Characterization of Biological Processes through Automated Image Analysis

Jens Rittscher

Visualization and Computer Vision Laboratory, GE Global Research, Niskayuna, New York, 12309; email: jens.rittscher@research.ge.com

Key Words
Biological imaging, live cell microscopy, cell segmentation and tracking, tissue analysis, phenotyping

Abstract
The systems-level analysis of complex biological processes requires methods that enable the quantification of a broad range of phenotypical alterations, the precise localization of signaling events, and the ability to correlate such signaling events in the context of the spatial organization of the biological specimen. The goal of this review is to illustrate that, when combined with modern imaging platforms and labeling techniques, automated image analysis methods can provide such quantitative information. The article attempts to review necessary image analysis techniques as well as applications that utilize these techniques to provide the data that will enable systems-level biology. The text includes a review of image registration and image segmentation methods, as well as algorithms that enable the analysis of cellular architecture, cell morphology, and tissue organization. Various methods that enable the analysis of dynamic events are also presented.
1. INTRODUCTION

Significant technical innovation is necessary to advance biomedical research in the postgenomic era. Today, more than 50 years after the discovery of the double-helical structure of DNA, we have access to a comprehensive, high-quality sequence of the human genome. In the spirit of the proposals put forward for mathematics by David Hilbert at the turn of the twentieth century, Collins and collaborators (1) formulate a series of grand challenges. Connecting genomics and biology is the first of the three challenges they discuss. Whereas the chemical structure of DNA is now well understood, the structure of the human genome is extraordinary complex, and its function is poorly understood. It is therefore necessary to determine how genome-encoded components function in an integrated manner to perform cellular and organismal functions. Collins et al. (1) stress the importance of understanding the organization of genetic networks and protein pathways to establish how they contribute to cellular and organismal phenotypes. The understanding of genes and pathways is directly linked to the grand challenge of translating genome-based knowledge into health benefits. Pharmaceuticals that are on the market today target fewer than 500 human gene products. The understanding of biological pathways provided by genomics will not only provide an enormous pool of currently untapped drug targets, but it should also contribute to the therapeutic design.

Roger Y. Tsien (2) makes the analogy that a genome sequence is at best some sort of telephone directory, perhaps a list of names of all citizens. Although it provides a lot of factual information, it
offers little information on how these citizens really live. Suppose, he continues, we were anthropologists from a different continent and had to understand how a particular community works. A group of spies would be helpful. Ideal spies would be natives, who are born naturally into the society and who keep us informed of the members’ whereabouts, their education, their partners, or the prevailing economic climate. The spies would help interrogate the different aspects of these biological processes that occur at multiple scales.

Molecular probes pay this crucial role. Fluorescent proteins (3) have enabled noninvasive imaging in living cells and organisms for reporter gene expression, protein trafficking, and other biochemical signals. Hundreds of small organic dyes are available commercially, and they have undergone industrial optimization of wavelength range, brightness, and photostability. The discovery of the green fluorescent protein (GFP) (4) from the jellyfish *Aequorea victoria* enabled a revolution in live cell imaging (5). The expression of GFP along or in most genetic fusions with other proteins results in visible fluorescence without requiring any other cofactors. If the gene for a particular protein is fused to the gene for GFP, a broad range of cellular proteins can be made fluorescent. Quantum dots are inorganic nanocrystals that fluoresce at sharp and discrete wavelengths depending on their size. They provide an additional technology for constructing molecular probes.

The ordered sequence of biochemical transactions, also referred to as signal transduction, initiates and regulates such processes. As a result, genes can be activated, the metabolism of a cell can be changed, or signaling proteins can change the metabolism of the cell. Cell proliferation and apoptosis or a stimulation of migration are other possible effects as summarized in Figure 1. Given the complexity of biological systems, researchers are often forced to study these steps in isolation. Molecular probes are used to monitor the activation of certain proteins. The combination of multiple probes then allows the interrogation of the signaling cascade. Phenotypical alterations, such as cell migration, can be observed through inspection of the image data, which were captured by a microscope.

Kitano (6) argues that we need to understand biology at the systems level. It is necessary to examine the structure and dynamics of cellular and organismal function rather than the characteristics of the isolated parts of a cell or organism. Eventually it will be possible to build detailed

---

**Figure 1**

Concept overview. Signal transduction events cause activation of genes, alterations in metabolism, cell proliferation and apoptosis, and cell migration. Events such as proliferation, apoptosis, and cell migration are phenotypical alterations that can be observed visually. Molecular probes are typically used to monitor signal transduction or the activation of enzymes.
models of cell regulation. Such models will, for example, impact drug discovery by helping identify feedback mechanisms that offset the effect of drugs and by predicting systematic side effects. A comprehensive set of quantitative data is required to conduct any type of systems-level analysis. The goal of this article is to illustrate that the combination of modern imaging techniques and image analysis methods enables (a) the quantification of a broad range of phenotypical alterations, (b) the localization and dynamics of certain signaling events, and (c) the correlation of multiple signaling events in the context of the spatial organization of the biological specimen. In a certain sense, these techniques and methods provide important building blocks for constructing a surveillance system for life at the molecular scale. Microscopic image analysis holds the promise of becoming an enabling technology for modern systems-biology research.

At the risk of being broad, this article is an attempt to provide a review of both necessary image analysis techniques and applications that utilize these techniques to provide the type of quantitative data that will enable systems-level biology. Rather than provide a comprehensive list of all the relevant work, I highlight certain aspects through examples with which I am personally familiar. The discussion of optical methods and related signal processing techniques, such as deconvolution methods and noise removal, are omitted from this article. Excellent reviews of these topics have already been published (see Reference 7 for an example). Where possible, the reader is referred to more complete survey articles. The application of image analysis techniques to life science applications is also the topic of a recently edited volume (8), which presents an in-depth discussion of relevant technologies and applications.

This article is structured into five major sections. The first three sections (Sections 2–4) provide an overview of image analysis methods that enable the quantification of phenotypical alterations as outlined in Figure 1. Methods for image registration and the automated segmentation of multichannel data are presented in Section 2. Image registration methods are commonly used to assemble a single data volume from the set of input images. Automated image segmentation techniques play a central role in the acquisition and analysis of the probe signal. In some sense, this section gives an overview of some of the basic image analysis tools that are needed.

The analysis of nuclear architecture, cell morphology, and aspects of tissue architecture are discussed in Section 3. Cellular morphology is generally regarded as an important manifestation of the physiological state of cells. Now, there is some growing evidence of links between gene function and nuclear architecture. Recent studies even demonstrate that activity-independent genome repositioning events are related to early state tumorigenesis. Changes in tissue architecture provide information on how local events, such as mitosis and cell death, relate to tissue-level properties.

The analysis of dynamic events at various different scales, which is the focus of Section 4, provides important clues about signaling events and developmental processes such as embryogenesis and cell differentiation. Methods relevant to the analysis of intracellular dynamics, cell tracking, and detection of cellular events are highlighted.

Sections 5 and 6 focus on studies that utilize automated image analysis. This set of examples illustrates how this technology can contribute to systems-biology research. I adapt the term in toto imaging, introduced by Megason & Fraser (9), to describe a set of projects that aim to analyze the overall development of a specific organ or the entire organism, such as the zebrafish (Danio rerio). I illustrate how imaging methods can be used to capture longitudinal information in a quantitative fashion. As the longitudinal study of developmental processes poses some specific challenges, developing techniques that allow the detail analysis of a large number of complex biological specimens in high throughput is an ambitious undertaking. For the purpose of this review, I refer to such applications as enterprise-level applications, which are presented in Section 6.
Finally, this review concludes by giving a brief overview on what software tools and data sets are openly available. In addition, I make an attempt to summarize what challenges the field needs to address to gain a broader acceptance of automated image analysis techniques.

2. IMAGE REGISTRATION AND SEGMENTATION

Image registration and segmentation are core image analysis tasks that are fundamental for any quantitative microscope image analysis of biological specimens. In a certain sense, they are the basic tools that allow us to assemble complex data sets and identify the set of objects, such as individual cells and their compartments, that form the basis for biologically relevant measurements. It would be overly ambitious to include a comprehensive overview of either of these two topics as part of this article. Instead, the following two subsections present a set of selected articles that highlight specific challenges and innovative solutions that are relevant.

2.1. Image Registration

In many cases, it is necessary to acquire a set of images and then assemble these to form a larger mosaic or three-dimensional data volume that will then represent the specimen of interest. In addition, it is often necessary to combine data from different channels or imaging modalities. Integrating and comparing such data sets requires image registration. Broadly speaking, image registration algorithms can be classified into intensity-based methods and feature-based methods. The goal of any registration algorithm, given two images, is to estimate a mapping that will bring the two images into alignment. This transformation can be a linear transformation, which includes rotation and scaling, that is global in nature. Nonlinear deformable transformations can locally warp the target image with a reference image and can, for example, account for local deformations of tissue sections. Hill et al. (10) present a comprehensive overview of medical image registration methods. Open source toolkits, provide, as discussed in Section 7, readily usable implementations for commonly used registration methods.

The image registration method that Magee et al. (11) developed for the reconstruction of 3D histopathology volumes from serial sections highlights a number of challenges other researchers have encountered. Because histopathology is typically applied after sectioning, the 3D volume is constructed through the registration of adjacent slices. Because of the high elasticity of thin issue sections, nonlinear deformations need to be taken into account. The proposed registration method works by performing dynamic programming separately on each pair of corresponding rows and each pair of corresponding columns. The resulting dense transform field, which can be very noisy, is then regularized by the fitting of a parametric transform in a way that minimizes least-squares errors. Their experimental results demonstrate a substantial error reduction when compared with traditional registration methods.

Anderson et al. (12) address the challenge of building large serial-section transmission electron microscopy volumes for the study of neural circuitry. In addition to the construction of the 3D volume itself, every individual layer consists of a mosaic that needs to be constructed from approximately 1000 individual tiles as is illustrated in Figure 2. In the first step, it is necessary to find an ordering of the random image tiles. Here, the Fourier transform properties are being used to make this problem computationally tractable. To construct the final volume, the authors developed an entire registration pipeline that includes the use of metadata, where possible, and nonlinear refinement of the transform. Addressing similar registration problems, Emmenlauer et al. (13) developed a comprehensive set of registration tools that enable the assembly of large 3D data sets.
Assembly of complex 3D data sets. Canonical fields of rabbit retina are being sectioned from the ganglion cell (GC) layer to the amacrine cell (AC) layer at 70 nm, and tiled mosaics are acquired for volume assembly. Figure courtesy of Anderson et al. (12).

2.2. Image Segmentation

The automatic quantification of protein expression profiles, as illustrated in Figure 3, motivates the necessity for an accurate and robust delineation of individual cells and their compartments through image segmentation. In general, signal quality, morphology, and packing density of the objects of interest determine the degree of difficulty of any image segmentation problem. In certain

Fundamental importance of image segmentation. In addition to functional protein markers, this tissue section has been labeled with a set of compartmental markers that aid the identification of the cell nuclei, the cytoplasm, and the cell membrane. Although the fluorescent labels aid the identification of the different cellular compartments, sophisticated segmentation methods are still necessary so that the compartments can be delineated. The segmentation mask then provides that basis for quantifying the different protein expression patterns.
circumstances, detecting the locations of strong signals by thresholding image intensities might be sufficient (14). The analysis of complex biological specimens (see, for example, Figures 2 and 3) does, however, require sophisticated segmentation algorithms that make effective use of data-driven features as well as prior assumptions about morphology and anatomy. Signal attenuation, which is a common problem in confocal microscopy, and low signal-to-noise ratios are challenges that need to be addressed.

Although different theoretical approaches—including watershed segmentation (15), variational approaches (16), level sets (17–19), and Markov random fields (20) and related approaches (21–23)—have been proposed and refined, image segmentation is still an area of active research. By automatically classifying well-segmented and poorly segmented cells, Hill et al. (24) were able to demonstrate that certain biologically relevant morphological changes can be captured only in cases when cells are well segmented. This section draws attention to the segmentation problems that are of high relevance to microscopy applications, describing the segmentation of images acquired using label-free methods, the segmentation of multichannel 3D data sets, and the analysis of histology images.

Having employed statistical learning techniques, Pan et al. (25) present an approach that captures cell appearances and intercell relationships. After an initial separation of background and foreground pixels, pixel locations that could potentially represent the center of a cell are identified through a mean shift–type algorithm (26). A discriminative model that takes into account the local gradient statistics, gradient magnitude, and orientation classifies each of the candidate pixel locations as either cell locations or background. Finally, the problem of oversegmentation is being addressed with the help of a grouping algorithm.

The reliable segmentation of densely packed structures in 3D data sets is certainly one of the most challenging segmentation tasks. Roysam and collaborators (27, 28) present a list of object morphologies that describe frequently observed structures in cell and tissue imagery. This list includes blobs (nucleus), tubes (vasculature, neurites), shells (nuclear membrane, cell membrane), foci or puncta, plates, and clouds (extracellular matrix); it forms the basis for their FARSIGHT segmentation framework. They propose to design specific segmentation algorithms for each of these object morphologies. By extending the idea of watershed segmentation into 3D, Lin et al. (29, 30) developed a method for detecting cell nuclei in fluorescent-labeled 3D image data sets. Li et al. (31) propose a three-stage algorithm to achieve a reliable segmentation of cell nuclei. These stages include a diffusion process for the purpose of denoising, gradient flow tracking, and grouping, as well as local thresholding. Khairy et al. (32) express the segmentation problem in terms of Bayesian inference and use data-driven Markov Chain Monte Carlo methods to fit the image model to data. Biologically inspired priors are an integral part of their approach.

In addition to the detection of nuclei, the segmentation of microvasculature, or generalized tubular structures, is of high relevance. Utilizing superellipsoids as a basic shape model that describes localized vessel segments, Tyrrell et al. (33) developed an algorithm that can extract extremely complex vessel structures and implicitly handle branching. A recursive scheme is used to update the superellipsoid model estimates at each point along the length of the vessel. Mayerich et al. (34) follow a similar approach but use the isosurface of the vascular network to refine the vascular model. By mapping the isosurface onto the extracted skeleton, they managed to correct misclassification and errors of the tracing step. The maximum-intensity projections of the capillary network in the cerebellum of the neocortex of a rat, which are shown in Figure 4, demonstrate that this method is well suited to study the microvasculature of the brain.

Addressing the challenge of imaging neuronal cells closely relates to addressing the challenge of segmenting microvasculature. Jurrus et al. (35) developed a carefully engineered algorithm
Figure 4
Visualization of complex vascular structure. Understanding microvasculature structures and their relationships to cells in biological tissue is an important and complex problem. Brain microvasculature in particular is known to play an important role in chronic diseases. The image shows examples of the neovasculature of a rat brain. Figure courtesy of D. Mayerich (34).

using Kalman snakes and optical flow computation to enable tracking of axons across large distances in volumes acquired by serial block-face scanning electron microscopy. Axon tracking is initialized with user clicks or initialized automatically through the watershed segmentation algorithm, which identifies axon centers. The researchers tracked multiple axons from slice to slice through a volume, updating the positions and velocities in the model and providing constraints to maintain smoothness between slices. Losavio et al. (36) propose a computational and experimental framework toward real-time functional imaging of neuronal cells; the framework consists of a number of steps that produce a 3D geometrical model used for functional simulations of the neuron cell as illustrated in Figure 5. The key steps include a frame-shrinkage denoising algorithm suitable for tubular structures, a dendritic-detection algorithm that is based on learning and predicting generalized 3D tubular models, and a novel skeletonization algorithm performed using a morphology-guided deformable model.

3. CELLULAR ORGANIZATION AND STRUCTURE

More recently, researchers have investigated the question of how gene function is integrated into the architectural framework of the cell nucleus and how structural elements of the nucleus affect nuclear processes such as gene expression and DNA repair. Efforts that address the analysis of nuclear architecture, specifically the localization of genes and proteins, are highlighted in Section 3.1. Recent studies that demonstrate the existence of activity-independent genome repositioning events in the early states of tumor formation indicate that the use of quantitative information at the subcellular scale could one day become routine clinical practice.

Cellular morphology is, in general, an important large-scale manifestation of the global, organizational, and physiological state of cells. This is governed by highly regulated biological processes and is controlled by interactions between the cytoskeleton, the membrane and membrane-bound proteins, and the extracellular environment. Morphological measurements have commonly been used as proxy measurements for the global cell status, including differentiation, cell cycle state, and apoptosis.
Figure 5
Guided functional imaging of neurons. (a) Morphological reconstruction obtained from structural imaging, (b) computational model, and (c) guided functional imaging. Figure courtesy of I.A. Kakadiaris (104).

3.1. Analysis of Nuclear Architecture

The architectural organization of the cell nucleus and the location of genes and proteins are determined by complex mechanisms, and they depend on multiple layers of control processes. The systematic and comprehensive analysis of subcellular location is therefore needed as part of systems-biology efforts. Currently, subcellular localization patterns are frequently determined by manual visual inspection. Extracting quantitative data in a repeatable and operator-independent manner requires automated analysis methods. Proteins are typically labeled with a fluorescent probe, and fluorescent microscopy is being used for image acquisition. By inserting the coding sequence of a GFP randomly into genomic DNA by a retroviral vector, researchers can label individual proteins in live cells.

Murphy and collaborators (37, 38) have investigated various classification methods to address the problem of detecting protein locations at a subcellular scale. They developed image analysis methods that quantitatively describe the subcellular location patterns in fluorescent microscopy images. Invariance with respect to position, rotation, and overall intensity are important design criteria. A taxonomy of descriptors that utilizes edge features, geometric features, texture descriptors, and the DNA distribution in the nucleus is being proposed. The feature sets are being calculated...
on 2D as well as 3D image data sets to detect proteins that were labeled by immunofluorescence. Machine learning techniques, such as support vector machine (SVM) classifiers (39), are used to detect and label the subcellular location patterns. The interesting aspect of this particular approach is the sole reliance on detection and classification techniques as opposed to segmentation methods. The reported results are promising and are well suited for larger-scale experiments.

Meaburn et al. (40, 41) present a systematic analysis of genome reorganization events during early tumorigenesis. To study how specific genes behave during normal and cancerous mammary differentiation, Meaburn and Misteli mapped the radial position of 11 cancer-related genes during normal acinar differentiation using interphase fluorescence in situ hybridization (FISH) (42). The quantitative analysis of the FISH signal distributions was enabled by an automated image analysis system (43) and is illustrated in Figure 6. A preprocessing step normalizes the data to handle nonuniform intensity variations. Accurate segmentation is achieved through multistage classification. When tested on manual segmentations, the system identified 99% of the FISH signals successfully and had a failure rate of 1%.

Studies first using a 3D mammary epithelial differentiation system (40) and subsequently using human tissue sections (41) demonstrate changes in spatial positioning of a set of cancer-associated genes during both normal breast epithelial differentiation and early tumorigenesis (see the graph in Figure 6). A set of 20 gene loci was visualized by FISH in a panel of 11 normal and 14 invasive carcinoma breast tissues. In both studies, the researchers successfully identified several genes that differentiate between normal and disease (see Figure 6). In addition, the cancer-specific repositioning events in tissues differ from those observed in the model system. Importantly, Meaburn et al. document that cancer tissues can be identified accurately through a comparison with a standardized normal gene distribution, which is critical for potential clinical use.

Using a simpler model organism, Berger et al. (44) studied the spatial organization of the genome inside eukaryotic cell nuclei. This spatial arrangement is not random, and it plays a central role in transcriptional regulation, DNA repair, and replication. They developed a computational imaging approach that captures a locus position by its distance from the nuclear center and by the angle from the axis passing through the nuclear center and the nucleolar centroid. Because of the absence of a third nuclear landmark, the angle around the central axis is left undefined. The resulting 2D histograms reveal more information than the one-dimensional radial distance histograms that were used previously.

Because optical microscopy studies have a diffraction-limited resolution of at best 250–500 μm and the presence of stochastic motion, gene localization can be done only probabilistically. A large population of cells—in this case, between 1000 and 5000 cells—are necessary to estimate the location information. Fluorescent labels are used to tag the single locus, the nuclear pore protein, the nuclear envelope, and the nucleolus. An automated module is used to detect interphase cells (i.e., cells that are in the G1 or S phase) to rule out substantial cell cycle effects. To aggregate the extracted positions into a single map, the researchers aligned the nuclear landmarks of the different nuclei through translation and rotations. Berger et al. (44) were able to produce location maps that clearly reveal small, probabilistic gene territories that are tied to the nuclear architecture. The authors suggest that this strong compartmentalization places strong constraints on the possible interaction of genes with one another. The statistical model would need to be extended to analyze more complex nuclei of higher eukaryotes.

3.2. Morphological Assessment of Cells

Changes in cell morphology and tissue structure still play an important role in the diagnosis of cancer. However, a large-scale analysis requires quantitative features that capture the cellular
Figure 6

Analysis of gene localization patterns. This figure demonstrates the ability to precisely detect the location of gene localization patterns in human tissues using interphase fluorescence in situ hybridization (FISH). (Top left) The 4′,6-diamidino-2-phenylindole (DAPI)-labeled nuclei (gray) in a thin section of normal human breast tissue and FISH labeling of the HES5 gene (red) and the HEY1 gene (green). (Top right) The automatic segmentation of all the nuclei in the image. (Bottom left) The automatically selected nuclei that are used for gene centeredness measurement. This method allows arbitrarily and differently shaped nuclei to be compared with one another. (Bottom right) Graph showing the output result comparing a cancer sample (blue) with a normal sample (green) for the HES5 gene. The graphs show cumulative probability distributions where 0 on the x axis corresponds to the edge of the nucleus and 1 on the x axis corresponds to the center of the nucleus, as defined by the Euclidean distance transform. The Kolmogorov–Smirnov test yields the result that the probability the two curves are the same is 0.000272. Figure courtesy of K. Meaburn (41).

morphology and tissue organization. Pincus & Theriot (45) present a comparison of quantitative methods for cell-shape analysis. Ideally, any method that analyzes cell shape should not add spurious information and should not discard true shape information. It should also capture shape variations that are biologically relevant. Finally, the numerical components should also be meaningful and interpretable on their own. A large feature set of shape descriptors is not desirable in case the numbers themselves are not biologically meaningful. The difficulty is that cells have shapes that are smooth and highly variable. Therefore, they do not provide many points that can be used as landmarks for registration.
At a minimum, such a feature set will include intrinsic measurements such as the area or volume and the length of the major and minor axes. Pixel-based representations, such as binary images, are not very suitable because they depend on both resolution and scale. Techniques that apply Fourier or Zernike analysis have been developed for the analysis of 2D shapes (46). These polynomials can be used to break down images into contributions from lower and higher frequencies. Principal component analysis (39), or PCA, is a classical method for examining the different modes of variation in the data. PCA-based methods have been applied with great success in medical imaging and computer vision (47–49). Whereas PCA decorrelates the data, independent component analysis (50), or ICA, can be used to unmix a data set into contributions from statistically independent components. Both Fourier coefficients and Zernike coefficients can be made rotationally invariant. More generally, global alignment of populations can be used to achieve rotational invariance. The results of the study (45) indicate that the PCA-based representations appear more suitable because they help extract biologically meaningful information.

The fact that shape spaces are not necessarily linear motivates the deformation-based nuclear morphometry framework (51). Standard linear methods, such as PCA, are often not sufficient to achieve highly accurate and robust shape representations. Rohde et al. (51) make use of the computational anatomy framework (52), which aims to quantify shape differences by analyzing the spatial transformations that map the different samples. In this context, the definition of a shape space is linked to a template image, such that every image in the given set can be composed through deformations of the template image. First, it is necessary to find the spatial transformation that matches two given images in an optimal sense. Second, the mean shape of the population, which will act as the template, needs to be determined.

The way spatial deformations that map one image or shape to another are modeled is central to this approach. These spatial deformations are described by diffeomorphisms, which are smooth and invertible functions. A cost function that takes into account the geodesic distance on the manifold of diffeomorphisms and the image difference measured in a least-square sense is used to find the optimal deformation map. The mean nuclear shape is also determined via an energy-minimization approach. The concept of multidimensional scaling (53) is then applied to find the set of coordinates in a Euclidian space that best preserves the idea of the geodesic distances. The proposed system has been tested on a set of 87 HeLa cell nuclei as well as HeLa cell nuclei expressing lamin modifications, which affect nuclear morphology. The experiments demonstrate that it is possible to establish a relationship between nuclear morphology and lamin concentration.

Combining morphological features with the quantitative assessment of protein expression holds the promise of providing a wealth of disease-relevant information. Rimm and collaborators (54, 55) make use of automated image segmentation techniques to develop a molecular-based method of quantitative protein assessment using immunohistochemistry. Their work is motivated by the need to evaluate and validate molecular biomarkers in histology tissue samples. To date, many U.S. Food and Drug Administration (FDA)-approved tests still suffer from high intraobserver and interobserver variation. Automated methods could help standardize such tests. The commercially available AQUA system integrates a set of algorithms that allows for the automated analysis of tissue microarrays, including the separation of tumor from stromal elements and the subcellular localization of signals.

In a first step, the system identifies a tumor-specific mask by thresholding the image of a marker that differentiates tumor from surrounding stroma and/or leukocytes. Subsequently, all out-of-focus information is removed through the combination of images that are taken at two different focal planes. With the help of compartmental markers, a segmentation algorithm assigns each pixel location to the nuclear, cytoplasmic, or membrane area. Based on the subcellular label map, the protein expression profile is added up. The study on the quantitative determination of expression...
of the prostate cancer protein α-methylacyl-CoA racemase (AMACR) (54) demonstrates that the AQUA system can identify prostate cancer from benign tissue in the given tissue microarray samples.

### 3.3. Tissue Architecture

The tendency to minimize surface energy or maximize space filling governs the spatial organization of many tissues. Proliferating epithelia only rarely follow this pattern owing to the effect of cell division. Gibson et al. (56) formulated a discrete Markov chain to capture the stochastic nature of cell proliferation. This model enables numerical simulation of what the stable state of the system will look like and effectively allows researchers to make a quantitative prediction. It turns out that the average mitotic cell possesses not six but rather seven sides. The results indicate that simple emergent mechanisms determine cell shape. Doyle et al. (57) propose a machine-learning approach to discriminate for an automated and quantitative grading of prostate biopsy specimens. They effectively infer the tissue organization from a set of training images. The proposed SVM classifier (39) uses features that capture local texture information as well as features that capture the local spatial organization of the tissue. By computing tessellations from the given histology sections, they derive a number of architectural features.

By taking intrinsic as well as associative measurements into account, Bjornsson et al. (28) demonstrate the ability to extract a comprehensive set of quantitative data that captures the complex structure of the tissue architecture. A set of ~100-μm thick slices of rat brain tissue were labeled with five different fluorescent probes and imaged using spectral confocal microscopy. In addition to identifying all the relevant entities and structures, the authors defined the associative measurements that quantify the relationship between the segmented objects. As a result of the automated multichannel segmentation, the system identifies cell nuclei and traces microglial processes, blood vessels, and astrocytes. After the classification of each cell type via supervised classification methods, different associative measurements are extracted. They include a nearest-neighbor graph of neurons, the cell vascular distance that quantifies the spatial relationships of cells and vasculature, and the regional cell distribution. The output is a detailed data representation that can be used for diverse, quantitative, hypothesis-driven studies.

### 4. DYNAMIC EVENTS

Computerized video time-lapse microscopy enables the monitoring of dynamics of tagged proteins inside a cell, the motion of individual cells, or entire cell populations over extended periods of time. Analyzing such data sets is a major challenge and constitutes a major bottleneck for the full exploitation of multidimensional microscopy sequences that document studies of biological object dynamics. Only recently it became possible to study the dynamics of subcellular signaling events within the living cell. Approaches that support such studies are presented in Section 4.1.

The ability to track cells has broad applicability and can, for example, be used to study cell migration in general, wound healing and repair, and population dynamics. The development of cell-tracking algorithms has, of course, benefited enormously from the ongoing research in computer vision. Section 4.2 gives an overview of this development and highlights some of the specific challenges associated with the problem of cell tracking. The cell-tracking information can then be utilized, as discussed in Section 4.3, to detect and analyze an array of biologically relevant events. Here, two closely related approaches that address the problem of estimating cell cycle phase are presented. Analyzing the tracking information itself to detect certain biological events is also highlighted.
4.1. Intracellular Dynamics

Using video microscopy, researchers can capture the dynamical behavior of tagged proteins within the living cell. The transportation of molecules between different compartments is guided by the cytoskeleton, which is made of actin filaments, intermediate filaments, and microtubules. It is now possible to collect high-frame-rate time-lapse imagery of living cells without disturbing the cell activity. A number of groups have addressed the challenge of tracing microtubules automatically. Hadjidemetriou et al. (58), for example, present an approach that allows the detection and tracking of the outer microtubule tips. Analyzing the trafficking patterns of the proteins along the microtubules is an extremely challenging problem. By making an analogy with traffic patterns on computer networks or road networks, Boulanger et al. (59) developed a physics-based approach that can mimic the traffic patterns of proteins within cells. Jaqaman et al. (60) utilize the technique of multiple-hypothesis tracking (61) to extract the trajectories of single particles. Given a set of detected particles between consecutive frames, the algorithm links the track segments and captures the particle merge and split events. Detecting the merge and split events allows the study of the association and dissociation of receptors on a single-molecule level. The optimal data association is estimated with the help of global optimization that uses linear programming.

Machacek et al. (62) introduce a framework for morphodynamic profiling, which enables the study of shape evolution of cells during migration by tracking arbitrarily complex boundary movements. Given a sequence of cell boundaries from time-lapse microscopy, the framework obtains protrusion and retraction rates by measuring the local boundary displacement. A level-set approach is used to model the boundary evolution between two consecutive time points, and the continuous curve evolution is used to enforce topological constraints on a set of boundary markers. The authors demonstrate that the quantitative information supports a hypothesis on what molecular mechanisms generate the different protrusion patterns.

4.2. Cell Tracking

Although each cell-tracking application has specific requirements, they share several core challenges. Each data set typically contains a large number of cells to be tracked. Unless otherwise stimulated, cells normally move randomly at modest speeds, but biologically relevant events such as mitosis (cell division) need to be accurately captured. Furthermore, because the field of view of the microscope is limited, some cells move in and out of the image or focal plane. As certain fluorescent dyes have toxic effects, only low concentrations of such dyes are used. Consequently, segmenting the image data is more challenging because of the low signal-to-noise ratio. Some experiments may be imaged every 3 or 5 min, which reduces the spatial overlap of the cells between adjacent frames.

Extensive research in the field of computer vision has resulted in powerful and versatile algorithms for visual tracking. The existing methods can roughly be divided into three groups: independent segmentation of individual frames followed by data association, model-based contour-evolution approaches (63), and stochastic filtering (64). Contour-evolution approaches, and in particular level-set approaches, offer some advantages because they can easily handle changes in topology. Yang et al. (65) and Padfield et al. (66) make use of this fact by evolving level sets on the spatiotemporal volume to effectively associate detections over time. In other approaches, such as that of Dufour et al. (67), levels sets are propagated over time via use of the contour from the previous image as an initialization. Although level-set methods handle topology changes effectively, additional heuristics are necessary to reinitialize level sets to accommodate fast motions or the appearance or disappearance of new cells.
In general, probabilistic approaches rely on strong model assumptions. The necessary model parameters are typically learned from training data sets. Mean-shift tracking (68) can, for example, be used to track objects through the use of a basic appearance model as suggested by Debeir et al. (69). Stochastic filters (70, 71) can be extremely powerful if the object motion can be modeled, and such approaches are used for cell tracking in studies by Li et al. and Genovesio et al. (72, 73). Sometimes it is difficult to capture the wide range of biological motions and events such as mitosis using a consistent stochastic motion model.

Algorithms that first segment the objects in each of the images and then associate these objects have also been shown to be effective for cell tracking. Al-Kofahi et al. (74) use linear programming on various matching hypotheses, but their approach does not capture cell occlusion and cells entering and leaving the field of view. Matching based on Euclidean distances is used by De Hauwer et al. (75) to associate targets across frames. Zhang et al. (76) use a minimum-cost flow network for globally assigning a small number of detections over time, during which the optimization is run multiple times as different hypotheses are tested; this flow network is computationally expensive, and the large number of hypotheses limits its application to problems with a small number of tracks. The failure modes that are typical for this category of algorithms suggest that standard matching techniques are not sufficient for associating entering and leaving cells and that mitosis and occlusion events need to be treated carefully.

Many of these approaches embed the problem in some standard tracking framework by extending and combining approaches and by introducing postprocessing steps. This can lead to complex and computationally intensive systems that require many parameters to tune. Padfield et al. (77) developed a mathematical framework that formulates the cell-tracking problem using a single model that is consistent, extensible, general, and efficient; in addition, it requires few parameters. The authors accomplished this by formulating the association problem in a graph-theoretic flow network framework (cell-tracking examples are shown in Figure 7). Modeling mitosis and merging events requires one-to-many and many-to-one associations, which are modeled using coupled minimum-cost flow, which enforces a coupling of the flow of certain edges. This extended framework then enables the implicit modeling of different cell behaviors including moving, splitting, merging, moving into the image, and moving out of the image.

4.3. Cellular Events

Cell cycle control and its links with cancer and biological responses to DNA damage comprise a dynamic area of contemporary biomedical research (78). The effective study of the progression, pausing, or stalling of cells through the cell cycle requires methods that can estimate the cell cycle state for each individual cell. Harder et al. (79, 80) introduce a computational framework to automatically segment, track, and classify cell nuclei into different mitotic phases by utilizing image features. The proposed computational framework for analyzing the duration of mitotic phases consists of four steps: segmentation, cell tracking, feature extraction, and classification into seven mitotic phases. An SVM classifier with a Gaussian radial basis function as kernel is used to classify the nuclei into seven mitotic phases. The system achieves an overall classification accuracy of more than 94%.

A set of dynamic fluorescent markers (81) enables cell motion and cell cycle transitions to be imaged over time in living cells without disturbing the cell cycle progression. Expression of these sensors in stable cell lines enables high-content analysis of cell cycle progression using real-time, fluorescent, time-lapse automated image analysis. Padfield et al. (66) developed automated image analysis methods that effectively estimate the cell cycle phase of individual cells using these markers. A visualization of the resulting temporal signature is shown in Figure 8. In these data
sets, the appearance of cells changes over time so that no single model can extract the nucleus reliably. Here, all images of the sequence are assembled to form a spatiotemporal or XVT volume. In certain phases of the cell cycle, the appearance of cells does not change. Hence it is possible to identify cells with the help of 3D segmentation methods. With the help of a phase-transition model, these cell tracks are linked to form a complete lineage tree.

Cohen et al. (82) address the problem of extracting biologically relevant information from cell-tracking data by summarizing the data in a meaningful way. All tracking information is summarized in one data structure, the attributed tracking graph, which stores the tracking information as well as all the associated measurements. An information theoretical distance measure is then used to compare object tracks. The meaningful summarization is captured via the gap statistic (83), which is an effective tool for estimating the number of clusters in a given data set. When applied to time-lapse microscopy data of cultured neural progenitor cells, the method correctly distinguished neurons from progenitors without requiring a fixative stain.

5. IN TOTO IMAGING

Monitoring the overall development of a specific organ or the entire organism is one of the most ambitious goals of systems-biology research. Here, imaging methods not only allow the
segmentation of individual cells but also permit the longitudinal study of individual cells in the context of the overall anatomy. Capturing such data sets requires careful experimental design and engineering throughout the entire processing pipeline. The practice typically includes labeling, image capture, image analysis, and problem-specific quantification. Megason and Fraser (9) introduce the term in toto imaging for combining all the necessary techniques into one consistent approach. For the purpose of this review, a set of related projects is presented in the context of in toto imaging. All of the following research studies rely on imaging methods to capture longitudinal data in a quantitative fashion.

Imaging methods play a critical role in elucidating the dynamic relationship between dendrite growth and synaptogenesis. Several in vivo studies have described the highly dynamic process of dendritic arbor elaboration. Today, little is known about the process by which dendritic cells create the intricate branching pattern that is unique to each specific neuronal class. Although
coverage of the underlying biology exceeds the scope of this article, how imaging methods are being used to study the dynamic formation of synapses in vivo is discussed. Liebling et al. (84) present an integrated solution that enables the study of genetic contributions to heart formation in zebrafish. They directly address the requirement of providing high spatial and temporal resolution to quantify force relationships and morphogenetic changes while the heart is beating. Acquiring data that will aid the understanding of the embryogenesis of zebrafish presents some additional challenges. As it is not possible to take a high-resolution image of the entire fish, a large number of subvolumes need to be acquired separately. The following sections present a brief summary of these applications.

5.1. In Vivo Imaging of Synapse Formation

The study of dendritic arbor-growth development plays an important role in neuroscience research. The factors that control the dendritic architecture directly affect neuronal function and circuit properties. A number of in vivo studies (85–87) indicate that during the first phase of circuit development, the growth phase, a paradoxically high rate of branch dynamics can be observed. In vivo time-lapse data indicate a rapid rate of branch additions and retractions. As neurons mature, the retractions decrease. Dynamically adding and retracting branches does not appear to be the most efficient way of growing an arbor. One hypothesis (86) is that branches could possibly test the environment for optimal synaptic partners. The question about the role synaptic activity plays in arbor growth is a field of active research. The fact that modern imaging techniques enable the longitudinal study of arbor growth is critical to these works.

Whereas the imaging techniques are a necessary first step for observing this phenomenon, automated image analysis techniques enable the quantitative study of these dynamic processes. Niell et al. (85) examined the dynamics of synapse formation in relation to arbor growth through time-lapse imaging of immature dendrite in the intact larval zebrafish at 20-min intervals for up to 24 h. The abundance of transient fine-terminal processes form a complex 3D structure. Niell et al. (85) report that these dynamic protrusions of neurons are enriched in actin and that they satisfy the common morphological, dynamic, and cytoskeletal criteria for dendritic spines. They refer to the newly formed protrusions as filopodia. The process of dendrite growth occurs by an interactive sequence of selective filopodial stabilization and punctum formation.

Quantitative measurements were obtained through the application of image analysis algorithms that measured branch length by tracking dendritic arbors through three dimensions. Individual puncta were identified as feature points and then tracked through the consecutive frames of a time-lapse video via custom software. The tracking information was then used to quantify the total lifetime as well as the number of puncta in each frame. The resulting quantitative information confirmed the hypothesis that dendritic filopodia play a role in probing the environment for synaptic contacts. Furthermore, the Neill et al. provide evidence that only a small fraction of filopodia persist as stable branches. The data (85) suggest that, as opposed to needing to pull in an axon through dense neuropil, dendrites essentially grow out to meet it via filopodial stabilization at the point of contact.

5.2. Monitoring Embryonic Heart Development

The motion induced by the fast cardiac contraction of the beating embryonic zebrafish heart is the main challenge that must be addressed. Liebling et al. (84) address this challenge by improving the microscope used for image acquisition and by developing specific software that is necessary to assemble the 4D image sequence. A newly developed slit-scanning confocal microscope allows the
image capture at 120 frames per second for a 512 × 512 pixel image. The conventional pinhole of a laser-scanning microscope is being replaced by a slit with adjustable width. This design enables the high-frame-rate imaging while retaining the ability to block out-of-focus light from regions above and below the plane of interest. Here, a highly efficient line-array charge-coupled device (CCD) is used for image capture. One can produce a full frame at a given depth z by scanning the line across the sample in the y direction.

The resulting 2D image slices are temporally out of phase with one another. By taking advantage of the heart’s periodic motion and the similarity of the image slices at adjacent optical sections, researchers can assemble the data into a 4D data set. In the first step, a time series of 2D slices is acquired at a fixed depth in the beating heart without any external gaiting signal. The stage is then moved to acquire a new time series at a different phase in the heart cycle. Through the assumption that pixel intensities in adjacent sections are similar, the shift between the given time sequences is estimated by minimizing the squared differences of the intensity values. Additional constraints are necessary to reconstruct sequences that include aperiodicities. The atrial and ventricular volumes are determined with the help of semiautomatic segmentation methods.

Liebling and collaborators used this system for collecting images of beating hearts of transgenic zebrafish embryos expressing GFP under control of a GATA-1 promoter. GFP is expressed strongly in blood cells and weakly in both the endocardial and myocardial layers of the heart. By making use of these labels, the researchers estimated 3D architecture and blood flow from the reconstructed sequences. The fully developed zebrafish heart has two chambers consisting of a single atrium and a single ventricle. The detailed quantitative analysis provided by Liebling and collaborators provided new insights into the appearance of functional valves. They observe, for example, a clear separation between the atrial and ventricular blood volume ~110 h after fertilization. Such a detailed analysis of the blood flow in the zebrafish heart has not been possible before.

5.3. Embryogenesis of Zebrafish

The study of the embryonic development of zebrafish (88) is the goal of the Digital Fish Project. Because the embryos and larvae of zebrafish are transparent, the entire embryo can be imaged through the use of optical microscopy. Megason and collaborators aim to develop image analysis methods that track all the cell movements, cell divisions, and cell deaths that occur during tissue formation. In addition, they plan to register and compare the data from different embryos. A genetic approach named FlipTraps enables the comprehensive spatiotemporal analysis of gene expression and function in the developing embryo (see also Figure 9). The FlipTrap cassette forms a functional fusion protein with a green fluorescent tag when inserted into an intron, the DNA region within a gene that is not translated into protein. These fusions generally do not disrupt the function of the trapped protein. This technology makes it possible to see what tissues and cells express a protein at different times of development. Because the fluorescent proteins are genetically encoded, the specimen itself can be made to produce the label.

To compare results between different experiments, Megason et al. propose a systematic and standardized method of image acquisition. Each embryo is mounted in a reproducible method. A series of optical sections is then captured at different focal planes to generate a z-stack. Image registration methods, similar to those discussed in Section 2.1, are necessary to montage the optical sections. The resulting data volumes, which can contain up to 100 z sections and 500 time points, can be up to 150 GB large. Segmentation and tracking methods are currently developed to form lineage trees. Finally, the group plans to identify the different cell types and tissue structures. One important aspect of the project is that all of the data set as well as the necessary tools are being made available to the public.
Figure 9
Fluorescent labels used in the Digital Fish Project. Confocal sections showing the (a) green channel and (b) red channel of the eye of a zebrafish embryo labeled with Histone-EGFP and membrane-localizing mCherry. (c) Nuclei are better resolved by subtracting the image in panel b from the one in panel a. (d–f) Histone-EGFP also marks mitosis as shown in three consecutive frames of a dividing cell, 4 min between frames. Figure courtesy of S. Megason.

6. ENTERPRISE-LEVEL APPLICATIONS

The design of algorithms, tools, and workflows that enable the longitudinal study of developmental processes is the focus of the projects presented as in toto imaging approaches. The development of techniques that allow the detailed analysis of a large number of biological samples in an automated fashion poses the challenges of scale, in terms of both sample size and derived data. Although the concept of high-throughput screening is certainly not new, existing high-throughput techniques should not be confused with emerging applications that aim to extract high-content data from a large number of complex biological samples. For the purpose of this review, I refer to these types of efforts as enterprise-level applications.

Although imaging and image analysis methods form a critical and enabling component of such applications, they comprise only one step in a complex workflow. A number of challenges need to be addressed to scale up sample preparation and sample handling. The resulting data sets are typically so large that a manual review of the data is prohibitive. It is necessary to develop computational approaches that mine the resulting data then summarize and extract the relevant information. If successful, efforts like these will become blueprints on how biology can be studied in a quantitative, data-driven fashion.

The length and scope of this article does not allow a complete overview of such enterprise-level projects. Rather than aim for completeness, I present each of the studies to highlight certain important aspects. The study of the tumor microenvironment is fundamental for understanding the orchestrated sequence of events that leads to tumor progression at a cellular as well as a molecular scale. Mosaliganti and collaborators (90) established a
workflow that enables the study of mammary gland ducts. Because mammary glands are potential sites for tumor initiation, they provide a suitable model to study the role of the tumor microenvironment.

The development of a technology platform that provides the ability to multiplex a large number of protein targets in a single tissue section is the goal of the molecular pathology program being carried out at General Electric’s Global Research Center. Advances in fluorescent staining, image analysis, and bioinformatics were necessary to implement this goal. This newly developed technology platform allows the analysis of more than 30 different protein markers in the context of the spatial organization of the given tissue section.

Previous sections of this article highlight tools and techniques that will lead to a better understanding of neuroanatomy. Here I summarize the study of the adult fly brain, which is in progress at the Howard Hughes Medical Institute’s Janelia Farm Research Campus.

6.1. Quantification of Structural Phenotypes in Tissues

In recent years, the approach of studying cancer by focusing on the tumor microenvironment (TME) has gained considerable importance. This is motivated, in part, by studies suggesting that tumor stroma strongly influences the initiation, progression, and metastasis of cancer. The communication or signaling events between the tumor cells and the surrounding cells in the microenvironment helps drive the process of tumor progression (89). A salient approach to studying these influences is understanding the spatial structure of cells in thick tissue sections from complex mammalian model organisms. The spatial organization of the cells, in turn, can be obtained via imaging sections of tumor tissue labeled with fluorescence markers. Here, determining the phenotype of individual cells and analyzing their spatial organization are critical requirements, and comprehensively characterizing the microstructure of the tissue layers of interest is necessary. The tissue layer is composed of multiple components or phases, each derived from different material distributions arranged to attain various spatial arrangements and material interfaces.

Mosaliganti et al. (90) present an imaging workflow that includes the acquisition, reconstruction, quantitative analysis, and visualization of sets of large serial sections obtained from a mouse placenta. In the first stage, large sets of histological slides are produced and digitized. A set of standard processing steps corrects for illumination and staining artifacts and detects cell nuclei. Subsequently, image registration methods are utilized to align the set of 2D images to assemble a 3D data volume. The segmentation process identifies regions corresponding to different tissue structures. One of the novel contributions of this work (90) is the use of tessellations to provide refined segmentations. Rather than trying to capture boundaries correctly, a more meaningful solution is to tessellate the image into regions that house an individual nucleus completely. Our methods iteratively alter the initial Voronoi tessellations to provide better nuclei separation by accounting for the overlap regions. Various (isosurface) views and volume renderings of the ductal architecture are shown in Figure 10b. The cell shapes are extrapolated across the slices to get a 3D rendition.

The resulting 3D structure and quantitative measurements on the specimen enable further modeling in systems-biology study. Whereas some of the algorithms presented here are optimized for characterizing phenotypical changes in the mouse placenta in gene knockout experiments, the architecture of the workflow enables the system to be easily adapted to countless biomedical applications, including our exploration of the organization of tumor microenvironment.
Figure 10
Enhanced visualization of duct sequences. (a) The segmented volume consisting of nuclei, extracellular matrix, lumen, and the duct. (b) True-color rendering showing the epithelial nuclei surrounding the duct. (c) Epithelial nuclei lying around a branching duct. A cutaway section reveals the underlying lumen. (d) Top view of the duct with the see-through hole. Figure courtesy of R. Machiraju (90).

6.2. Multiplexing through Sequential Staining

The detection of one or more protein expression profiles on cells or tissues in situ opens up the possibility of studying signal pathways involved in both normal development and diseases, as well as the effect of drug treatment. These techniques, sometimes referred to as expression proteomics, hold particular promise in the fields of disease biomarker discovery, toxicology, and drug-target evaluation. Immunohistochemistry (IHC)—more specifically immunofluorescence—is the technique of staining tissue with fluorescence-labeled antibodies against a particular protein,
and it provides important protein expression profiling in situ. It allows detection of the location of protein expression and detection of the proteins' spatial relationship in the context of tissue morphology. This cannot be achieved through the use of flow cytometry or western blot because they require the disaggregation of the tissue.

The multiplexing technology developed by GE Global Research has significantly advanced the state of the art of IHC by allowing the analysis of a large number of proteins in the context of tissue morphology. It employs a proprietary chemical bleaching technique between antibody staining rounds; one key risk is that this chemical application may impact the integrity of the antigen, thus reducing antibody affinity or making the antigen no longer recognizable by the antibody. Therefore, to address this risk, researchers will compare the antibody staining intensity with the tissue before and after treatment of the chemical agent to ensure no difference.

This sequential staining technique (91) enables high-throughput detection of multiple proteins in the same tissue samples. Automated image analysis methods enable, as shown in **Figure 11**, the registration of all images acquired after each round of staining and the analysis of the expression profile at the subcellular level. Currently under development is a set of tools that analyze the subcellular protein expression patterns; these tools allow utilization of data for basic research studies as well as preclinical applications. The data generated from a first study by Ginty et al.

---

**Figure 11**

Molecular pathology. The technique allowing the multiplexing of a large number of biomarkers in the same tissue is enabled through a set of innovations that include fluorescent labeling, image analysis algorithms, and bioinformatics. These innovations help summarize the detailed protein expression information from a single tissue or an entire study. The figure illustrates the schematic workflow of image processing procedures being used.
(91) indicate that the relative subcellular distribution of Met, as measured by the newly developed automated image analysis, may be a valuable biomarker for estimating colon cancer prognosis. The group is currently working with pharmaceutical partners on a series of preclinical studies.

6.3. Optical Imaging of the Fly Brain

Advances in neural imaging and molecular tools are providing exciting new methods for investigating the development and function of the nervous system. Owing to its small size but behavioral complexity, the brain of the adult fruit fly (Drosophila melanogaster) is now largely accepted as an important model for brain function. The goal of this particular research is the development of a statistical atlas of the Drosophila brain. In Drosophila, development of the P[GAL4] enhancer-trap system, which is a biochemical method used to study gene expression and function in organisms, has been of particular significance. Not only does it mark cells for visualization but it also provides a means for manipulating them in the living organism. This technology makes it possible to visualize specific axonal processes in the brain (92). For the purpose of this atlas development, synaptic neuropils in the brain tissue are stained using the monoclonal antibody nc82, which visualizes the entire shape of the brain. Multiphoton laser-scanning microscopy is used to directly acquire 3D image stacks of the entire brain of the fly.

Statistical atlases are commonly used to model the anatomy of a specific organ. As opposed to a single reference image, multiple instances are being used to learn a statistical representation of the anatomy. There has been extensive research in the field of medical imaging to develop the necessary tools and techniques to both model and analyze organ morphology. Conceptually, this 3D statistical neuronal atlas of the Drosophila brain is similar to the digital cell atlas of C. elegans (93) that Long et al. built previously. This endeavor is, however, significantly more challenging. The Drosophila brain is estimated to have 100,000–150,000 neurons, significantly more than the total number of cells of the C. elegans worm, which is approximately 1000. Once completed, this atlas will include the statistics of neuronal distributions, projections, and connections (see Figure 12). This statistical information can then be used to model the wiring of neurons in

![Figure 12](image-url)
the fly brain and to detect the associations between the neuronal distributions and animal behaviors.

Registration and segmentation are tasks that need to be performed as a first step in this workflow and related ones. For the processing of a large amount of samples, robustness is an important criterion. Lam et al. (94) evaluate different approaches. Segmentations that were obtained using level-set methods (95) suffered from oversegmentation, even after morphological filtering for hole filling. Lam et al. (94) improved segmentation performance by formulating the segmentation problem as a quest to find a deformable cutting surface, which is defined in a 3D mesh. They proposed a suitable energy function that controls the localization of the cutting surface. Validation was performed on 200 real 3D image volumes from two different sources, and the overall segmentation accuracy of the deformable model segmentation was 85%.

7. OPEN CHALLENGES

The techniques and studies reviewed in this article provide evidence that sophisticated image analysis does not only accelerate biology research by automating certain tasks, such as cell counting and cell tracking, but it also enables a type of data-driven research in biology that would otherwise not be possible. In comparison with the field of biology at large, the biomedical imaging community is still very small. This does, of course, raise the question of whether the discussed methods will be accepted and adopted by the biological research community. This process will depend on the usability, availability, applicability, and robustness of image analysis algorithms and tools. To conclude this article, I discuss these issues and outline some of the open challenges.

7.1. Open Source + Open Data = Open Science

Today, the benefits of open source software are widely acknowledged. Two extremely successful examples of open source software tools that are relevant to biomedical imaging are undoubtedly ImageJ (96) and the Insight Segmentation and Registration Toolkit (ITK) (97). Whereas ImageJ offers a good user interface that includes only a limited set of image analysis functions, ITK provides a set of carefully implemented data structures and algorithms. Importantly, all algorithms, which include various image registration and image segmentation methods, have been thoroughly tested and validated.

A number of research groups have already made a great effort in making their software publicly available. The Open Microscopy Environment (98–100) includes a set of tools that support the archival, visualization, and annotation of microscopy images and metadata. The tools are designed to interact with other existing open source and commercial offerings. Carpenter and collaborators have created CellProfiler (101), an open source, modular system for high-content, high-throughput screening. It includes methods for illumination correction, cell segmentation, and single-cell measurements. Some of the methods for the segmentation of nuclei and the tracing of vessels in 3D multichannel images that were reviewed in Section 2.2 have been made available as part of the FARSIGHT project (28). With a system named V3D (http://penglab.janelia.org/proj/v3d/), Hanchuan Peng provides a fast and powerful 3D image and surface rendering tool that also includes some 3D image analysis and metadata management capabilities.

Although image acquisition is a routine task for common medical imaging and computer vision applications, it is an important milestone in the research process of many microscopy applications that are reviewed in this article. Consequently, access to some of those data sets is in fact limited. In the Digital Fish Project (see Section 5), Megason et al. (9) specifically address this problem.
by making both the data set and the developed tools publicly available. The field would greatly benefit if—as is discussed in the following section—the associated metadata that would enable algorithm validation and comparison could also be provided.

7.2. Constancy of Purpose as a Secret of Success

As is the case with any new or emerging technology, there is the question whether it will be in a position to deliver on its promise. As suggested by the title of this section heading (taken from a Benjamin Disraeli quotation), developing an awareness of some of the important long-term factors and challenges will help ensure acceptance of image analysis methods. Scientific rigor requires a thorough validation of any of the proposed methods and the ability to reproduce experimental results. Although the topic of algorithm validation is not discussed in sufficient detail in this review, it is of paramount importance. Obtaining validation data that is of significantly high quality is a tremendously challenging task. Even in the case of spheroid tumor models, how the data should be interpreted is not often clear. The high variance in sample quality poses additional challenges. The process of validation is cumbersome because input from multiple domain experts is required to ensure that the automatically generated result is correct. In many cases, computing a consensus because the true interpretation of the data is not known is necessary, as formally introduced in the context of image segmentation (102).

In addition, it is often not clear which particular image registration, segmentation, or tracking algorithm is well suited for the given data type. Today, most data sets are being analyzed by experts trained in the field. Many researchers who are not familiar with image analysis algorithms and their mathematical foundations will simply be overwhelmed. It would be helpful if the decision on what type of algorithm should be used, or what particular parameter setting should be used, could be made automatically as suggested by A. Plant (personal communication).

Establishing a set of publication guidelines could help ensure the reproducibility of results. In addition, it would facilitate an effective comparison of related methods. The proteomics community agreed on a set of standards (103) regarding what information must be published about a given experiment. Formulating a similar set of guidelines would potentially help raise the level of acceptance of quantitative studies that were enabled through image analysis methods.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED


