_x}$ to a reference image $\mathbf{R} \in \mathbb{R}^{P_y \times P_x}$ such that some distance metric is minimized, where $Q = Q_y Q_x$ is the number of pixels in the target image and $P = P_y P_x$ is the number of pixels in the reference image. The transformation $\varphi : \Omega_{\mathbf{R}} \to \Omega_{\mathbf{T}}$ is a mapping of points from the reference image domain $\Omega_{\mathbf{R}} \subset \mathbb{R}^2$ to points in the target image domain $\Omega_{\mathbf{T}} \subset \mathbb{R}^2$.

Registration problems are typically categorized as being either *intensity-based* or *feature-based* depending on the type of distance metric they employ [166, 173, 207, 230, 253]. Intensity-based distance metrics compare the grayscale intensity values of the images or regions within the images; whereas, feature-based metrics evaluate the distance between a set of corresponding image feature points. Since there are many fewer feature points than grayscale values in an image, each iteration of a feature-based method is often much faster to compute than an iteration of an intensity-based method. For reasonably-sized images, however, intensity-based methods tend to have greater accuracy.

3.2.1 Intensity-Based Registration

The reference image $\mathbf{R}$ and target image $\mathbf{T}$ are represented as sets of discrete grayscale values corresponding to grid points along the vertical and horizontal image axes. In order to evaluate the intensity at an off-grid coordinate point, we require an interpolation function (e.g., linear, polynomial, or spline) that models the properties of the underlying image scene. Let $I_{\mathbf{R}} : \Omega_{\mathbf{R}} \to \mathbb{R}$ be an interpolation function that produces a grayscale value for every coordinate point in the reference image domain. Similarly, let $I_{\mathbf{T}} : \Omega_{\mathbf{T}} \to \mathbb{R}$ be an interpolation function for the target image. With these interpolators, we can write the intensity-based registration problem as

$$\min_{\varphi} \int_{\Omega_{\mathbf{R}}} D_I\big(I_{\mathbf{R}}(\mathbf{x}), I_{\mathbf{T}}(\varphi(\mathbf{x}))\big) d\mathbf{x}, \text{ subject to } \varphi \in \mathbb{M}, \tag{3.1}$$

where $\mathbf{x} \in \Omega_{\mathbf{R}}$ is a coordinate point, $D_I : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a distance function, and $\mathbb{M}$ is a set of admissible transformations (which is often restricted to smooth and invertible functions). Registration is an ill-posed problem that is unlikely to have a unique solution in realistic scenarios [60]. As such, it is necessary to restrict the set of feasible transformations to $\mathbb{M}$. Also, note that the transformed target image, denoted $\mathbf{T}_{\varphi} \in \mathbb{R}^{P_y \times P_x}$, can be generated by evaluating $I_{\mathbf{T}}(\varphi(\mathbf{x}))$ for all of the grid points in the reference image. Therefore, Eq. 3.1 can be interpreted as finding the transformation that minimizes the intensity differences between $\mathbf{R}$ and $\mathbf{T}_{\varphi}$.

There are many distance functions that can be used in the intensity-based registration framework, and the appropriate metric largely depends on the types of images being registered. For instance, a direct intensity comparison, such as the squared difference $D_{I,SD}(r,t) = \frac{1}{2}(r-t)^2$, can be used for unimodal images because they have similar structures and intensity profiles. Whereas, indirect intensity gradient comparisons, such as mutual information [175], are best for multimodal images with different intensity profiles. By letting $\mathrm{Pr}_{\mathbf{R}}$ and $\mathrm{Pr}_{\mathbf{T}}$ be the marginal probability distributions for the reference and target image intensities, respectively, and letting $\Pr_{\mathbf{R},\mathbf{T}}$ be the joint probability distribution of combined intensities, the mutual information can be written as

$$D_{I,MI}(r,t) = \Pr_{\mathbf{R},\mathbf{T}}(r,t) \log \frac{\Pr_{\mathbf{R},\mathbf{T}}(r,t)}{\Pr_{\mathbf{R}}(r) \Pr_{\mathbf{T}}(t)}. \tag{3.2}$$

In this context, the joint and marginal probability distributions are the combined and individual histograms of image intensities, respectively. Other popular distance metrics include cross-correlation, normalized cross-correlation, and normalized gradient field [158].

3.2.2 Feature-Based Registration

Instead of using the grayscale intensity values to determine the alignment of two images, feature-based methods use the distance between a set of corresponding feature points. A *feature point*, or landmark, is an automatically or manually selected location of interest (e.g., corner, edge, intersection, curve apex, region, or fiducial marker) that exists in both the reference and target images. Although manually selected feature points are still common, there are now a variety of automatic feature detectors available [132, 194, 219] (see Table 3.1). Automatic feature detectors use a three step approach to finding feature points: (1) candidate feature points are identified in the reference image, (2) candidate feature points are identified in the target image, and (3) corresponding feature points are determined by matching the candidate reference and target feature points. The matching process typically compares candidate feature points based on the properties of their local image regions.

Let's denote the indices for the set of k corresponding feature points as $\mathbb{K} = \{1, \ldots, k\}$ and the two feature point functions as $F_{\mathbf{R}}$ and $F_{\mathbf{T}}$. The functions $F_{\mathbf{R}} : \mathbb{K} \to \Omega_{\mathbf{R}}$ and $F_{\mathbf{T}} : \mathbb{K} \to \Omega_{\mathbf{T}}$ map an index to the corresponding feature point coordinates in the reference and target image domains, respectively. As such, we can

Table 3.1: Common feature detectors

Detector	Feature	Citation
Harris	corner	Harris & Stephens [87]
MinEigen	corner	Shi & Tomasi [201]
Features from Accelerated Segment Test (FAST)	corner	Rosten & Drummond [187]
Binary Robust Invariant Scalable Keypoints (BRISK)	corner	Leutenegger et al. [127]
Scale Invariant Feature Transform (SIFT)	blob	Lowe [141]
Speeded-Up Robust Features (SURF)	blob	Bay et al. [15]
Oriented FAST and Rotated BRIEF (ORB)	blob	Rublee et al. [190]
Accelerated-KAZE (AKAZE)	blob	Alcantarilla et al. [6]
Fast Retina Keypoint (FREAK)	blob	Alahi et al. [4]
Learned Invariant Feature Transform (LIFT)	blob	Yi et al. [245]
Maximally Stable Extremal Regions (MSER)	region	Matas et al. [150]

formulate the feature-based registration problem as

$$\min_{\varphi} \int_{\mathbb{K}} D_F\big(\varphi(F_{\mathbf{R}}(y)), F_{\mathbf{T}}(y)\big)dy, \text{ subject to } \varphi \in \mathbb{M}, \tag{3.3}$$

where $y \in \mathbb{K}$ is a feature point index and D_F is a distance function. A wide variety of distance metrics can be used in Eq. 3.3, but a common choice is the squared Euclidean distance, defined as $D_{F,E}(\mathbf{r}, \mathbf{t}) = \frac{1}{2}\|\mathbf{r} - \mathbf{t}\|_2^2$. For the parametric transformations discussed in the next section, Eq. 3.3 has an analytical solution [11, 226, 203]; however, the existence of incorrect correspondence often requires a numerical solution. In practice, most feature-based registration problems rely on the iterative Random Sample Consensus (RANSAC) method to find the best transformation [61].

3.2.3 Parametric Transformation

We have seen that the registration problem can be formulated using an intensity-based or feature-based framework; however, we have not yet examined the transformation

φ. Regardless of whether we use an intensity-based or feature-based approach, Modersitzki [158] points out that φ can be viewed as being either parametric or nonparametric. In parametric registration, φ is parameterized by the vector $\mathbf{a} = (a_1, \dots, a_d)^T$. Since $\mathbf{a}$ has d elements, the problem is said to have d degrees of freedom. While more degrees of freedom generally allows for a better alignment between the transformed target image and the reference image, it comes at the cost of greater computational complexity.

The affine transform is perhaps the most commonly used transformation in parameterized registration problems. In two-dimensions, affine transforms have $d = 6$ degrees of freedom which allows for *rotation*, *dilation* (i.e., uniform scaling), *transvection* (i.e., shearing), *translation*, or any combination of these actions[6]. The typical form of an affine transformation is

$$\begin{aligned} \varphi(\mathbf{a}, \mathbf{x}) &= \begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix} \\ &= \mathbf{A}\mathbf{x} + \mathbf{b}, \end{aligned} \tag{3.4}$$

where $\mathbf{b}$ performs the translation and $\mathbf{A}$ performs the rotation, dilation, and transvection. Note that all linear transforms and their combinations can be written as simplified versions of the affine transform. Table 3.2 includes a list of a few examples of affine transforms—the most common being the isometry (i.e., rigid) and similarity transforms.

Affine transforms are said to have a *global* property because all coordinate points undergo the same transformation defined by $\mathbf{A}$ and $\mathbf{b}$. However, not all parametric models are global. If more degrees of freedom are required for accurate registration, φ can be formulated such that different transformations are applied to coordinate points from different regions. In this scheme, the function φ is referred to as a deformation

---

[6] Affine transforms also allow for reflection; however, this property is often not necessary or desired in registration problems.

Table 3.2: Common types of affine transformations

	d	$\varphi(\mathbf{a},\mathbf{x})$	Action
Rotation	1	$\begin{bmatrix} \cos(a_1) & \sin(a_1) \\ -\sin(a_1) & \cos(a_1) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$	rotate
Dilation (scaling)	1	$\begin{bmatrix} a_1 & 0 \\ 0 & a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$	scale
Transvection (shearing)	2	$\begin{bmatrix} 1 & a_2 \\ a_1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$	shear
Translation	2	$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$	translate
Isometry (rigid)	3	$\begin{bmatrix} \cos(a_1) & \sin(a_1) \\ -\sin(a_1) & \cos(a_1) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_2 \\ a_3 \end{bmatrix}$	rotate, translate
Similarity	4	$a_1 \begin{bmatrix} \cos(a_2) & \sin(a_2) \\ -\sin(a_2) & \cos(a_2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}$	scale, rotate, translate

and it works by applying a displacement to every coordinate point. An example of a *local* parametric model is the *basis spline*, or B-spline, registration [191]. The concept behind B-spline registration is to locally deform a target image using a mesh of $m \times n$ control points represented by the parameter vector $\mathbf{a}$. Since each control point has a vertical and horizontal component, B-spline deformation has $d = 2mn$ degrees of freedom. The B-spline deformation can be written as

$$\varphi(\mathbf{a},\mathbf{x}) = \mathbf{x} + \sum_{i=1}^{m} \sum_{j=1}^{n} B_{i,j}(\mathbf{x}) \begin{bmatrix} a_{i+m(j-1)} \\ a_{mn+i+m(j-1)} \end{bmatrix}, \tag{3.5}$$

where $B_{i,j}$ denotes the Cox-de Boor recursive spline function (typically of degree four) [40, 45]. The "locality" of the deformation is dependent on the density of the mesh. As more control points are added, smaller deformations are possible.

3.2.4 Nonparametric Transformation

Nonparametric image registration problems, occasionally called nonlinear registration, generally use a transformation of the form

$$\varphi(\mathbf{x}) = \mathbf{x} + u(\mathbf{x}), \tag{3.6}$$

where $u : \Omega_{\mathbf{R}} \to \mathbb{R}^2$ is a displacement function. Rather than optimizing over a set of parameters, the goal of nonparametric schemes is to find a separate displacement vector for every point in $\Omega_{\mathbf{R}}$. Since the nonparametric registration problem has as many degrees of freedom as there are coordinate points, the constraints in Eq. 3.1 and Eq. 3.3 are often replaced by regularizers to ensure a reasonable deformation φ. To avoid deformation artifacts (e.g., holes, folds, kinks, etc.), the nonparametric registration problem typically follows some physically motivated model such as elastic body, viscous fluid flow, diffusion, among others [158, 207]. Even still, great care must be taken when using nonparametric registration as the target image can easily become over-deformed.

3.3 Reconstruction

Just as Heard used manual registration to reconstruct three-dimensional wax models from serial sections, today we use the numerical registration methods discussed in the previous section to reconstruct three-dimensional digital models from a set of digital images (see Ware & LoPresti [234] for a historical perspective on reconstruction). Modern automated microtomes, such as the Automatic Tape Collection Ultra-Microtome (ATUM) invented by Bullen [23], have removed many of the large-scale slice-to-slice inconsistencies; however, registration is still necessary to remove fine-scale cutting, mounting, staining, and imaging artifacts for precise three-dimensional models. The question of how best to extend pairwise registration methods to the reconstruction of serial sections remains an open and active area of research [157, 173].

As we have discussed, registration aligns a target image to a reference image. When extending registration to three-dimensional reconstruction, we must determine which serial section will define the reference coordinate system. An obvious option is to adopt the framework described by Levinthal & Ware [129] whereby the first serial section is oriented as desired and each subsequent section is registered to the previously registered section. In other words, reconstruction is performed by registering all serial sections to the coordinate system defined by the first section. The primary disadvantage of this approach is the propagation and accumulation of registration error. Any errors made in registering the first sections will be carried through to all remaining sections. Pichat et al. [173] points out that the propagation of error can be minimized by registering each section to the center of the stack of serial sections. In this way, any registration errors will only affect half of the serial sections. There also exist automated methods for choosing the optimal reference section from a stack [13]. Alternatively, Pichat et al. [174] proposed a graph-based approach to reduce error propagation by finding the composition of transformations with the least error. Unfortunately, this method requires that each section is registered not just to its direct neighbors but also to second and third neighbors.

The amount of registration error depends largely on the registration framework (intensity-based or feature-based) and the transformation. Although transformations with many degrees of freedom are better able to align two images, they are susceptible to the banana reconstruction problem (see Cooper et al. [37], Malandain et al. [146] and references therein). The *banana reconstruction problem* refers to the loss of displacement information as a result of *in sectio* imaging that can cause a banana-shaped object to be reconstructed into a cylinder. Variations of the banana reconstruction problem have been known to scientists for nearly a century. In 1951, Heard wrote:

> Unless there is an exact record of the dimensions of the block or a precise measurement made of the mounted sections before the paraffin is removed, there is no way to check the amount of compression and therefore no means of knowing the size to which the image should he rectified. [88]

The solution proposed by Heard [88] was to produce a ground truth displacement image for every section by photographing the sample surface before each cut of the microtome. This process allowed Heard to register each serial section to its ground truth and maintain the correct displacement relative to the sample's medial axis. Similar methods acquire ground truth volumes using *in toto* volumetric imaging [218]. Although they do not provide ground truth information, fiducial marks are a common method for avoiding the banana reconstruction problem. Fiducial marks, described as early as 1900 [238], are corresponding points that exist in all serial sections and can be used to reconstruct the three-dimensional volume without shifts along the sample's medial axis. Lastly, Colchester et al. [34] showed that registering the bottom surface of one section to the top surface of the next section provides enough correspondence to avoid reconstruction artifacts. Unfortunately, this method requires twice the amount of images.

3.3.1 Extensions to Gigapixel Volumetric Data

The introduction of high-throughput whole slide imaging (WSI) systems have expanded the use of serial sections in histology and pathology [22, 57, 58, 71, 151, 182]. WSI microscopes are automated platforms for generating high-resolution images of histology sections by stitching together many smaller high-magnification images (e.g., 40–60 ×). A single section can range from tens to hundreds of millimeters in length and width. This means that a WSI system using a 0.25 µm/pixel resolution would generate an image with up to 150 gigapixels. As such, each serial section image can easily occupy tens of gigabytes of storage space. Images from WSI microscopes typically utilize a special multi-resolution file format (e.g., SVS, VMS, NDPI, SCN, MRXS, etc.) in order to handle the large storage size and facilitate loading and viewing.

Three-dimensional reconstruction of gigapixel serial sections is a difficult problem because their size precludes any method that requires an image to be entirely loaded into memory [10, 33, 168, 174, 197, 242]. Although some registration methods have been developed for gigapixel images [38, 140, 182, 193, 199, 206], there are no online methods for dynamically reconstructing subvolumes of interest from serial gigapixel images. To address this, we describe a multiresolution scaling and propagation method for serial image registration.

Our method was inspired by the observation that clinicians are often only interested in small subvolumes of data that exhibit disease hallmarks. When using existing methods, accessing such subvolumes required the prior registration of the entire series of gigapixel images. In addition to being computationally expensive, some methods may fail if the image dimensions are too large. Therefore, our method computes the registration at a manageable resolution during an offline stage and then performs a scaling and propagation to extract subvolumes of interest on-the-fly for any desired resolution. In this way, we minimize the computational burden and focus the analysis on the tissue subvolumes of interest to domain experts.

3.4 Subvolume Registration from Gigapixel Serial Sections

Reconstruction from gigapixel serial sections is an out-of-core problem that is typically approached by performing the registration subproblem in pieces. However, registration at the native gigapixel resolution can be time consuming, and the fusion of piecewise transformations can be complex. For many clinical applications, we note that registration at the native resolution is often not necessary. Furthermore,

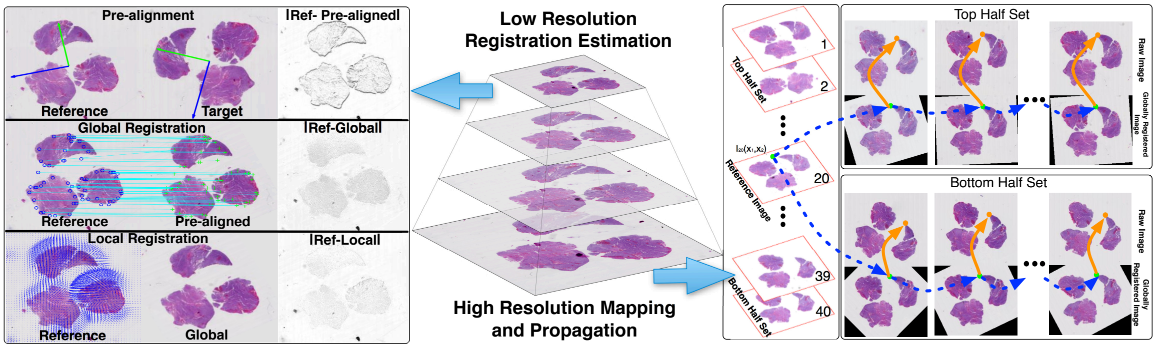


Figure 3.3: An overview of the two-stage subvolume reconstruction method consisting of offline registration at low resolution (left panel) and online composition of scaled transforms (right panel).

clinicians typically only require access to small subvolumes for closer examination. Therefore, we present a two-stage registration method that (1) performs pairwise registration at a low resolution and (2) reconstructs the requested subvolume by first scaling all transformations to the appropriate resolution and propagating reference coordinate points by composition of the pairwise transformations (see Figure 3.3). In this section, we describe the methods used for each stage and evaluate the performance.

3.4.1 Registration

Clinical tissue samples are typically collected, processed, and serially sectioned for diagnostics purposes, and the histological procedures used are optimized for throughput rather than three-dimensional reconstruction. In other words, the orientation and consistency between serial sections is not preserved or recorded as with automated microtomy or blockface imaging. Nevertheless, pathologists and clinicians can greatly benefit from having access to accurately reconstructed subvolumes. In order to properly align such sections, we employ a three-step registration including (1) principal axis alignment, (2) global feature-based registration, and (3) local B-spline registration (see left panel of Figure 3.3). Registration is computed between each sequential pair of grayscale images within the stack. The resolution used for registration is chosen based on the scale of tissue features and the memory limitations of the system. In general, image downsampling ranges from 1/64–1/256 × the native resolution.

Alignment

To correct for the random orientation of tissue sections across sequential images, each image pair is roughly aligned by their principal axes [8]. Conceptually, the principal axes of a tissue section represent the major and minor axes of the ellipse that has been fitted to the tissue contours, and the origin of the axes is the centroid of the tissue contours. The polylines defining the tissue contours are determined by first isolating the primary tissue components with an adaptive global threshold. As histology sections typically have high contrast, simple Otsu thresholding is sufficient for most situations. The p points defining the contour polylines are denoted by the set $\mathbb{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_p\}$, where $\mathbf{c}_i = (c_{i,1}, c_{i,2})^T$ is a single polyline point. We next determine the origin of the section by computing the centroid of the tissue contours, denoted as $\boldsymbol{\mu} = \frac{1}{p}\sum_{i=1}^{p} \mathbf{c}_i$.

The problem of finding the principal axes is equivalent to the problem of fitting an ellipse to the tissue contours $\mathbb{C}$ which can be accomplished by matrix diagonalization of the contour scatter matrix. The scatter matrix, denoted $\mathbf{S} \in \mathbb{R}^{2\times 2}$, is determined by $\mathbf{S} = \sum_{i=1}^{p}(\mathbf{c}_i - \boldsymbol{\mu})(\mathbf{c}_i - \boldsymbol{\mu})^T$. Cholesky factorization of $\mathbf{S}$ gives the eigenvectors $\mathbf{V} = (\mathbf{v}_1, \mathbf{v_2})$, where $\mathbf{v_1}, \mathbf{v_2} \in \mathbb{R}^2$. In this context, the eigenvectors $\mathbf{v_1}$ and $\mathbf{v_2}$ represent the major and minor axes of the fitted ellipse, respectively.

The process described above is performed separately for the reference and target images to give the reference axes $\mathbf{V_R}$, target axes $\mathbf{V_T}$, reference origin $\boldsymbol{\mu}_\mathbf{R}$, and target origin $\boldsymbol{\mu}_\mathbf{T}$. Since the axes may be 180° opposed, the rotation matrix is either $\boldsymbol{\Pi}_+ = \mathbf{V_R}\mathbf{V_T}^T$ or $\boldsymbol{\Pi}_- = -\mathbf{V_R}\mathbf{V_T}^T$. We take the rotation direction associated with the smaller difference between the reference image and the aligned target image defined as

$$\boldsymbol{\Pi}_{\mathbf{P}} = \underset{\boldsymbol{\Pi} \in \{\boldsymbol{\Pi}_+, \boldsymbol{\Pi}_-\}}{\arg\min} \sum_{\Omega_{\mathbf{R}}} \big(I_{\mathbf{R}}(\mathbf{x}) - I_{\mathbf{T}}(\boldsymbol{\Pi}\mathbf{x} + (\boldsymbol{\mu}_{\mathbf{R}} - \boldsymbol{\Pi}\boldsymbol{\mu}_{\mathbf{T}}))\big)^2. \tag{3.7}$$

Recall that $I_{\mathbf{R}} : \Omega_{\mathbf{R}} \to \mathbb{R}$ and $I_{\mathbf{T}} : \Omega_{\mathbf{T}} \to \mathbb{R}$ are interpolation functions for the reference and target images, respectively. Since principal axis alignment consists of a rotation and translation, it can be thought of as a parameterized rigid transformation defined as

$$\begin{aligned} \varphi_{\mathbf{P}}(\boldsymbol{\Pi}_{\mathbf{P}}, \boldsymbol{\mu}_{\mathbf{P}}, \mathbf{x}) &= \boldsymbol{\Pi}_{\mathbf{P}}\mathbf{x} + \boldsymbol{\tau}_{\mathbf{P}} \\ &= \begin{bmatrix} \boldsymbol{\Pi}_{\mathbf{P}} & \boldsymbol{\tau}_{\mathbf{P}} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \\ &= \mathbf{P}\hat{\mathbf{x}} \end{aligned} \tag{3.8}$$

where $\boldsymbol{\tau}_{\mathbf{P}} = \boldsymbol{\mu}_{\mathbf{R}} - \boldsymbol{\Pi}_{\mathbf{P}}\boldsymbol{\mu}_{\mathbf{T}}$ and $\mathbf{0}$ is a vector of zeros. The resulting principal axis transformed target image $\mathbf{T}_{\varphi_{\mathbf{P}}}$ is generated by evaluating $I_{\mathbf{T}}(\mathbf{P}\hat{\mathbf{x}})$ for all grid points in $\Omega_{\mathbf{R}}$ (see Figure 3.4). An important limitation of principal axis alignment is that it fails when contours have rotational symmetry. Since the fitted ellipse of a rotationally symmetric object is a circle, the major and minor axes cannot be uniquely determined.

Global Registration

The principal axis rigid transformation roughly aligns each serial section to a defined reference section. However, an additional global registration is required to fine tune the rotation and translation parameters as well as correct for scaling artifacts resulting from slight magnification differences. In general, adding a degree of freedom for scaling is not advised for most tissue samples because it can result in the sample being reconstructed into a cylinder. However, sections containing multiple tissue pieces are less susceptible to this problem because any scaling also increases the relative distance between each tissue piece.

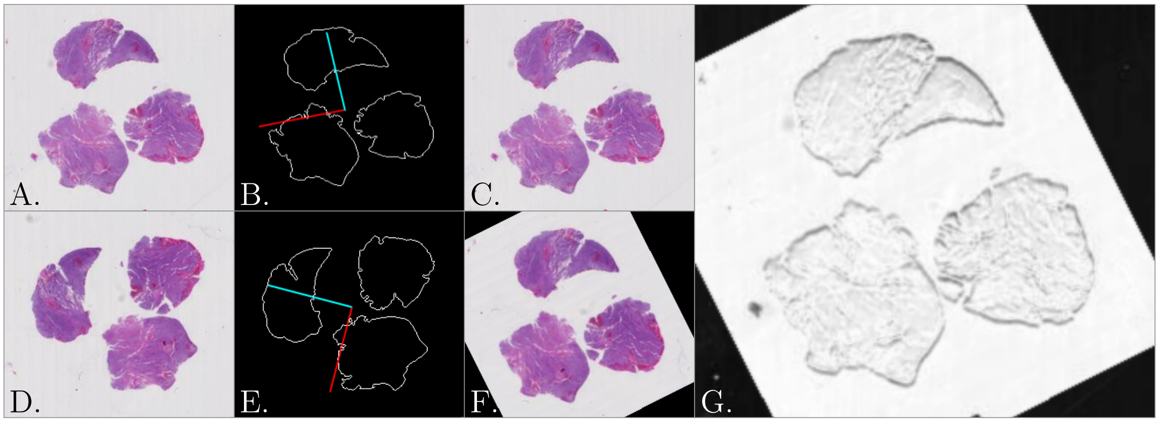


Figure 3.4: Principal axis alignment of a reference image (A-C) and a target image (D-F) using tissue contours. The reference (A) and target (D) images are randomly oriented as a result of sectioning. The origins and principal axes of the tissue contours are determined for the reference image (B) and target image (E). Aligning the origins and principal axis of the target image to the reference image (C) produces the aligned target image (F). The quality of principal axis alignment (G) is shown as the absolute difference between the reference (C) and aligned target (F) images.

For global similarity transformation, we detect prominent tissue components in the aligned serial sections using a features detector. Since tissue samples do not contain hard lines or large contiguous regions, the SURF blob detector was chosen over the alternative corner or region detectors. The SURF detector uses three filter octaves and four octave layers; however, the filter sizes should be adjusted to fit the feature sizes at the downsampled resolution. Next, correspondence between reference and target landmarks are determined by matching local descriptors. With the corresponding features, we determine the global similarity transform using a generalized RANSAC method called Maximum Likelihood Estimation Sample Consensus (MLESAC) [221]. The global transform is written as

$$\begin{aligned}\varphi_{\mathbf{G}}(s, \mathbf{\Pi}_{\mathbf{G}}, \boldsymbol{\tau}_{\mathbf{G}}, \mathbf{x}) &= s\mathbf{\Pi}_{\mathbf{G}}\mathbf{x} + \boldsymbol{\tau}_{\mathbf{G}} \\ &= \begin{bmatrix} s\mathbf{\Pi}_{\mathbf{G}} & \boldsymbol{\tau}_{\mathbf{G}} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \\ &= \mathbf{G}\hat{\mathbf{x}},\end{aligned} \tag{3.9}$$

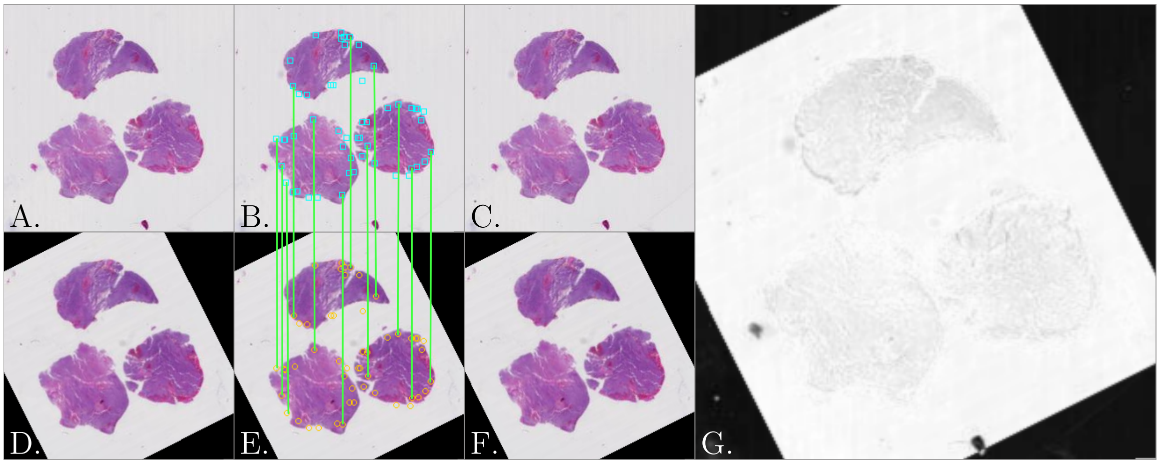


Figure 3.5: Global registration of a reference image (A-C) and a aligned target image (D-F) using SURF features and a similarity transform. The target image (D) has been roughly aligned to the reference image (A). Corresponding SURF features are identified between the reference (B) and target (E) images. MLESAC is used to find the similarity transform that registers the target image (F) to the reference image (C). The quality of global registration (G) is shown as the absolute difference between the reference (C) and transformed target (F) images.

where s, $\mathbf{\Pi_G}$, and $\boldsymbol{\tau}_\mathbf{G}$ are the global scaling parameter, rotation matrix, and displacement vector, respectively. The pairwise global transforms are used to register each aligned section to a given reference image (see Figure 3.5).

Local Deformation

Global registration is followed by a nonrigid registration to compensate for local tissue deformations. The local deformation is computed using the B-spline transform described in Eq 3.5 with a 32×32 grid of control points. The optimal local deformation is found by simultaneously maximizing the normalized mutual information and minimizing transformation energy (see Figure 3.6) [14, 191]. We denote the local deformation as the displacement function $\varphi_L : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ that maps a globally registered reference coordinate to a local deformed target coordinate. Care must be taken when using a free form deformation because over-deformation artifacts can arise. However, over-deformation is expected to be minimal at downsampled resolutions.

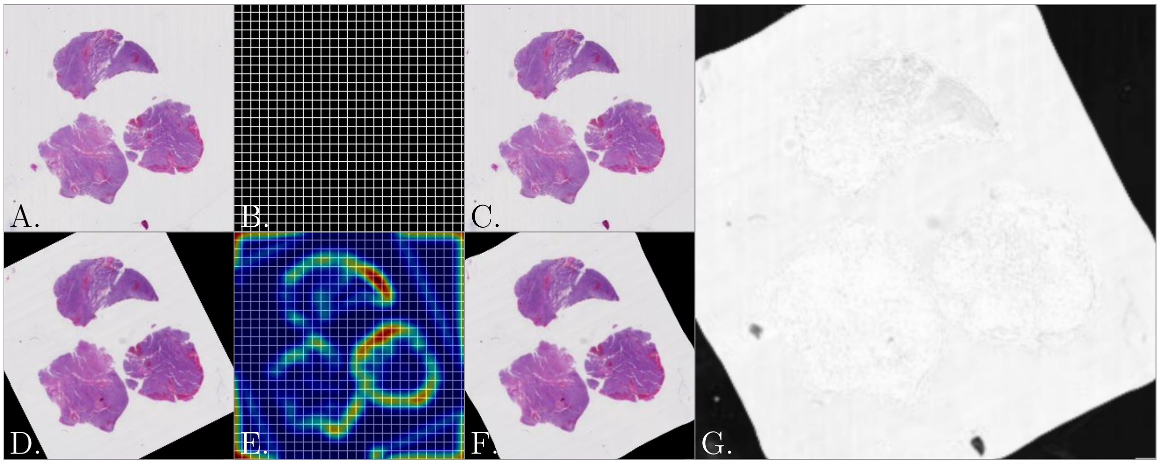


Figure 3.6: Local registration of a reference image (A-C) and a globally registered target image (D-F) using a B-spline transform. The target image (D) has been globally registered to the reference image (A). A B-spline mesh (B) is used to determine the deformation between the reference image and the target image (E). The B-spline deformation is used to register the target image (F) to the reference image (C). The quality of local registration (G) is shown as the absolute difference between the reference (C) and deformed target (F) images.

3.4.2 Reconstruction by Composition of Scaled Transforms

The offline three-step registration process generates three sets of pairwise registrations. Since each subsequent registration step was performed on the result of the previous step, the composition of transforms necessary for final reconstruction is not obvious. In order to extract registered subvolumes of interest, we must first scale each pairwise transformation to the appropriate resolution. To this point, the transformations were determined based on a defined grid size. Higher resolution images are produced by using points on a finer grid.

To simplify the problem, we combine the rigid transformation from alignment with the similarity transformation from feature-based registration. Since both steps are formulated as affine transformations, we can denote the combined global transformation as

$$\begin{aligned}\varphi_{\mathbf{G},\mathbf{P}}(\mathbf{G},\mathbf{P},\mathbf{x}) &= \begin{bmatrix} s\boldsymbol{\Pi}_{\mathbf{G}} & \boldsymbol{\tau}_{\mathbf{G}} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Pi}_{\mathbf{P}} & \boldsymbol{\tau}_{\mathbf{P}} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \\ &= \mathbf{GP}\hat{\mathbf{x}}. \end{aligned} \tag{3.10}$$

Since the displacement terms of $\mathbf{G}$ and $\mathbf{P}$ are not scale-invariant, they must be scaled appropriately for every desired resolution. Let α denote the magnification factor between resolutions. We can construct the scaled combined global transformation and its inverse as

$$\begin{aligned}\bar{\mathbf{G}} &= \begin{bmatrix} s\boldsymbol{\Pi}_{\mathbf{G},\mathbf{P}} & \alpha\boldsymbol{\tau}_{\mathbf{G},\mathbf{P}} \\ \mathbf{0} & 1 \end{bmatrix} \\ \bar{\mathbf{G}}^{-1} &= \begin{bmatrix} s\boldsymbol{\Pi}_{\mathbf{G},\mathbf{P}}^T & -\alpha s\boldsymbol{\Pi}_{\mathbf{G},\mathbf{P}}^T\boldsymbol{\tau}_{\mathbf{G},\mathbf{P}} \\ \mathbf{0} & 1 \end{bmatrix}\end{aligned} \tag{3.11}$$

where $\boldsymbol{\Pi}_{\mathbf{G},\mathbf{P}} = \boldsymbol{\Pi}_{\mathbf{G}}\boldsymbol{\Pi}_{\mathbf{P}}$ and $\boldsymbol{\tau}_{\mathbf{G},\mathbf{P}} = \boldsymbol{\tau}_{\mathbf{G}} + \boldsymbol{\tau}_{\mathbf{P}}$. Note that when $\alpha = 1$, $\bar{\mathbf{G}} = \mathbf{GP}$. For convenience, we denote the scalable global transformation as $\varphi_{\bar{\mathbf{G}}}(\alpha,\mathbf{G},\mathbf{P},\mathbf{x}) = \bar{\mathbf{G}}\hat{\mathbf{x}}$, where $\hat{\mathbf{x}} = (x_1, x_2, 1)^T$.

With the scaled transformations defined, we now turn to the problem of mapping points in the reference domain to points in any target domain using a cascade of pairwise transformations. To facilitate discussion, let us denote the i^{th} original image, globally registered image, and locally deformed image as $\mathbf{I}_i$, $\bar{\mathbf{I}}_i$, and $\tilde{\mathbf{I}}_i$, respectively. Assuming that there are N serial images in the imaging volume, we set image i^* as the reference image for the whole volume, where $i^* = \left\lfloor \frac{N}{2} \right\rfloor$. We can therefore denote $\varphi_{\bar{\mathbf{G}},i}$ and $\varphi_{\bar{\mathbf{G}},i}^{-1}$ as the scalable global transformations from $\mathbf{I}_i$ to $\mathbf{I}_{i+1}$ and from $\mathbf{I}_{i+1}$ to $\mathbf{I}_i$, respectively. Similarly, let $\varphi_{L,i}$ and $\varphi_{L,i}^{-1}$ be the nonrigid deformation from $\bar{\mathbf{I}}_i$ to $\bar{\mathbf{I}}_{i+1}$ and from $\bar{\mathbf{I}}_{i+1}$ to $\bar{\mathbf{I}}_i$, respectively.

As the nonrigid registration at the low resolution is applied to the globally registered image sequence and only local alignment transformations for all adjacent image pairs are available, the method for composing the pairwise transformations is as follows:

Step 1: As globally registered images $\bar{\mathbf{I}}_i$ have the same dimensions as the reference image $\mathbf{I}_{i^*}$ where $i < i^*$, we first map an arbitrary coordinate $(x_1^{i^*}, x_2^{i^*}) \in \Omega_{\mathbf{I}_{i^*}}$ to globally registered image $\bar{\mathbf{I}}_{i^*-1}$ space by nonrigid mapping φ_{L,i^*-1}^{-1} (see blue arrows in Figure 3.3).

Step 2: The coordinate $(x_1^{i^*-1}, x_2^{i^*-1}) \in \Omega_{\bar{\mathbf{I}}_{i^*-1}}$ can be mapped to $\Omega_{\mathbf{I}_{i^*-1}}$ using the global transformation $\varphi_{\bar{\mathbf{G}},i^*-1}^{-1}$ (see orange arrows in Figure 3.3).

Step 3: Repeat Step 1 and 2 to propagate φ_L^{-1} and $\varphi_{\bar{\mathbf{G}}}^{-1}$ for aggregated transformation from $\mathbf{I}_{i^*}$ to any arbitrary image in the volume. A similar procedure with the noninverted transformations is applied when $i > i^*$.

With the above method for composing transformations, the online extraction of registered subvolumes requires that we perform the above steps only for the subvolume coordinates. The algorithmic description of this registration propagation is presented in Algorithm 3. The algorithm for subvolume reconstruction was written in MATLAB and uses the OpenSlide [77] library to handle and access gigapixel image data. A version of this method is currently deployed on the *Digital Pathology* web portal for use in reconstructing subvolumes for pathology research [112].

3.4.3 Evaluation

Evaluating the performance and robustness of large-scale registration problems remains a challenge. Rohlfing [183] has shown that tissue overlap and similarity scores are not always reliable; however, ground truth data for serial gigapixel images in not readily available. Furthermore, the type of corresponding *in toto* volumetric ground truth images suggested by Pichat et al. [173] are often not possible for small tissue samples. Therefore, we evaluated our method by qualitatively inspecting the reconstructed subvolumes and by quantifying the propagation of error resulting from

Algorithm 3: Subvolume Reconstruction by Composition of Scaled Transforms

```
Input  : original serial images $\{\mathbf{I}_i, i = 1, \ldots, N\}$
Output: registered serial images $\{\tilde{\mathbf{I}}_i, i = 1, \ldots, N\}$
Set reference image index $i^* = \left\lfloor \frac{N}{2} \right\rfloor$;
Compute $\varphi_{\bar{\mathbf{G}}}$ and $\varphi_L$ by alignment, global registration, and local deformation;
for $i \in (i^* - 1, i^* - 2, \ldots, 1)$ do
    $j \leftarrow i^* + 1$;
    for $(x_1, x_2) \in \mathrm{ROI} \subseteq \Omega_{\mathbf{I_{i^*}}}$ do
        while $j \geq i$ do
            Apply $\varphi_{L,j}^{-1}$;
            $j \leftarrow j - 1$;
        end
        while $j \geq i$ do
            Apply $\varphi_{\bar{\mathbf{G}},j}^{-1}$;
            $j \leftarrow j - 1$;
        end
    end
end
for $i \in (i^* + 1, i^* + 2, \ldots, N)$ do
    $j \leftarrow i^* + 1$;
    for $(x_1, x_2) \in \mathrm{ROI} \subseteq \Omega_{\mathbf{I_{i^*}}}$ do
        while $j \geq i$ do
            Apply $\varphi_{L,j}$;
            $j \leftarrow j + 1$;
        end
        while $j \leq i$ do
            Apply $\varphi_{\bar{\mathbf{G}},j}$;
            $j \leftarrow j + 1$;
        end
    end
end
```

scaling and composing transformations generated at non-native resolutions.

We first tested our method for subvolume reconstruction on a set of forty serial hematoxylin and eosin stained tissue sections from a glioblastoma (GBM) biopsy with an overall image volume of $12.930 \times 15.279 \times 0.200\,\mathrm{mm}$. Each GBM serial section was cut to $5\,\mu\mathrm{m}$ thick and digitized at $0.2265\,\mu\mathrm{m/pixel}$ resolution, resulting in a 150 gigavoxel volume. From the entire image volume, we extracted a registered $928 \times 928 \times 200\,\mu\mathrm{m}$, or 670 megavoxel, subvolume by performing alignment, global registration, and local deformation at a $1/256\times$ downsampled resolution and mapping to the native gigapixel resolution.

First, registration accuracy was examined visually for tissue edge discontinuity in distinct cross-sections through the low resolution image volume in Figure 3.7. Each cross-section plane was uniquely colored in the three-dimensional view of the registered volume. Cross-section locations were randomly selected and colored to match the $y \times z$ cross-section views. Scaling in the z dimension was increased to aid visual inspection of contour discontinuities. We observed the progressive improvement in registration for rigid alignment, similarity transformation, and B-spline deformation. The subvolume montage view of a small tissue region of interest in Figure 3.8 showed a high correspondence across serial sections, suggesting the efficacy of our proposed method.

We quantitatively evaluated the propagation of error as a result of transformation scaling and composition by computing the registration accuracy for ten synthetic image volumes. For each such synthetic ground truth volume, we replicated a single section from the GBM data set fifty times at $1/256\times$, $1/128\times$, and $1/64\times$ downsampling. We intentionally produced random global transformations and local deformations to the $1/256\times$ synthetic volumes to simulate distortions introduced by tissue processing. Global transformations were constructed using sheer, translation, and rotation values drawn from $\mathcal{N}(0, 0.001)$, $\mathcal{N}(0, \rho/3)$, where ρ is the smallest im-

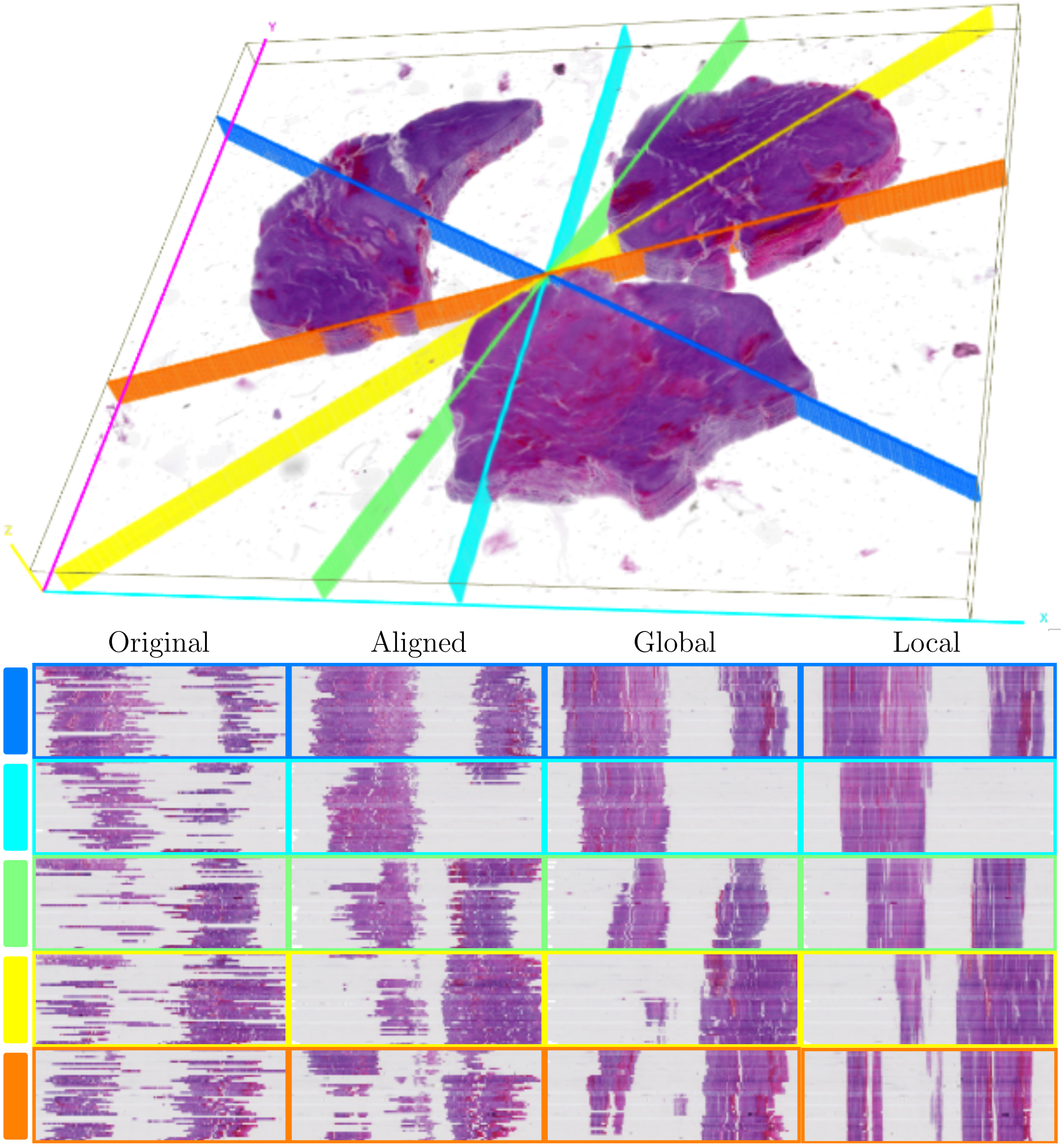


Figure 3.7: Qualitative evaluation of registration accuracy using random cross-sections. Each row in the bottom panel displays the $y \times z$ cross-sections at each registration stage for the corresponding cutting plane in the three-dimensional rendered volume at the top.

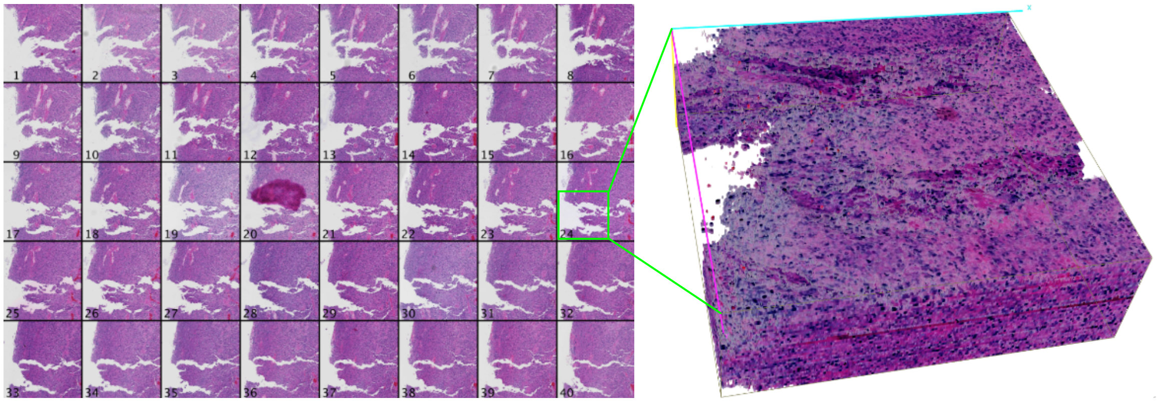


Figure 3.8: Results of high resolution mapping and propagation for a 4096 × 4096 pixel region of serial GBM sections. A montage of the registered region is presented on the left, and a three-dimensional rendered volume corresponding to the green box is shown on the right.

age dimension, and $\mathcal{U}(0, 360)$, respectively. We extrapolated a 4×4 matrix, subject to $\mathcal{N}(0, 4)$, to appropriate image dimensions for generating local deformations. The global and local deformations were scaled to produce identical distortions for the $1/128\times$ and $1/64\times$ ground truth volumes.

Registration was performed for each distorted synthetic volume at $1/256\times$ downsampling, and the average registration accuracy was computed for alignment, global, and local registration. The registration accuracy for each synthetic volume was measured by the background overlap between the synthetic ground truth and deformed volume (see Wang et al. [232] for additional detail). The average registration accuracies at $1/256\times$ resolution described the distortion correction power at each sequential registration step. To explore the error in multi-resolution mapping, we scaled the deformations at $1/256\times$ resolution and applied them to the $1/128\times$ and $1/64\times$ resolutions. The registration accuracy for each step at each resolution is presented in Table 3.3. We noticed that sequential alignment, global, and local registration can correct most simulated distortions, and there was no substantial increase in error introduced by scaling between resolutions. The observed degradation from aligned to global registration could have resulted from the extra degree of freedom for scaling

Table 3.3: Registration accuracy for each processing step at three different levels: Mean±Std ($n = 50$).

Downsampling	Aligned	Global	Local
$1/256\times$	0.938 ± 0.005	0.903 ± 0.020	0.989 ± 0.001
$1/128\times$	0.934 ± 0.005	0.900 ± 0.019	0.974 ± 0.001
$1/64\times$	0.931 ± 0.005	0.897 ± 0.018	0.965 ± 0.001

in the similarity transform estimation, given our synthetic data is not substantially scaled by simulated distortions.

3.5 Discussion

Volumetric data is becoming an integral component of many research disciplines. As clinicians and pathologists turn to digital WSI technology, there is a growing need for algorithms that can process large amounts of related image data. In this chapter, we proposed a two stage registration method that performs offline registration at a downsampled resolution and reconstructs subvolumes of interest online at any desired resolution. Both qualitative and quantitative evaluation results demonstrate the promise of our registration method for reconstruction from serial gigapixel whole slide histology images. We note that our method presents several limitations. Namely, registration accuracy will always be limited to the resolution of the downsampled images, and repeated requests for the same subvolume requires recomputing the same composition of transformation. For practical purposes, we do not believe that these limitations will greatly impose on clinical and research use via the *Digital Pathology* web portal.

In the future, we plan to extend this work with iterative registration corrections across multiple resolutions to further improve concordance across serial sections. Moreover, we intend to incorporate human-annotated landmarks to extend the quantitative assessment.

Chapter 4

Temporal Data

Sociologist and science historian Hannah Landecker describes the use of visual media in the sciences as "part of the exploration of the role of time in experimentation" [116]. Drawings, photographs, and models have allowed scientists to freeze particular moments during their experiments, but it was not until the twentieth century that scientists were truly able to manipulate time to gain new insights into physical phenomena. The introduction of film and video into the laboratory was heralded by three Victorian-era scientists: physiologist Étienne-Jules Marey, astronomer Jules Janssen, and photographer Eadweard Muybridge [80]. In 1874, Janssen developed the *photographic revolver*—a device that was able to capture nearly fifty images in one minute. Soon after, Marey and Muybridge leveraged Janssen's technology to great effect in the study of motion. Marey referred to the technique of capturing sequential photographs over time as *choronophotography.* Chronophotographic images were initially displayed as arranged sets of photographs, but technical improvements later allowed these sequential views to be overlayed in a single image. For the first time, scientists had a mechanism to not only freeze a moment, but to see how motion progressed in time.

As acquisition and projection methods continued to develop, chronophotography gave way to cinematography. Although much attention has been given to cinematography from the perspective of entertainment and the arts, film played an equally important role in shaping the practices and communication of science (see works by Canales [25], Gaycken [69], Gouyon [80], Landecker [117], Stramer & Dunn [212]). Some of the earliest uses of film were to record and enhance the observation of microscopic specimen. The field of microcinematography, or photomicrography, was pioneered by the French doctor Jean Comandon. As Comandon wrote in his 1932 article "Cinematography and the Science of Nature":

> Only *microcinematography* is able to preserve the traces of phenomena that occur in the preparation. Like the retina of an eye which never tires, the film follows, for an extended time, all the changes that take place; moreover, the cinematograph is, like the microscope itself, an instrument of research, while one acts on the visual space, the other acts on time, *condensing* or *spreading* movements by *accelerating or slowing them*; it thus reduces their speed to a magnitude that is more easily perceptible, which even reveals to us that which we do not suspect.[1][36]

Comandon viewed film as an extension of the eye that could manipulate the scale of time in order to expose the secret motions of cells. Throughout the twentieth century and into the twenty-first, film and video played an integral role across many scientific disciplines. Video and temporal imaging is used today to communicate protocols [170], observe cell processes [98], visualize ecosystems [161], measure climate change [3], and track animal behavior [235]. With the decreasing cost of digital video hardware and increased commercial use of video systems, analytical methods for

---

[1] Original text: "La microcinématographie est seule capable de conserver des traces des phénoménes qui se passent dans la préparation. Comme la rétine d'un œil jamais fatigué, le film suit, pendant un temps prolongé, tous les changements qui s'opérent; bien plus, le cinématographe est, ainsi, que le microscope lui-même, un instrument de recherches, tandis que celui-ci agit sur l'espace visuel, celui-là agit sur le temps, en y condensant ou en y étalant les mouvements par l'accéléré ou le ralenti; il ramène ainsi leur vitesse à une grandeur qui les rend plus facilement perceptibles, qui, même, les révéle là où nous ne les soupçonnions pas." [36].

processing large amounts of video data are becoming vital for modern science.

This chapter explores the use of temporal data in scientific research with an emphasis on large data sets. We begin by discussing the different modes of video and temporal imaging from chronophotography to multi-location imaging. We then examine common approaches to video analysis including object detection, recognition, and tracking. Lastly, we present and evaluate a method for tracking insect movement across multiple locations and show how this technique can be used to monitor the behavior of honey bees.

4.1 Temporal Imaging

Chronophotography, although not a form of cinematography or videography, was the first attempt to capture and study movements over time. Early chronophotographs consisted of several photos arranged to show the progression of movement. In later years, the development of the stroboscopic flash allowed scientists to capture motion by taking multiple exposures on the same piece of film. The superposition of multiple poses in a single image made it possible to precisely measure the displacement caused by some action. Although chronophotography refers to all photographic methods that allow the study of motion, modern techniques are generally categorized by whether they compress or expand time (see Figure 4.1). *Time-lapse photography* compresses time by capturing photos at a slow frequency; whereas, *high-speed photography* expands time by capturing photos in rapid succession.

The progression from static chronophotographs to dynamic films grew from work in animation and projection. It was quickly realized that chronophotographs taken by Myubridge and his contemporaries could be made into zoetropes that would recapitulate the original motion of an object. A *zoetrope* is a device that gives the illusion of motion by quickly displaying an array of images one after the other. In 1879,

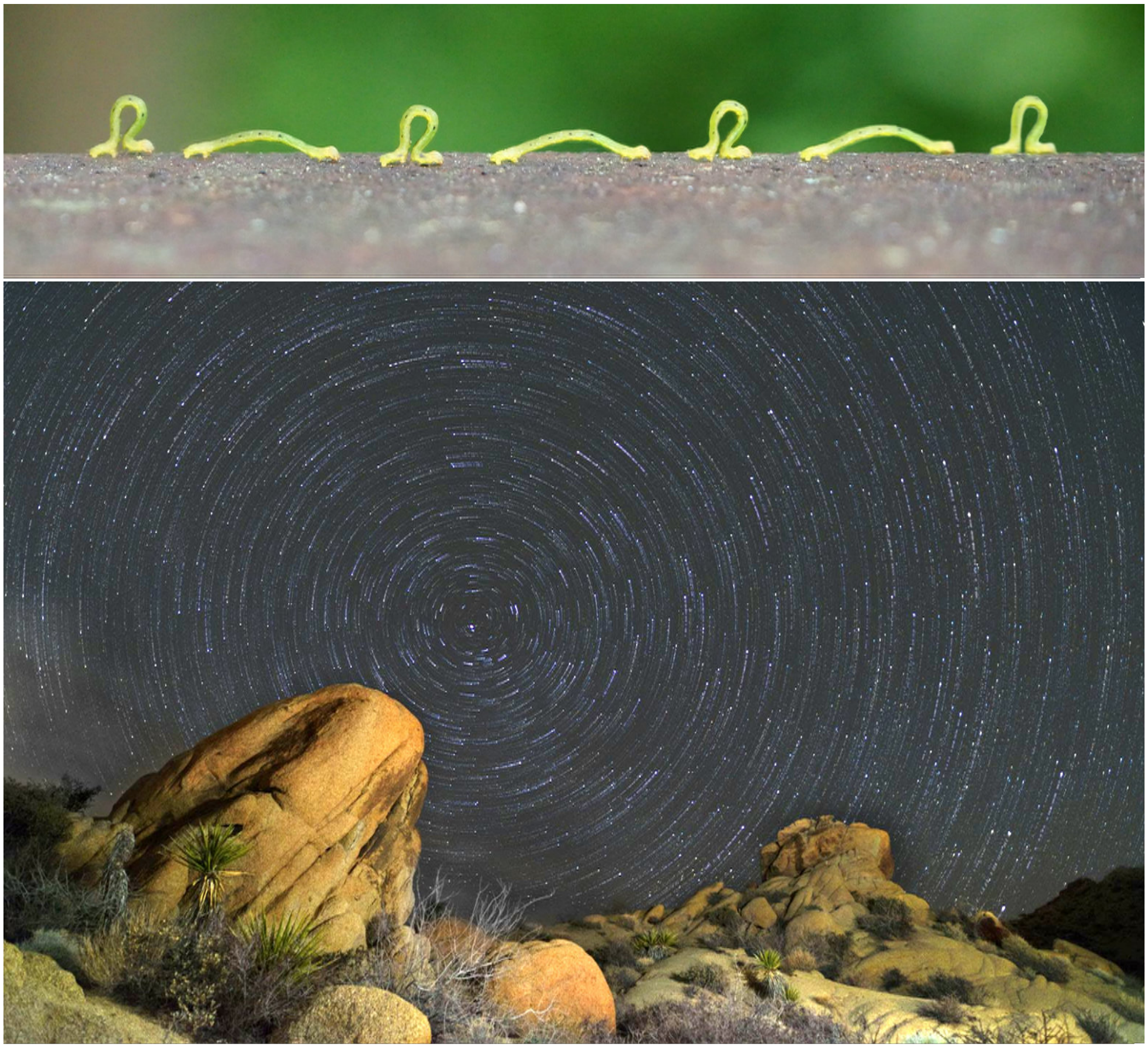


Figure 4.1: Chronophotographs displaying the locomotion of an inchworm (top) and the rotation of the Earth (bottom). By overlaying a series of photographs taken in rapid succession, the relatively quick movements of an inchworm can be decomposed into a sequence of poses. Similarly, the slow motion of the Earth relative to the night sky can be observed by overlaying a series of photographs taken over a long period of time. The resulting star trails form arcs around the triple star system Polaris, also called the North Star or Pole Star. Depending on the speed and duration of acquisition, chronophotography either expands or compresses time so that the extreme motion of an object can be observed in a single image.

Muybridge developed a specialized zoetrope-style projector called the Zoopraxiscope to animate his chronophotographs [89]. Both the zoetrope and modern movies rely on visual persistence to give the sensation of continuous motion. As Michael Faraday wrote in his 1931 article "On a Peculiar Class of Optical Deceptions":

> The eye has the power, as is well known, of retaining visual impression for a sensible period of time; and in this way, recurring actions, made sufficiently near to each other, are perceptibly connected, and made to appear as a continued impression. [56]

This visual persistence, which was measured by Patrice d'Arcy to be about 130 ms [43], partially dictates the *flicker fusion rate*—the frequency at which intermittent light appears to be continuous. By displaying images at a frequency equal or greater than the flicker fusion rate, the once static chronophotographs became motion pictures.

Landecker notes that time plays three important roles in scientific research:

> There is the time of experiment, the time of recording, and the time of demonstration, and these parameters can be manipulated in relation to one another in order to see new phenomena or to see well-known phenomena in a new way. [116]

Time-lapse imaging uses a low recording frequency and high viewing frequency to highlight motion that occurs over long timescales; whereas, *high-speed imaging* uses a high recording frequency and low viewing frequency to slow the motion of fast moving objects. Modern time-lapse systems require automated software to trigger and expose each image and auxiliary equipment to stabilize and protect the camera for the duration of the experiment. To achieve high-speed imaging, the exposure time must be sufficiently low as to avoid motion blur and acquire enough images over the short duration of the experiment. Therefore, high-speed systems need elaborate lighting and triggering mechanisms to ensure that enough photons are available to properly expose each image.

In addition to their ability to compress and expand time, films and videos allow scientists to analyze a phenomena multiple times without having to physically rerun an experiment. As video-capable imaging sensors become smaller and cheaper, their use in real-time monitoring (i.e., videos in which the times of experimentation, recording, and demonstration are all equal) has greatly increased [47, 212]. Moreover, the transition from film to digital video has opened the door to advanced video analysis. Nearly all video analysis is focused on describing the motion of objects through detection and tracking algorithms. In the next section we explore the general categories and methods for object detection and tracking with an emphasis on their applications in ecology.

4.2 Object Tracking

Object tracking is the problem of measuring the trajectory of an image element as it moves through a scene. Although the term most specifically refers to the process of following an object, *object tracking* is also used in reference to the entire workflow of object detection, recognition, and tracking. The object detection and recognition subproblems aim to locate image elements of interest and classify them against a set of known objects types (e.g., cat, dog, sign, human, etc.). Object tracking has received a great deal of attention over the past two decades (see reviews by Cox [39], Joshi & Thakore [101], Trucco & Plakas [222], Yang et al. [244], Yilmaz et al. [246]), and particular applications, such as surveillance, have been extensively explored [101, 144, 159, 160]. Despite these efforts, object tracking remains a challenging problem. Videos are an imperfect two-dimensional projection of a three-dimensional scene. The tracking of an object is impeded by occlusion, image noise, low frame rates, illumination changes, and the relative motion between the sensor and the object. Moreover, large volumes of streaming and/or multi-view video data, as in surveillance and monitoring, makes it difficult to achieve suitable processing times for real-time or batch analysis.

At the highest level, tracking algorithms can be separated by whether they track a single or multiple objects. For the purposes of this work, however, we will focus on the generalized multiple object tracking methods. Luo et al. [144] divides these algorithms into *detection-based tracking* (DBT) and *detection-free tracking* (DFT)[2]. In DBT, object tracks are built by linking together objects that have been automatically detected in multiple video frames. DFT, on the other hand, requires that objects be manually initialized. Local region descriptors for each manually-initialized object are used to track its motion across the scene.

Both DBT and DFT can be performed in an online or offline manner. *Online* tracking is suitable for real-time video analysis, but requires that trajectories be extended in time without information from future video frames. As a result, tracking is more difficult in the beginning frames because there is less information about the objects being tracked. *Offline* tracking can be used whenever real-time tracking data is not required. Since videos are processed in batch, offline tracking can leverage information about an object across all video frames. However, this additional information comes at the cost of requiring more storage and faster processing hardware than online algorithms. Several benchmarks have been established to evaluate specific applications of both online and offline tracking algorithms [130, 204, 241].

Although the majority of object tracking methods have focused on applications in areas such as surveillance [94], action recognition [176], and particle tracking [155], inexpensive video sensors have opened the doors to the remote monitoring of animals [47, 114, 161, 237]. Videos taken from remote camera traps, monitoring stations,

---

[2] Yang & Nevatia [243] originally termed "detection-based tracking" and "detection-free tracking" as "association-based tracking" and "category-free tracking," respectively.

aerial drones, and animal-borne cameras are used to study animal behavior as it relates to social dynamics, communication, foraging, disease transmission, and more.

4.2.1 Applications in Ecology

Measurement of animal movement is key to understand ecological and evolutionary processes such as dispersal, population and metapopulation dynamics, disease transmission, and gene flow [225]. Many of these studies are central to conservation efforts [55]; however, we still have a limited understanding of animal movement in most applications. For example, there are very few diseases for which we have a well-characterized empirical understanding of spatial transmission dynamics driven by host movement. This is true even for diseases important to public health [181, 233].

Existing categories of methods for recording animal trajectories include:

- direct human observation involving marked [180] and unmarked [75] individuals;
- trace-based methods with visible trail markers such as powdered dyes [1];
- active and passive electronic tags including radio tracking [2], harmonic radar [167], GPS tags [179], and RFID [110];
- biomarkers including stable isotopes [189];
- and image-based methods including camera traps, video tracking [47], and fingerprinting methods [113, 171].

These methods balance trade-offs between cost, accuracy, reliability, tracking area, ability to distinguish individuals, tracking capacity, and behavior-altering impediments. The low-cost and scalability of image-based methods has made them an attractive compromise between the accuracy of radio tracking and the cost of trace-based methods.

A limited number of methods exist for measuring movement for small organisms such as insects because they require the use of miniaturized components that are susceptible to false negative detection. Within the existing repertoire, even fewer methods offering consistent individual-level resolution are affordable, scalable and operable in the field [24, 109, 156, 224, 235]. Therefore, we developed a method that combines automated video capture and a graphical interface to quantify motion dynamics of insects from discrete locations by video analysis of inexpensive ($\ll$ \$0.01 per tag) and lightweight tags attached to individual insects. This method differs from similar optical tag tracking methods in that the tags are human-readable and the setup does not require specialized materials [28, 41, 235]. We have deployed consumer-grade digital cameras for video capture, similar to work by Steen [209], with simple weatherproofed enclosures, keeping the cost of the entire system low (see Appendix B). The key developmental component is a video analysis and graphical editing software that identifies tags in video frames, assembles these discrete tags into tracks of the same insect moving through scenes, infers the tag identifier by digit recognition, and provides a user-friendly graphical environment for editing tracking data.

4.3 Graphical Insect Tracking Environment

We have developed an analysis pipeline and graphical editor, called the GRAPHical Insect Tracking Environment (GRAPHITE), for end-to-end processing of video data [184]. GRAPHITE[3] is a modular set of functions with a graphical interface entirely written in MATLAB. The software consists of a video preprocessor, tag detector, digit reader, track builder, and a processing interface (see Figure 4.2). The user can choose to initiate the entire set of analysis routines as a single pipeline or access each module independently for a tailored analysis of a particular video. Batch processing

---

[3] GRAPHITE is open source and available at https://github.com/brossetti/graphite.

is performed offline using a parallel framework. As such, GRAPHITE is capable of handling large volumes of video data.

Our system is designed for use in field settings, in contrast to other video-based methods that are used in laboratory studies with predefined tracking areas [156, 165]. Compared with typical camera traps [188], our system can distinguish individuals via human-readable tags with unique numeric identifiers. Moreover, using a video-based method overcomes three limitations of existing RFID technology. First, it enables easier access to location-based data by monitoring more colonies at a fraction of the cost of comparable RFID systems. Second, small RFID tags require close proximity to a detector for accurate tracking; whereas, our system only requires that insects pass within the visible range of the camera. Lastly, our method can be scaled-up to track more insects at more tracking locations with minimal additional expense.

We show proof-of-concept of this method by tracking honey bees (*Apis mellifera*) at the entrance of beehives (see Appendix B for the experimental setup). We expect that our method can be deployed to track other types of uniquely marked insects (approximately 3 mm long or larger) moving at discrete locations. In particular, our method is most readily applied to central-place foragers with small nest, colony, or roost entrances relative to the animal's size. It would also be straightforward to deploy our method in studies with bait stations and feeders, such as artificial flowers and pollinator feeding stations for honey bees [79] and social stingless bees [97]. Although free-ranging animal movements could also be tracked, this would be more challenging than the present study. Limitations to our method are three-fold. First, organisms need to be tagged. This necessitates prior capture as well as knowledge about which individuals are expected to be seen at each camera location. Second, the system is not expected to be as effective with solitary animals. Third, GRAPHITE is an offline processing tool and requires more storage than online tracking solutions. Even with these limitations in mind, we believe that our low-cost experimental setup and user-

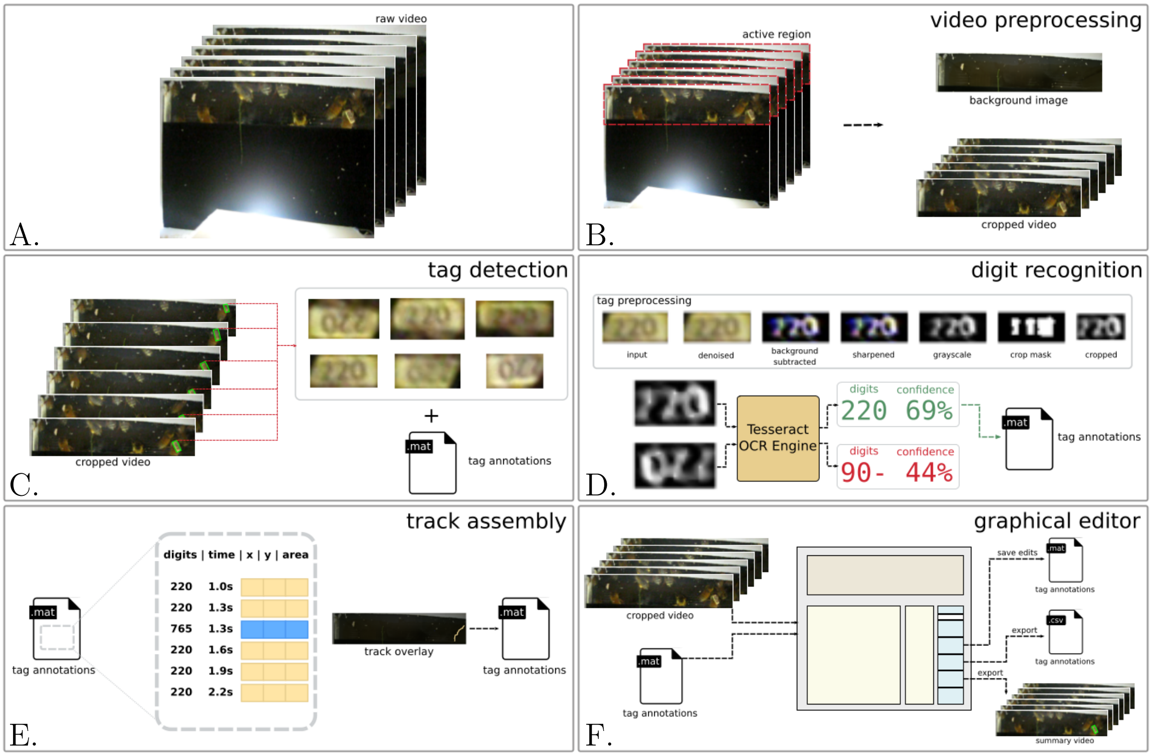
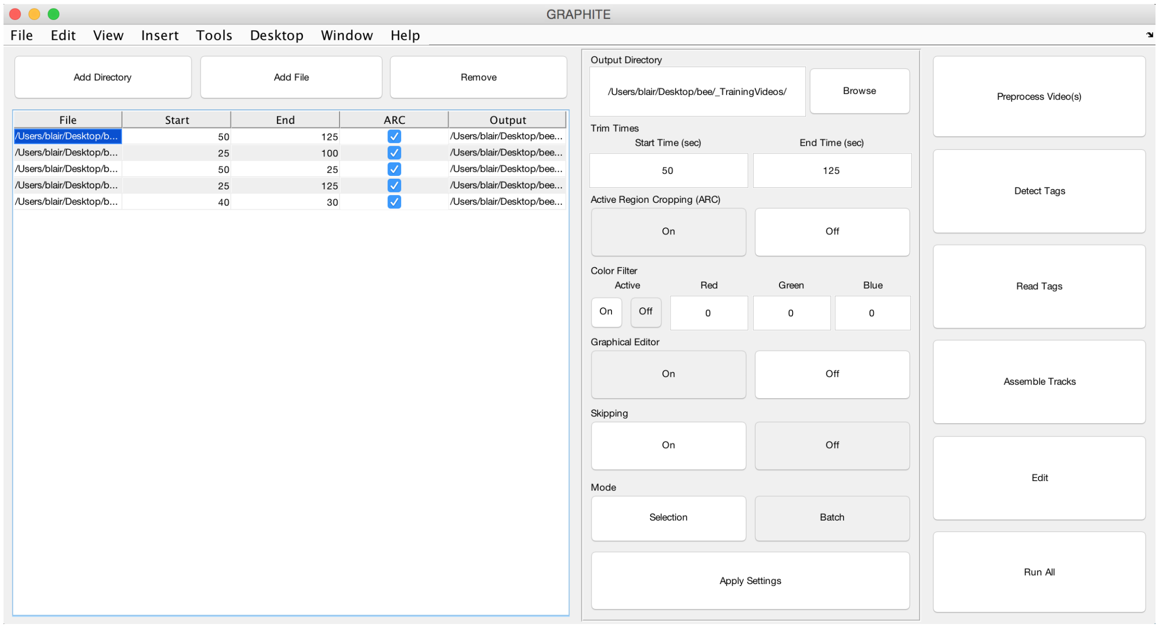


Figure 4.2: The GRAPHITE workflow consists of five modules: video preprossesor, tag detector, digit reader, track builder, and graphical editor. The video preprocessor accepts a raw video (A) as input and generates a background image and, optionally, a cropped video file as output (B). Each frame of the cropped video is searched for tag regions, and the resulting individual tag images are extracted, saved, and logged in an annotation file (C). Tag regions are preprocessed and sent to the Tesseract OCR engine for digit recognition (D). The orientation with the highest average confidence is chosen as the correct orientation, and digit recognition results are appended to the annotation file. Tag data from different frames are linked as tracks based on their spatial locations and sizes (E). The graphical editor allows individual and global changes to tag data stored in the annotation file, and it allows users to export the annotation data along with a summary video of tracked animals (F).



File	Start	End	ARC	Output
/Users/blair/Desktop/b...	50	125	☑	/Users/blair/Desktop/bee...
/Users/blair/Desktop/b...	25	100	☑	/Users/blair/Desktop/bee...
/Users/blair/Desktop/b...	50	25	☑	/Users/blair/Desktop/bee...
/Users/blair/Desktop/b...	25	125	☑	/Users/blair/Desktop/bee...
/Users/blair/Desktop/b...	40	30	☑	/Users/blair/Desktop/bee...

Output Directory: /Users/blair/Desktop/bee/_TrainingVideos/

Start Time (sec): 50

End Time (sec): 125

Red: 0

Green: 0

Blue: 0

Figure 4.3: The main GRAPHITE interface allows users to select videos for batch processing (left), adjust tracking parameters (middle), and execute specific modules (right). In case of disruptions, GRAPHITE remembers which modules have already been run for each video file and it will skip steps as appropriate when the analysis is resumed.

friendly graphical software has the potential to contribute to a spectrum of insect movement studies.

GRAPHITE is a DBT system that works by detecting tags across all video frames and linking them together into tracks. The entire system is tied together with an intuitive graphical interface that provides access to critical analysis parameters (see Figure 4.3). In the sections below we examine each of the five GRAPHITE modules: (1) video prepressesor, (2) tag detector, (3) digit reader, (4) track builder, and (5) graphical editor. Video tutorials describing the batch processing functionality and graphical editor interface are included as supplemental data.

4.3.1 Video Preprocessing Module

The first module is a video preprocessor that prepares each video for tracking (see Figure 4.2 A-B). The module allows the user to crop videos in the temporal and spatial

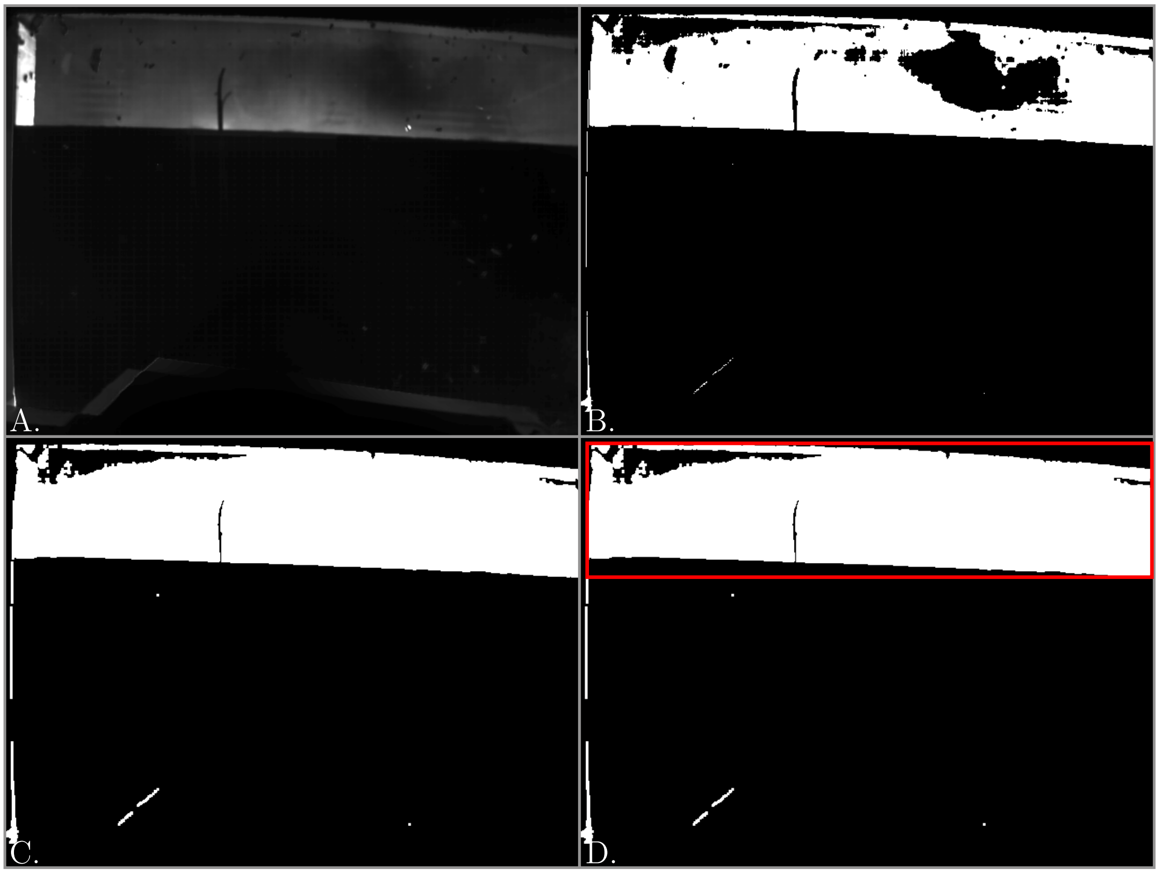


Figure 4.4: Active region detection and cropping from a raw video file. A map of the active areas is determined by the pixel-wise variance over all frames where higher intensity values represent a greater deviation from the background, indicating motion events (A). Global adaptive thresholding is used to separate active (white) from static (black) regions in the activity map (B). The binarized image is cleaned by morphological operations including dilation (C). A bounding box is determined for the largest active region (outlined in red), and the associated coordinates are used to crop the raw video (D).

dimensions to reduce the processing time. The user can specify trimming times to remove frames from the beginning and/or end of each video. Trimming is useful for discarding videos segments in which the camera is being positioned or removed from the camera housing. The spatial dimensions can also be cropped to remove areas that fall outside the camera viewport, or active region, as described in Appendix B. Enabling active region cropping reduces the search space for subsequent modules and can lead to faster tag detection (see Figure 4.4).

Since motion events correspond to pixels with high variance, the active region is located by computing the pixel-wise variance across the trimmed video frames. The resulting variance image provides a map of activity where brighter pixels correspond to higher levels of motion. The activity map is segmented by an adaptive threshold, such as Otsu, into active foreground and static background regions. The binary map is cleaned by a series of morphological operations to more clearly define the active region, and the bounding box coordinates are determined for the largest contiguous region. These coordinates are used to crop the raw video frames for subsequent analysis.

The most important function of the video preprocessor is to generate a static background image for use in tag detection. The background image should represent all of the static visual components in the scene. A possible method for generating a background image is to capture a still frame of the empty scene; however, acquiring a background image in this way is not practical since bees are constantly entering and exiting the beehive. Instead, we generate a grayscale background image by calculating the mean pixel-wise intensity over all frames of the trimmed and cropped videos sequence. Since the field-of-view is known to be fixed, this method produces a robust background image that accurately captures the properties of a given scene (i.e., camera position, obstructions, and lighting condition). We note, however, that using the pixel-wise mean intensity to represent the background is not advised when scene elements can change over the course of a single video.

4.3.2 Tag Detection Module

As a DTB system, GRAPHITE operates by separately detecting tags in each video frame (see Figure 4.2 C). In a given video frame, a moving object can either be a tagged or untagged insect. Since the tags and insects are different colors (white and yellow, respectively), these two objects can be separated using a color filter. To

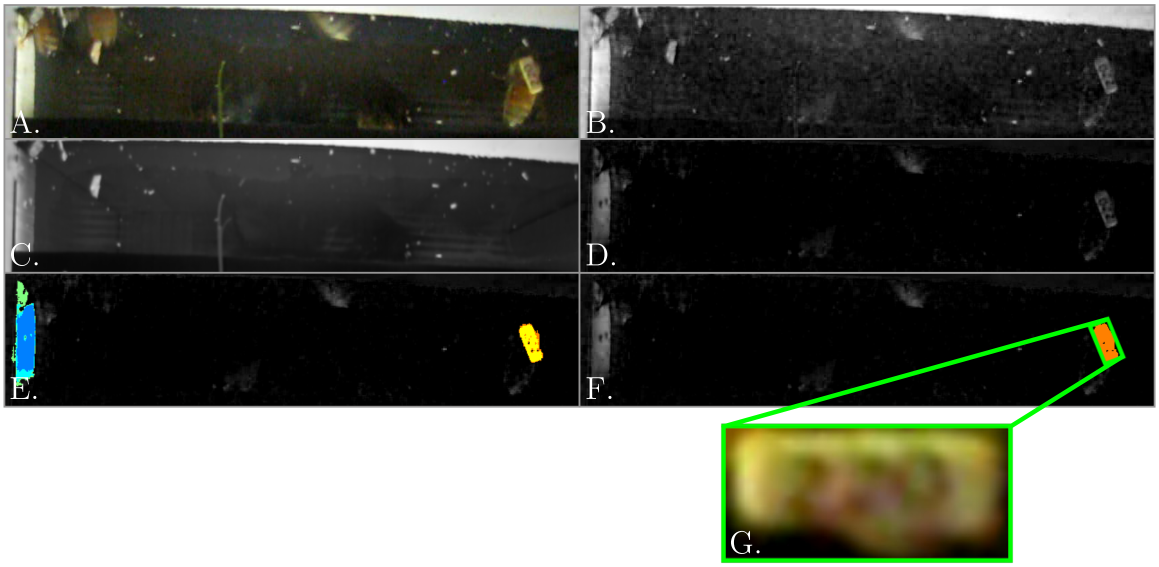


Figure 4.5: Tag detection from a single video frame. A full-color video frame (A) is color filtered and converted to grayscale (B). The grayscale background image (C) is subtracted from the color-filtered image to produce a motion map (D). The MSER feature detector identifies five contiguous areas with stable intensity labeled in orange, yellow, green, cyan, and blue (E). MSERs are filtered by solidity, aspect ratio, and eccentricity to remove non-tag regions and a minimum bounding rectangle (MBR), shown in green, is fit to the remaining orange tag regions (F). The MBR coordinates are used to extract and rotate the tag region (G).

accommodate different color schemes, the user can input any RGB triplet specifying the color to be removed. Both the video sequence and RGB triplet are converted to hue-saturation-value colorspace. The specified color is filtered by setting the value of all pixels with a hue within 15° of the specified color to zero. Finally, the filtered frames are converted to grayscale for subsequent processing.

Locating motion in a scene requires that we identify image elements that deviate from the reference background image. Therefore, the background image generated during video preprocessing is subtracted from each color-filtered video frame to produce a motion map. Contiguous regions representing possible tags are detected in each motion map using a Maximally Stable Extremal Region (MSER) feature detector. MSERs represent appropriately colored moving objects with a size ranging from 300 to 3,000 pixels in area (see Figure 4.5 A-E).

To this point, the tag detection module has focused on the detection of all putative tag regions. The module proceeds by discarding non-tag regions, or false positive. Since the tag dimensions are known *a priori*, we use shape measurements including solidity, aspect ratio, and eccentricity to conservatively filter MSERs that do not represent tags. Solidity is the ratio of the region area to the convex hull area; aspect ratio is the ratio of the minor-axis length to the major-axis length of the region's fitted ellipse; and eccentricity is the distance between the fitted ellipse foci and the major-axis. In the case where the MSER feature detector finds overlapping areas, only the smallest region by size is retained (see Figure 4.5 F).

Although filtering by physical attributes removes most non-tag regions, additional steps are required to further reduce the number of false positives. In our solution, each potential tag region is fitted with a minimum-area bounding rectangle (MBR). The MBR coordinates are used to rotate and crop the region from the full-color video frame. Each cropped region is resized to 60×30 pixel rectangle and represented by a histogram of oriented gradients (HOG) feature vector. A HOG feature vector is a series of one-dimensional histograms describing the edge orientations within each 4×4 pixel image patch. By encapsulating the shape components found within the image, we classify cropped regions as *tag* or *non-tag* using a two-class support vector machine (SVM). The SVM was trained on HOG features from 4093 false tag images and 880 positive tag images. Any remaining tag regions are passed onto the next module for digit recognition.

4.3.3 Digit Recognition Module

Digit recognition from natural images has been an area of intensive research [78, 102, 247, 251]. The digit recognition module uses the Tesseract optical character recognition (OCR) engine [205] to identify digits in tag images (see Figure 4.2 D). Each extracted tag image is first preprocessed to enhance the contrast of digit characters.

Tag image preprocessing begins with channel-wise wavelet denoising and a rolling-ball background subtraction to correct for uneven illumation [210]. Wavelet denoising uses a discrete stationary wavelet transform to remove noise in the image frequency domain without excessive edge blurring. The rolling-ball background is generated by a morphological open operation on each color channel with a 5 pixel radius spherical structuring element. Each channel is then histogram-normalized and sharpened before conversion to grayscale.

As any remaining marks within border regions can result in incorrect digit recognition, we limit analysis to the digit containing region. We multiply the column sum and row sum of the tag image to produce a map of the digit region. This map is binarized and the bounding box coordinates for the foreground digit region are recorded. If a digit region is not found, the tag is marked as a false positive and removed from the following analysis.

Preprocessed tag images are passed to the Tesseract OCR engine for digit prediction. The OCR engine was trained on more than 100 examples for each preprocessed digit. As tag images can be in two possible orientations (right-side up and upside down), digit predictions are made for both orientations. Digits with the highest three confidence levels are retained as the predictions for each orientation, and the highest average confidence level is used to indicate the correct orientation.

4.3.4 Track Assembly Module

To this point, each video frame has been analyzed independently to detect and read tags. Relating frame-wise tag data into tracks is necessary to achieve an interpretation of bee activity (see Figure 4.2 E). For this purpose, the trajectory of each tagged bee is assembled by linking tag data from frames based on centroid (x, y) location and tag size (area). The x-y-area feature vectors for sequential tag images are compared with a nearest neighbor algorithm. The Euclidean distance between each feature

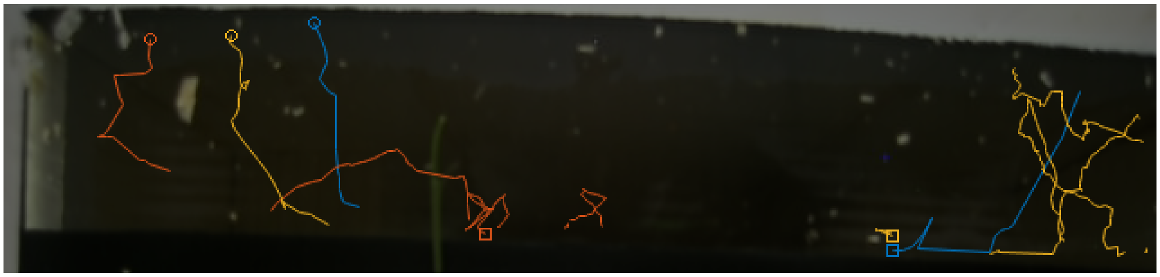


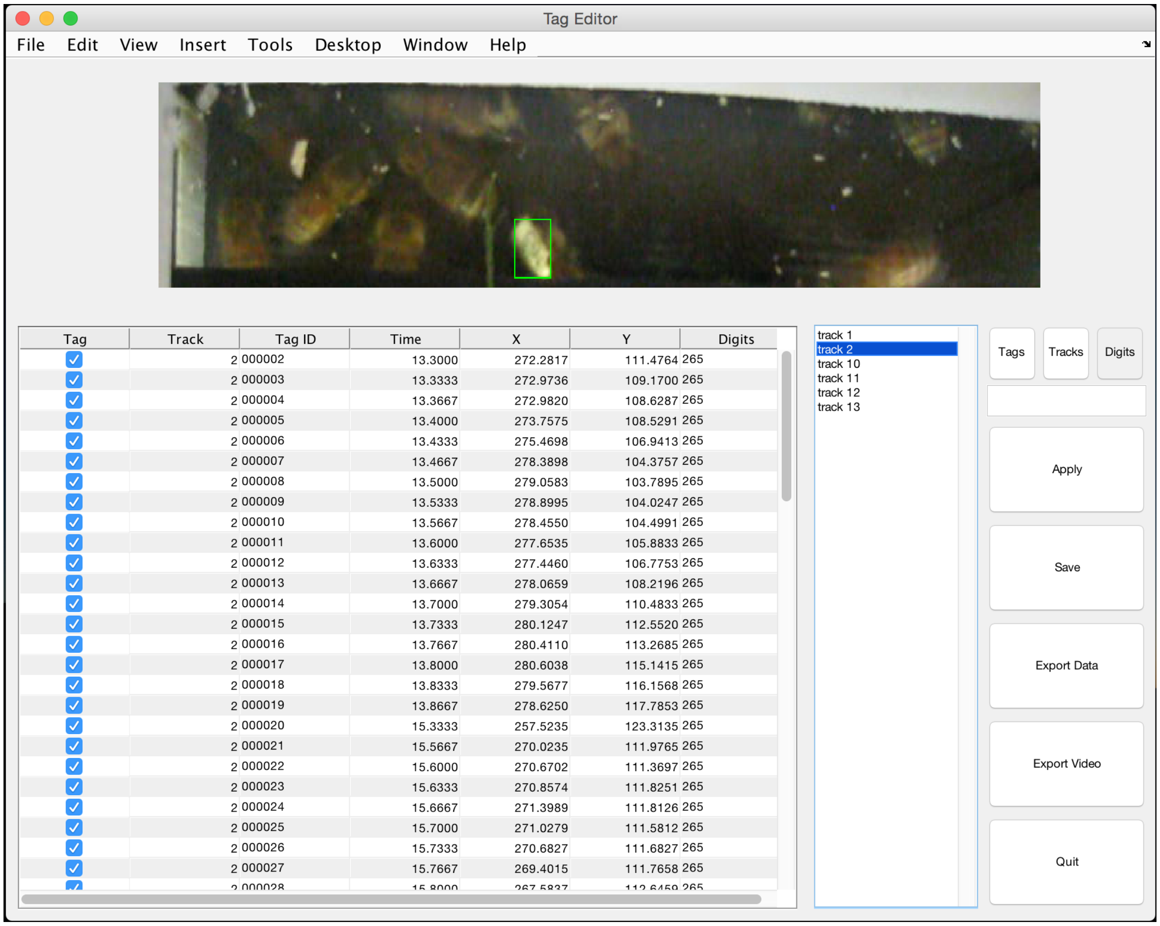
Figure 4.6: The tracks for three tagged bees are overlaid onto the background image derived from the 50 minute video. Each color (red, yellow, and blue) represents the path of one tagged bee. Squares indicate the points of first observation, and circles denote the points of last observation. Gaps in the tracks indicate an occlusion of at least one second. The tracks show that all three bees exited the beehive (top represent the beehive opening).

vector is used to match a tag in one frame to a single tag in an adjacent frame. The track assembly algorithm tolerates gaps between matches of half a second to account for momentary occlusions. Matched tags are linked together into tracks with unique track identification numbers to represent tagged bee motion paths (see Figure 4.6).

4.3.5 Graphical Editor

We set the software parameters to provide robust results for our experimental setup, but conditions can vary tremendously even within a single study. Therefore, a full-featured graphical editor is provided to allow users to correct errors in the automatically generated tag data (see Figure 4.2 F). The editor is designed to provide users with easy access to critical tag data, including tag digits, track identifiers, and false positive status.

The editor presents users with two tabular windows (see Figure 4.7). The first window allows users to select one or more tracks. Once tracks are selected, the second window displays all tags included in those tracks. Selecting a tag will display tag-related video frames with a green bounding box around the tag of interest. All other detected tags within frames are bounded in yellow. Edits can be made for individual tags or even groups of tags for enhanced efficiency.



Tag	Track	Tag ID	Time	X	Y	Digits
☑	2	000002	13.3000	272.2817	111.4764	265
☑	2	000003	13.3333	272.9736	109.1700	265
☑	2	000004	13.3667	272.9820	108.6287	265
☑	2	000005	13.4000	273.7575	108.5291	265
☑	2	000006	13.4333	275.4698	106.9413	265
☑	2	000007	13.4667	278.3898	104.3757	265
☑	2	000008	13.5000	279.0583	103.7895	265
☑	2	000009	13.5333	278.8995	104.0247	265
☑	2	000010	13.5667	278.4550	104.4991	265
☑	2	000011	13.6000	277.6535	105.8833	265
☑	2	000012	13.6333	277.4460	106.7753	265
☑	2	000013	13.6667	278.0659	108.2196	265
☑	2	000014	13.7000	279.3054	110.4833	265
☑	2	000015	13.7333	280.1247	112.5520	265
☑	2	000016	13.7667	280.4110	113.2685	265
☑	2	000017	13.8000	280.6038	115.1415	265
☑	2	000018	13.8333	279.5677	116.1568	265
☑	2	000019	13.8667	278.6250	117.7853	265
☑	2	000020	15.3333	257.5235	123.3135	265
☑	2	000021	15.5667	270.0235	111.9765	265
☑	2	000022	15.6000	270.6702	111.3697	265
☑	2	000023	15.6333	270.8574	111.8251	265
☑	2	000024	15.6667	271.3989	111.8126	265
☑	2	000025	15.7000	271.0279	111.5812	265
☑	2	000026	15.7333	270.6827	111.6827	265
☑	2	000027	15.7667	269.4015	111.7658	265
☑	2	000028	15.8000	267.5837	112.6459	265

Figure 4.7: The graphical user interface for editing tag data automatically generated by the prior modules.

After edits are made, users can export the annotations as either an Excel or CSV file. For an intuitive overview of the data, users can export a summary video that contains annotated video segments of each tag track (see supplemental video). Each track is represented by an MBR in a unique color and with tag digits displayed.

4.3.6 Evaluation

GRAPHITE has a human-in-the-loop design in which a user screens potential tracks that are automatically generated by the video analysis pipeline that preemptively minimizes the false negative detection rate. Monitoring was performed for 90 colonies at six apiaries resulting in 1,339 video files with a cumulative duration of approximately

12,000 hours. The false negative rate was determined by manually reviewing a random sample of 600 one-minute videos segments. A set of 362 segments were randomly sampled from the 181 videos containing detected tags and 238 segments were randomly sampled from the remaining 1,157 videos. On average, a one-minute segment was reviewed for approximately 20 % of the videos where a tag was not detected. This review resulted in a false negative rate of 0 %.

GRAPHITE detected 1,160,145 potential tag regions in 181 of these videos. Potential tag regions were manually reviewed with the graphical editor. 6,766 tags were identified (representing 450 tracks from 229 bees) resulting in a false positive rate of 99.4 %. Despite the expectedly high false positive rate, the pipeline reduced the manual screening time by over 1000 × (from 12,000 hours to 11 hours) without missing any tagged bees. In addition, false positives were mostly grouped into a small number of tracks that were quickly reviewed and removed in bulk.

4.4 Discussion

GRAPHITE offers a low-cost, end-to-end animal movement tracking environment with a user-friendly graphical interface. We demonstrate the efficacy of the developed software with specific application to tracking tagged bees. The accessible and minimal hardware requirements along with the highly automated and flexible processing modules allow for many different experimental setups with various model organisms. This flexibility allows capabilities beyond video tracking software with no means to identify individuals traversing different tracking stations [109, 224].

A major advantage of this method is its ability to track individual insect movements in a low-cost field setting, as opposed to average movement rates that routine techniques such as powdered dye provide. Individual variation in movement can have large consequences for the ecology and evolution of species [18, 19]. For example, in infectious disease studies, certain individuals may be more likely to move and thereby have greater contact rates than other individuals. Highly mobile and connected individuals could thereby have major impacts on disease transmission, and in some cases act as superspreaders [138].

In future work, the GRAPHITE digit reading module can be upgraded to more efficient learning engines. In addition, corrections made via the editor can be fed back into machine learning models to iteratively improve the accuracy of digit recognition. The SVM classifier used to remove non-tag regions during tag detection could also benefit from the same feedback mechanism.

Chapter 5

Conclusion

“Seeing is believe, but measuring is knowing,” has been the unofficial motto of the bioimage informatics field for many years, and no other tool has been more valuable in the measurement and analysis of biological images than the computer. In their 1966 paper “The Analysis of Cell Images,” Prewitt & Mendelsohn wrote:

> [M]odern large capacity, high speed data facilities at last provide the ability to manipulate the hitherto unmanageable quantities of optical information contained within all but the simplest images. With the basic materials for achieving automation via mensuration finally at hand, attention has been turned towards generating and evaluating methods for extracting meaning from quantitative optical information. [177]

This perspective still rings true today. As hardware for generating big biological data has continued to mature, the focus of modern bioimage informatics research has been the development of scalable algorithms for the storage, manipulation, and processing of image data. In this work we presented three such methods spanning all five image dimensions: space (x, y, z), color (λ), and time (t). These methods represent stepping stones on the way to robust and versatile processing of bioimages of large volume, high velocity, and wide variety.

In Chapter 2, we presented a novel NMF approach, called SSASU, that produces clean unmixing results for contaminated spectral micrographs. High-dimensional spectral images remain a powerful imaging modality for the analysis of complex communities, and SSASU is a practical solution for decomposing real-world data. In addition, we provided a numerical method for the estimation of endmembers from reference images. This method increases the scalability of spectral image analysis by removing the need for manual segmentation or the use of thresholding algorithms. We note that, like other NMF methods, SSASU does not guarantee a unique solution. Therefore, future work in spectral unmixing should focus on leveraging additional *a priori* information to reduce the solution space. Promising avenues in this area include the pixel-purity assumption and statistical approaches that account for the endmember variation contained in reference images.

The problem of volumetric reconstruction described in Chapter 3 is of great interest in modern bioimage informatics. The widespread use of three-dimensional imaging systems including confocal and lightsheet microscopes, block facing scanning electron microscopes, and serial whole slide scanners have amplified the need for accurate and efficient registration methods. We proposed a two-stage registration method that performs offline registration and online subvolume reconstruction. This method is ideally suited for use with interactive systems (e.g., web portals and virtual slide viewers) that allows researchers and clinicians to request specific subvolumes for further analysis. As mentioned, a limitation of this work is that reconstruction accuracy will always be limited to the resolution used for registration. Future work will use a multilevel registration approach in which the transformations are modified based on information from higher resolutions.

We described in Chapter 4 the GRAPHITE software for interactive video-based animal tracking. A major advantage of GRAPHITE is its ability to use low-cost, human-readable tags for monitoring animal motion across many discrete locations.

The minimal hardware requirements and flexible processing modules allow for many different experimental setups with various model organisms. GRAPHITE's modularity also allows for easy upgrades to different portions of the analysis pippeline. For instance, GRAPHITE's digit recognition module could be replaced by a more advanced convolutional neural network to improve reliability in tag identification. In addition, it is possible to use the human annotations collected by the editor to iteratively improve the accuracy of the neural network throughout an experiment.

In closing, and as we consider our next steps in bioimage informatics, it seems fitting to consider again the epigraph at the start of this dissertation:

> We will begin by asking if the prevailing notion of what computer systems can achieve in biological image-processing may not be overly modest. This question seems to us fundamental, for such an underestimate might produce not only a limitation of objectives, but also a restriction of the class of problems considered amenable to computer processing. [134]

Appendix A

Spectral

A.1 Sample Preparation

To evaluate the proposed method, a set of seven reference samples, ten test samples, and one no-probe control was prepared. The bacteria *Leptotrichia buccalis* was used as the biological target for generating reference samples for each of the seven fluorophores: DY-415, DY-490, ATTO 520, ATTO 550, Texas Red-X, ATTO 620, and ATTO 655 (Dyomics GmbH; ATTO-TEC GmbH; Thermo Fisher Scientific Inc.). *L. buccalis* cells were cultured, fixed, and then separately hybridized using custom fluorophore-conjugated oligonucleotide probes (biomers.net GmbH) as decribed by Mark Welch et al. [147].

Test samples consisted of biofilms that were collected from the dorsum of the tongue, chemically fixed, and hybridized using a set of probes specific to different taxa of bacteria (see Table 2.1). Two samples were taken from each of five human subjects (A-E). An additional biofilm sample was collected from subject D, chemically fixed, and hybridized without a fluorophore to generate a no-probe control sample (used to measure autofluorescence). After hybridization, the reference samples, test samples, and no-probe control were mounted on slides as per Mark Welch et al. [147].

A.2 Imaging and Preprocessing

All spectral micrographs were acquired on a Zeiss LSM 880 using a 63 ×/1.4 NA Plan-Apochromat objective. Point scanning was performed simultaneously for the 405 nm, 488 nm, 561 nm, and 633 nm lasers using the 405 and 488/561/633 dichroic mirrors. Spectral data was collected from a range of 410–696 nm using 8.6 nm steps. Each reference image was acquired using dimensions of 512 × 512 pixel at 0.263 µm/pixel resolution. Test images were acquired at 2048 × 2048 pixel and down-sampled to 1024 × 1024 pixel at a resolution of 0.220 µm/pixel.

The use of dichroic mirrors blocked the detection of emitted light near the excitation wavelengths. Since these dark bands contained little information, they were removed from the reference and test images prior to any analysis. In total, 6 of the 32 spectral bands were removed.

Appendix B

Temporal

B.1 Experimental Setup

Tags were designed to be durable in the outdoor environment, easily visible, lightweight, and low-cost. Each tag consisted of a unique three-digit number that was inkjet-printed (7.5 pt font) on white card stock (Neenah Exact Index, `Item# 40508`) with UV-resistant ink (PrintPayLess Black UV-Resistant Dye Ink). An inverted color scheme can be used for light-colored insects. Tags were punched from the card stock and trimmed to a final size of 2.5 × 6 mm. The tags were then sprayed with a UV-resistant coating (Krylon UV-Resistant Clear Acrylic Coating Spray, `Item# 1305`) and a waterproof coating (Scotchgard Outdoor Water Shield, `Item# 5019-6`).

We recorded bee movement for a total of 90 colonies at six apiaries managed by the University of Georgia. To ensure that bees were correctly tagged with their respective colony and queen, brood frames were moved to an enclosed environment one day prior to tagging. We tagged newly emerged worker bees as they are not yet able to fly or sting. A unique tag was secured to the thorax of newly emerged bees using a waterproof glue (Titebond III, `Item# 1411`). The glue was allowed to become tacky and then applied to the bees using a wood toothpick. The tag was then affixed and held briefly to set. All tags were oriented with the rightmost number towards the head of the bee. Ethical considerations must be given to the tagging of sensitive or threatened species and the impact of tagging on the tracked animals.

A camera housing was temporarily mounted to the entrance of each beehive for monitoring tagged bees that exited and entered the apiaries (see Figure B.1). Each camera housing was $10 \times 15 \times 15.5\,\mathrm{cm}$ with a lower landing that extended 8.5 cm from the front face. Bees could pass through a $100 \times 8\,\mathrm{mm}$ opening at the front of the camera housing. The camera compartment was separated from the passage by a 3 mm thick OPTIX acrylic sheet. The camera-facing-side of the acrylic sheet above the entrance was painted black except for an 18 mm viewport strip for video recording. The entrance-facing-side of the acrylic sheet was treated with a lubricant (3-IN-ONE Dry Lubricant, `Item# 3IO-DL-00`) to inhibit bees from walking in an orientation that obscured the tags. Lighting was provided by a 1.5 W battery-powered LED (LouisaStore Portable Pocket LED Card Light, `Item# B00PIU26T0`) located within the camera compartment. Modifications to the camera housing can be made to accommodate alternative experiments and organisms provided the camera retains clear *en face* view of tags.

Videos were recorded on Canon PowerShot SD1100 IS model cameras (30 fps; 640×480 pixel; automatic white balance; macro mode). Video duration ranged from 45 minutes to one hour depending on the battery. The camera was mounted in the camera compartment on a wooden shelf 106 mm above the acrylic sheet. Frame-by-frame tracking was restricted to the viewport area; however, integration of data from multiple camera housings allows low-resolution tracking of tagged insects across sites.

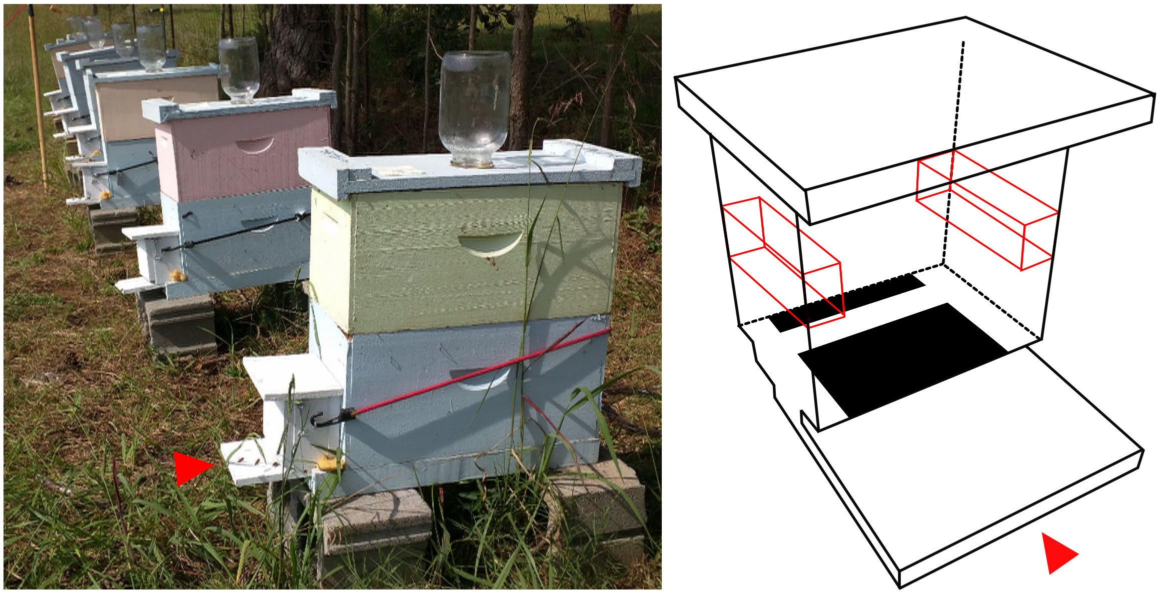


Figure B.1: A camera housings was attached to the front entrance of each beehive as shown on the left. Each camera housing was constructed with a lower landing, viewport, and camera shelf as diagrammed on the right. The red arrows point to the lower landing.

Bibliography

[1] Adler, L. & Irwin, R. (2006). Comparison of pollen transfer dynamics by multiple floral visitors: Experiments with pollen and fluorescent dye. *Annals of Botany*, *97*(1), 141–150.

[2] Aebischer, N., Robertson, P., & Kenward, R. (1993). Compositional analysis of habitat use from animal radio-tracking data. *Ecology*, *74*(5), 1313–1325.

[3] Ahn, Y. & Box, J. (2010). Glacier velocities from time-lapse photos: Technique development and first results from the extreme ice survey (eis) in greenland. *Journal of Glaciology*, *56*(198), 723–734.

[4] Alahi, A., Ortiz, R., & Vandergheynst, P. (2012). Freak: Fast retina keypoint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 510–517)., Providence, Rhode Island, USA. IEEE.

[5] Alam, F., Rahman, S., Ullah, S., & Gulati, K. (2018). Medical image registration in image guided surgery: Issues, challenges and research opportunities. *Biocybernetics and Biomedical Engineering*, *38*(1), 71–89.

[6] Alcantarilla, P., Nuevo, J., & Bartoli, A. (2013). Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *Proceedings of the British Machine Vision Conference 2013*, (pp. 13.1–13.11)., Bristol, UK. BMVA Press.

[7] Alfeld, M. & de Viguerie, L. (2017). Recent developments in spectroscopic imaging

techniques for historical paintings–a review. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *136*, 81–105.

[8] Alpert, N., Bradshaw, J., Kennedy, D., & Correia, J. (1990). The principal axes transformation–a method for image registration. *Journal of Nuclear Medicine*, *31*(10), 1717–1723.

[9] Arena, E., Rueden, C., Hiner, M., Wang, S., Yuan, M., & Eliceiri, K. (2017). Quantitating the cell: Turning images into numbers with ImageJ. *WIREs Developmental Biology*, *6*(2), e260.

[10] Arganda-Carreras, I., Fernández-González, R., Muñoz-Barrutia, A., & Ortiz de Solorzano, C. (2010). 3D reconstruction of histological sections: Application to mammary gland tissue. *Microscopy Research and Technique*, *73*(11), 1019–1029.

[11] Arun, K., Huang, T., & Blostein, S. (1987). Least-squares fitting of two 3-D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-9*(5), 698–700.

[12] Attota, R. (2018). Through-focus or volumetric type of optical imaging methods: A review. *Journal of Biomedical Optics*, *23*(7), 070901.

[13] Bagci, U. & Bai, L. (2010). Automatic best reference slice selection for smooth volume reconstruction of a mouse brain from histological images. *IEEE Transactions on Medical imaging*, *29*(9), 1688–1696.

[14] Bardinet, E., Cohen, L., & Ayache, N. (1996). Tracking and motion analysis of the left ventricle with deformable superquadrics. *Medical Image Analysis*, *1*(2), 129–149.

[15] Bay, H., Tuytelaars, T., & Van Gool, V. (2006). Surf: Speeded up robust features.

In *9th European Conference on Computer Vision*, (pp. 404–417)., Graz, Austria. Springer.

[16] Bhargava, R. & Madabhushi, A. (2016). Emerging themes in image informatics and molecular analysis for digital pathology. *Annual Review of Biomedical Engineering*, *18*, 387–412.

[17] Bioucas-Dias, J., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(2), 354–379.

[18] Bolnick, D., Amarasekare, P., Araújo, M., Bürger, R., Levine, J., Novak, M., Rudolf, V., Schreiber, S., Urban, M., & Vasseur, D. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, *26*(4), 183–192.

[19] Bolnick, D., Svanbäck, R., Fordyce, J., Yang, L., Davis, J., Hulsey, C., & Forister, M. (2003). The ecology of individuals: Incidence and implications of individual specialization. *The American Naturalist*, *161*(1), 1–28.

[20] Born, G. (1883). Die plattenmodellirmethode. *Archiv für mikroskopische Anatomie*, *22*(1), 584–599.

[21] Bouvier, T. & Del Giorgio, P. (2003). Factors influencing the detection of bacterial cells using fluorescence in situ hybridization (FISH): A quantitative review of published reports. *FEMS Microbiology Ecology*, *44*(1), 3–15.

[22] Bueno, G., Fernández-Carrobles, M., Deniz, O., & García-Rojo, M. (2016). New trends of emerging technologies in digital pathology. *Pathobiology*, *83*(2-3), 61–69.

[23] Bullen, A. (2019). The automatic tape collection ultramicrotome (ATUM). In *Biological Field Emission Scanning Electron Microscopy*, volume 2 (pp. 485–494). Wiley Online Library.

[24] Campbell, J., Mummert, L., & Sukthankar, R. (2008). Video monitoring of honey bee colonies at the hive entrance. *Visual observation & analysis of animal & insect behavior, ICPR*, *8*, 1–4.

[25] Canales, J. (2015). Dead and alive: Micro-cinematography between physics and biology. *Configurations*, *23*(2), 235–251.

[26] Chang, Y., Foley, P., Azimi, V., Borkar, R., & Lefman, J. (2016). Primer for image informatics in personalized medicine. *Procedia Engineering*, *159*, 58–65.

[27] Chen, B.-C., Legant, W., Wang, K., Shao, L., Milkie, D., Davidson, M., Janetopoulos, C., Wu, X., Hammer, J., Liu, Z., et al. (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, *346*(6208), 1257998.

[28] Chen, C., Yang, E.-C., Jiang, J.-A., & Lin, T.-T. (2012). An imaging system for monitoring the in-and-out activity of honey bees. *Computers and Electronics in Agriculture*, *89*, 100–109.

[29] Chen, C. & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, *275*, 314–347.

[30] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209.

[31] Chi, M., Plaza, A., Benediktsson, J., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, *104*(11), 2207–2219.

[32] Cichocki, A., Zdunek, R., Phan, A., & Amari, S. (2009). *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation.* John Wiley & Sons.

[33] Cifor, A., Bai, L., & Pitiot, A. (2011). Smoothness-guided 3-D reconstruction of 2-D histological images. *Neuroimage*, *56*(1), 197–211.

[34] Colchester, A., Ourselin, S., Zhu, Y., Bardinet, E., He, Y., Roche, A., Al-Sarraj, S., Nailon, B., Ironside, J., & Ayache, N. (2000). 3-d reconstruction of macroscopic optical brain slice images. In *Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 95–105)., Pittsburgh, Pennsylvania, USA. Springer.

[35] Cole, R., Thibault, M., Bayles, C., Eason, B., Girard, A.-M., Jinadasa, T., Opansky, C., Schulz, K., & Brown, C. (2013). International test results for objective lens quality, resolution, spectral accuracy and spectral separation for confocal laser scanning microscopes. *Microscopy and Microanalysis*, *19*(6), 1653–1668.

[36] Comandon, J. (1932). Le cinématographie et les sciences de la nature. *Le Cinéma, des origines à nos jours*, 313–322.

[37] Cooper, L., Huang, K., Sharma, A., Mosaliganti, K., & Pan, T. (2006). Registration vs. reconstruction: Building 3-d models from 2-d microscopy images. In *Proceedings of the 2006 International Workshop on Multiscale Biological Imaging, Data Mining and Informatics*, (pp. 57–58)., Santa Barbara, California, USA. BMC.

[38] Cooper, L., Huang, K., & Ujaldon, M. (2011). Parallel automatic registration of large scale microscopic images on multiprocessor cpus and gpus. In *2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*, (pp. 1367–1376)., Anchorage, Alaska, USA. IEEE.

[39] Cox, I. (1993). A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, *10*(1), 53–66.

[40] Cox, M. (1972). The numerical evaluation of b-splines. *IMA Journal of Applied Mathematics*, *10*(2), 134–149.

[41] Crall, J., Gravish, N., Mountcastle, A., & Combes, S. (2015). BEEtag: A low-cost, image-based tracking system for the study of animal behavior and locomotion. *PLoS One*, *10*(9), e0136487.

[42] Danuser, G. & Waterman-Storer, C. (2006). Quantitative fluorescent speckle microscopy of cytoskeleton dynamics. *Annual Review of Biophysics and Biomolecular Structure*, *35*, 361–387.

[43] d'Arcy, C. (1765). Mémoire sur la durée de la sensation de la vue. *Histoire de l'Académie Royale des Sciences*, 439–451.

[44] Darrodi, M. M., Finlayson, G., Goodman, T., & Mackiewicz, M. (2015). Reference data set for camera spectral sensitivity estimation. *JOSA A*, *32*(3), 381–391.

[45] de Boor, C. (1971). Subroutine package for calculating with B-splines. Technical report, Los Alamos Scientific Laboratory.

[46] de Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., Le Montagner, Y., et al. (2012). Icy: An open bioimage informatics platform for extended reproducible research. *Nature Methods*, *9*(7), 690–696.

[47] Dell, A., Bender, J., Branson, K., Couzin, I., de Polavieja, G., Noldus, L., Pérez-Escudero, A., Perona, P., Straw, A., Wikelski, M., et al. (2014). Automated image-based tracking and its application in ecology. *Trends in Ecology & Evolution*, *29*(7), 417–428.

[48] Demchenko, Y., Grosso, P., de Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In *2013 International Conference on Collaboration Technologies and Systems*, (pp. 48–55)., San Diego, California, USA. IEEE.

[49] Descartes, R. (1644). *Principia philosophiae.* Amstelodami: Ludovicum Elzevirium.

[50] Diebold, F. (2019). On the origin(s) and development of "big data": The phenomenon, the term, and the discipline. `https://www.sas.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf`. Accessed May 30, 2019.

[51] Donoho, D. & Stodden, V. (2003). When does non-negative matrix factorization give a correct decomposition into parts? In *Advances in Neural Information Processing Systems 16*, (pp. 1141–1148)., Vancouver, British Columbia, Canada. NIPS.

[52] Drumetz, L., Chanussot, J., & Jutten, C. (2016). Variability of the endmembers in spectral unmixing: Recent advances. In *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, (pp. 1–5)., Los Angeles, California, USA. IEEE.

[53] Du, J., Li, W., Lu, K., & Xiao, B. (2016). An overview of multi-modal medical image fusion. *Neurocomputing*, *215*, 3–20.

[54] El Gamal, A. (2002). Trends in CMOS image sensor technology and design. In *Technical Digest of International Electron Devices Meeting*, (pp. 805–808)., San Francisco, California, USA. IEEE.

[55] Fahrig, L. (2007). Non-optimal animal movement in human-altered landscapes. *Functional Ecology*, *21*(6), 1003–1015.

[56] Faraday, M. (1931). On a peculiar class of optical deception. *The Journal of the Royal Institution of Great Britain*, *1*, 205–223.

[57] Farahani, N., Braun, A., Jutt, D., Huffman, T., Reder, N., Liu, Z., Yagi, Y., & Pantanowitz, L. (2017). Three-dimensional imaging and scanning: Current and future applications for pathology. *Journal of Pathology Informatics*, *8*, 36.

[58] Farahani, N., Parwani, A., & Pantanowitz, L. (2015). Whole slide imaging in pathology: Advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International*, *7*, 23–33.

[59] Filler, A. (2010). The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, and DTI. *Internet Journal of Neurosurgery*, *7*(1), 5–35.

[60] Fischer, B. & Modersitzki, J. (2008). Ill-posed medicine-an introduction to image registration. *Inverse Problems*, *24*(3), 034008.

[61] Fischler, M. & Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.

[62] Freeman, H. & Glass, J. (1969). On the quantization of line-drawing data. *IEEE Transactions on Systems Science and Cybernetics*, *5*(1), 70–79.

[63] Fu, X., Huang, K., Sidiropoulos, N., & Ma, W.-K. (2019). Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications. *IEEE Signal Processing Magazine*, *36*(2), 59–80.

[64] Gagnier, K. & Shipley, T. (2016). Visual completion from 2D cross-sections: Implications for visual theory and STEM education and practice. *Cognitive Research: Principles and Implications*, *1*(1), 9.

[65] Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

[66] Gao, L. & Smith, R. (2015). Optical hyperspectral imaging in microscopy and spectroscopy–a review of data acquisition. *Journal of Biophotonics*, *8*(6), 441–456.

[67] Garini, Y. & Tauber, E. (2013). Spectral imaging: Methods, design, and applications. In *Biomedical Optical Imaging Technologies* (pp. 111–161). Springer.

[68] Garini, Y., Young, I., & McNamara, G. (2006). Spectral imaging: Principles and applications. *Cytometry Part A: The Journal of the International Society for Analytical Cytology*, *69A*(8), 735–747.

[69] Gaycken, O. (2011). the swarming of life: Moving images, education, and views through the microscope. *Science in Context*, *24*(3), 361–380.

[70] Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., & Plaza, A. (2017). Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 37–78.

[71] Ghaznavi, F., Evans, A., Madabhushi, A., & Feldman, M. (2013). Digital imaging in pathology: Whole-slide imaging and beyond. *Annual Review of Pathology: Mechanisms of Disease*, *8*, 331–359.

[72] Gillis, N. (2012). Sparse and unique nonnegative matrix factorization through data preprocessing. *Journal of Machine Learning Research*, *13*(1), 3349–3386.

[73] Gillis, N. (2014). The why and how of nonnegative matrix factorization. In *Regularization, Optimization, Kernels, and Support Vector Machines* (pp. 257–291). Chapman & Hall/CRC.

[74] Gillis, N. & Glineur, F. (2010). Using underapproximations for sparse nonnegative matrix factorization. *Pattern Recognition*, *43*(4), 1676–1687.

[75] Gòmez, J. M. (2003). Spatial patterns in long-distance dispersal of Quercus ilex acorns by jays in a heterogeneous landscape. *Ecography*, *26*(5), 573–584.

[76] Gomez-Marin, A., Paton, J., Kampff, A., Costa, R., & Mainen, Z. (2014). Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, *17*(11), 1455–1462.

[77] Goode, A., Gilbert, B., Harkes, J., Jukic, D., Satyanarayanan, M., et al. (2013). OpenSlide: A vendor-neutral software foundation for digital pathology. *Journal of Pathology Informatics*, *4*(1), 27.

[78] Goodfellow, I., Bulatov, Y., Ibarz, J., Arnoud, S., & Shet, V. (2013). Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*.

[79] Gould, J. (1975). Communication of distance information by honey bees. *Journal of Comparative Physiology*, *104*(2), 161–173.

[80] Gouyon, J.-B. (2016). Science and film-making. *Public Understanding of Science*, *25*(1), 17–30.

[81] Grusche, S. (2014). Basic slit spectroscope reveals three-dimensional scenes through diagonal slices of hyperspectral cubes. *Applied Optics*, *53*(20), 4594–4603.

[82] Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, *13*, 49–52.

[83] Haeckel, E. & Uschmann, G. (1984). *Ernst Haeckel: Biographie in briefen*. Prisma Verlag.

[84] Han, J. & Burgess, K. (2009). Fluorescent indicators for intracellular pH. *Chemical Reviews*, *110*(5), 2709–2728.

[85] Harmany, Z., Fereidouni, F., & Levenson, R. (2017). Spectral unmixing methods and tools for the detection and quantitation of collagen and other macromolecules in tissue specimens. In *Fibrosis* (pp. 491–509). Springer.

[86] Harris, A. (2006). Spectral mapping tools from the earth sciences applied to spectral microscopy data. *Cytometry Part A*, *69*(8), 872–879.

[87] Harris, C. & Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*, (pp. 147–151)., Manchester, UK. BMVC.

[88] Heard, O. (1951). Section compression photographically rectified. *The Anatomical Record*, *109*(4), 745–755.

[89] Herbert, S. (Ed.). (2004). *Eadweard Muybridge: The Kingston Museum bequest.* Projection Box.

[90] Heylen, R., Parente, M., & Gader, P. (2014). A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), 1844–1868.

[91] Hilbert, M. & López, P. (2011). The worlds technological capacity to store, communicate, and compute information. *Science*, *332*(6025), 60–65.

[92] His, W. (1870). Beschreibung eines Mikrotoms. *Archiv für mikroskopische Anatomie*, *6*(1), 229–232.

[93] Hopwood, N. (1999). “Giving Body” to embryos: Modeling, mechanism, and the microtome in late nineteenth-century anatomy. *Isis*, *90*(3), 462–496.

[94] Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, *34*(3), 334–352.

[95] Huang, K., Sidiropoulos, N., & Swami, A. (2014). Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE Transactions on Signal Processing*, *62*(1), 211–224.

[96] Huang, S., Zhao, Y., & Qin, B. (2015). Two-hierarchical nonnegative matrix factorization distinguishing the fluorescent targets from autofluorescence for fluorescence imaging. *BioMedical Engineering Online*, *14*(1), 116–135.

[97] Hubbell, S. & Johnson, L. (1978). Comparative foraging behavior of six stingless bee species exploiting a standardized resource. *Ecology*, *59*(6), 1123–1136.

[98] Inoué, S. (2013). *Video microscopy*. Springer Science & Business Media.

[99] Jonkman, J., Brown, C., & Cole, R. (2014). Quantitative confocal microscopy: Beyond a pretty picture. In *Methods in Cell Biology*, volume 123 (pp. 113–134). Elsevier.

[100] Jordan, G., Deeb, S., Bosten, J., & Mollon, J. (2010). The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of Vision*, *10*(8), 12–31.

[101] Joshi, K. & Thakore, D. (2012). A survey on moving object detection and tracking in video surveillance system. *International Journal of Soft Computing and Engineering*, *2*(3), 44–48.

[102] Jung, K., Kim, K., & Jain, A. (2004). Text information extraction in images and video: A survey. *Pattern recognition*, *37*(5), 977–997.

[103] Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, *74*(7), 2561–2573.

[104] Karatas, O. & Toy, E. (2014). Three-dimensional imaging techniques: A literature review. *European Journal of Dentistry*, *8*(1), 132–140.

[105] Keshava, N. (2003). A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, *14*(1), 55–78.

[106] Keshava, N. & Mustard, J. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*, *19*(1), 44–57.

[107] Ketcham, R. (2005). Computational methods for quantitative analysis of three-dimensional features in geological specimens. *Geosphere*, *1*(1), 32–41.

[108] Khan, M., Khan, H., Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern trends in hyperspectral image analysis: A review. *IEEE Access*, *6*, 14118–14129.

[109] Kimura, T., Ohashi, M., Okada, R., & Ikeno, H. (2011). A new approach for the simultaneous tracking of multiple honeybees for analysis of hive behavior. *Apidologie*, *42*(5), 607–617.

[110] Kissling, W., Pattemore, D., & Hagen, M. (2014). Challenges and prospects in the telemetry of insects. *Biological Reviews*, *89*(3), 511–530.

[111] Kneen, M., Farinas, J., Li, Y., & Verkman, A. (1998). Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophysical journal*, *74*(3), 1591–1599.

[112] Kong, J. (2019). Digital pathology. `http://dp.gsu.edu/`.

[113] Kühl, H. & Burghardt, T. (2013). Animal biometrics: Quantifying and detecting phenotypic appearance. *Trends in Ecology & Evolution*, *28*(7), 432–441.

[114] Kumar, S. & Singh, S. (2016). Visual animal biometrics: Survey. *IET Biometrics*, *6*(3), 139–156.

[115] Lambert, T. (2019). FPbase: A community-editable fluorescent protein database. *Nature Methods*, *16*(4), 277–278.

[116] Landecker, H. (2006). Microcinematography and the history of science and film. *Isis*, *97*(1), 121–132.

[117] Landecker, H. (2011). Creeping, drinking, dying: The cinematic portal and the microscopic world of the twentieth-century cell. *Science in Context*, *24*(3), 381–416.

[118] Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. Application Delivery Strategies 949, META Group Inc.

[119] Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, *6*, 1–40.

[120] Laurberg, H., Christensen, M., Plumbley, M., Hansen, L., & Jensen, S. (2008). Theorems on positive data: On the uniqueness of NMF. *Computational Intelligence and Neuroscience*, *2008*.

[121] Laurberg, H. & Hansen, L. (2007). On affine non-negative matrix factorization. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, (pp. 653–656). IEEE.

[122] Lawson, C. & Hanson, R. (1995). *Solving least squares problems.* SIAM.

[123] Lee, D. & Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791.

[124] Lee, D. & Seung, H. (2000). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13*, (pp. 556–562)., Denver, Colorado, USA. NIPS.

[125] Leong, F., Brady, M., & McGee, J. (2003). Correction of uneven illumination (vignetting) in digital microscopy images. *Journal of Clinical Pathology*, *56*(8), 619–621.

[126] Lerner, J., Lockett, S., & Zucker, R. (2006). Spectral imaging. *Cytometry Part A*, *69A*(8), 711–946. special issue.

[127] Leutenegger, S., Chli, M., & Siegwart, R. (2011). BRISK: Binary robust invariant scalable keypoints. In *2011 International Conference on Computer Vision*, (pp. 2548–2555)., Barcelona, Spain. IEEE.

[128] Levenson, R., Fornari, A., & Loda, M. (2008). Multispectral imaging and pathology: Seeing and doing more. *Expert Opinion on Medical Diagnostics*, *2*(9), 1067–1081.

[129] Levinthal, C. & Ware, R. (1972). Three dimensional reconstruction from serial sections. *Nature*, *236*(5344), 207–210.

[130] Li, P., Wang, D., Wang, L., & Lu, H. (2018). Deep visual tracking: Review and experimental comparison. *Pattern Recognition*, *76*, 323–338.

[131] Li, Q., He, X., Wang, Y., Liu, H., Xu, D., & Guo, F. (2013). Review of spectral imaging technology in biomedical engineering: Achievements and challenges. *Journal of Biomedical Optics*, *18*(10), 100901.

[132] Li, Y., Wang, S., Tian, Q., & Ding, X. (2015). A survey of recent advances in visual feature detection. *Neurocomputing*, *149*, 736–751.

[133] Lichtman, J., Pfister, H., & Shavit, N. (2014). The big data challenges of connectomics. *Nature neuroscience*, *17*(11), 1448–1454.

[134] Lipkin, L., Watt, W., & Kirsch, R. (1966). The analysis, synthesis, and description of biological images. *Annals of the New York Academy of Sciences*, *128*(3), 984–1012.

[135] Liu, Y., Pu, H., & Sun, D.-W. (2017). Hyperspectral imaging technique for evaluating food quality and safety during various processes: A review of recent applications. *Trends in Food Science & Technology*, *69*, 25–35.

[136] Livesu, M., Ellero, S., Martínez, J., Lefebvre, S., & Attene, M. (2017). From 3D models to 3D prints: An overview of the processing pipeline. *Computer Graphics Forum*, *36*(2), 537–564.

[137] Livet, J., Weissman, T., Kang, H., Draft, R., Lu, J., Bennis, R., Sanes, J., & Lichtman, J. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, *450*(7166), 56–62.

[138] Lloyd-Smith, J., Schreiber, S., Kopp, P., & Getz, W. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, *438*(7066), 355–359.

[139] Lohr, S. (2013). Searching for origins of the term 'big data'. *New York Times, New York edition*, B4.

[140] Lotz, J., Olesch, J., Müller, B., Polzin, T., Galuschka, P., Lotz, J., Heldmann, S., Laue, H., González-Vallinas, M., Warth, A., et al. (2016). Patch-based nonlinear image registration for gigapixel whole slide images. *IEEE Transactions on Biomedical Engineering*, *63*(9), 1812–1819.

[141] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

[142] Lu, G. & Fei, B. (2014). Medical hyperspectral imaging: A review. *Journal of Biomedical Optics*, *19*(1), 010901.

[143] Lugmayr, A., Stockleben, B., Scheib, C., Mailaparampil, M., Mesia, N., & Ranta, H. (2016). A comprehensive survey on big-data research and its implications—what is really 'new' in big data?–it's cognitive big data! In *Proceedings of the 20th PACIS*, (pp. 248–262)., Chiayi, Taiwan. AIS.

[144] Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., & Kim, T.-K. (2014). Multiple object tracking: A literature review. *arXiv preprint arXiv:1409.7618*.

[145] Ma, W.-K., Bioucas-Dias, J., Chan, T.-H., Gillis, N., Gader, P., Plaza, A., Ambikapathi, A., & Chi, C.-Y. (2014). A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*, *31*(1), 67–81.

[146] Malandain, G., Bardinet, E., Nelissen, K., & Vanduffel, W. (2004). Fusion of autoradiographs with an mr volume using 2-D and 3-D linear transformations. *NeuroImage*, *23*(1), 111–127.

[147] Mark Welch, J., Rossetti, B., Rieken, C., Dewhirst, F., & Borisy, G. (2016). Biogeography of a human oral microbiome at the micron scale. *PNAS*, *113*(6), E791–E800.

[148] Mashey, J. (1997). Big data and the next wave of infraS-tress. Computer Science Division Seminar, University of California, Berkeley.

[149] Mashey, J. (1999). Big data and the next wave of infrastress problems, solutions, opportunities. In *Proceedings of the 1999 USENIX Annual Technical Conference*, Monterey, California, USA. USENIX.

[150] Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, *22*(10), 761–767.

[151] McCann, M., Ozolek, J., Castro, C., Parvin, B., & Kovacevic, J. (2015). Automated histology analysis: Opportunities for signal processing. *IEEE Signal Processing Magazine*, *32*(1), 78–87.

[152] McNamara, G., Gupta, A., Reynaert, J., Coates, T., & Boswell, C. (2006). Spectral imaging microscopy web sites and data. *Cytometry Part A: The Journal of the International Society for Analytical Cytology*, *69*(8), 863–871.

[153] McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B., Karhohs, K., Doan, M., Ding, L., Rafelski, S., Thirstrup, D., et al. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biology*, *16*(7), e2005970.

[154] Megjhani, M., Correa de Sampaio, P., Leigh Carstens, J., Kalluri, R., & Roysam, B. (2017). Morphologically constrained spectral unmixing by dictionary learning for multiplex fluorescence microscopy. *Bioinformatics*, *33*(14), 2182–2190.

[155] Meijering, E., Dzyubachyk, O., & Smal, I. (2012). Methods for cell and particle tracking. In *Methods in Enzymology*, volume 504 (pp. 183–200). Elsevier.

[156] Mersch, D., Crespi, A., & Keller, L. (2013). Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science*, *340*(6136), 1090–1093.

[157] Miranda, K., Girard-Dias, W., Attias, M., de Souza, W., & Ramos, I. (2015). Three dimensional reconstruction by electron microscopy in the life sciences: An introduction for cell and tissue biologists. *Molecular Reproduction and Development*, *82*(7-8), 530–547.

[158] Modersitzki, J. (2009). *FAIR: Flexible algorithms for image registration*. SIAM.

[159] Moeslund, T. & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, *81*(3), 231–268.

[160] Moeslund, T., Hilton, A., & Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, *104*(2-3), 90–126.

[161] Moll, R., Millspaugh, J., Beringer, J., Sartwell, J., & He, Z. (2007). A new view of ecology and conservation through animal-borne video systems. *Trends in Ecology & Evolution*, *22*(12), 660–668.

[162] Neher, R., Mitkovski, M., Kirchhoff, F., Neher, E., Theis, F., & Zeug, A. (2009). Blind source separation techniques for the decomposition of multiply labeled fluorescence images. *Biophysical journal*, *96*(9), 3791–3800.

[163] Newton, I. (1665). Of colours. MS Add. 3975, Cambridge University Library, Cambridge, UK.

[164] Newton, I. (2014). A letter of Mr. Isaac Newton, professor of the mathematicks in the University of Cambridge; containing his new theory about light and colors. *Philosophical Transactions of the Royal Society of London*, *6*(80), 3075–3087.

[165] Noldus, L., Spink, A., & Tegelenbosch, R. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*, *33*(3), 398–414.

[166] Oliveira, F. & Tavares, J. (2014). Medical image registration: A review. *Computer Methods in Biomechanics and Biomedical Engineering*, *17*(2), 73–93.

[167] Osborne, J., Clark, S., Morris, R., Williams, I., Riley, J., Smith, A., Reynolds,

D., & Edwards, A. (1999). A landscape-scale study of bumble bee foraging range and constancy, using harmonic radar. *Journal of Applied Ecology*, *36*(4), 519–533.

[168] Ourselin, S., Roche, A., Subsol, G., Pennec, X., & Ayache, N. (2001). Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing*, *19*(1), 25–31.

[169] Paatero, P. & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, *5*(2), 111–126.

[170] Pasquali, M. (2007). Protocol videos: The implications for research and society. *EMBO Reports*, *8*(8), 712–716.

[171] Pérez-Escudero, A., Vicente-Page, J., Hinz, R., Arganda, S., & de Polavieja, G. (2014). idTracker: Tracking individuals in a group by automatic identification of unmarked animals. *Nature methods*, *11*(7), 743–748.

[172] Phillips, M., Pinto, D., & Dobbie, I. (2018). SPEKcheck-fluorescence microscopy spectral visualisation and optimisation: A web application, javascript library, and data resource. *Wellcome Open Research*, *3*.

[173] Pichat, J., Iglesias, J., Yousry, T., Ourselin, S., & Modat, M. (2018). A survey of methods for 3D histology reconstruction. *Medical Image Analysis*, *46*, 73–105.

[174] Pichat, J., Modat, M., Yousry, T., & Ourselin, S. (2015). A multi-path approach to histology volume reconstruction. In *2015 IEEE 12th International Symposium on Biomedical Imaging*, (pp. 1280–1283)., Brooklyn Bridge, NY, USA. IEEE.

[175] Pluim, J., Maintz, J., & Viergever, M. (2003). Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, *22*(8), 986–1004.

[176] Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, *28*(6), 976–990.

[177] Prewitt, J. & Mendelsohn, M. (1966). The analysis of cell images. *Annals of the New York Academy of Sciences*, *128*(3), 1035–1053.

[178] Qureshi, R., Uzair, M., Khurshid, K., & Yan, H. (2019). Hyperspectral document image processing: Applications, challenges and future prospects. *Pattern Recognition*, *90*, 12–22.

[179] Recio, M., Mathieu, R., Denys, P., Sirguey, P., & Seddon, P. (2011). Lightweight GPS-tags, one giant leap for wildlife tracking? an assessment approach. *PLoS One*, *6*(12), e28225.

[180] Ricketts, T. (2001). The matrix matters: Effective isolation in fragmented landscapes. *The American Naturalist*, *158*(1), 87–99.

[181] Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science*, *316*(5829), 1298–1301.

[182] Roberts, N., Magee, D., Song, Y., Brabazon, K., Shires, M., Crellin, D., Orsi, N., Quirke, R., Quirke, P., & Treanor, D. (2012). Toward routine use of 3D histopathology as a research tool. *The American Journal of Pathology*, *180*(5), 1835–1842.

[183] Rohlfing, T. (2012). Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. *IEEE Transactions on Medical Imaging*, *31*(2), 153–163.

[184] Rossetti, B., Dynes, T., Brosi, B., de Roode, J., & Kong, J. (2018). GRAPHITE: A graphical environment for scalable in situ video tracking of moving insects. *Methods in Ecology and Evolution*, *9*(4), 956–964.

[185] Rossetti, B., Wang, F., Zhang, P., Teodoro, G., Brat, D., & Kong, J. (2017). Dynamic registration for gigapixel serial whole slide images. In *2017 IEEE 14th International Symposium on Biomedical Imaging*, (pp. 424–428). IEEE.

[186] Rossetti, B., Wilbert, S., Mark Welch, J., Borisy, G., & Nagy, J. (2019). Semi-blind sparse affine spectral unmixing of autofluorescence-contaminated micrographs. *bioRxiv*, 529008.

[187] Rosten, E. & Drummond, T. (2005). Fusing points and lines for high performance tracking. In *2005 IEEE International Conference on Computer Vision*, volume 2, (pp. 1508–1515)., Beijing, China. IEEE.

[188] Rowcliffe, J. & Carbone, C. (2008). Surveys using camera traps: Are we looking to a brighter future? *Animal Conservation*, *11*(3), 185–186.

[189] Rubenstein, D. & Hobson, K. (2004). From birds to butterflies: Animal movement patterns and stable isotopes. *Trends in Ecology & Evolution*, *19*(5), 256–263.

[190] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*, (pp. 2564–2571)., Barcelona, Spain. IEEE.

[191] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., & Hawkes, D. (1999). Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, *18*(8), 712–721.

[192] Rueden, C., Schindelin, J., Hiner, M., DeZonia, B., Walter, A., Arena, E., & Eliceiri, K. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, *18*(1), 529–555.

[193] Ruiz, A., Ujaldon, M., Cooper, L., & Huang, K. (2009). Non-rigid registration

for large sets of microscopic images on graphics processors. *Journal of Signal Processing Systems*, *55*(1-3), 229–250.

[194] Salahat, E. & Qasaimeh, M. (2017). Recent advances in features extraction and description algorithms: A comprehensive survey. In *2017 IEEE International Conference on Industrial Technology*, (pp. 1059–1063)., Toronto, Ontario, Canada. IEEE.

[195] Sansoni, G., Trebeschi, M., & Docchio, F. (2009). State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, *9*(1), 568–601.

[196] Schermelleh, L., Ferrand, A., Huser, T., Eggeling, C., Sauer, M., Biehlmaier, O., & Drummen, G. (2019). Super-resolution microscopy demystified. *Nature Cell Biology*, *21*(1), 72–84.

[197] Schmitt, O., Modersitzki, J., Heldmann, S., Wirtz, S., & Fischer, B. (2007). Image registration of sectioned brains. *International Journal of Computer Vision*, *73*(1), 5–39.

[198] Schröck, E., Du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M., Ning, Y., Ledbetter, D., Bar-Am, I., Soenksen, D., et al. (1996). Multicolor spectral karyotyping of human chromosomes. *Science*, *273*(5274), 494–497.

[199] Schwier, M., Böhler, T., Hahn, H., Dahmen, U., Dirsch, O., et al. (2013). Registration of histological whole slide images guided by vessel structures. *Journal of Pathology Informatics*, *4*(2), 10.

[200] Shaw, G. & Burke, H. (2003). Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, *14*(1), 3–28.

[201] Shi, J. & Tomasi, C. (1993). Good features to track. Technical report, Cornell University.

[202] Sivarajah, U., Kamal, M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286.

[203] Skrinjar, O. (2006). Point-based registration with known correspondence: Closed form optimal solutions and properties. In *Third International Workshop on Biomedical Image Registration*, (pp. 315–321)., Utrecht, Netherlands. Springer.

[204] Smeulders, A., Chu, D., Cucchiara, R., Calderara, S., Dehghan, A., & Shah, M. (2013). Visual tracking: An experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(7), 1442–1468.

[205] Smith, R. (2007). An overview of the Tesseract OCR engine. In *9th International Conference on Document Analysis and Recognition*, volume 2, (pp. 629–633)., Curitiba, Paranà, Brazil. IEEE.

[206] Song, Y., Treanor, D., Bulpitt, A., Wijayathunga, N., Roberts, N., Wilcox, R., & Magee, D. (2014). Unsupervised content classification based nonrigid registration of differently stained histology images. *IEEE Transactions on Biomedical Engineering*, *61*(1), 96–108.

[207] Sotiras, A., Davatzikos, C., & Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, *32*(7), 1153–1190.

[208] Specht, E., Braselmann, E., & Palmer, A. (2017). A critical and comparative review of fluorescent tools for live-cell imaging. *Annual Review of Physiology*, *79*(1), 93–117.

[209] Steen, R. (2016). Diel activity, frequency and visit duration of pollinators in focal plants: in situ automatic camera monitoring and data processing. *Methods in Ecology and Evolution*, *8*(2), 203–213.

[210] Sternberg, S. (1983). Biomedical image processing. *Computer*, *16*(1), 22–34.

[211] Stockman, A. & Sharpe, L. (2000). The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, *40*(13), 1711–1737.

[212] Stramer, B. & Dunn, G. (2015). Cells on film–the past and future of cinemicroscopy. *Journal of Cell Science*, *128*(1), 9–13.

[213] Struck, R., Cordoni, S., Aliotta, S., Pérez-Pachón, L., & Gröning, F. (2019). Application of photogrammetry in biomedical science. In *Biomedical Visualisation* (pp. 121–130). Springer.

[214] Stytz, M., Frieder, G., & Frieder, O. (1991). Three-dimensional medical imaging: Algorithms and computer systems. *ACM Computing Surveys*, *23*(4), 421–499.

[215] Suetens, P. (2017). *Fundamentals of medical imaging*. Cambridge university press.

[216] Sun, Y., Wallrabe, H., Booker, C., Day, R., & Periasamy, A. (2010). Three-color spectral FRET microscopy localizes three interacting proteins in living cells. *Biophysical Journal*, *99*(4), 1274–1283.

[217] Suvarna, K., Layton, C., & Bancroft, J. (2018). *Bancroft's theory and practice of histological techniques*. Elsevier Health Sciences.

[218] Sze, G., De Armond, S., Brant-Zawadzki, M., Davis, R., Norman, D., & Newton, T. (1986). Foci of MRI signal (pseudo lesions) anterior to the frontal horns:

histologic correlations of a normal finding. *American Journal of Roentgenology*, *147*(2), 331–337.

[219] Tareen, S. & Saleem, Z. (2018). A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In *2018 International Conference on Computing, Mathematics and Engineering Technologies*, (pp. 1–10)., Sukkur, Pakistan. IEEE.

[220] Tong, L., Zhou, J., Qian, Y., Bai, X., & Gao, Y. (2016). Nonnegative-matrix-factorization-based hyperspectral unmixing with partially known endmembers. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(11), 6531–6544.

[221] Torr, P. & Zisserman, A. (2000). MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, *78*(1), 138–156.

[222] Trucco, E. & Plakas, K. (2006). Video tracking: A concise survey. *IEEE Journal of Oceanic Engineering*, *31*(2), 520–529.

[223] Truong, K. & Ikura, M. (2001). The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo. *Current Opinion in Structural Biology*, *11*(5), 573–578.

[224] Tu, G., Hansen, M., Kryger, P., & Ahrendt, P. (2016). Automatic behaviour analysis system for honeybees using computer vision. *Computers and Electronics in Agriculture*, *122*, 10–18.

[225] Turchin, P. (1998). *Quantitative analysis of movement: Measuring and modeling population redistribution in animals and plants.* Sinauer Associates Sunderland.

[226] Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(4), 376–380.

[227] Valm, A., Cohen, S., Legant, W., Melunis, J., Hershberg, U., Wait, E., Cohen, A., Davidson, M., Betzig, E., & Lippincott-Schwartz, J. (2017). Applying systems-level spectral imaging and analysis to reveal the organelle interactome. *Nature*, *546*(7656), 162–167.

[228] Valm, A., Mark Welch, J., Rieken, C., Hasegawa, Y., Sogin, M., Oldenbourg, R., Dewhirst, F., & Borisy, G. (2011). Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *PNAS*, *108*(10), 4152–4157.

[229] Valm, A., Oldenbourg, R., & Borisy, G. (2016). Multiplexed spectral imaging of 120 different fluorescent labels. *PLoS One*, *11*(7), e0158495.

[230] Viergever, M., Maintz, J., Klein, S., Murphy, K., Staring, M., & Pluim, J. (2016). A survey of medical image registration—under review. *Medical Image Analysis*, *33*, 140–144.

[231] Waller, L. & Tian, L. (2015). Computational imaging: Machine learning for 3d microscopy. *Nature*, *523*(7561), 416–417.

[232] Wang, C., Gosno, E., & Li, Y. (2015). Fully automatic and robust 3D registration of serial-section microscopic images. *Scientific Reports*, *5*, 15051.

[233] Wang, L., Wang, Z., Zhang, Y., & Li, X. (2013). How human location-specific contact patterns impact spatial transmission between populations? *Scientific Reports*, *3*.

[234] Ware, R. & LoPresti, V. (1975). Three-dimensional reconstruction from serial sections. In *International Review of Cytology*, volume 40 (pp. 325–440). Elsevier.

[235] Wario, F., Wild, B., Couvillon, M., Rojas, R., & Landgraf, T. (2015). Auto-

matic methods for long-term tracking and the detection and decoding of communication dances in honeybees. *Frontiers in Ecology and Evolution*, *3*, 103.

[236] Waters, J. (2009). Accuracy and precision in quantitative fluorescence microscopy. *Journal of Cell Biology*, *185*(7), 1135–1148.

[237] Weinstein, B. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, *87*(3), 533–545.

[238] Wilson, J. (1910). Improved methods of utilising organised structures as directing marks for plastic reconstruction, and other notes on microscopical technique. *Zeitschrift fur Wissenschaftliche Mikroscopie*, 227–234.

[239] Woolfe, F., Gerdes, M., Bello, M., Tao, X., & Can, A. (2011). Autofluorescence removal by non-negative matrix factorization. *IEEE Transactions on Image Processing*, *20*(4), 1085–1093.

[240] Wu, B., Klatzky, R., & Stetten, G. (2012). Mental visualization of objects from cross-sectional images. *Cognition*, *123*(1), 33–49.

[241] Wu, Y., Lim, J., & Yang, M.-H. (2013). Online object tracking: A benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition 2013*, (pp. 2411–2418)., Portland, Oregon, USA. IEEE.

[242] Xu, Y., Pickering, J., Nong, Z., Gibson, E., Arpino, J.-M., Yin, H., & Ward, A. (2015). A method for 3D histopathology reconstruction supporting mouse microvasculature analysis. *PLoS One*, *10*(5), e0126817.

[243] Yang, B. & Nevatia, R. (2012). Online learned discriminative part-based appearance models for multi-human tracking. In *12th European Conference on Computer Vision*, (pp. 484–498)., Florence, Italy. Springer.

[244] Yang, H., Shao, L., Zheng, F., Wang, L., & Song, Z. (2011). Recent advances and trends in visual tracking: A review. *Neurocomputing*, *74*(18), 3823–3831.

[245] Yi, K., Trulls, E., Lepetit, V., & Fua, P. (2016). LIFT: Learned invariant feature transform. In *14th European Conference on Computer Vision*, (pp. 467–483)., Amsterdam, Netherlands. Springer.

[246] Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys*, *38*(4).

[247] Yin, X.-C., Zuo, Z.-Y., Tian, S., & Liu, C.-L. (2016). Text detection, tracking and recognition in video: A comprehensive survey. *IEEE Transactions on Image Processing*, *25*(6), 2752–2773.

[248] Zack, G., Rogers, W., & Latt, S. (1977). Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry*, *25*(7), 741–753.

[249] Zankel, A., Wagner, J., & Poelt, P. (2014). Serial sectioning methods for 3D investigations in materials science. *Micron*, *62*, 66–78.

[250] Zare, A. & Ho, K. (2014). Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, *31*(1), 95–104.

[251] Zhu, Y., Yao, C., & Bai, X. (2016). Scene text detection and recognition: Recent advances and future trends. *Frontiers of Computer Science*, *10*(1), 19–36.

[252] Zimmermann, T., Marrison, J., Hogg, K., & O'Toole, P. (2014). Clearing up the signal: Spectral imaging and linear unmixing in fluorescence microscopy. In *Confocal Microscopy* (pp. 129–148). Springer.

[253] Zitová, B. & Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, *21*(11), 977–1000.