Databases and ontologies

Bisque: a platform for bioimage analysis and management

Kristian Kvilekval[1,*], Dmitry Fedorov[1,*], Boguslaw Obara[1,2], Ambuj Singh[1,3] and B. S. Manjunath[1]

[1]Center for Bio-Image Informatics, Electrical and Computer Engineering Department, University of California Santa Barbara, CA, USA, [2]Oxford e-Research Centre and Oxford Centre for Integrative Systems Biology, University of Oxford, Oxford, UK and [3]Department of Computer Science, University of California Santa Barbara, CA, USA

ABSTRACT

Motivation: Advances in the field of microscopy have brought about the need for better image management and analysis solutions. Novel imaging techniques have created vast stores of images and metadata that are difficult to organize, search, process and analyze. These tasks are further complicated by conflicting and proprietary image and metadata formats, that impede analyzing and sharing of images and any associated data. These obstacles have resulted in research resources being locked away in digital media and file cabinets. Current image management systems do not address the pressing needs of researchers who must quantify image data on a regular basis.

Results: We present Bisque, a web-based platform specifically designed to provide researchers with organizational and quantitative analysis tools for 5D image data. Users can extend Bisque with both data model and analysis extensions in order to adapt the system to local needs. Bisque's extensibility stems from two core concepts: flexible metadata facility and an open web-based architecture. Together these empower researchers to create, develop and *share* novel bioimage analyses. Several case studies using Bisque with specific applications are presented as an indication of how users can expect to extend Bisque for their own purposes.

Availability: Bisque is web based, cross-platform and open source. The system is also available as software-as-a-service through the Center of Bioimage Informatics at UCSB.

Contact: kris@cs.ucsb.edu; fedorov@ece.ucsb.edu

Supplementary information: The supplementary material is available at *Bioinformatics* online, including screen shots, metadata XML descriptions and implementation details.

1 INTRODUCTION

Current research in biology is increasingly dependent on conceptual and quantitative approaches from information sciences, ranging from theory through models to computational tools (Brent, 2004). Ready availability of new microscopes and imaging techniques has produced vast amounts of multidimensional images and metadata. The introduction of new models, measurements and methods has produced a wealth of data using image-based evidence (Peng, 2008).

Researchers dependent on images as experimental evidence face the daunting task of managing, analyzing and sharing images in addition to gaining and providing access to analysis methods and results (Andrews *et al.*, 2002). We describe a new platform (Bisque :Bio-Image Semantic Query User Environment) to assist researchers to effectively deal with image-based data-driven science. Bisque provides facilities for management, organization and custom analysis of images and metadata, as well as providing a unique web-based platform for sharing of data and results. Below we review several issues that motivated the development of Bisque.

The key to properly interpreting biological images is the experimental and image-related metadata captured during experimental preparation and imaging. Biology labs employ diverse experimental procedures (Tuan and Lo, 2000) and continually invent new procedures and preparations resulting in unique local workflows (Schilling *et al.*, 2008). While several laboratory image database systems have been developed, database schema rigidity has often become problematic as requirements evolve due to changes in experimental protocols and required analyses.

In addition to the data management problem, researchers are increasingly dependent on automated or semiautomated image analysis (Swedlow *et al.*, 2003) due to large amounts of images involved in modern biological analysis. New measurements, analysis and statistics have become increasingly complex and challenging (Carpenter, 2007). Novel analysis techniques and results may require changes to the underlying data model (Swedlow *et al.*, 2006).

As computational image analysis is further integrated into the scientific process, accurate tracking of experimental results becomes a primary concern (NIH, 2009). Maintenance of original data, while ensuring accurate tracking of result data is fast becoming a requirement. In order to ensure accurate provenance, result data need to be reliably marked by a tamper-resistant system in which the analysis and the resultant data become fixed once added to the system.

While funding agencies require access to published data for open science requirements (NIH, 2009), many researchers also realize the importance of data sharing and collaboration which can lead to better data validation, increased knowledge discovery and cross-domain collaborations (Vannier and Summers, 2003). Even though researchers are usually willing to share data once published, strict security must be in place for works in progress. Achieving the goals

---

*To whom correspondence should be addressed.

of strict security and ease of sharing has proved challenging and in many cases sharing of data has suffered.

The Bisque environment aims at solving the aforementioned challenges. It was designed to be flexible enough to accurately model the broad range of acquisition and analysis workflows as well as experimental procedures that exist across different research labs. Image analysis is at the core of Bisque, and the system provides a smooth integration of internal and external tools. Bisque provides support for secure image storage, analysis and provenance management while giving users control over sharing of images and analysis results. The system architecture is scalable, reconfigurable and extensible with core functions implemented as web services.

The Bisque system is unique in its embrace of web-based technologies to drive open access at several levels. All managed items, such as images, datasets, metadata, analysis and services are given unique URLs allowing access by a variety of tools and platforms. All system services and user-extended analyses are available from any webbrowser, permitting novel uses including mash-ups (Belleau *et al.*, 2008) and multilanguage scripting. The Bisque system is thus uniquely placed to take advantage of the emerging computational resources such as cluster and cloud computing.

The Bisque system may be used as a software-as-a-service giving biologists easy access to their data from any location. It allows textual and graphical annotation creation and 5D image and annotation visualization. The system enables sharing (with format conversion tools) and collaboration. Several generic image analysis tools are also available and new modules added to the system become available to users instantaneously (no installation cost).

Larger laboratories with specific analysis requirements, using Bisque, gain many key elements that enable constructive collaboration between biologists and computer scientists, where the data need to be captured, stored, analyzed, interpreted and easily accessed as part of the collaborative research workflow. Web-enabled custom image analysis algorithms can be rapidly implemented utilizing existing image services, such as format conversions, metadata retrieval and storage and preexisting analysis modules.

In summary, Bisque provides a platform for *management*, *sharing* and *deployment* of images and custom analyses. The article's primary contribution is in providing a comprehensive introduction of Bisque for the bioinformatics community, its core functionalities, associated challenges and solutions in building such a system.

2 RELATED WORK

Many image management tools have been built for specialized domains, most notably the medical domain. One of the first projects to investigate high-throughput microscopy image analysis was the Open Microscopy Environment (OME) (Goldberg *et al.*, 2005), which was developed at Massachusetts Institute of Technology (MIT) and National Institutes of Health (NIH). OME is an open source project designed for storage of large amounts of high throughput imaging and respective microscope acquisition data. The OMERO project, a Java-based derivative of OME, has improved the performance of the original OME server and provides a rich client application.

Another early system, PSLID (Murphy, 2005), provides both management of datasets and analysis. The focus of the PSLID effort

has been the creation of biologically specific image features for the purpose of classification, search and comparison.

The eXtensible Neuroimaging Archive toolkit (XNAT) (Marcus *et al.*, 2005) is an extensible platform for secure management and exploration of neuroimaging data. XNAT supports workflows including metadata validation through an online approval process.

While the above efforts do allow extensions for new experimental protocols and analyses, these extensions require database and data modeling expertise. This limits the adoption of image databases with extensive contextual metadata that is needed for reinterpretation and analysis of previous studies.

Bisque extends the current state-of-the-art image databases with a flexible user-defined data model that allows experimental and image metadata to change rapidly. This permits easier analysis integration as each analysis may also need to extend the data-model. Recent research (Damme *et al.*, 2007; Hunter *et al.*, 2008; Reid and Edwards, 2009) has shown that user-created data models or 'Folksonomies' can be a powerful tool in creating communities and consensus which can be eventually solidified into ontologies.

Bisque also provides users with trusted provenance of analysis results by controlling access unlike standalone image analysis applications. Bisque's open system design allows elements to be reused by other websystems and tools through common web protocols. Finally, the low barrier to entry, a web browser, allows users to access most of the Bisque functionality without installation issues.

Bisque has demonstrated compatibility with existing image analysis tools. ImageJ (ImageJ, 2009) is a popular tool for image analysis. Many plugins have been successfully used to extend its functionality. Bisque provides plugins to interact with ImageJ allowing images to be imported and exported from/to Bisque servers. The Insight Toolkit (ITK) has been used extensively in the medical image analysis domain. Bisque integrates with ITK in order to build special purpose modules and includes an example ITK pipeline. CellProfiler (CellProfiler, 2009) is a specialized Matlab package for quantifying changes in biological images. It allows a user to create processing pipelines and develop custom modules.

Bisque is not a general purpose image-processing environment (Matlab, IDL, OpenCV), although it provides many services (reading, converting, resizing) needed while developing biologically related image-processing routines. We propose and demonstrate that design choices including universal accessibility, metadata flexibility and system extensibility are unique elements that enable better development and dissemination of research data and analysis results.

3 INTERFACE OVERVIEW

Bisque provides an online resource for management and analysis of 5D biological images. In addition to image collection management, the system facilitates common biological workflows typical of biological images: imaging, experimental annotation, repeated analysis and presentation of images and results (Supplementary Figure 1).

3.1 Ingestion of images and metadata

Image and metadata ingestion is the first step in using our database and analysis system. Bisque offers a wide array of ingestion techniques: web browser based, tool based and script based.

The *Digital Notebook* application (Section 5.4) allows users to rapidly annotate and upload groups of images with textual and graphical metadata. For complete control, scripts may be written in various programming languages allowing complete customization of image and metadata upload process (Section 3.7 in Supplementary Material).

3.2 Annotation with textual and graphical metadata

Metadata plays a vital role in interpretation, querying and analysis of images. For example, an experimental condition may highlight differences between images, image resolution may be needed to detect abnormally sized cells and experimenter comments may be handy in finding images of interest. There are many different types of metadata that may be associated with an image such as textual annotations, image-based object outlines and statistical distributions. In order to store the many types of annotations, we use 'tagging' or hierarchical name–value pairs. Flexible and hierarchical tagging is a key feature of Bisque that enables the user to rapidly model diverse experimental information (Supplementary Figure 2).

Graphical annotations (see Section 4.1) allow for complex and hierarchical 5D visual outlines with custom and biologically meaningful types; additional information is attached via textual tags.

3.3 Organization and search

Images and metadata are organized with tags (name–value pairs) associated with an image. Typically, users locate images of interest by browsing through collections or by searching with specific queries. The system has an integrated web image browser (Supplementary Figure 4), that allows browsing, filtering, sorting and ordering of images. It also allows searching based on tag–value pairs with support for search expressions. The image organizer (Supplementary Figure 5) allows for advanced sorting by way of hierarchical tag ordering. For example, it is possible to choose *image date* as the primary sorting tag and then *species* and *antibodies*. The number of levels is unlimited and available tags are dynamically loaded into the image organizer.

3.4 Analysis

Bisque provides an extensible analysis system that allows both internal and external analysis and visualization tools. Internal analysis methods are executable through the web, providing ease of use while transferring the computation load to the Bisque servers. Furthermore, Bisque facilitates 'web applications', custom user interfaces built around an analysis problem and its results encapsulating a predefined workflow. An external analysis refers to a tool remotely accessing data within the Bisque system that may utilize special hardware or other local features not available through the web. External tools are also useful for rapid prototyping of analysis before creating an internal version. Results produced by either internal or external tools maybe stored back into the system in the form of tags, graphical objects and/or images.

3.5 Visualization and sharing

As mentioned earlier, metadata in Bisque can take many forms: text, objects of interest, user annotations or another web-based file (e.g. associated publication in PDF). Textual and graphical markup viewing and editing is available in the web 5D image viewer (Supplementary Figure 3). The viewer is used for image and object browsing, ground-truth acquisition and statistical summaries of biological objects. Additionally, it allows for various visualization options such as channel mapping, image enhancement, projections and rotations. A tag type may be interpreted by a specific visualization widget. For example, a vector of numbers of type *histogram* may be plotted using the histogram–widget instead of the default tag editor.

The system also supports external visualization tools that can make use of unique hardware that is available only on the client such as high-speed graphics. The BioView3D (see Sec. 5.4) provides rapid visualization of 3D images along with graphical annotations.

Biological image sharing has often been difficult (Andrews *et al.*, 2002) due to proprietary formats. In Bisque, sharing images, metadata and analysis results can be performed through the web. The system contains an export facility that allows conversions of image formats, application of a variety of image-processing operations and export of textual or graphical annotations as XML, CSV or to Google Docs.

4 ARCHITECTURE

The global architecture (Fig. 1) of the system reflects the primary concerns of accessibility, extensibility, reconfigurability and scalability. The functionalities of the system are fundamentally separated into services. The core services include image storage and management, metadata management and query, analysis execution and client presentation. The *Image Service* (IS) provides simple image processing functions such as formatting, slicing and resizing. The IS is built on top of the *Blob Service*, which provides secure storage and access to files. The *Data Service* (DS) is responsible for image metadata storage and querying. This service provides searchable annotations for images as well as analysis results. The *Analysis Service* (AS) manages the integration and execution of integrated analysis. The *Client Service* handles communication with the user and is responsible for providing a unified view of the other core services.

Each Bisque service (Blob, Image, Metadata, Analysis) is constructed of one or more *servers*. Each group of servers can be
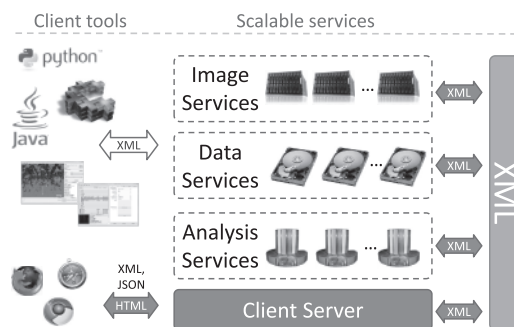

Fig. 1. Bisque is implemented as a scalable and modular web service architecture. *Image Servers* store and manipulate images. Data servers provide flexible metadata storage. Extension servers house executable modules. The client server seamlessly integrates services across storage and processing hardware. Communication between various components is performed in a RESTful manner through HTTP requests carrying XML and JSON. Web clients use dynamic web components built with AJAX.

deployed on a single machine or distributed across a cluster in order to scale to expected load. For example, multiple *Blob Servers* can be configured for increased image storage. Services communicate using standard web protocols allowing servers to leverage preexisting utilities for increased performance.

Thus, we view the Bisque system as a group of web services, each providing a simple service: client presentation, data storage and query, simple image processing and analysis. This system organization is referred to as Service-Oriented-Architecture or 'Micro-Apps' (MicroApps, 2009) depending on the context.

4.1 Common architectural themes

While Bisque has been constructed as a set of independent services, there is a need to coordinate in several areas of cross-cutting concerns. These areas include heterogeneous metadata, security, access methods and management of provenance.

4.1.1 Access methods All services and modules are accessible via standard web access methods (HTTP/S). This permits a wide variety of tools, from web browsers to custom analysis applications, to interact with our system. Most Bisque services are implemented using the RESTful design pattern architecture (Pautasso *et al.*, 2008) that exposes resources through URIs. Resources are manipulated by using the common HTTP methods. Among many benefits attributed to RESTful pattern (Fielding and Taylor, 2002) are scalability through web caches and the use of client side state and processing resources. Bisque services exchange metadata tag documents. Developers extending or customizing the system need an understanding of web protocols for data transfers and tag documents for service manipulations.

4.1.2 Flexible metadata Metadata for samples, experimental methods, imaging techniques and images is highly varied and may change rapidly as new hypotheses, experiments and methods are introduced. Experimental models need to be continually refined and adjusted. Bisque uses a flexible metadata model based on tag documents. A tag is a named field with an associated value. The tags themselves may be nested, and include values that are strings, numbers, other documents or lists. Tags may also contain bookkeeping attributes such as type, owner, timestamp and provenance. Very few restrictions are placed on the structure of the tag documents: tag names may be repeated, tag types are free form but may be interpreted by services.

Tag documents allow many items associated with the scientific workflow to be modeled trivially including sample preparation, experimental conditions, imaging parameters and analysis results. In a general system, it is impossible to know the particulars of any specific workflow. However, the metadata system permits users to define the particulars of their specific experimental workflows and metadata needs. In fact, the use of tag documents allows experimenters to modify their metadata as experiments progress without necessitating changes to the system's internals.

Bisque supports an extensible metadata format for graphical annotations that has a number of graphical primitives and can be extended by new object types and object properties in a simple and reliable manner. Predefined graphical primitives include point, polyline, polygon, rectangle, square, circle, ellipse and label. Graphical properties of these primitives are defined via vertices. Each vertex contains coordinates in a 5D image space and all coordinates are image centric. Any graphical type that has not been explicitly predefined (i.e. is not a graphical primitive) is treated as a complex type. Using this object-based approach, users can construct biologically meaningful types. A detailed description along with several practical examples of the flexible metadata documents can be found in the Section 2 of the Supplementary Material.

4.1.3 Security management Any system that contains sensitive material must ensure that access is both controlled and secure. Scientific images and results, while eventually shared, are often highly sensitive and therefore must be protected from unauthorized access. Bisque ensures secure access for images, metadata and results through the use of secure web protocols and server-based authentication. Each service request is authenticated by the authentication service. Each image and tag document is marked with an owner (creator) and a set of simple permissions. Read and write accesses are always granted to the owner, while for others, objects are marked either public or private. Private items can never be accessed by anyone except the owner. Analysis results are always under the permission of the user who ran the analysis and which has complete control over dissemination.

4.1.4 Provenance management There are fundamentally two types of provenance, best-effort and trusted. Best-effort leaves the control over data provenance for source data, methods and results to the scientists and the tools used. For example, the widely used HDF format allows provenance metadata to be embedded. However, to be accurate and trustworthy, provenance data must be automatic and outside the direct control of the experimenter. Bisque maintains two facilities for trusted provenance. First, its blob server is a write-once, read-only server ensuring data stored there is rarely deleted and user tamperproof. Secondly, the metadata server marks all data with the user operation or activity that produced it.

For users of the system, provenance tracking is automatic and outside of their direct control. Users can track the provenance of their resultant data back to the original image and metadata. For example, the chain of results from summary statistics to original image collection can be found. Provenance can also be used to reapply updated analysis methods (especially when software errors are found). Users can also use provenance information to purge the system of bad results caused by discovered errors. For example, if a sample or an image is discovered to be tampered with or corrupted, the image itself and all results based on the image may be marked for removal.

5 CORE SERVICES AND TOOLS

In this section, we will describe only the most essential services and tools, leaving the screen shots, listing of additional services and other details for the Sections 1 and 3.3 in Supplementary Material.

5.1 Blob and IS

Bisque contains an image server designed to work with different image types to perform many common image operations. The *Image Server* (IS) (shown in the service layer of Fig. 1) is built upon a *Blob Server* which is responsible for storing and retrieving binary objects or files. It provides a *write-once/read-only* service. This measure was taken in order to ensure provenance for the complete database.

The IS operates on 5D images that potentially may have X, Y, Z, time (T) and channel (C) coordinates.

The core functionalities of the IS can be categorized into three groups: (i) Decoding and encoding (including many proprietary microscope formats), (ii) Extracting embedded metadata and (iii) Applying image-processing operations (such as resize, crop and remap). This server has a pluggable encoder/decoder interface and supports a broad range of image formats (including proprietary).

A single IS request may contain multiple subrequests to be executed. The subrequests are pipelined to create a chain of suboperations and each suboperation is cached to avoid excess network transport where possible, and to avoid recomputation of costly operations such as resizing of a large image stack. A detailed description of formats, operations and requests can be found in Section 3.2 of the Supplementary Material.

5.2 Flexible metadata service

The DS provides storage of textual and graphical annotations based on the flexible data model, described in Section 4.1.2. Data servers typically handle storage and query requests for other Bisque clients, e.g. a web browser, a stand-alone application or another Bisque service.

Typical searches include finding an object tagged with a certain tag name, value or field combination. Other queries include finding all values of a certain tag or all tags with a certain name. Queries return tag documents containing references to objects or full objects depending on the view parameters requested. These documents can potentially be very large and DS may decide to stop returning requested values providing a link for a continued request. Detailed requests and example responses are available in Section 3.1 of the Supplementary Material.

5.3 AS and module extensions

A core facility of Bisque is the ability to integrate analysis into the system. While offline or disconnected analysis of downloaded data is possible, there are many benefits to providing a mechanism for integrated analysis including analysis directories, common infrastructure utilization, possibility of data reuse and result provenance tracking. Bisque supports several types of integrated modules, ranging from external tools to fully integrated web-based applications. Integration effort can range broadly depending on the sort of interaction required.

The module service provides access to analysis modules and offers registration and execution services. As with other services, ASs can be parallelized using multiple *Module Servers* (MS). Requests are sent by the client in form of tag documents to the MS for execution. The MS interacts with the other servers (image, data) and uses module engines to actually execute analysis modules. An engine is a wrapper that creates web accessible entry points for a specific programming language. Engines have been implemented to support modules developed in Matlab, Python, Java and C++.

External modules are applications interacting with Bisque resources by means of HTTP requests to read/write images and metadata. Any environment that can fetch internet resources and parse XML can be used to create a module. Module integration can be further simplified by using the BisqueAPI developed for Matlab, Python, Java and C++. The API hides most of the communications and parsing details. It should be noted that external modules, by

design, leave all the user interface construction to be done on the client side. *Digital Notebook* and *BioView3D* (Section 5.4) are such examples.

Internally integrated modules are constructed in a similar fashion, with an addition of a module description tag document (see Supplementary Material, Section 7.2). This document allows MS to instantiate a correct engine and provide required parameters. The formal input and output parameters are passed to the running module, e.g. as function arguments in Matlab or as command line arguments to an executable. This means that a module developer does not necessarily have to interact with the Bisque system directly, unless some advanced access is necessary.

Since all internal modules are web instances, they need to generate a web page at some point. The Bisque MS is able to generate a basic HTML interface based solely on the module description. Advanced modules may provide their own HTML interface, simplified by calls to standard Bisque JavaScript widgets allowing selection of an image, export and visualization of resultant data. *3D nuclei detector* (Section 6.1) is an example of such a web-application. The advantage of creating internal modules is in the server side execution that can take advantage of powerful resources such as cluster or cloud computing infrastructures.

Prior to module execution, MS creates a module execution record (MEX), a tag document that contains values of all formal inputs and will contain formal outputs upon module termination. Module receives a URL for its MEX and can use it for active communication, e.g. progress output or interactive input.

Formal inputs are passed directly into the module, although images are passed as URLs to their image objects. The module is free to make HTTP requests to fetch the image directly from the IS or to use BisqueAPI to receive pixels in a native format, e.g. BisqueAPI for Matlab will return a native matrix for a plane of an image. Any 5D image can be fetched in parts (planes, channels, etc.) and in desired formats (pixel bit depth, compression, etc.). Details of module development can be found in Section 7 and the listing of currently available modules in Sections 3.4, 3.5 and 3.6 of the Supplementary Material.

5.4 External tools: annotation and visualization

The Microscopist's *Digital Notebook* is a cross-platform image annotation and Bisque database ingestion tool. It was originally created for offline rapid annotation of multiple images directly on the microscope controlling computer. Several refinement iterations has taken place based on direct feedback from the biologists.

The built-in graphical annotation editor allows users to create graphical annotations by outlining objects of interest on 5D images. Graphical annotations are highly customizable and provide an extensive platform for ground truth acquisition. Users may create templates of complex graphical objects with assigned biological meaning and reuse them for annotations. Furthermore, simple statistics of lengths, areas, etc. of graphical annotations can be generated for selected images in both pixels and physical units (Sections 5.1 in Supplementary Material).

BioView and *BioView3D* are cross-platform applications for visualization of biological imagery and its metadata. They support many biological image formats and live channel mapping. *BioView* features a remote control interface for very large displays. *BioView3D* is a high speed and versatile 3D image stack

visualization system based on modern graphics hardware supporting OpenGL. Tool features include fly-path and live video export as well as visualization of Bisque's graphical annotations (see Sections 5.2 and 5.3 in Supplementary Material).

6 BISQUE ANALYSIS AND CASE STUDIES

The Bisque platform is used to create problem-solving environments for specific biological studies. These case studies are based on the basic Bisque functionality that provided facilities for microscope image decoding and access, metadata management, result visualization and statistics. Each study has, at its core, a growing number of integrated image analysis, pattern recognition and data mining modules. In many cases, a custom analysis will need to be specifically developed until better general methods are discovered.

The Bisque system includes a growing set of image analysis modules developed in an interdisciplinary research effort between biologists and computer scientists. Some modules will be introduced in the following case studies, whereas their in-depth description can be found in the referenced papers.

For each study, an analysis module was developed and used to identify objects or quantify changes in sequences of images. The results of the analysis are stored in the system and used by the standard visualization and statistics modules. Bisque provides many services to ease the construction of novel analyses. The IS allows manipulation of proprietary formatted microscope images with common tools. The metadata service provides uniform access to embedded metadata, while the module sevice allows access to previously written modules.

Novel or modified techniques being developed for quantitative image analysis require validation prior to operational use. Ground truth collection is an essential part of validation and its acquisition and ingestion into the system is aided by the Digital Notebook. Once in the system, manual annotations can be compared against automated techniques and Bisque contains many generic validation modules for various graphical objects. Additionally, Bisque facilitates the incorporation of new validation techniques using module interface.

Most of the images and their annotations, described in the following case studies, are publicly available for researchers developing novel analysis techniques and can be accessed through Bisque hosted at the Center for Bio-Image Informatics (CBI) at UCSB. In addition, the center is leading an effort to standardize validation techniques involved in quantitative biological image analysis by hosting a 'Bio-Segmentation Benchmark' which provides several public image datasets along with associated ground truth in a flexible tag document format (Sections 6.2 and 3.6 in Supplementary Material).

6.1 *Arabidopsis* studies at Meyerowitz lab, Caltech

Processes of development in the *Arabidopsis thaliana* and, specifically, the meristem are of significant interest to plant biologists (Gordon *et al.*, 2007). Recent research focuses on computational modeling of cells and their patterns in the developing meristem and simulation of developmental processes under different conditions (Heisler and Jönsson, 2007).

Meyerowitz's group is simulating the computational models of cell division and pattern formation on 3D cellular templates extracted from laser scanning confocal stacks. Models are initialized from centroids of cell nuclei present in the images. The evolution predicted by the model is also validated by the nuclei centroids detected in consecutive images.

The meristem CLSM imagery routinely contains 1500 nuclei in 3D and can take an expert several days to annotate all centroids in a single image. The same detection (Fig. 2) takes 5 min using an integrated *3D nuclei detector* module (Obara *et al.*, 2008). We have validated the 3D nuclei detection by visually inspecting results using *BioView3D* and by comparing automated results with several images manually annotated by multiple experts. Experts have used the *Digital Notebook* application and produced point centroids in 3D. Moreover, we have combined multiple ground-truth annotations in order to obtain 'gold-standard' ground-truth and verified that automated annotations are comparable to it in quality and offer precision and recall >85%. Additionally, the automated method offers reproducible results. The *3D nuclei detector* module (Sections 4.1 in Supplementary Material) was initially developed specifically for this study, but due to its generic nature and implementation it was successfully used in several other projects. These include the study of Alzheimer's influence to hippocampal region CA1 in mice and a study of the growth patterns in Ascidian worm embryos. This module is available as a web-based application providing a specially developed user interface combining documentation with workflow elements (see Supplementary Material for details).

Furthermore, Shapiro *et al.* under the aegis of Sigmoid project (Sigmoid, 2009) have integrated a Mathematica powered modeling software with the Bisque web-based system. The modeling is based on quantitative estimation of WUS (transcription factor WUSCHEL) expression from green fluorescent protein (GFP) intensities in segmented cells (Jönsson *et al.*, 2005). The simulation is implemented as an external service presenting a specialized user interface to access images and annotations stored in the Bisque system. In order to achieve this, Sigmoid researchers used the Bisque tag document format and achieved integration in <2 weeks, utilizing preexisting facilities for image access, nuclei detection and web-based visualization.

6.2 Microtubule studies at Feinstein and Wilson labs, UCSB

Microtubules carry out several essential functions within cells, playing an important role in both intracellular cargo transport (moving vesicles, granules, organelles and chromosomes via special attachment proteins) and chromosome segregation at mitosis. Structurally, they are linear polymers of tubulin with diameter of ~25 nm (Feinstein and Wilson, 2005). The neural microtubule-associated protein *tau* binds directly to microtubules and regulates their dynamic behavior (growing and shortening). In addition to being required for normal development and function of the nervous system, tau is associated with neurodegenerative diseases, including Alzheimer's disease and certain cancers (Jordan and Wilson, 2004).

To obtain a quantitative description of dynamic microtubule behavior, cells are injected with fluorophore-labeled tubulin (or alternatively engineered to express GFP-tubulin fusion protein) and visualized with the Time-Lapse Fluorescence microscopy (Fig. 3a).
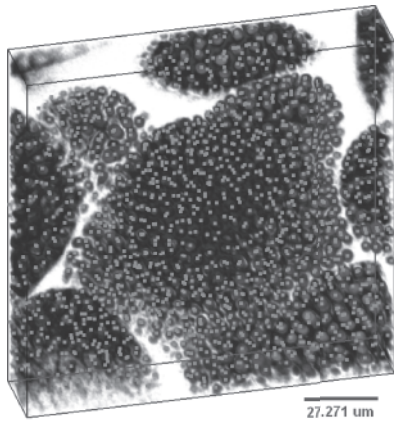
Fig. 2. A 3D CLSM image of the *Arabidopsis* shoot meristem with overlayed nuclei centroids automatically detected by *3D nuclei detector* and visualized by *BioView3D*.
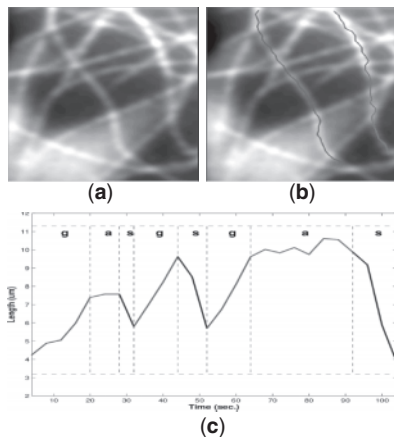

Fig. 3. Analysis workflow used to measure how tau molecules regulate microtubule dynamics: (**a**) original image, (**b**) exemplary traces, (**c**) microtubule life history plot computed for several time points: dynamic events are labeled as **g** for growth, **s** for shortening and **a** for attenuation.

These images are typically very noisy due to the small size of microtubules. Individual microtubules are then tracked or traced using manual or semiautomatic approaches (Sargin *et al.*, 2007). The tracking/tracing results are then analyzed to produce dynamicity statistics (Altinok *et al.*, 2007b).

The *Digital Notebook* application allows scientists to manually track the position of microtubule tips in time, outline their bodies and eventually upload all annotations for analysis. To achieve this, we have created templates that describe these complex graphical objects and allow their creation within *Digital Notebook*.

The manual tracking is tedious and requires several hours per time lapse. Thus, annotating dozens of images for a typical experiment may take weeks. Additionally, manual tip positions are non-reproducible and have a significant degree of uncertainty afflicting intra-investigator dynamicity statistics (Altinok *et al.*, 2007a).

When the image quality permits, the semiautomatic tracing is available via an integrated and specially developed *microtubule tracking* module (Fig. 3b). This Matlab-based module is accessible

as a web-based application and provides a specifically developed user interface that integrates documentation and the workflow (Sections 4.2 in Supplementary Material). The tracing is initialized using the Bisque image viewer (browser based) by drawing a line over a segment of a microtubule indicating its position and the direction of the tip. This graphical annotation is requested by the tracking module as one of the input parameters in the module definition tag document. An example annotation of tip tracking and body tracing can be found in the Supplementary Material (Section 2.1.1 The tracking/tracing results in the form of polylines in time are used in calculating the dynamicity statistics using an integrated and also specially developed *microtubule dynamics analysis* module (Fig. 3c).

In conclusion, previous studies of microtubule dynamics have been severely constrained by data acquisition and analysis procedures that were tedious, labor intensive and had a significant degree of uncertainty. Our newly developed analytical tools enable investigators to acquire more and better quality data in a semiautomated manner. Currently, we host >2500 annotated *in vivo* and *in vitro* images in different conditions. These capabilities markedly improve data quality and reliability, thereby significantly promoting progress on biological studies of microtubule dynamics and their regulation. Further, given more abundant and higher quality data, our procedures will allow the use of statistics strategies to address novel questions regarding the behavior of populations of microtubules that have not been approachable previously.

6.3 Retinal studies at Fisher lab, UCSB

The retina is composed of several neuronal layers, where each layer has a different structure. Changes in the integrity of the layers, such as deformations of the layer boundaries and cell densities, serve as an index of retinal function after injury. Hence, measures such as the cell count, the layer thickness and changes in protein distribution in each layer are used in quantifying the effects of retinal detachment or injury and thus are of interest to retina biologists.

CBI currently hosts a collection of >6000 retinal images acquired in >15 different conditions from different species including human, mouse, cat, etc. These images are used in several experiments conducted in collaboration between image-processing scientists and retinal biologists. As a result of this collaboration, Bisque offers a number of specific image-processing modules available for retinal biologists (Fisher *et al.*, 2006).

In particular, the Bisque system is used to verify that the number of photoreceptors decrease in response to a retinal detachment. Images of interest for each experiment are selected using query system based on textual annotations. The inner (INL) and outer (ONL) nuclear layers within images of interest are automatically segmented using a specifically developed *retinal area layers segmentation* (Bertelli *et al.*, 2007; Vu *et al.*, 2007) module, producing polygons marked as ONL and INL by tagging graphical objects. Then, the centroids of the photoreceptor nuclei in these layers are detected using the generic *2D nuclei detector* module (Byun *et al.*, 2006a), producing graphical centroid locations. Additional details regarding 2D/3D nuclei detector module can be found in the Supplementary Material. Finally, the average number of photoreceptor nuclei per square milli meter of the INL and ONL is calculated from these annotations using a simple statistics module.

Ultimately, the researchers have been able to show that after 3 days of detachment, the number of photoreceptors is reduced by 8% of control, whereas the average number of nuclei within the INL reduced by only 1% (Byun *et al.*, 2006a). This was the first time that quantitative analysis verified an earlier qualitative prediction of cell survival in the inner retina in response to detachment.

Similarly, detected nuclei centroids are used to quantify layer distortions providing comparisons of thickness and local density between these two conditions (Byun *et al.*, 2006b). This study also required the development of a new Matlab-based module.

7   DISCUSSION AND CONCLUSION

Quantitative measurement of images is fundamental for discovery of biological processes. Research using new microscopes, imaging methods and analysis have been challenged by the increasing amounts of data to be stored, organized and analyzed. Bisque provides secure, extensible and web-based platform for 5D image analysis and management for growing image collections ranging from hundreds to hundreds of millions of images. Key features for practitioners are a flexible user-defined data model and facilities to create and integrate custom analysis modules.

Biological images cannot be properly interpreted without context metadata. Bisque supports a flexible data model allowing researchers and laboratories flexibility in annotating, linking, searching and measuring images. Annotations can be created in many different forms such as textual, numeric, graphical and external files.

The platform combines image analysis and secure provenance for tracking results. Bisque supports custom image and metadata analysis integration and provides supports for both external tool access and internal web-based applications.

Bisque represents a dynamic, accessible and extensible framework for bioimage analysis. It can access and be combined with other systems due to its web-based architecture. The use of web-based technologies allow unprecedented opportunities of simple data and analysis sharing for bioimage practitioners. The system is available as software-as-a-service through the Center of Bioimage Informatics at UCSB (Bisque, 2009) or through source code for local deployment. Details can be found in the Supplementary Material for those interested in using or installing a Bisque instance (Section 6). Performance evaluation is also available in section 8 in Supplementary Material.

Conflict of Interest: none declared.

REFERENCES

Altinok,A. *et al*. (2007a) Model based dynamics analysis in live cell microtubule images. *BMC Cell Biology*, **8**(Suppl. 1), S4.

Altinok,A. *et al*. (2007b) Segmenting filamentous biological structures by automated tracing. In *Society for Neuroscience Annual Meeting*, San Diego, CA, USA.

Andrews,P.D. *et al*. (2002) To 5D and beyond: quantitative fluorescence microscopy in the postgenomic era. *Traffic*, **3**, 29–36.

Belleau,F. *et al*. (2008) Bio2rdf: towards a mashup to build bioinformatics knowledge systems. *J. Biomed. Inform*. **41**, 706–716.

Bertelli,L. *et al*. (2007) A variational approach to exploit prior information in object-background segregation: application to retinal images. In *IEEE International Conference on Image Processing*, Vol. 6, pp. 61–64, San Antonio, TX, USA.

Bisque (2009) Bisque: Bio-Image Semantic Query User Environment. Center for Bio-Image Informatics, University of California, Santa Barbara, CA, USA. Available at http://bioimage.ucsb.edu/bisque/ (last accessed date January 11, 2010).

Brent,R. (2004) A partnership between biology and engineering. *Nat. Biotechnol*.. **22**, 1211–1214.

Byun,J. *et al*. (2006a) Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images. *Mol. Vision*, **12**, 949–960.

Byun,J. *et al*. (2006b) Quantifying structural distortions in retinal tissue before and after injury. In *Workshop on Multiscale Biological Imaging, Data Mining & Informatics*, Santa Barbara, CA, USA.

Carpenter,A.E. (2007) Software opens the door to quantitative imaging. *Nat. Methods*. **4**, 120–121.

CellProfiler (2009) Cell image analysis software. Broad Institute of MIT and Harvard, Cambridge, MA, USA. Available at http://www.cellprofiler.org/ (last accessed date January 11, 2010).

Damme,C.V. *et al*. (2007) Folksontology: an integrated approach for turning folksonomies into ontologies. In *Bridging the gap between semantic Web and Web 2.0 (ESWC 2007)*, pp. 57–70, Innsbruck, Austria.

Feinstein,S.C. and Wilson,L. (2005) Inability of tau to properly regulate neuronal microtubule dynamics: a loss-of-function mechanism by which tau might mediate neuronal cell death. *Biochim. Biophys. Acta*., **1739**, 268–279.

Fielding,R.T. and Taylor,R.N. (2002) Principled design of the modern web architecture. *ACM Trans. Internet Technol*.. **2**, 115–150.

Fisher,S.K. *et al*. (2006) Image informatics tools for the analysis of retinal images. In *Invest Ophthalmol Vision Science. Association for Research in Vision and Ophthalmology (ARVO 2006)*, pp. 4225, Fort Lauderdale, FL, USA.

Goldberg,I. *et al*. (2005) The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol*.. **6**, R47.

Gordon,S.P. *et al*. (2007) Pattern formation during de novo assembly of the Arabidopsis shoot meristem. *Development*. **134**, 3539–3548.

Heisler,M.G. and Jönsson,H. (2007) Modeling meristem development in plants. *Curr. Opin. Plant Biol*.. **10**, 92–97.

Hunter,J. *et al*. (2008) Harvana: harvesting community tags to enrich collection metadata. In *8th ACM/IEEE-CS joint conference on Digital libraries*, pp. 147–156, Pittsburgh, PA, USA.

Image,J. (2009) ImageJ. National Institute of Mental Health, Bethesda, MA, USA. Available at http://rsbweb.nih.gov/ij/ (last accessed date January 11, 2010).

Jönsson,H. *et al*. (2005) Modeling the organization of the WUSCHEL expression domain in the shoot apical meristem. *Bioinformatics*. **21**, 232–240.

Jordan,M.A. and Wilson,L. (2004) Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer*. **4**, 253–265.

Marcus,D.S. *et al*. (2005) The Extensible Neuroimaging Archive Toolkit (XNAT): an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*. **5**, 11–34.

MicroApps (2009) MicroApps. Available at http://microapps.org/ (last accessed date January 11, 2010).

Murphy,R.F. (2005) Location proteomics: a systems approach to subcellular location. *Biochemical Soc. Trans*.. **33**, 535–538.

NIH (2009) NIH data sharing policy. National Institute of Health. Available at http://grants1.nih.gov/grants/policy/data_sharing/ (last accessed date January 11, 2010).

Obara,B. *et al*. (2008) Automatic nuclei detection and dataflow in Bisquik system. In *Workshop on Bio-Image Informatics: Biological Imaging, Computer Vision and Data Mining*, Santa Barbara, CA, USA.

Pautasso,C. *et al*. (2008) Restful web services vs. "big" web services: making the right architectural decision. In *17th International World Wide Web Conference (WWW2008)*, pp. 805–814, Beijing, China.

Peng,H. (2008) Bioimage informatics: a new area of engineering biology. *Bioinformatics*. **24**, 1827–1836.

Reid,R. and Edwards,P. (2009) Ourspaces - a social semantic web environment for eScience. In *Spring Symposium on Social Semantic Web: Where Web 2.0 meets Web 3.0 (AAAI 2009)*, Stanford, CA, USA.

Sargin,M.E. *et al*. (2007) Tracing curvilinear structures in live cell images. In *IEEE International Conference on Image Processing*, Vol. 6, pp. 285–288, San Antonio, TX, USA.

Schilling,M. *et al*. (2008) Sharing images. *Standardizing Exp. Protocols*. **19**, 354–359.

Sigmoid (2009) Sigmoid: a database of cellular signaling pathways and their models. Bioinformatics and Genomics at the University of California Irvine, Beckman Institute of the California Institute of Technology, Bioengineering Department at Johns Hopkins University, USA. Available at http://www.iplantcollaborative.org/grand-challenges/about-grand-challenges/past-projects (last accessed date January 13, 2010).

Swedlow,J.R. *et al*. (2003) Informatics and quantitative analysis in biological imaging. *Science*. **300**, 100–102.

Swedlow,J.R. *et al*. (2006) Modelling data across labs, genomes, space and time. *Nat. Cell Biol.*. **8**, 1190–1194.

Tuan,R. and Lo,C. (eds) (2000) *Developmental Biology Protocols*, Volume 137 of *Methods in Molecular Biology*. Humana Press, Totowa, NJ, USA.

Vannier,M.W. and Summers,R.M. (2003) Sharing images. *Radiology*. **228**, 23–25.

Vu,N. *et al*. (2007) Retina layer segmentation and spatial alignment of antibody expression levels. In *IEEE International Conference on Image Processing*, Vol. 2, pp. 421–424, San Antonio, TX, USA.