

## Data and text mining

## Bioimage informatics: a new area of engineering biology

Hanchuan Peng

Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA

Received on June 6, 2008; revised on July 1, 2008; accepted on July 2, 2008

Advance Access publication July 4, 2008

Associate Editor: Jonathan Wren

## ABSTRACT

In recent years, the deluge of complicated molecular and cellular microscopic images creates compelling challenges for the image computing community. There has been an increasing focus on developing novel image processing, data mining, database and visualization techniques to extract, compare, search and manage the biological knowledge in these data-intensive problems. This emerging new area of bioinformatics can be called 'bioimage informatics'. This article reviews the advances of this field from several aspects, including applications, key techniques, available tools and resources. Application examples such as high-throughput/high-content phenotyping and atlas building for model organisms demonstrate the importance of bioimage informatics. **The essential techniques to the success of these applications, such as bioimage feature identification, segmentation and tracking, registration, annotation, mining, image data management and visualization, are further summarized, along with a brief overview of the available bioimage databases, analysis tools and other resources.**

**Contact:** pengh@janelia.hhmi.org**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

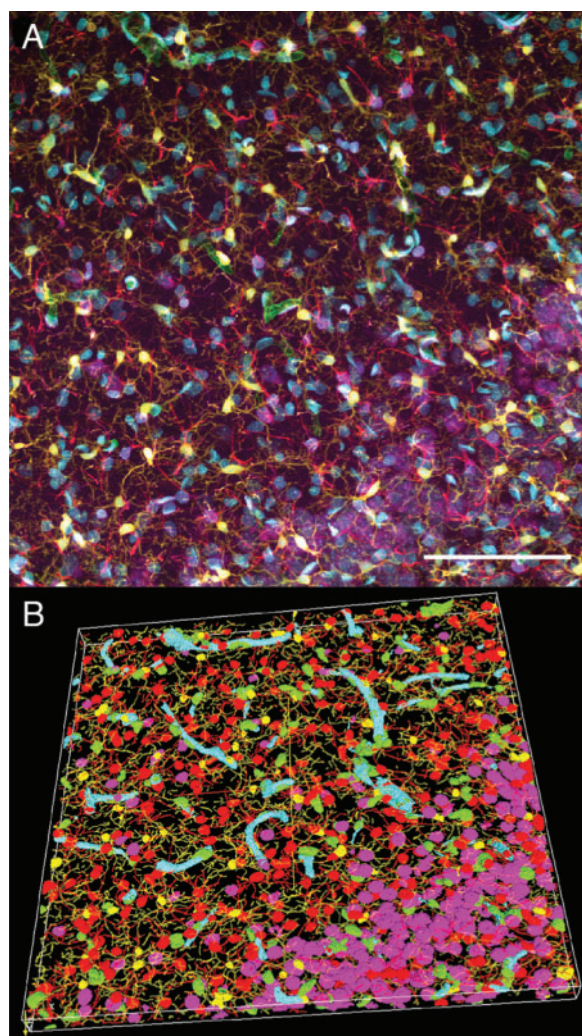
## 1 INTRODUCTION

In the last several decades, numerous biomedical imaging techniques were developed, ranging from the whole organism level (millimeter resolution) down to the single molecule level (nanometer resolution) (Murphy, 2001; Tsien, 2003). Some of the most widely used biological imaging methods include **confocal or two-photon laser scanning microscopy** (LSM) (Pawley, 2006), **scanning or transmission electron microscopy** (EM) (Bozzola and Russell, 1999), etc. Novel imaging techniques such as PALM (Betzig *et al.*, 2006), STORM (Rust *et al.*, 2006), STED (Hell, 2003) that far surpass the resolution of conventional optical microscopes currently can pinpoint the location of individual proteins that are only several nanometers apart. Along with the dramatic advances of many related techniques such as image signal digitization and storage, biological tissue labeling [e.g. green fluorescent proteins (GFP) and enhanced GFP (EGFP) (Heim *et al.*, 1995; Shimomura *et al.*, 1962), Dronpa (Ando *et al.*, 2004), Brainbow combinatorial labeling (Livet *et al.*, 2007)], the number of biological images (e.g. cellular and molecular images, as well as medical images) acquired in digital forms is growing rapidly. Large bioimage databases such as Allen Brain Atlas (Lein *et al.*, 2007) and the Cell Centered Database (CCDB; Martone *et al.*, 2002) are becoming available. These image data could involve (1) **two-dimensional (2D) or 3D spatial information**, (2) multiple

colors which may correspond to various molecular reporters, (3) **4D spatio-temporal information for developing tissues or moving cells**, (4) various co-localized biological signals such as mRNA expression levels of different genes (Lein *et al.*, 2007; Long *et al.*, 2007b; Peng *et al.*, 2007) or (5) other screening experiments related to RNA interference (RNAi), chemical compounds, etc. (Echeverri and Perrimon, 2006; Moffat *et al.*, 2006; Sepp *et al.*, 2008). Analyzing these images is critical for biologists to seek answers to many biological problems, such as differentiating cancer cell phenotypes (Long *et al.*, 2007a), categorization of neurons (Jefferis *et al.*, 2007), etc.

The deluge of complicated biological and biomedical images poses significant challenges for the image computing community. As a natural extension of the existing biomedical image analysis field, an emerging new engineering area is to develop and use various image data analysis and informatics techniques to **extract, compare, search and manage the biological knowledge** of the respective images. This new field can be called bioimage informatics. However, due to the great complexity and information content in bioimages, such as the very high density of cells (e.g. astrocytes, microglia, neurons) intertwined together (Fig. 1A), or very rapid microtubule growing process in a 4D movie of live cells, it is very challenging to directly apply existing medical image analysis methods to these bioimage informatics problems. **Special techniques such as those developed in the EARSIGHT project (Roysam, 2008) will be necessary to analyze these complicated image objects** (Fig. 1B). In addition, usually a single biological image stack has a large size (several hundreds of megabytes or even several gigabytes) and several color channels. The objects of interest in such an image, for instance the 3D structures of neurons, could have dramatic variations of morphology and intensity variations from image to image. It is yet not uncommon that thousands of images need to be automatically analyzed in a high-throughput way, in terms of the number of hours or days, but not months or years of manual work. All these difficulties make **it necessary to develop novel bioimage informatics algorithms and systems, especially from three aspects: image processing and mining, image database and visualization.**

Many studies of bioimage informatics are either underway or have been done over the last few years. Several very successful workshops (e.g. [bioimageinformatics.org](http://bioimageinformatics.org)) were organized to discuss the latest developments of this field. The goal of this essay is to briefly review the advance of bioimage informatics from the angles of applications, key techniques, available tools and resources. First, in Section 2 several application studies on the high-through biology, model organisms, etc., are introduced. Further, in Section 3 the desired computational techniques, including bioimage feature identification,



**Fig. 1.** (A) Maximum projection of a 5-channel confocal 3D image of a 100  $\mu\text{m}$  thick section of rat hippocampus. Red: GFAP-labeled astrocytes; green: EBA-labeled blood vessels; yellow: Iba1-labeled microglia; cyan: CyQuant-labeled cell nuclei; purple: NeuroTrace-labeled Nissl substance; scale bar = 50  $\mu\text{m}$ . (B) 3D rendering (with a similar color scheme) of the segmented and classified cells produced using the **FARSIGHT** techniques for (A). Image courtesy of Badrinath Roysam (Bjornsson *et al.*, 2008)

segmentation, registration, annotation, mining, indexing, retrieval and visualization, are discussed. In Sections 4 and 5 the available tools and resources are summarized. While in this short article, it is difficult to include all the important work, and to explain the details of the discussed applications and computing methods (such as their biological objectives, challenges and findings), I hope that the presented facts and links can be helpful for both researchers in this field and general audiences who may have interests in learning the basic ideas of bioimage informatics.

## 2 APPLICATIONS

Just like many other engineering fields, bioimage informatics is **application-driven**, as one can see from the following non-exclusive instances.

### 2.1 High-throughput and high-content analysis of cellular phenotypes

Large-scale screening of cellular phenotypes, at whole-cell or sub-cellular levels, is of importance for determination of gene functions, delineating cellular pathways, drug discovery and even cancer diagnosis. The CellProfiler system (Carpenter *et al.*, 2006; Lamprecht *et al.*, 2007) was developed to screen cellular images rapidly and gather information such as number of cells, size and other morphological features of cells, per-cell protein levels, cell cycle distribution, etc. This system has been used to detect various cell phenotypes, such as *Drosophila* Kc167 cells, whose images are often textured and clumpy, and human HT29 cells, which are smooth and elliptical. Intelligent human-computer interface and content-based image retrieval relevance feedback were also used to enable high-content screening of *Drosophila* (fruit fly) neurons (Hong, 2006; Lin *et al.*, 2007). Analysis of the morphological signatures of cells was used to study signaling pathways related to cell protrusion, adhesion and tension (Bakal *et al.*, 2007).

For high-resolution intracellular analysis, 3D protein location patterns associated with a number of subcellular organelles and components such as nucleus, nucleolus, mitochondria, cytoskeleton, etc., can be described and classified using fluorescence image features, such as Haralick textures features and Zernike moments (Murphy *et al.*, 2003). Spatial patterns may also be considered in clustering analysis and used for prediction of breast cancers (Long *et al.*, 2007a). More systematic descriptions, such as generative models for subcellular locations of proteins, can provide information for systems biology study (Zhao and Murphy, 2007).

### 2.2 Atlas building for model organisms

Bioimage informatics methods were used to study widely used model organisms, such as mouse (Dorr *et al.*, 2008; Lein *et al.*, 2007; Ng *et al.*, 2007), fruit fly (Luengo Hendriks *et al.*, 2006; Peng and Myers, 2004; H. Peng *et al.*, unpublished data), *Caenorhabditis elegans* (Liu *et al.*, 2008; Long *et al.*, 2007b), zebrafish (Megason *et al.*, 2007), etc. One very important aspect is to build various digital atlases of these organisms, and further integrate the respective anatomical and ontological knowledge into databases.

Allen Brain Atlas (Lein *et al.*, 2007) integrates the genome-wide RNA *in situ* hybridization (ISH) gene expression information of 20 000 mouse genes. Besides a manually generated reference atlas, the Anatomic Gene Expression Atlas (AGEA) is an interactive 3D atlas of the adult mouse brain based on ISH gene expression images. AGEA is based on approximately 4000 coronal gene sets, which allows anatomic specification and browsing based on 3D spatial coordinates and expression threshold control. With the pixel resolution at  $\sim 25 \mu\text{m}$ , Allen Brain Atlas provides very useful information for studies close to the cellular level.

Single-cell analysis for an entire animal is useful for understanding the cell functions, such as the neuronal circuit mapping based on 3D cellular images of a brain. This task is possible if the cells have unique identities, indicated by the stereotypy of their 3D locations, 3D morphology, birth orders (lineages), gene expression patterns or other functional properties. Several systems do have these distinct properties. In *C.elegans*, each cell has a unique lineage and identity. A recent development is the building of the single-cell atlas for the L1 stage of *C.elegans* (Long *et al.*,

2007b). It is based on a series of bioimage-processing and mining techniques including *C.elegans* worm body straightening (Peng *et al.*, 2008a), nuclei segmentation (Long *et al.*, 2007c), annotation and cell identification (Long *et al.*, 2008; Peng *et al.*, 2008b) and atlas modeling. With this atlas, systematic and high-throughput analysis of gene expression at the truly single-cell level, instead of clusters of cells, becomes feasible (Liu *et al.*, 2008). Several other pieces of similar work are underway for different systems, e.g. a fruit fly adult brain (H. Peng *et al.*, unpublished data).

## 2.3 Understanding the dynamic processes in cells and living organisms

For intracellular processes, the microtubule, one class of the cytoskeleton polymers that is constantly assembled and disassembled, receives much attention in studies of various cell functions, e.g. cell division. By imaging GFP fused to the distal ends of microtubules, it is possible to analyze the different dynamic patterns of microtubules, such as the velocity and acceleration, for mutants or under other conditions. Computationally, the microtubule growing, shortening and other dynamic patterns can be tracked in time-lapse microscopy images, via mixture analysis of hidden Markov models (Altinok *et al.*, 2006; Altinok *et al.*, 2007), minimum shared decomposition of directed graphs derived from the microtubule spots (Swidan *et al.*, 2007), particle filtering (Smal *et al.*, 2008), multiscale tip and body model (Jiang *et al.*, 2005), detecting individual segments and linking (Danuser *et al.*, 2000; Hadjidemetriou *et al.*, 2004; Meijering *et al.*, 2006). Hierarchical, agglomerative clustering analysis of various yeast mutants based on kinetochore microtubule dynamics was also reported (Jaqaman *et al.*, 2007).

For developmental biology, visualizing how genes are expressed in living organisms allows us to gain insight in the interactions of gene products. For developing zebrafish embryos, *in toto* imaging based on time-lapse, LSM were used to track cells in the four dimensions of space and time (Megason *et al.*, 2008). Image analysis methods were developed to read out quantitative, cell-based protein expression patterns and transcriptional expression patterns *in vivo*. The *in toto* imaging analysis approach is suitable for studying animal development from a systems biology perspective. For cases where it is difficult to directly observe how 3D spatial patterns of gene expression change over time, manifold learning can be used to computationally reconstruct the 4D spatio-temporal developmental dynamics of these patterns. For developing fruit fly embryos, spatial registration and comparison of 3D gene expression patterns were developed and conjugated with an approximation algorithm of the Traveling Salesman problem, to reconstruct the developing dynamics of genes such as *ftz* and *snail* (Peng *et al.*, 2005a).

## 2.4 Reconstruction of 3D neuronal structures and the wiring diagram of a brain

For neuroscience, there have been a lot of efforts on tracing and reconstruction of 3D structures of neurons, based on optical and electron microscopy images. *Neurolucida* (Glaser and Glaser, 1990), a pioneering software package in this sort, permits users to digitally trace neuronal structures in images. Many automated approaches were developed recently. Directional kernels were used to exploratorily search neuronal topology in confocal images (Al-Kofahi *et al.*, 2002, 2003). A repulsive force-based snake

model was proposed to segment axons in 2D images and then track them in 3D confocal images of transgenic mice that express fluorescent protein (Cai *et al.*, 2006). A graph cut method was used to segment neuronal structures in electron micrographs (Vu and Manjunath, 2008). The convolutional neural network was used to reconstruct the nanometer scale image objects from scanning electron microscopy (SEM) images (Jain *et al.*, 2007). Several automated 3D reconstruction software packages for optical and EM images were also built (e.g. Maack *et al.*, 2007; Y. Mishchenko, personal communication). The FARSIGHT project (Fig. 1), which targets integrating the automated 3D segmentation and tracing algorithms for astrocytes, microglia, neurons, etc., uses a systematic divide and conquer strategy for associative bioimage analysis (Bjornsson *et al.*, 2008; Roysam *et al.*, 2008). Thousands of reconstructed neurons have also been organized into publicly available databases, such as [NeuroMorpho.org](http://NeuroMorpho.org) (Ascoli, 2006). Along with a number of on-going projects on categorizing the types of neuronal structures, or mapping the neuronal circuits, these resources will provide very valuable information to understand and manipulate neuronal circuits.

One of the most exciting challenges in science is to understand how a brain works. The reverse-engineering approach to tackle this problem needs to reconstruct either the anatomical wiring diagram of the brain of an animal (e.g. a fruit fly's brain with 100 000 or so neurons), or the functional wiring diagram of this brain, or both. The aforementioned 3D neuron tracing techniques, as well as image segmentation and neuron classification methods are needed to identify neurons and study their wirings based on electron, optical or functional imaging (e.g.  $\text{Ca}^{2+}$ ) data.

## 2.5 Joint analysis using both bioimage informatics and other bioinformatics methods

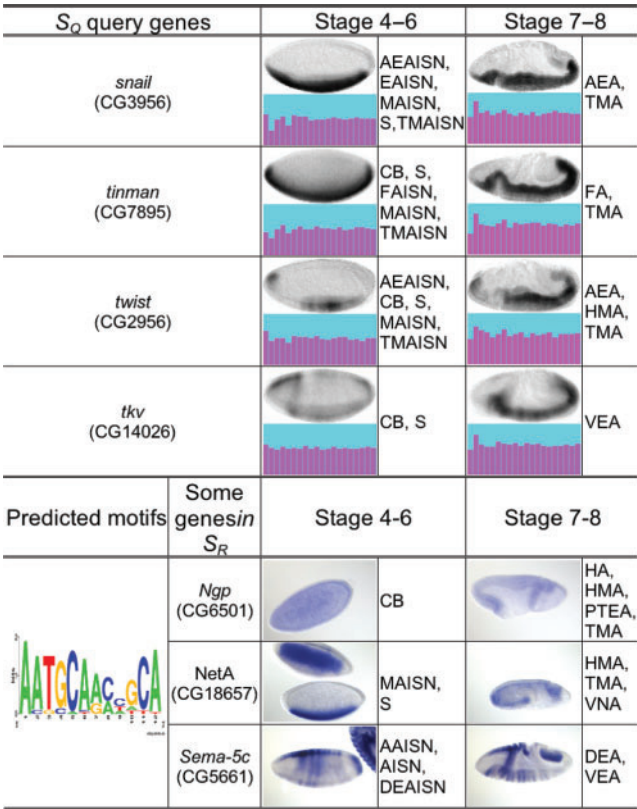
Bioimage informatics techniques can also be paired with conventional bioinformatics methods. For example, clustering embryonic gene expression patterns of fruit fly can be conjugated with comparative genomics approach to predict sequence motifs that may have regulatory functions (Fig. 2) (Peng *et al.*, 2007).

Besides the above examples, several bioimage informatics applications (e.g. functional genomics) have also been discussed in recent articles such as Megason and Fraser (2007) and Meijering *et al.* (2006).

## 3 CRITICAL TECHNIQUES

In order to cope with the complexity in bioimage data, a number of image analyses, machine learning and data mining techniques are needed. Data management and visualization techniques are also required in most bioimage informatics applications. Notably, some particular problems, such as tracking of fibrous microtubule or neuronal structures, may be tackled using different methods, e.g. segmentation versus classification. Therefore, I only review the basic categories of key techniques, but explain very briefly or ignore those more complicated combinations of these basic categories, such as various techniques for modeling. Due to the length limitation, I will also have to skip the signal-processing techniques for biomedical images, such as attenuation correction, deconvolution (Heintzmann, 2007), mixture model estimation, etc., as well as techniques that may be used for general scientific computing but not limited





**Fig. 2.** Clustering analysis of embryonic *in situ* mRNA gene expression patterns of fruit fly genes and its utility in assisting prediction of the regulatory sequence motifs. Based on clustering the eigen-embryo profiles (purple-cyan plot) of representative gene expression patterns, four genes in  $S_Q$  are detected to be co-expressed genes. This prediction is consistent with their known gene regulation relationship for fly mesoderm patterning. Further,  $S_Q$  can be used to predict sequence motifs. The motif example shown is detected using the entire upstream regions of the homologous genes in eight fly species *D.melanogaster*, *D.simulans*, *D.yakuba*, *D.erecta*, and *D.ananassae*, *D.pseudobscura*, *D.virilis*, and *D.mojavensis*, along with three randomly selected example genes in the subsequent genome-wide motif scanning results. BDGP (fruitfly.org) ISH images (in blue) and annotations are also shown, without image cropping or orientation correction. Short terms of annotations: AAISN, amnioserosa anlage in statu nascendi; AISN, anlage in statu nascendi; AEA, anterior endoderm anlage; AEAISN, anterior endoderm anlage in statu nascendi; CB, cellular blastoderm; DEA, dorsal ectoderm anlage; DEAISN, dorsal ectoderm anlage in statu nascendi; EAISN, endoderm anlage in statu nascendi; FA, foregut anlage; FAISN, foregut anlage in statu nascendi; HMA, head mesoderm anlage; HA, hindgut anlage; MAISN, mesoderm anlage in statu nascendi; PTEA, posterior endoderm anlage; S, subset; TMA, trunk mesoderm anlage; TMAISN, trunk mesoderm anlage in statu nascendi; VEA, ventral ectoderm anlage; VNA, ventral neuroderm anlage. Original image source: (Peng *et al.*, 2007)

to bioimage informatics, such as supercomputing with particular computer architecture (Rao *et al.*, 2007).

3.1 Feature extraction and selection

Image features are the fundamental description of pixels/voxels and all higher level objects. Useful image features can correspond to statistical, geometrical, morphological properties and frequency of

image pixels and regions, as well as the topological relationship of multiple image objects. Almost all bioimage-related studies rely on recognizing certain image features. For instance, **points, edges, curves, corners, ridges, textures have been considered in analyzing (e.g. tracking) dynamic fluorescence images** (Dorn *et al.*, 2008).

One way to extract features is based on domain knowledge, as seen in the analyses of fruit fly embryogenesis *in situ* mRNA gene expression patterns. **Local features based on Gaussian mixture model decomposition can be utilized to describe and compare gene expression patterns** (Peng and Myers, 2004). **Global decomposition based on eigen-embryo analysis can be used for clustering these patterns** (Peng *et al.*, 2006). **Wavelet features that capture both global and local frequency properties of these patterns can be used to recognize these gene expression patterns and thus enable automatic annotation** (Peng *et al.*, 2007; Zhou and Peng, 2007). Other useful features, such as those obtained via independent component analysis (Pan *et al.*, 2006) and invariant moments (Gurunathan *et al.*, 2004), were also proposed. Another way for effective features extraction is to consider as many image transformations as possible, and thus generate a rich set of image features. For instance, Murphy *et al.* (2003) considered many features such as texture and moments to characterize the 3D protein location patterns associated with major subcellular organelles and structures. The WND-CHARM system (Orlov *et al.*, 2008) of multipurpose bioimage classification uses compound image features. Five types of features, including pixel statistics, textures, polynomial decompositions, high contrast features (e.g. object number, spatial distribution, size, shape, etc.), and standard image transforms (Fourier, wavelet, Chebyshev) were produced. These features together were used to classify image patterns. One problem with the rich feature set is that it may contain redundant features, which will degrade the performance of classifiers. The minimum-redundant maximum-relevant (mRMR) feature selection algorithm (Ding and Peng, 2005; Peng *et al.*, 2005b) has been used to determine an optimal set of least redundant features, yielding significantly improved recognition accuracy of gene expression patterns (Zhou and Peng, 2007).

3.2 Segmentation

Image segmentation is one of the most basic processing steps in many bioimage informatics applications. While the goal is simply to segment out the meaningful objects of interest in the respective image, this task is non-trivial in many cases. Very complicated cases also exist due to problems such as a low signal-noise ratio and a big variability of image objects. Remarkably, bioimage segmentation strongly depends on the features used. For example, for chromatin composition, texture features can be used, whereas for nuclear morphology, the concavity features may be considered.

Practically speaking it seems intuitive to categorize image segmentation methods for molecular and cellular images based on the overall shape of an image object. One class of segmentation problems is to segment globular objects such as nuclei/cells in 2D or 3D images of cell-based assay, where nuclear compartment may be fluorescently labeled for localization of molecules. Several widely used methods, e.g. **globular-template-based segmentation, watershed segmentation, Gaussian mixture model estimation and active contour/snake methods**, which can be further improved by considering different shape or intensity cues of the objects (Cong and Parvin, 1999; Han *et al.*, 2007; Lin *et al.*, 2003, 2005;

Long *et al.*, 2007c; Parvin *et al.*, 2002). Gradient information will also provide useful cues in some cases (Li *et al.*, 2007). Model-based merging was considered to reduce the over-segmentation (Lin *et al.*, 2005; Long *et al.*, 2007c). Note that sometimes the globular object segmentation could be very tricky, due to the irregular stains of the objects. For example, for a DAPI-stained nucleus, its nucleolus (or nucleoli) may not be stained. As a result, the nucleus will appear to be hollow. This requires special processing such as hole filling before applying the watershed (Long *et al.*, 2007c). Watershed segmentation has also been used for EM image segmentation where the object morphology is irregular and very complicated (Y. Mishchenko, personal communication).

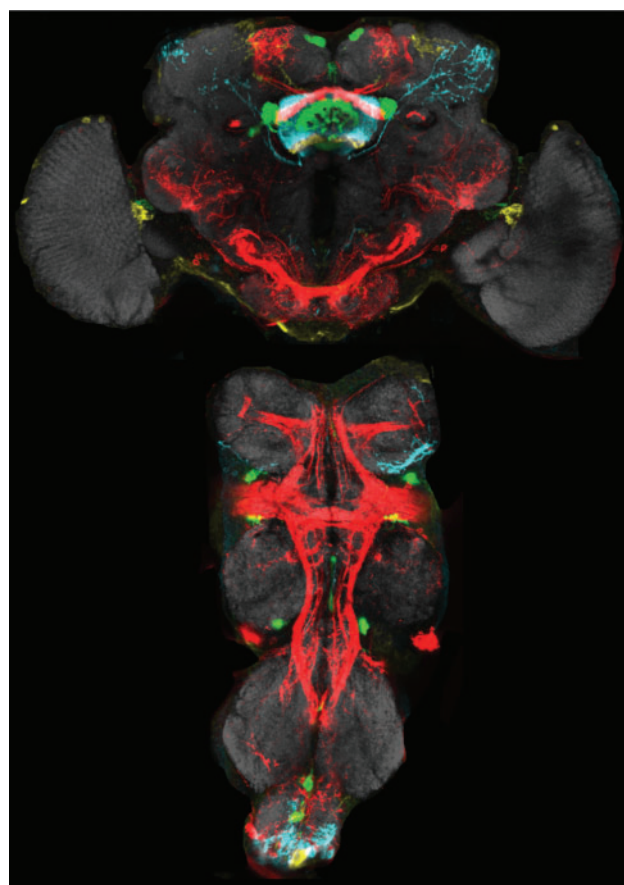
Non-globular object segmentation is often more complicated. One problem of interest is the tracing of neurons in optical images. Some of the latest developments were discussed earlier in Section 2.4. Generally, local search and fitting methods, such as the directional kernels (Al-Kofahi *et al.*, 2002, 2003) have been found effective. Some of these techniques have been commercialized in neuroanatomical analysis software such as Neurolucida (<http://www.mbfbioscience.com>). Other available tools include the ImageJ plugin NeuronJ (Meijering *et al.*, 2004), NeuriteTracer (Longair, 2008).

Image object tracking in fluorescent time-lapse images is another well-studied topic that relies on image segmentation. Many pieces of related work were discussed in Section 2.3.

### 3.3 Registration

Bioimage registration is essential in many applications that need to compare multiple image subjects of different conditions. Quantitative measurements and visualization of comparing patterns in the registered images can be done directly in a ‘standard’ space. Image registration was used in applications such as building the brain atlases (Carson *et al.*, 2005; Ng *et al.*, 2007; Toga and Thompson, 2001), comparison of neuron morphology and gene expression patterns in fruit fly (Ahmmed *et al.*, 2005; Jefferis *et al.*, 2007; H. Peng *et al.*, unpublished data), cardiac imaging of Zebrafish embryos (Liebling *et al.*, 2005), standardization of *C.elegans* images (Peng *et al.*, 2008a). Figure 3 shows one example of the 3D registered fruit fly nervous system, where different GAL4 neuronal patterns highlighted in different colors are mapped into a ‘standard’ space (H. Peng *et al.*, unpublished data). Many of the 2D and 3D image registration methods proposed for medical image analysis, such as the mutual information registration (Volla and Wells, 1997), spline-based elastic registration (Rohr *et al.*, 2003), invariant moment feature-based registration (Shen and Davatzikos, 2002), cone-beam registration (Miller, 2006; Zollei *et al.*, 2005), etc., can be extended to align the molecular and cellular images. However, due to the great complexity and variation of patterns, the big volume of images, (e.g.  $2048 \times 2048 \times 300$  pixels), and a low signal-noise ratio, 3D bioimage registration remains very challenging in general.

Image registration will also help to produce a panoramic scene of the 2D or 3D images that correspond to tiles of tissues. This is often called montaging or tiling. In serial EM, many physical sections are generated for imaging. Each section may also be imaged as many overlapping tiles. Hence, there are two alignment problems: first, stitching all corresponding tiles into a complete single picture, and second, aligning adjacent sections if they have different orientations and deformations (e.g. stretch, shear, compression)



**Fig. 3.** Maximum projection of 3D registered and overlaid neuronal patterns of multiple fruit fly central complexes (top) and thoracic ganglia (bottom), each with a different GAL4 line (Peng *et al.*, unpublished data). Red: a205; Green: EB1; Cyan: NP2320; Yellow: NP6510; gray: NC82-labeled neuropil. Raw confocal images were produced by Julie Simpson and Phuong Chung.

introduced during sample preparation stages such as sectioning and fixation/dehydration/embedding. The first alignment problem can be solved via maximizing the cross-correlation of overlapping regions of neighboring tiles. The second alignment problem can be solved via finding a global 2D affine transformation for adjacent sections, followed by slight local non-linear deformation. Many previous tutorials provide the details (Szeliski, 2006).

Sometimes registration needs to be considered in the domain of extracted image objects, besides aforementioned pixel-domain image alignment. For fruit fly blastoderm embryos, each nucleus can be described using a point in the 3D space. Point cloud registration method was used to generate a virtual fruit fly embryo (Fowlkes *et al.*, 2008). The fairly broad expression patterns of the reference markers, such as the transcriptional factor *evenskipped* which is expressed as seven stripes around an embryo, and the non-trivial variation of the number of nuclei (in the  $\pm 10\%$  range), make it difficult to achieve the single nucleus accuracy for the registered point clouds. For *C.elegans* and the embryonic central nervous system of fruit fly, both the single-cell-level automatic cell recognition technique (Long *et al.*, 2008) and 3D annotation tool WANO (Peng *et al.*, 2008b) have been developed to determine the

identities of cells/nuclei and produce digital point-cloud atlases at single-cell/nucleus resolution.

### 3.4 Clustering, classification and annotation

Many applications such as phenotyping cells and determination of subcellular locations of proteins require the pattern clustering and classification techniques (Arif and Rajpoot, 2007; Chen *et al.*, 2006; Newberg and Murphy, 2008). Multiresolution classification of HeLa cells was proposed (Chebira *et al.*, 2007). Graph-partition-based clustering, such as the minimum-spanning-tree-cut (Peng *et al.*, 2006), was used to group potentially *in situ* mRNA expression patterns of co-regulated genes and thus to detect sequence motifs (Peng *et al.*, 2007). Pattern classification can also help other processing and analysis tasks, for instance the watershed segmentation and grouping of over-segmented objects (Lin *et al.*, 2005; Long *et al.*, 2007c). Automatic determination of cell identities (Long *et al.*, 2008) is also developed, which uses both the absolute 3D location of cells and their relative location patterns to determine the identities of cells. This technique is essential for both high-throughput measuring gene expression level at the single-cell level and manipulating single cells based on optogenetic methods. Cell identity tracking can also be combined with temporal information, as shown in the work to trace lineage of dividing embryonic cells of *C.elegans* (Bao *et al.*, 2006).

Annotation of bioimage objects converts the image content information to concrete semantically meaningful information that is usually texts and can be conveniently organized and searched. This task is often accomplished manually, such as the anatomical and ontological annotation of the gene expression patterns collected for about 5000 fruit fly genes in BDGP database ([www.fruitfly.org](http://www.fruitfly.org)). Automatic annotation of bioimage patterns has begun to be studied (Peng *et al.*, 2007; Zhou and Peng, 2007). Bioimage patterns could correspond to many (e.g. 100 or more) anatomical and ontological annotation terms. Thus this problem can be formulated as pattern classification with hundreds of mutually non-exclusive classes, which falls outside of the framework of conventional multiclass classification that involves a much smaller number (e.g. 10) of mutually exclusive classes. This challenging annotation problem can be solved via parallel classifiers, each performing a bi-classification to indicate if a specific target annotation term should be assigned to the image pattern or not (Zhou and Peng, 2007).

### 3.5 Indexing and retrieval

Currently there are two ways to access the bioimage data in databases. The prevailing method is to provide and organize the text descriptors. These metadata are indexed and thus searchable. They serve as the proxy to find the real image data. Existing relational database indexing and searching techniques can be used. Comparison of biological image patterns is often complicated due to the lack of standards in nomenclature; therefore, it will be a big advantage if annotations stored in a bioimage database are organized based on the controlled/standard ontological vocabulary. The web-based annotation system for fruit fly gene expression patterns in BDGP (Tomancak *et al.*, 2002) provides a set of controlled ontological words used by the curator to assign to an image displayed. Of note, techniques of biomedical ontology and semantic web techniques ([www.semanticweb.org](http://www.semanticweb.org)) can be naturally blended

into bioimage databases. New ontology systems were introduced, e.g. subcellular anatomy of the nervous system (Larson *et al.*, 2007).

The second way is to enable content-based access of the image data in term of raw and processed data. Comparing image patterns requires aforementioned feature extraction, selection and data clustering and classification methods. Various distance metrics, such as Euclidean distances and the earth mover's distance (EMD) (Peleg *et al.*, 1989), can be considered. Recent work (Ljosa *et al.*, 2006) shows that a multiresolution LB-index approach can be used to index the EMD scores. Lower bounds were derived to compute EMD at various resolutions. This approach led to faster similarity query than conventional methods for a database of fluorescent confocal retina images consisting of microglial cells and blood vessels. Query and retrieval on the probability density functions, which may be modeled by adaptive-piecewise-linear approximations, have been developed (Ljosa and Singh, 2007).

### 3.6 Visualization

Bioimage visualization is a subfield of the general scientific data visualization. The widely used techniques for both the original and processed bioimages are volume, surface, flow visualization. Tools for interactive processing and visualization of images for protein surfaces, retinal optical coherence tomographic data and gene expression images of early stage fruit fly embryogenesis were recently developed (Stadt *et al.*, 2007). Scalable volume visualization was used to study cell lineage and gene expression of developing *C.elegans* embryos (Cedilnik *et al.*, 2007). On the other hand, immersive visualization systems, where a user walks into the data volume/model, may enable one to analyze the data like playing a video game. A few systems, such as NCMIR's *ATLAS in silico* system that utilizes CalIT2's 100-million-pixel autostereographic display (West, 2007), the ImmersaDesk™ system (Ai *et al.*, 2005), etc., support such immersive visualization, which requires virtual reality methods.

## 4 AVAILABLE TOOLS

Many tools have been developed for various aspects of the above techniques as well as applications. Some popular tools are summarized below.

### 4.1 Image formats and I/O

Microscope vendors use different file formats to store their raw image data. In addition, users may add customary metadata/tags to the raw or processed images. It is very useful to be able to read, write and convert different file formats. ImageJ (<http://rsb.info.nih.gov/ij/>, Abramoff *et al.*, 2004), empowered by a number of free codes/plugins contributed by volunteers, has the ability to read and write a number of bioimage file formats, such as the Bio-rad PIC file, Zeiss LSM file and others.

Embedding the reading/writing engines of different bioimage formats in one's own code is also desirable. One useful standalone Java library for importing/exporting various bioimage data is Bio-Formats (<http://www.loci.wisc.edu/ome/formats.html>). It can be used in ImageJ, Matlab, etc. With the ability to parsing both pixels and metadata for a large number of formats, it finds a range of applications in the bioimage informatics. Sometimes these images, such as the Zeiss LSM files, are variants of the TIFF

images, therefore can be handled using the open-source libtiff library (<http://www.libtiff.org/>).

## 4.2 Image analysis tools

ImageJ (<http://rsb.info.nih.gov/ij/>, Abramoff *et al.*, 2004) is a Java-based cross-platform tool for biomedical image processing and measurement. A number of image analysis toolboxes such as fluorophore tracking, filament detection, etc., were developed by various groups (Unser, 2008). ImageJ is not only useful for daily use to view and small-scale analysis of images, but can also be deployed to run large-scale analysis in batch.

ITK ([www.itk.org](http://www.itk.org), Yoo *et al.*, 2002) provides a number of image segmentation and registration functions. In this category similar tools include the Matlab image-processing toolbox and other third party toolbox such as the DIPImage toolbox ([www.diplib.org](http://www.diplib.org)).

More and more sophisticated bioimage analysis tasks need tools to perform heavy duty image tasks such as 3D registration of animals' brains, 3D automatic neuron tracing, etc. Several projects are currently underway, such as V3D (H. Peng *et al.*, unpublished data), which tries to integrate a suite of convenient 3D image segmentation, registration, standardization and visualization tools to improve the efficiency of the workflow. Several labs have begun to use the alpha test version of V3D to study the fruit fly nervous systems at embryonic, larval and adult developmental stages. ZFIQ (Zebrafish Image Quantitator) (Liu *et al.*, 2008) is another toolkit, which provides a set of image analysis tools for quantitative, reproducible and accurate interpretation of zebrafish imaging data. Cell-ID (Gordon *et al.*, 2007), an open-source cell finding and tracking package, was developed first for yeast cells, can be used for other regularly shaped cells as well. Other useful analysis packages include CellProfiler (Carpenter *et al.*, 2006; Lamprecht *et al.*, 2007), STARRYNITE (Bao *et al.*, 2006), Neuron Image Quantitator ([neuroniq.cbi-platform.net](http://neuroniq.cbi-platform.net)) and those listed at the NCMIIR site (<http://ncmir.ucsd.edu/downloads/software>).

## 4.3 Database and annotation tools

OME (Open Microscopy Environment) ([openmicroscopy.org](http://openmicroscopy.org)) (Swedlow *et al.*, 2003) is a microscopic image and metadata management system. It is divided into several parts, the OME server, which implements image-based analysis or cellular localization and phenotypes, as well as an OME-XML schema language, and OMERO, which is a suite of java-based tools for data storage, management and annotation.

The UCSB Bisque system (<http://dough.ece.ucsb.edu/bisque/>) provides an integrated online environment for users to upload, search, edit and annotate images. It also includes a few analysis and visualization modules.

Several other systems can also build images database and manage tens of thousands of images and associated metadata entries in a scalable way; examples include XNAT (Extensible Neuroimaging Archive Toolkit, [www.xnat.org](http://www.xnat.org)) (Marcus *et al.*, 2007), Biotrue CDMS ([www.biotrue.net](http://www.biotrue.net)) and Axiope e-CAT ([www.axiope.com](http://www.axiope.com)).

Annotating segmented image objects in 3D is another interesting topic. One tool available is WANO (Peng *et al.*, 2008b, <http://research.janelia.org/peng/proj/wano/index.html>), a QT-based cross-platform 3D annotator, which provides a spreadsheet of all segmented 3D-image objects linked to both the 3D view of the raw image and that of the segmentation mask. WANO enables a user to

quickly add or edit the annotations such as cell names/properties in images, as well as editing the segmentation results such as adding or removing segmented objects. This tool has been used to build digital atlases of *C.elegans* and fruit fly (Long *et al.*, 2007b).

## 4.4 Visualization tools

For visualization of multidimensional multicolor images, such as confocal image stacks, commercially available products include Amira (Mercury), Volocity (Improvision), etc. Free visualization tools include Voxel (www.nephrology.iupui.edu/imaging/voxel/), Chimera (www.cgl.ucsf.edu/chimera/), Volume Rover (cvcweb.ices.utexas.edu/software/), many ImageJ plugins, etc. Blender (<http://www.blender.org/>) is often considered in rendering models.

For displaying and browsing large 2D/3D image set such as the stitched EM sections, each of which could easily exceed the size 100 000 pixels by 100 000 pixels, some tools such as Zoomify (<http://www.zoomify.com/>) and HDView (Microsoft) can be used to build the atlas view of a big image, similar to the Google map.

To develop visualization systems, many studies have relied on VTK ([www.vtk.org](http://www.vtk.org)), which provides multilanguage interfaces to a rich set of visualization functions. For heavy-duty visualization tasks such as large volume rendering, people may consider using OpenGL or even GPU programming directly. For building of cross-platform GUI, QT (<http://trolltech.com/products/qt>) and Java are often considered.

# 5 OTHER RESOURCES

## 5.1 Bench test datasets

There are a number of bioimage databases available for various model organisms, including for example: the Allen Brain Atlas database ([www.brain-map.org](http://www.brain-map.org)) with genome-wide *in situ* gene expression patterns for the mouse brain; the interactive and multiresolution database for scanned and annotated images of serial sections of both primate and non-primate brains (Brainmaps.org); the BDGP database ([www.fruitfly.org](http://www.fruitfly.org)) containing *in situ* embryogenesis gene expression patterns of about 5000 fruit fly genes; the GFP expression pattern database for *C.elegans* ([gfpworm.org](http://gfpworm.org)) and the ZFin FishNet ([www.fishnet.org.au](http://www.fishnet.org.au), Bryson-Richardson, 2007) that is a 3D database of zebrafish development from the early embryo to adult.

For different disciplines, there are also many established databases, such as the 3D neuronal structure database Neuromorpho ([neuromorpho.org](http://neuromorpho.org)) (Ascoli, 2006), which arranges the neuronal structures based on animal species, brain regions, neuron types, research labs, etc., and also provides useful neuron structure measuring, comparison and visualization tools. Similarly useful databases include CCDB ([ccdb.ucsd.edu](http://ccdb.ucsd.edu)), which provides a venue for sharing and mining cellular and subcellular data derived from light and electron microscopy, including correlated imaging. CCDB provides the raw data, reconstructed and segmented data for download and includes 2D images and animations. Another interesting database is PSLID ([pslid.cbi.cmu.edu](http://pslid.cbi.cmu.edu)), a database of protein subcellular location images. This database collects 2D through 5D fluorescence microscope images, annotations and derived features in a relational schema. There are also efforts to establish some general bench test datasets. Some authors have contributed data for the OME bench test database currently with



about 10 datasets ([ome.grc.nia.nih.gov/iicbu2008/](http://ome.grc.nia.nih.gov/iicbu2008/)). The Biomedical Informatics Research Network (BIRN, [www.nbirn.net](http://www.nbirn.net)) is a multisite collaboration to facilitate data sharing of different labs; biomedical images and associated metadata of various animal models are available for downloading.

## 5.2 Conferences, special issues and books

There is an increasing interest for research meetings in this new area. The 2005 Bioimage Informatics meeting was held at Stanford University ([bioimageinformatics.org](http://bioimageinformatics.org)). The 2008 meeting at UC Santa Barbara attracted about 150 frontier researchers in this field. The upcoming conference in 2009 will be held at Janelia Farm Research Campus, Howard Hughes Medical Institute. Many other events include workshops on Microscopic Image Analysis with Applications in Biology ([miaab.org](http://miaab.org)), several workshops related to bioimage analysis in the annual IEEE ISBI conferences ([biomedicalimaging.org](http://biomedicalimaging.org)), NIST workshop on 2D/3D image content representation, analysis and retrieval ([www.nist.gov](http://www.nist.gov)), etc.

There are several special issues of journals and books on the topics of bioimage informatics, molecular and cellular image analysis, etc. BMC Cell Biology published a special issue in 2007 (<http://www.biomedcentral.com/1471-2121/8?issue=S1>), including nine papers covering new image analysis and mining algorithms, data visualization, biological applications, enabling supercomputing techniques, and computer vision and machine learning methods to solve other biology problems. It also includes a short summary of the bioimage informatics challenges (Auer *et al.*, 2007), including the demand for bioimage informatics techniques, the need of multiscale imaging, collaboration and communication between biologists and engineers, common bioimage informatics problems and bench test datasets and modeling. Other special editions include for example the IEEE Transactions on Image Processing 2005 special issue on Molecular and Cellular Bioimaging (edited by Murphy, R., Meijering, E. and Danuser, G.), etc. Artech Publishing House is going to publish a book on the Microscopic Image Analysis for Life Science Applications in 2008 (edited by Rittscher, J., Machiraju, R. and Wong, S.).

## 6 DISCUSSIONS AND CONCLUSION

While in the earlier sections, the molecular and cellular images are emphasized, many of the techniques can be used for other biological image or video data. Characterizing the behaviors of living animals in videos relies on a similar set of tracking techniques to phenotyping and tracking microtubule activities. Recent developments include tracking of *C.elegans*, fruit fly, mouse and fish (Armstrong, 2005; Branson and Belongie, 2005; Fontaine *et al.*, 2007, 2008; Fry *et al.*, 2003; Geng *et al.*, 2004; R. Kerr, personal communication; Roussel *et al.*, 2007; Tsechpenakis *et al.*, 2007). Other examples, include the analysis of gel and microarray images (Angulo and Serra, 2003; Jung and Cho, 2002; White *et al.*, 2005; Young *et al.*, 2004), etc.

Remarkably bioimage computing methods are also demanded to improve the quality and throughput of novel digital imaging techniques, e.g. the super-resolution PALM (Betzig *et al.*, 2006) and correlative microscopy (Grabenbauer *et al.*, 2005; Robinson *et al.*, 2001). It is also possible to adaptively acquire fluorescence microscopic images with consideration of image classification accuracy (Merryman and Kovačević, 2005).

The ultimate evaluation standard of bioimage informatics is how these computational techniques can be used to enhance our understanding of the biological entities and ability to solve the respective problems. With a number of new computing tools and databases that are increasingly shared by different research labs, this new engineering biology field will see a boom in the coming years.

## ACKNOWLEDGEMENTS

I thank Fuhui Long, Yuriy Mishchenko, Ting Zhao and the Associate Editor Jonathan Wren for all the suggestions, comments and criticisms that help improve the article significantly, Margaret Jefferies for improvement of the technical writing. I also thank Badrinath Roysam for providing Figure 1, Julie Simpson and Phuong Chung for generating the raw images used for Figure 3, and the anonymous reviewers for suggesting several references.

*Conflict of Interest:* none declared.

## REFERENCES

- Abramoff, M.D. *et al.* (2004) Image processing with ImageJ. *Biophoto. Int.*, **11**, 36–42.
- Ahammad, P. *et al.* (2005) Joint nonparametric alignment for analyzing spatial gene expression patterns in *Drosophila* imaginal discs. *IEEE CVPR 2005*, **2**, 20–25.
- Ai, Z. *et al.* (2005) Reconstruction and exploration of three-dimensional confocal microscopy data in an immersive virtual environment. *Comput. Med. Imaging Graph.*, **29**, 313–318.
- Al-Kofahi, K. *et al.* (2002) Rapid automated three-dimensional tracing of neurons from confocal image stacks. *IEEE Trans. Inf. Technol. Biomed.*, **6**, 171–187.
- Al-Kofahi, K. *et al.* (2003) Median based robust algorithms for tracing neurons from noisy confocal microscope images. *IEEE Trans. Inf. Technol. Biomed.*, **7**, 302–317.
- Altinok, A. *et al.* (2006) Activity analysis in microtubule videos by mixture of hidden Markov models. *IEEE CVPR*, **2**, 1662–1669.
- Altinok, A. *et al.* (2007) Model based dynamics analysis in live cell microtubule image. *BMC Cell Biol.*, **8** (Suppl. 1), S4.
- Amanda, M. *et al.* (2005) Automated microarray image analysis toolbox for MATLAB. *Bioinformatics*, **21**, 3578–3579.
- Ando, R. *et al.* (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science*, **306**, 1370–1373.
- Angulo, J. and Serra, J. (2003) Automatic analysis of DNA microarray images using mathematical morphology. *Bioinformatics*, **19**, 553–562.
- Arif, M. and Rajpoot, N. (2007) Classification of potential nuclei in prostate histology images using shape manifold learning. *IEEE Int. Conf. Machine Vision*, 113–118.
- Ascoli, G. (2006) Mobilizing the base of neuroscience data: the case of neuronal morphologies. *Nat. Rev. Neurosci.*, **7**, 318–324.
- Auer, M. *et al.* (2007) Development of multiscale biological image data analysis: review of 2006 international workshop on multiscale biological imaging, data mining and informatics, Santa Barbara, USA (BII06). *BMC Cell Biol.*, **8** (Suppl. 1), S1.
- Bakal, C. *et al.* (2007) Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science*, **316**, 1753–1756.
- Bao, Z. *et al.* (2006) Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA*, **103**, 2707–2712.
- Betzig, E. *et al.* (2006) Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, **313**, 1642–1645.
- Bjornsson, C.S. *et al.* (2008) Associative image analysis: a method for automated quantification of 3D multi-parameter images of brain tissue. *J. Neurosci. Methods* (in press).
- Bozzola, J. and Russell, L.D. (1999) *Electron Microscopy*. 2nd edn. Jones & Bartlett Publishers.
- Branson, K. and Belongie, S. (2005) Tracking multiple mouse contours (without too many samples). In *Proceedings of the IEEE CVPR 2005*. pp. 1039–1046.
- Bryson-Richardson, R. *et al.* (2007) FishNet: an online database of zebrafish anatomy. *BMC Biol.*, **5**, 34.
- Burgess, H.A. and Granato, M. (2007) Modulation of locomotor activity in larval zebrafish during light adaptation. *J. Exp. Biol.*, **210**, 2526–2539.
- Cai, H. *et al.* (2006) Repulsive force based snake model to segment and track neuronal axons in 3D microscopy image stacks. *NeuroImage*, **32**, 1608–1620.



- Carpenter, A.E. *et al.* (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.*, **7**, R100.
- Carson, J.P. *et al.* (2005) A digital atlas to characterize the mouse brain transcriptome. *PLoS Comp. Biol.*, **1**, e41.
- Cedilnik, A. *et al.* (2007) Integration of information and volume visualization for analysis of cell lineage and gene expression during embryogenesis. *Proc. SPIE*, **6809**.
- Chebir, A. *et al.* (2007) A multiresolution approach to automated classification of protein subcellular location images. *BMC Bioinformatics*, **8**, 210.
- Chen, X. *et al.* (2006) Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE Trans. Biomed. Eng.*, **53**, 762–766.
- Cong, G. and Parvin, B. (1999) Model based segmentation of nuclei. In *IEEE CVPR'99*. pp. 23–25.
- Danuser, G. *et al.* (2000) Tracking differential interference contrast diffraction line images with nanometre sensitivity. *J. Microsc.*, **198**, 34–53.
- Ding, C. and Peng, H. (2005) Minimum redundancy feature selection from microarray gene expression data. *J. Bioinform. Comput. Biol.*, **3**, 185–205.
- Dorn, J.F. *et al.* (2008) Computational processing and analysis of dynamic fluorescence image data. *Methods Cell Biol.*, **85**, 497–538.
- Dorr, A.E. *et al.* (2008) High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult C57Bl/6J mice. *Neuroimage* (in press).
- Echeverri, C.J. and Perrimon, N. (2006) High-throughput RNAi screening in cultured cells: a user's guide. *Nat. Rev. Genet.*, **7**, 373–384.
- Fontaine, E. *et al.* (2007) Model-based tracking of multiple worms and fish. In *ICCV Workshop on Dynamical Vision*.
- Fontaine, E. *et al.* (2008) Automated visual tracking for studying the ontogeny of zebrafish swimming. *J. Exp. Biol.*, **211**, 1305–1316.
- Fowlkes, C. *et al.* (2008) A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm. *Cell*, **133**, 364–374.
- Fry, S. *et al.* (2003) The aerodynamics of free-flight maneuvers in *Drosophila*. *Science*, **300**, 495–498.
- Gelasca, E.D. *et al.* (2008) Evaluation and benchmark for biological image segmentation. In *Proceedings of the IEEE ICIP 2008*. San Diego, CA.
- Geng, W. *et al.* (2004) Automatic tracking, feature extraction and classification of *C. elegans* phenotypes. *IEEE Trans. Biomed. Eng.*, **51**, 1811–1820.
- Giuliano, K. *et al.* (2003) Advances in high content screening for drug discovery. *Assay Drug Dev. Technol.*, **1**, 565–577.
- Glaser, J.R. and Glaser, E.M. (1990) Neuron imaging with Neurolucida – a PC-based system for image combining microscopy. *Comput. Med. Imaging Graph.*, **14**, 307–317.
- Glory, E. and Murphy, R.F. (2007) Automated subcellular location determination and high throughput microscopy. *Dev. Cell*, **12**, 7–16.
- Gordon, A. *et al.* (2007) Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nat. Methods*.
- Grabenbauer, M. *et al.* (2005) Correlative microscopy and electron tomography of GFP through photooxidation. *Nat. Methods*, **2**, 857–862.
- Gurunathan, R. *et al.* (2004) Identifying spatially similar gene expression patterns in early stage fruit fly embryo images: binary feature versus invariant moment digital representations. *BMC Bioinformatics*, **5**, 202.
- Hadjidemetriou, S. *et al.* (2004) Automatic quantification of microtubule dynamics. In *Proceedings of the IEEE ISBI 2004*.
- Han, J. *et al.* (2007) Segmentation of mammosphere structures from volumetric data. In *IEEE ISBI 2007*. pp. 524–527.
- Heim, R. *et al.* (1995) Improved fluorescence. *Nature*, **373**, 663–664.
- Heintzmann, R. (2007) Estimating missing information by maximum likelihood deconvolution. *Micron*, **38**, 136–144.
- Hell, S.W. (2003) Toward fluorescence nanoscopy. *Nat. Biotechnol.*, **21**, 1347–1355.
- Heward, J.A. *et al.* (2005) flyTracker: real-time analysis of insect courtship. In *Proceedings of Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research*, August 30–September 2 2005. Wageningen, The Netherlands, pp. 409–410.
- Hong, P. (2006) Interactive analysis of high-content cellular images via relevant feedback. In *2006 Workshop on Multiscale Biological Imaging, Data Mining and Informatics*. Santa Barbara, CA, USA.
- Jain, V. *et al.* (2007) Supervised learning of image restoration with convolutional networks. In *ICCV 2007*.
- Jaqaman, K. *et al.* (2007) Phenotypic clustering of yeast mutants based on kinetochore microtubule dynamics. *Bioinformatics*, **23**, 1666–1673.
- Jefferis, G.S. *et al.* (2007) Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell*, **128**, 1187–1203.
- Jiang, M. *et al.* (2005) Automated extraction of microtubules and their plus-ends. In *IEEE Workshop on Applications of Computer Vision*. pp. 336–341.
- Jung, H.-Y. and Cho, H.-G. (2002) An automatic block and spot indexing with k-nearest neighbors graph for microarray image analysis. *Bioinformatics*, **18**, S141–S151.
- Lamprecht, M.R. *et al.* (2007) CellProfiler: free, versatile software for automated biological image analysis. *Biotechniques*, **42**, 71–75.
- Larson, S.D. *et al.* (2007) A formal ontology of subcellular neuroanatomy. *Front. Neuroinform.*, **1**, 3.
- Lein, E. *et al.* (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, **445**, 168–176.
- Li, G. *et al.* (2007) 3D cell nuclei segmentation based on gradient flow tracking. *BMC Cell Biol.*, **8**, 40.
- Liebling, M. *et al.* (2005) Four-dimensional cardiac imaging in living embryos via postacquisition synchronization of nongated slice sequences. *J. Biomed. Opt.*, **10**, eid 054001.
- Lin, C. *et al.* (2007) Intelligent interfaces for mining large-scale mri-hcs image databases. In *IEEE 7th International Conference on Bioinformatics and Biomedical Engineering*. Boston, MA, USA.
- Lin, G. *et al.* (2003) A hybrid 3-d watershed algorithm incorporating gradient cues & object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry*, **56A**, 23–36.
- Lin, G. *et al.* (2005) Hierarchical, model-based merging of multiple fragments for improved 3-D segmentation of nuclei. *Cytometry*, **63A**, 20–33.
- Liu, T. *et al.* (2008) ZFIQ: a software package for zebrafish biology. *Bioinformatics*, **24**, 438–439.
- Liu, X. *et al.* (2008) Molecular signatures and gene expression at the single cell level in *C. elegans*. *Stanford University Technical Report*.
- Livet, J. *et al.* (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, **450**, 56–62.
- Ljosa, V. and Singh, A.K. (2007) APLA: indexing arbitrary probability distributions. In *Proceedings of the 23rd International Conference on Data Engineering (ICDE)*.
- Ljosa, V. *et al.* (2006) Indexing spatially sensitive distance measures using multi-resolution lower bounds. In *Proceedings of the 10th International Conference on Extending Database Technology*. pp. 865–883.
- Long, F. *et al.* (2007a) Phenotype clustering of breast epithelial cells in confocal images based on nuclear protein distribution analysis. *BMC Cell Biol.*, **8** (Suppl.1), S3.
- Long, F. *et al.* (2007b) A 3D digital cell atlas for the first larval stage of *C. elegans* hermaphrodite. *HHMI JFRC Technical Report*.
- Long, F. *et al.* (2007c) Automatic segmentation of nuclei in 3D microscopy images of *C. elegans*. In *Proceedings of the IEEE ISBI 2007*. pp. 536–539.
- Long, F. *et al.* (2008) Automatic recognition of cells (ARC) for 3D images of *C. elegans*. In *Lecture Notes in Computer Science: Research in Computational Molecular Biology*. Springer, Berlin, Heidelberg, pp. 128–139.
- Longair, M. (2008) Available at <http://homepages.inf.ed.ac.uk/s9808248/imagej/tracer/>.
- Luengo Hendriks, C.L. *et al.* (2006) 3D morphology and gene expression in the *Drosophila* blastoderm at cellular resolution I: data acquisition pipeline. *Genome Biol.*, **7**, R123.
- Maack, N. *et al.* (2007) 3D reconstruction of neural circuits from serial EM images. *31st Göttingen Neurobiology Conf.*, **31**, 1195.
- Marcus, D.S. *et al.* (2007) The extensible neuroimaging archive toolkit (XNAT): an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, **5**, 11–34.
- Martone, M.E. *et al.* (2002) A cell centered database for electron tomographic data. *J. Struct. Biol.*, **138**, 145–155.
- Megason, S. and Fraser, S. (2007) Imaging in systems biology. *Cell*, **130**, 784–795.
- Megason, S.G. *et al.* (2008) The digital fish project – in toto imaging and fliptraps for digitizing development. *FASEB J.*, **22**, 253.3.
- Meijering, E. *et al.* (2004) Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry*, **58A**, 167–176.
- Meijering, E. *et al.* (2006) Tracking in molecular bioimaging. *IEEE Signal Proc. Mag.*, **46**–53.
- Merryman, T.E. and Kovačević, J. (2005) An adaptive multirate algorithm for acquisition of fluorescence microscopy data sets. *IEEE Trans. Image Proc.*, **14**, 1246–1253.
- Miller, E. (2006) Data driven image models through continuous joint alignment. *IEEE Trans. Pattern Anal. Mach. Intell.*, **28**, 236–250.
- Moffat, J. *et al.* (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, **124**, 1283–1298.
- Murphy, D.B. (2001) *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley-Liss Inc.
- Murphy, R.F. *et al.* (2003) Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. *J. VLSI Sig. Proc.*, **35**, 311–321.

- Newberg,J. and Murphy,R.F. (2008) A framework for the automated analysis of subcellular patterns in human protein atlas images. *J. Proteome Res.*, **9**.
- Ng,L. *et al.* (2007) Neuroinformatics for genome-wide 3-d gene expression mapping in the mouse brain. *IEEE/ACM Trans. Comput. Biol.Bioinform.*, **4**, 382–393.
- Pan,J.Y. *et al.* (2006) Automatic mining of fruit fly embryo images. In *Proceedings of the 12th ACM SIGKDD 2006*.
- Parvin,B. *et al.* (2002) BioSig: an imaging bioinformatic system for studying phenomics. *Computer*, **35**, 65–71.
- Parvin,B. *et al.* (2007) Iterative voting for inference of structural saliency and localization of subcellular structures. *IEEE Trans. on Image Process.*, **16**, 615–623.
- Pawley,J.B. (ed.) (2006) *Handbook of Biological Confocal Microscopy*. 3rd edn. Springer, Berlin.
- Peleg,S. *et al.* (1989) A unified approach to the change of resolution: space and gray-level. *IEEE Trans. Pattern Anal. Mach. Intell.*, **11**, 739–742.
- Peng,H. and Myers,E.W. (2004) Comparing in situ mRNA expression patterns of *Drosophila* embryos. In *Proceedings of the RECOMB 2004*. San Diego, USA, pp. 157–166.
- Peng,H. *et al.* (2005a) Reconstructing a developmental time series of 3D gene expression patterns in *Drosophila* embryos. 2005 *Drosophila* Meeting, San Diego, CA. Available at [http://research.janelia.org/peng/papersall/docpdf/2005\\_FlyMeeting\\_poster.pdf](http://research.janelia.org/peng/papersall/docpdf/2005_FlyMeeting_poster.pdf).
- Peng,H. *et al.* (2005b) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**, 1226–1238.
- Peng,H. *et al.* (2006) Clustering gene expression patterns of fly embryos. In *Proceedings of the IEEE ISBI 2006*. pp. 1144–1147.
- Peng,H. *et al.* (2007) Automatic image analysis for gene expression patterns of fly embryos. *BMC Cell Biol.*, **8** (Supp. 1), S7.
- Peng,H. *et al.* (2008a) Straightening *Caenorhabditis elegans* images. *Bioinformatics*, **24**, 234–242.
- Peng,H. *et al.* (2008b) WANO: a 3D bioimage annotation system. *HHMI JFRC Technical Report*.
- Robinson,J.M. *et al.* (2001) Correlative fluorescence and electron microscopy on ultrathin cryosections: bridging the resolution gap. *J. Histochem. Cytochem.*, **49**, 803–808.
- Rohr,K. *et al.* (2003) Spline-based elastic image registration, integration of landmark errors and orientation attributes. *Comput. Vis. Image Underst.*, **90**, 153–168.
- Roussel,N. *et al.* (2007) A computational model for *C. elegans* locomotory behavior: application to multi-worm tracking. *IEEE Trans. Biomed. Eng.*, **54**, 1786–1797.
- Roysam,B. *et al.* (2008) The FARSIGHT project: associative multi-dimensional image analysis methods for optical microscopy. In Rittscher,J. *et al.* (eds) *Microscopic Image Analysis for Life Science Applications*. Artech Publishing House.
- Rust,M. *et al.* (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods*, **3**, 793–796.
- Sepp,K. *et al.* (2008) From flies to mice: identification of neural outgrowth genes using genome-wide RNAi. *PLoS Genet.*
- Shen,D. and Davatzikos,C. (2002) HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging*, **21**, 1421–1439.
- Shi,J. and Malik,J. (2000) Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, **22**, 888–905.
- Shimomura,O. *et al.* (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, *Aequorea*. *J. Cell Comp. Physiol.*, **59**, 223–239.
- Smal,I. *et al.* (2008) Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: application to microtubule growth analysis. *IEEE Trans. Med. Imaging*.
- Swedlow,J.R. *et al.* (2003) Informatics and quantitative analysis in biological imaging. *Science*, **300**, 100–102.
- Swidan,F. *et al.* (2007) MAD: minimum shared decomposition of DAGs for multitarget tracking. *HHMI JFRC Technical Report*.
- Szeliski,R. (2006) Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.*, **2**, 1–104.
- Toga,A.W. and Thompson,P.M. (2001) The role of image registration in brain mapping. *Image Vis. Comput.*, **19**, 3–24.
- Tomancak,P. *et al.* (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.*, **3**, research0088.1-0088.14.
- Tsechpenakis,T. *et al.* (2007) Tracking *C. elegans* populations in fluid environments for the study of different locomotory behaviors. In *Proceedings of the MIAAB 2007*.
- Tsien,R.Y. (2003) Imaging imaging's future. *Nat. Rev. Mol. Cell Biol.*, **4**, SS16–SS21.
- Unser,M. (2008) Advanced image processing and analysis using ImageJ. In Davos,G.R. (ed.) *8th European Light Microscopy Initiative Meeting*, May 27–30, 2008, Switzerland.
- Viola,P. and Wells,W.M. (1997) Alignment by maximization of mutual information. *Int. J. Comput. Vis.*, **24**, 137–154.
- Vu,N. and Manjunath,B.S. (2008) Graph cut segmentation of neuronal structures from transmission electron micrographs. In *Proceedings of the IEEE ICIP 2008*, October 2008. San Diego, CA.
- West,R.G. (2007) ATLAS in silico. In *ACM SIGGRAPH 2007 Art Gallery*, August 05–09, 2007. San Diego, California, 225 pp.
- Yang,Q. and Parvin,B. (2003) Harmonic cuts and regularized centroid transform for localization of subcellular structures. *IEEE Trans. Bioeng.*, **50**, 469–476.
- Yoo,T.S. *et al.* (2002) Engineering and algorithm design for an image processing API: a technical report on ITK – the insight toolkit. In Westwood,J. (ed.) *Proceedings of the Medicine Meets Virtual Reality*. IOS Press, Amsterdam, pp. 586–592.
- Young,N. *et al.* (2004) GelScape: a web-based server for interactively annotating, manipulating, comparing and archiving 1D and 2D gel images. *Bioinformatics*, **20**, 976–978.
- Zhao,T. and Murphy,R.F. (2007) Automated learning of generative models for subcellular location: building blocks for systems biology. *Cytometry*, **71A**, 978–990.
- Zhou,J. and Peng,H. (2007) Automatic recognition and annotation of gene expression patterns of fly embryos. *Bioinformatics*, **23**, 589–596.
- Zollei,L. *et al.* (2005) Efficient population registration of 3D data. In *ICCV Workshop on Computer Vision for Biomedical Image Applications: Current Techniques and Future Trends*.