Chapter 1
Introduction to Big Data Microscopy Experiments

1.1 Image Processing Pipeline

An image is an array of picture elements (called pixels) arranged in columns and rows. At every column and row, the pixel has one or more values. For example, images acquired by phase contrast microscopes have single-value pixels, while images collected using bright field or Raman spectroscopy microscopes have multiple-value pixels. Microscope images can be acquired over time to form a video or over multiple z-depths to form a 3D volume. These videos and 3D volumes are still images, each represented by an array of pixels arranged in columns, rows, and either time frames or z-stacks. The image acquisition process is referred to as imaging (see Fig. 1.1) and yields images of a specimen of interest (also denoted as sample). The imaging instrument of interest in this book is a microscope, and the series of computational steps applied to acquired images is the web processing pipeline.

Image processing
Image processing refers to algorithms that take images as inputs and return images as outputs.[1] Image processing typically performs mathematical operations on images to eliminate imaging artifacts, enhance image content, integrate multiple images into the same coordinate system, prepare images for information extraction, or any combination of these operations. For instance, flat field correction eliminates imaging artifacts due to illumination inhomogeneities introduced during imaging. Gaussian filtering enhances image regions buried in Gaussian noise added during imaging by microscope digital circuitry. Segmentation extracts the locations of objects of interest contained in a single image (frequently denoted as one field of view or FOV in microscopy). Image stitching integrates multiple images with a partial spatial overlap to create one large image containing objects of interest spanning a large spatial area.

[1] http://www.coe.utah.edu/~cs4640/slides/Lecture0.pdf

P. Bajcsy et al., *Web Microanalysis of Big Image Data*,
https://doi.org/10.1007/978-3-319-63360-2_1

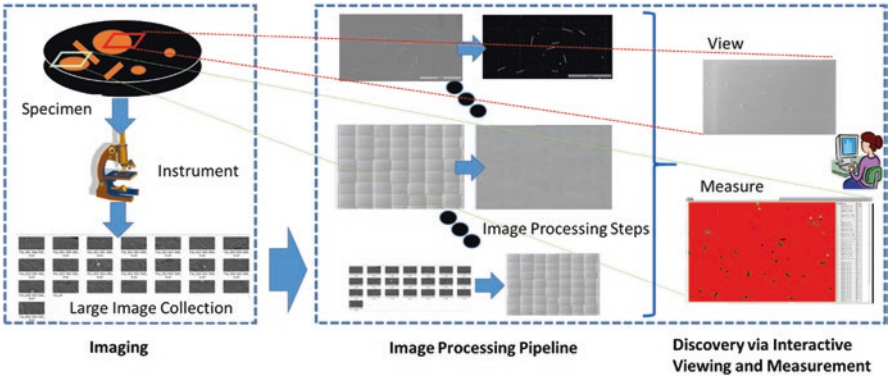

Fig. 1.1 An illustration of web image processing pipeline as a mechanism for enabling discovery via interactive viewing and measurements of very large collections of images

Image processing pipeline

The many applications of image processing include visualization, information restoration, image retrieval, registration, pattern measurement, object detection, and object recognition.[2] We focus on image processing applications that enable discovery over very large image collections via viewing and quantitative measurements of objects of interest. Scientific discovery is characterized by a workflow (also denoted as a pipeline) of image processing steps. The pipeline is designed by a scientist and depends on the discovery method. Image processing pipelines require human input not only when chaining computational steps but also when selecting computational parameters, choosing information measurements to extract and visualize data, and exploring the semantic meaning derived from image measurements.

Web image processing pipeline

With advancements in microscopy imaging, one experiment can yield large quantities of images that are beyond the processing capabilities of typical personal computers. These processing capabilities include off-line and interactive computations that are critical for making discoveries. We refer to these quantities of images as big image data. In addition, the path to a discovery requires frequent sharing of big image data, intermediate large-scale measurements, and explorations by multiple researchers with varying expertise. These realizations lead us to web image processing pipelines where the processing capabilities (hardware and software) can be scaled to the size of the image data and computational time requirements, while the data sharing and collaborative discovery are supported by the distributed access via web browsers.

[2] http://www.engineersgarage.com/articles/image-processing-tutorial-applications

1.2 Web Image Processing Pipeline

In the context of this book, a web image processing pipeline consists of a client-server system and the algorithms that are executed either on the client side or on the server side.

Client-server system
A client-server system can perform computation steps either off-line or interactively (on-demand). For simplicity, the term client can be understood as a web browser running on a researcher's computer or another device. The web browser allows the researcher to view images via computer/device displays and make measurements in the browser environment using the underlying hardware. The term server can be viewed as a web-networked computer (or multiple computers) capable of storing large image collections, serving the images to multiple clients, and handling requests for uploading, computing, and downloading. Data sharing and collaborative discovery is facilitated by the ability of multiple web clients to communicate with a server from any web-networked geographical location. The ability of a server to distribute storage and computational requests from many clients during peak usage to other available computational resources provides the web image processing pipeline with the ability to scale with data size and computational requests. A client-server-based web image processing pipeline requires software algorithms running across diverse web browsers using a server environment.

Client software
Software algorithms running in a web browser are written in JavaScript and are integrated with web technologies such as Hypertext Markup Language (HTML) and Cascading Style Sheets (CSS). While HTML is the language for creating web pages, CSS is the language for describing the style of an HTML document. Unfortunately, existing web browsers do not support the same features of JavaScript, HTML, and CSS languages. Thus, some web image processing steps and interactive features are supported only in a subset of browsers. To address the interoperability, open standards for the web features are critical for the long-term growth of the web and are being addressed by the World Wide Web Consortium (W3C). For example, the W3C HTML Working Group prepared the 5th revision of HTML (HTML5) in 2014 which is used in the web image processing pipeline described in the next chapters.

Server software
The server-side software must communicate with clients and must run user-requested computational algorithms. This communication can be mediated by software such as Apache Tomcat that allows servers and browser-based clients to exchange information similar to a conversation between people. Computational algorithms that are written in multiple programming languages must be seamlessly integrated into the communication between clients and servers. For example, a communication request to stitch a set of overlapping microscopy image tiles into a single image can require the launching of a stitching algorithm with input images and input parameters and the retrieval of the stitching output (i.e., translation vector per

image tile). This assumes that the stitching algorithm has been compiled for the server operating system (OS), can access the input data and output storage, and can be executed in user mode on the underlying hardware [1]. The execution of the same algorithm on distributed and heterogeneous environments requires the addition of software that will distribute the data, schedule the computations, and collect the results.

Client-server communication
The communication and data exchange channel is an important aspect of client-server software design. In practice, this channel is almost always limited by the network bandwidth which has implications for the latency and interactivity of on-demand computations. To overcome latency, especially for very large image collections, images are frequently compressed and use special representations for fast retrieval. For example, large gigapixel images can be represented by a tiled multi-resolution pyramid that is created by iteratively down-sampling the original image size by one half. Each down-sampled image is then tiled into a predefined size (e.g., 256 pixels × 256 pixels). This representation allows for the transmission of smaller more manageable subregions of a gigapixel image (i.e., set of pyramid tiles) for a view or computation.

1.3 Big Data Microscopy Experiments

We will focus on big data microscopy experiments in cell biology and materials science. As mentioned above, these microscopy experiments can yield large image datasets that are beyond the processing capabilities of a typical personal computer. These images are acquired as collections of individual microscope fields of view (FOV). For a fixed acquisition rate, the big image collection originates from imaging a large spatial area at high magnification over extended time and from using many spectral channels (imaging modalities). The image collection also grows with the number of specimens that are sampled from a large specimen bank (e.g., cells from a vial). These "replicate" measurements aim at establishing statistical reproducibility, as well as at understanding response functions under many treatments.

Examples of big data microscopy experiments in cell biology
In cell biology, as well as in histopathology, microscopy imaging provides raw image data about cells and tissues. The research frequently involves studying either population statistics or characteristics of individual cells. The image of an entire specimen is preferable to spot checking randomly selected FOVs for population-level statistics. The spot-checking approach fails to capture rare events or can bias the biological interpretation of cells. There is a need to collect enough data to capture these heterogeneities. Furthermore, in cell therapy applications, it is possible that each cell matters and must be imaged and inspected for quality. Big data microscopy experiments in cell biology are becoming more and more frequent, while the computational solutions for dealing with acquired big data are not yet

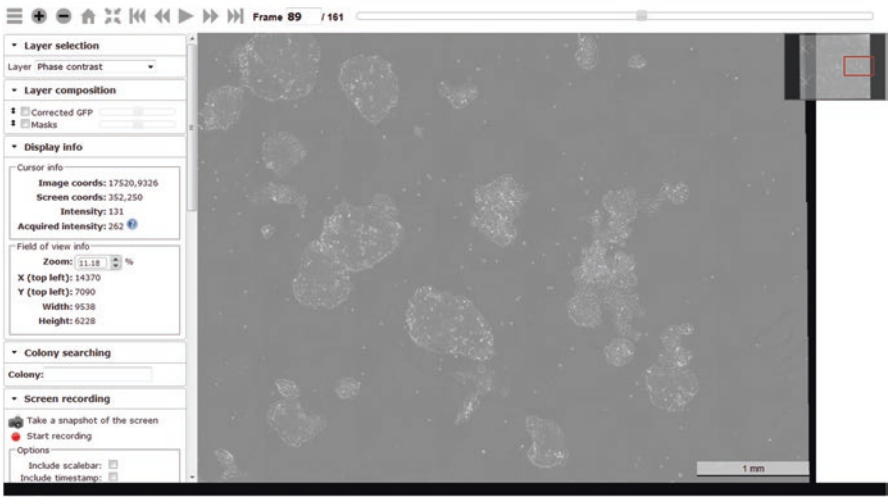

Fig. 1.2 Phase microscopy images of cell colonies imaged over 396 (18 × 22) overlapping FOVs and over a period of 5 days (161 time points at 45 min acquisition rate). The data is available online (https://isg.nist.gov/deepzoomweb/data/stemcellpluripotency)

available. An example of such a big data experiment focusing on cell colony growth and heterogeneity of pluripotency marker is presented in Fig. 1.2.

Examples of big data microscopy experiments in materials sciences
In materials science, microscopy imaging provides raw images about materials and their properties. The research goals include studying properties of materials (ceramics, metals, or polymers), investigating mixtures of materials (compounds, alloys, composites), looking for occurrence of rare particles, or discovering physical and chemical properties of metallic elements and their interactions. The key challenge of these studies lies with the material specimens being much larger than their elements and requiring the imaging of a large area at high spatial resolution. Furthermore, the materials consist of heterogeneous elements that are unevenly distributed and hence spatial sampling must provide sufficient information about the heterogeneity. Figure 1.3 presents an example of a big data experiment focusing on the shape and distribution of aerosolized nanoparticles.

Differences between two application domains
The key difference between the cell biology and materials science application domains lies in the experimental constraints. In order to study living cells, one must choose cell specimen preparation and imaging modalities that do not harm cells and do not change their behavior. Thus, the preferred imaging modality for cell biology is optical microscopy (e.g., bright field, phase contrast, or differential interference contrast). Other microscopy imaging modalities are used to understand cell structure at multiple scales but cannot be used for studying living cells since cells are either fixed

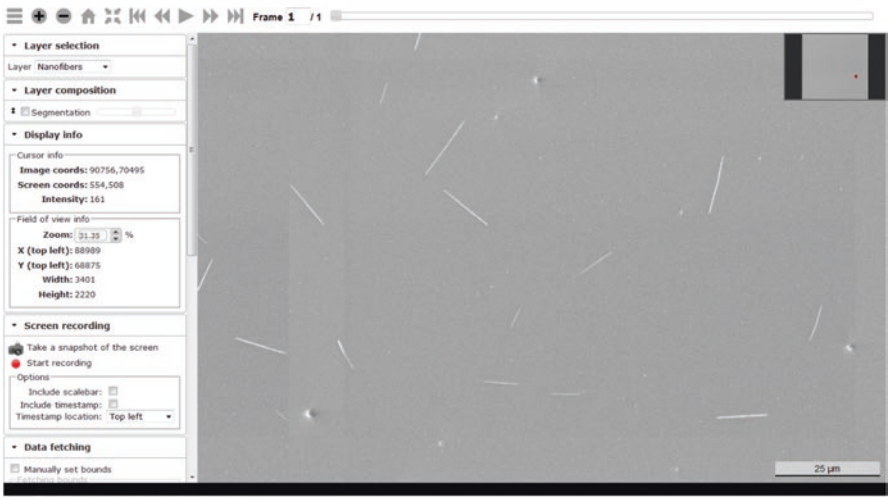

Fig. 1.3 Images of aerosolized carbon nanotubes collected on a Si wafer imaged over 2360 overlapping FOVs. The data is available online (https://isg.nist.gov/deepzoomweb/data/materialparticlesdistribution)

or destroyed during imaging. Any steerable experiments with living cells also require near real-time processing of images to provide feedback to the experimental configuration (e.g., selected field of view). In order to discover and design new materials in materials science, one is typically more concerned with high spatial and temporal image resolutions and less concerned with destroying the specimen during experiments. In other words, studying material dynamics does not assume that a specimen is unchanged during an imaging experiment while studying living cells assumes that a specimen did not change due to the measurement interrogation during imaging.

1.4 Motivation of Big Data Microscopy Experiments

Measurements at multiple spatial scales
One of the fundamental motivations for conducting big data microscopy experiments is the desire to perform ensemble "bulk" and individual object microscopic measurements. Figure 1.4 illustrates the characteristics of the two types of measurements (top) and the benefits for scientists from big data microscopy experiments (bottom). The measurement problems lie in the large differences between the scales of a specimen and the scale of an observed phenomenon (i.e., cells and subcellular structures are much smaller than the entire Petri dish).

Complexity of studied phenomena
Another motivation for collecting big data is the discovery of complex governing laws, for example, those of biological cells and tissues. Bio- and material

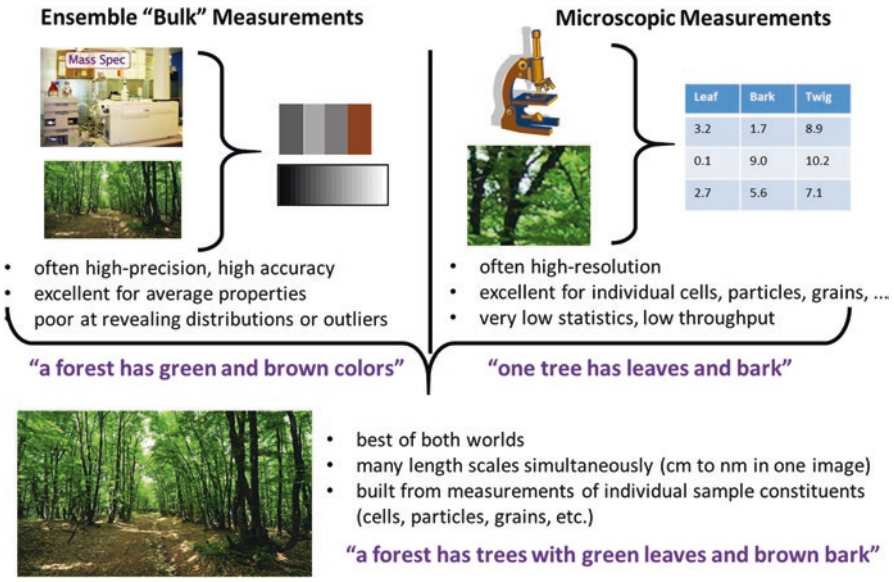

Fig. 1.4 Comparison of today's characteristics of ensemble "bulk" and microscopic measurements (*top*). If solutions for processing big data experiments become available, then a new type of measurement science is enabled

complexity is characterized by many factors, and interactions, and by many temporal and spatial scales. This complexity is related to whole cell and particle responses to various treatments and environments while measuring morphology, function, and spatial and temporal distributions. To study such complex systems, one must collect cell data over the huge space of factors, using many imaging modalities and at many temporal and physical spatial scales leading to big data microscopy experiments.

In addition, the inherent complexity requires leveraging the multidisciplinary expertise of geographically distributed team members and hence the acquired data must be accessible by all team members.

Statistical properties of measurements

By observing stochastic processes changing as a function of many factors, scientists are frequently deriving conclusions from a limited number of observations (i.e., statistical samples of the underlying stochastic process). This raises questions about (a) how many observations to collect in order to derive conclusions with high statistical confidence and (b) which observation to collect in order to acquire representative statistical samples. If statistical samples are chosen based on assumptions about a specimen that do not hold, then validity and reproducibility of data-driven inferences and models (scientific results) are in jeopardy. There is thus a need to run big data experiments to collect as much data as possible to validate appropriate sampling and verify confidence in the conclusions. The statistical benefits of so-called "census" measurements are high and include increased robustness and statistical significance of measurements at the cost of infrastructure investments [2].

Provenance trail of measurements
In scientific discovery from big image data, scientists often struggle with the reproducibility of their analyses. Lack of reproducibility can be attributed to the large number of processing steps associated with algorithmic versions, parameter settings, operating system dependencies, and hardware-specific configurations. It is difficult to keep all intermediate results, software versions, and metadata about configurations hence the analyses may be performed on multiple computers by several scientists. These metadata, describing the path from imaging to reported results (referred to as a provenance trail), are critical for the transparency of published work and traceability of measurements. One of the benefits of a web image processing pipeline is that it can capture the provenance trail. By connecting intermediate results and metadata files using web hyperlinks, scientific results become transparent and traceable.

From scientific research to practice
In addition to meeting the needs of a scientist, big data experiments can have large benefits for the manufacturing sector with quality control and quality assurance. For example, with some luck, rare events during bio-manufacturing of cell therapies can be observed by random spot-checking. However, quantifying the frequency of the rare occurrence requires a well-designed big data experiment, automated processing, and the determination of an overall probability distribution function (PDF). If a rare event has a critical impact, then industries using high-throughput imaging and performing quality assessment may be interested in web image processing pipelines.

Summary of goals of big data microscopy imaging experiments
In a summary, the combination of big data microscopy imaging experiments and web image processing pipelines enable the user:

- To understand multi-scale relationships of phenomena
- To study complex governing laws of phenomena
- To improve statistical properties of measurements
- To facilitate transparency and traceability of research results
- To enable reliable quality control and quality assurance

The ultimate goal for many scientists is also to transition scientific results to bio-manufacturing environments. This transition is much easier if scientific results are accessible and traceable. The web image processing pipeline described in this book is viewed as a tool to make the transition easier. It is also viewed as an open-source component of a cloud-based laboratory information management system.[3] In research labs, it is viewed as an infrastructure that enhances access to expertise and measurement tools.

[3] https://appexchange.salesforce.com/listingDetail?listingId=a0N30000008YhTtEAK

1.5 Range of Applications Leveraging Image Processing Pipelines

Applications at the cell level
Live cell microscopy imaging has applications for developing stem cell therapies, for advancing regenerative medicine, and for designing drugs. Whether the cells are administered to the body to benefit the recipient (cell therapy) or the cells are regenerated into tissues and organs to restore normal functions (regenerative medicine),[4] they must be inspected for quality. Microscopy imaging with image processing provides one of the quality assurance tools. Image processing pipelines can be used to monitor cell responses to a large space of treatments during the design process for new therapies, ultimately serving as a replacement for the highly manual and less statistically significant visual spot-checking.

Applications at the tissue level
A web image processing pipeline can help with analyzing histopathology images. These images are collected from biopsies and pose a challenge due to the volume of data, complexity of image content, and a desired short turnaround time on measurements. In comparison to live cell microscopy images, histopathology images are typically stained with hematoxylin and eosin (H&E) and imaged as color images. There is an entire branch of digital pathology devoted to analyzing histopathology images [3]. It is conceivable that digital pathology combined with tailored medical treatment to the individual characteristics of each patient would play a role in an emerging approach to disease treatment and prevention referred to as precision medicine.

Applications at the organ level
Microscopy imaging and image processing have been used to study the brain. Due to the unprecedented size of brain imaging datasets, this application requires sharing data, algorithms, and results, as well as joint collaborations across multiple funding agencies and research institutions. Web image rendering and annotation systems have been the key in the Human Connectome[5] project funded by federal agencies, FlyEM[6] project funded by the Howard Hughes Medical Institute (HHMI) Janelia Farm, and Allen Brain Atlas[7] funded by the Allen Institute for Brain Science.

Materials science applications
A web image processing pipeline can be a tool for material scientists to study images and make measurements of nanoscale particles, for example, that can cause health hazards for exposed humans or failures in engines and building materials. In addition to safety- and quality-related static particle measurements, dynamic measurements are needed to prevent chemical reactions during operations, such as detection of destabilizing chemicals in lithium ion batteries used frequently in cell phones and laptops [4].

---

[4] http://stemcellassays.com/2011/12/distinction-cell-therapy-regenerative-medicine/

[5] http://www.openconnectomeproject.org/ and https://www.humanconnectome.org/

[6] https://www.janelia.org/project-team/flyem

[7] http://brain-map.org/

Other applications generating large size image collections
While this book is focused on microscopy imaging, a web image processing pipeline can be applied to satellite and airborne imaging in geospatial information systems (GIS), as well as to telescopic imaging in astronomy. To adjust to specific measurement objectives of each application domain, image processing steps (algorithms) must be specifically designed, and can be easily integrated into the web pipeline architecture.

1.6 Challenges of Big Data Microscopy Experiments

Given the examples of big data experiments, web image processing pipelines can play a significant role in the transition from big data to knowledge, enabling technologies for real-time control of experiments (i.e., steerable experiments). Extracting knowledge from big image data is challenged by its size and complexity. Steerable experiments are additionally challenged by the limited time for processing, when the imaging microscope requires human input for changing FOVs based on the data it is acquiring. The challenges of big data experiments related to scale, complexity, and speed suggest that web image processing pipelines might offer a viable solution. They hold the promise of managing large image collections and utilizing all investments to acquire big data, sharing the data, providing access for geographically distributed teams, and distributing and accelerating computations that have time-critical aspects. Next, we describe the three main challenges: scale, complexity, and speed.

Data scale
To illustrate changes in data acquisition rates that directly contribute to large image scales, Fig. 1.5 and Table 1.1 summarize the acquisition rates of several microscopes at National Institute of Standards and Technology (NIST). While there are already cameras on the market that can acquire images at a rate of 1 terabyte (TB) per 3 min, our laptops and office/lab desktops are not ready to process at this rate. One can see that one microscope can generate 15 TB in 45 min which is equivalent to all the data in the American Library of Congress that have been gathered since 1800 (i.e., more than 158 million items in 460 languages). The amount of data being generated is currently doubling roughly every 18 months,[8] and the challenge of working at this scale is not going to go away.

Image content complexity
Another property of big data experiments is the complexity of studied phenomena that goes beyond a single discipline, a single expertise, and many times a single institution. The phenomena in cell biology and materials science require a variety of geographically distributed inputs to design image models of multimodal time-dependent TB-sized images over multiple scales. The complexity of image content is illustrated in Fig. 1.6.

[8] http://www.datanami.com/2015/08/18/beware-the-dangers-of-dark-data/

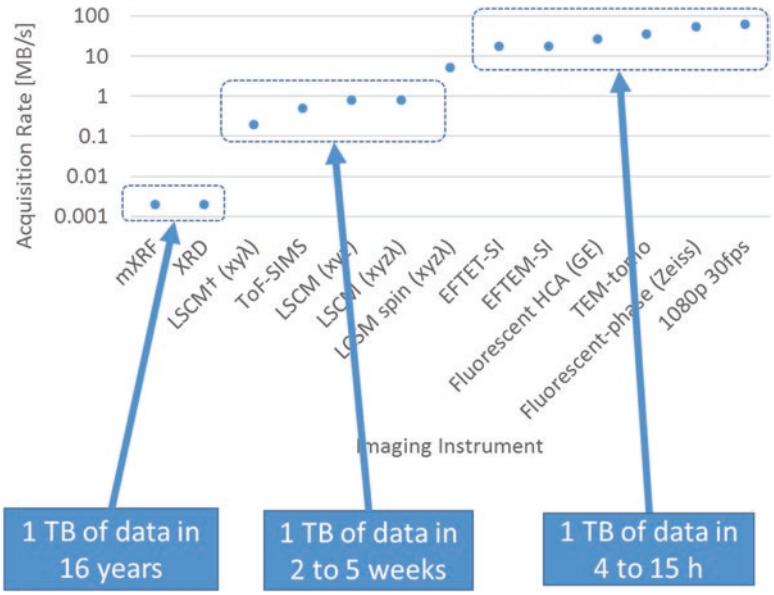

Fig. 1.5 Examples of acquisition data rates by a variety of microscopes

Table 1.1 Microscope abbreviations used in Fig. 1.5

mXRF	Micro X-ray fluorescence spectrometry
XRD	X-ray diffraction imaging
LSCM	Laser scanning confocal microscopy
ToF-SIMS	Time-of-flight secondary ion mass spectrometry
EFTET-Si	Energy-filtered transmission electron tomography-spectral imaging
EFTEM-Si	Energy-filtered transmission electron microscopy
TEM-tomo	Transmission electron microscopy-tomography
1080p 30fps	Wilco Imaging's second generation WIL-HD1080p with 1/3″ Complementary Metal–Oxide–Semiconductor (CMOS) Panasonic sensor, 1080 horizontal lines (progressive scan) and 1920 vertical lines per image (1080p), and 30 frames per second (fps), also denoted as High-Definition (HD) Serial Digital Interface (SDI) camera
Fluorescent HCA (GE)	High content analysis, General Electric (GE) InCell 2200
LSCM spin (xyzλ)	Spinning disk laser scanning confocal microscopy
Fluorescent-phase (Zeiss)	Zeiss Axiovert 200 M fluorescence/live cell imaging microscope

Given TB-sized videos with gigapixel frames, automation of measurements is inevitable. Automated analyses depend on designing models at macro (centimeter) to micro (nanometer) spatial scales over time and imaging modalities. The design of models for algorithms (i.e., segmentation and tracking) relies on limited human visual inspection or other reference measurements. The challenges of this type of complexity require solutions to become interactive so that a computer and a human can work in a tandem.

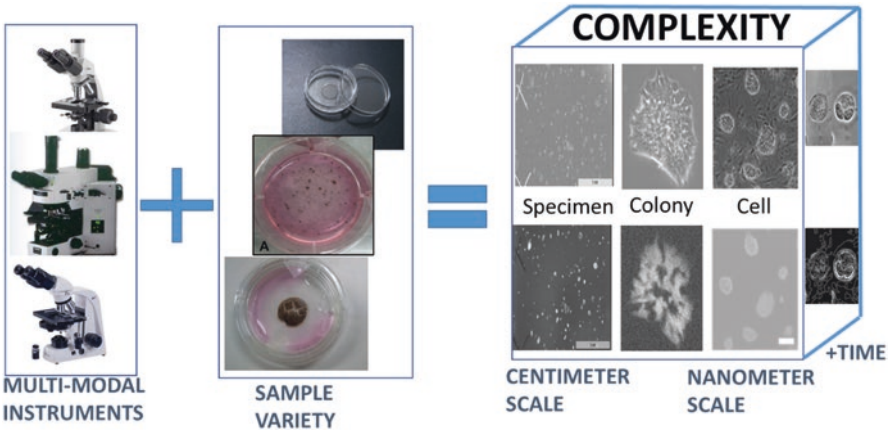

Fig. 1.6 Complexity of image content when imaging cell colonies using multimodal microscopes, over a variety of specimens, in space and time

Speed
Finally, the speed of image processing must also be considered. For instance, if a doctor is depending upon microscope images to monitor a cell therapy being delivered to a patient, and processing that data takes longer than a transition in the state of those cells, there is a risk in delivering the cell therapy. The current computational times on a personal computer are limited to its hardware specifications. For example, after acquiring 2 TB of data in 2 min, it would take (33 to 66) min to move the data over a 1 Gbit/s network, more than 250 laptops with 8 gigabytes (GB) Random Access Memory (RAM) to load 2 TB of data, and about 33 min to perform integer multiplication on an Intel Pentium processor with 3 GHz clock speed.

1.7 Considerations Before and After Digital Images Are Acquired

Specimen condition and image quality
Although this book is focused on the web image processing pipeline after images are acquired, the success of image processing depends very much on specimen preparation and microscope modality and its parameters and calibration. For example, in live cell experiments, a team must make a trade-off between image quality measured by signal-to-noise ratio (SNR) and the amount of illumination that can harm living cells. Higher illumination intensity and acquisition rate is harmful to cells but produces higher SNR images, which makes it easier to detect cells by image processing. Thus, one must compromise to collect meaningful measurements. Furthermore, live cells are fed on a regular basis by exchanging their media. If changing cell media perturbs the imaging configuration, then image registration and cell tracking become difficult.

Microscope configuration
While configuring a microscope, there are also several trade-offs that affect successful image processing. Due to the limited bandwidth for writing to a computer disk, one might have to compromise on the acquisition rate defined by intensity dynamic range (bits per pixel), spatial resolution, and number of channels per time interval. In addition, a traditional microscope is limited by the maximum speed of its motorized stage. Many times, the stability of the specimen and imaging instrument over time is also important for understanding image quality. Finally, the sensitivity of image processing results to microscope calibration has been known and duly noted, for instance, in the case of image stitching [5].

Computational technology
After preparing a specimen, configuring a microscope, and acquiring thousands of FOVs, there are several technology decisions, with related financial costs, to ensure that the necessary computational infrastructure is in place to allow the scientist to explore and obtain measurements for the acquired large data sets. First, raw digital images cannot be visually inspected and analyzed without preprocessing small FOVs (image tiles) into a large FOV. Traditional software libraries used for bioimage processing are not designed for assembling, viewing, and analyzing very large images on desktops and pose a constraint for scientists on extracting useful information from big images. If traditional software libraries for bioimage processing are run in a computer cloud (computer cluster), then they will likely not utilize all cluster nodes and will be limited by the RAM of each cluster node. To overcome this computational constraint, there is a need to facilitate the transition from desktop computing to distributed computing and/or hide this computational aspect from users by designing client-server systems.

Image sharing
Another decision for principal investigators is the allocation of resources between experimental data acquisition, data analyses to learn from acquired data, and data sharing to increase the overall work impact and receive credit for the experimental and data science work. Given significant investments needed for conducting big data experiments, one needs infrastructure for big data analyses and sharing that can help principal investigators to lower costs of needed information technology work. This favorable allocation of resources toward experimental data acquisition can be achieved by reusing open-source solutions and organizing acquired image collections for data sharing. Ideally, big image collections and their analyses can be curated and stored in a data repository and cross-referenced in the corresponding scientific publications.

Collaborative visual exploration and modeling
A principal investigator must decide between the financial cost of a powerful shared server to be used by all collaborators for remote server-based measurements and the lost time and efficiency of having the collaborators download big image data to their local computer for exploration. If the financial resources are available for a powerful server or for provisioning a cloud deployment, then the client-server system can

deliver remote "anytime and anywhere" access and customizable, scalable, and flexible computational tools for explorations and modeling.

As an added benefit, server-based image measurements allow for building mechanisms for gathering computational provenance into web systems. The benefit of having computational provenance for traceability of intermediate results and shared images is yet another aspect of web systems that is counterweighted by the cost of software development, hardware to host data, and labor to maintain the web system. Thus, one must weight all aspects of an open source web image processing pipeline including collaborative modeling, data sharing, scalability of computational resources, and computational provenance, against the costs of hardware, additional software, and labor to maintain the web system.

1.8 Enabling Reproducible Science from Big Data Microscopy Experiments

Enabling reproducible science from big data microscopy experiments is a tall order and a lofty goal. We are attempting to make a useful contribution toward this goal by presenting software, data, theory, and practical usage of an open-source web system. This chapter outlined the value of web image processing pipelines in terms of multi-scale understanding of phenomena, studying complex phenomena, improving statistical significance, preparing for high-throughput and time-critical processing, and facilitating data sharing, collaboration, and traceability of research results.

Usefulness of web image processing pipeline

A web image processing pipeline can be viewed as a useful data science infrastructure tool to enable discoveries and facilitate publications from big data experiments. Its benefits to researchers and practitioners lie in increasing scientific productivity and in delivering reproducible results. From a discovery perspective, Table 1.2 summarizes the role of web image processing pipelines in the transition from today's microscopic measurements to tomorrow's macro-to-microscopic measurements. From a publication perspective, advances in the understanding of cell responses under drug treatments can be made more rapidly if an investigator does not have to be concerned about "big data" issues such as configuring computers (the web system is already deployed in the cloud), adjusting software to scale with data size (the software is not limited by the specifications of a local computer), keeping track of intermediate results, guessing algorithmic parameters instead of interactively selecting them, and preparing published materials by linking them with individual data points. A publication with such traceable results also allows reviewers to be more productive and objective because they can now verify results derived from big data collections. This is almost impossible when the data sets are too large and require excessive computation.

Table 1.2 Summary of the role of web image processing pipeline in the transition from today's microscopic measurements to tomorrow's macro- to microscopic measurements from big data microscopy experiments

Role of web image pipeline	Today's microscopic measurements	Tomorrow's macro- to microscopic measurements
Acquire	Small field of view Single length scale	Large field of view Multiple length scales
Measure	Qualitative measurements Image algorithms Questionable traceability Serial processing	Quantitative measurements Well-characterized image algorithms Traceable measurements Parallel processing
Analyze	Spot-checking: Micro-scale measurements Unable to detect rare events Estimated population statistics with large error	Relationship across micro−/macroscales Rare event detection Accurate population statistics
Share and collaborate	Limited data sharing Lack of reproducibility Disconnect between publications and data	Sharing via web Traceability of measurements and analyses Publications linked to data

References

1. Zhou, J.: Getting the most out of your image-processing pipeline. *Techonline*, pp. 1–12, 30-Oct-2007
2. Robinson, C.G., et al.: Automated infrastructure for high-throughput Acquisition of Serial Section TEM image volumes. Microsc. Microanal. **22**(S3), 1150–1151 (2016)
3. Deroulers, C., Ameisen, D., Badoual, M., Gerin, C., Granier, A., Lartaud, M.: Analyzing huge pathology images with open source software. Diagn. Pathol. **8**, 92 (2013)
4. Schalek, R., et al.: Imaging a 1 mm 3 volume of rat cortex using a MultiBeam SEM. Microsc. Microanal. **22**(S3), 582–583 (2016)
5. Chalfoun, J., Majurski, M., Blattner, T., Keyrouz, W., Bajcsy, P., Brady, M.: MIST accurate and scalable microscopy image stitching method with stage Modeling and error minimization. Nat. Sci. Rep. **7**, 1 (2017). DOI: 10.1038/s41598-017-04567-y