THE CENTER FOR INTEGRATIVE BIOMEDICAL COMPUTING: ADVANCING BIOMEDICAL SCIENCE WITH OPEN SOURCE

Joshua Cates, David Weinstein and McKay Davis

Scientific Computing and Imaging Institute
University of Utah

ABSTRACT

The Scientific Computing and Imaging Institute at the University of Utah has created the new Center for Integrative Biomedical Computing (CIBC) whose mission is to produce high performance image analysis, simulation, and visualization software in support of biomedical research. Software at the CIBC is developed in close collaboration with scientists to satisfy real needs within the community. We use an agile software process that is fully open source and integrates the best of the open-source tools for biomedical image analysis with our own expertise in image processing, scientific visualization, and modeling. This paper describes some of the research collaborations driving our work, our software development process, and some initial results at the Center.

1. INTRODUCTION

The field of biological imaging is expanding quickly. This trend stems, in part, from recent developments in imaging, the advent of new medical imaging modalities and the increased performance of many existing ones. This expansion also results from new scientific and biological applications for such data, which are more concerned with mapping and statistical characterization rather than the conventional clinical paradigms of detection and diagnosis. The quantitative analysis of biological images is now often applied to populations of subjects in order to test scientific hypotheses. Despite the extensive use of computers in the biosciences, there is still a relatively low level of interaction between the research leaders in biomedicine and those in computational science. As a result, computer technology tends to lag behind the needs of the biosciences. At the same time, most bioscientists do not have the expertise required to create cutting edge computational software. In response to the need for new computational tools and for increased interaction between computational scientists and bioscientists, as well as other priorities identified by the National Institutes of Health (NIH) underscoring the essential and central role of computing in biomedical research [1], the Scientific Computing and Imaging Institute at the University of Utah has created the Center for Integrative Biomedical Computing (CIBC). Funded by the NIH National Center for Research Resources, our mission at the CIBC is to produce high performance open-source software for use in the biomedical research community, with a focus on image analysis, multi-scale tissue modeling and simulation, and scientific visualization.

At the CIBC we use a collaborator-driven approach to development: software is designed to address specific needs of practicing scientists who work closely with us throughout the development process. Though our software is conceived in the context of specific collaborative research projects, our larger goal is to disseminate our work to the biomedical community as a whole. We believe in an open-source model of development and are committed to producing software tools that are compatible with other biomedical software initiatives. To this end, we are implementing a rigorous software process and conform to open-source principles of development and licensing. We are also avid *consumers* of open-source, combining our own expertise with the best of the existing technologies for imaging and visualization.

This paper describes some of the work underway at the CIBC. The next section is a brief overview of some of our major scientific collaborations and the role that our Center plays in each. Our software process is outlined in Section 3, which covers issues of design, development, and dissemination. Finally, we present some initial results of our work: software that supports image segmentation and analysis, model building for simulation, and cutting-edge volume and tensor-field visualization tools.

2. BIOMEDICAL RESEARCH COLLABORATIONS

At the CIBC, collaborations with practicing biomedical scientists allow us to develop imaging, simulation and visualization software that has an immediate and measurable application to biomedicine. Our various collaborations are managed as projects with clearly defined goals, regular internal review, and regular communication. An important goal for some projects is to secure independent funding and free CIBC resources for new collaborations. The Center has many collaborative projects, each with different priorities and levels of resource commitment. This section describes several of our major collaborations their associated research goals.

The laboratory of Dr. Mario Cappechi at the University of Utah's Eccles Institute of Human Genetics is investigating the phenotypic expression of specific, induced genetic abnormalities in mice, a model that has been shown to provide insight into the ontogeny of congenital human disease. Conventional analysis of mouse skeletal structure requires sacrificing the research animal and a labor-intensive process of skeleton preparation and physical inspection under a dissecting microscope. Many tens or even hundreds of specimens are often required for a meaningful statistical analysis, which represents an enormous investment of time and money. The goal of the CIBC collaboration with the Cappechi lab is to develop a faster, non-invasive protocol for skeletal analysis that uses semi-automated image processing of three-dimensional micro-CT rather than hand measurements of prepared skeletal specimens. We are developing a set of image segmentation, measurement and visualization tools for quantitative morphometry that allow us to experiment with new metrics such as the analysis of bone *shape* that would not be possible with prepared skeletal specimens. Furthermore, we expect that our tools will allow for more precise and repeatable measurements for length, density and volume, and therefore give insight into genetic alter-

ations that have previously been described as pleiotropic (partially penetrant) or that have been written off as minor effects.

The National Center for Microscopy and Imaging Research (NCMIR), headed by Professor Mark Ellisman at the University of California San Diego, is an NCRR NIH Biomedical Technology Research Center established to develop advanced, computer-aided microscopy for acquisition of structural and functional data. The goal of the CIBC collaboration with NCMIR is to develop image analysis algorithms and software to help biomedical researchers understand structural and functional relationships within cells and tissues through a range of scales from macromolecular complexes to organelles and multi-component structures like synapses. Several of our other collaborators have identified the extraction and analysis of cell geometry from microscopy data as a significant hurdle in their research, so the tools we develop with NCMIR are expected to be immediately useful to other researchers. We feel that microscopy data has not received the same level of attention from the image processing community as many of the more clinically oriented imaging modalities and offers unique and important challenges. With the NCMIR group, for example, we are developing tools for quantitative and subjective analysis of the structural properties of dendritic spines, a challenging project that requires segmentation of reconstructed electron microscope tomography volumes and research into new volume visualization techniques.

Several of our principal collaborators are conducting research in the area of bioelectric field modeling, an area of research which is also the focus of several of the CIBC faculty. Together with Dr. Craig Henriquez from Duke University, the CIBC is exploring the feasibility of creating discrete bidomain models at a cellular level in order to study the effects of tissue structure on the propagation of action potentials in cardiac tissue. A similar modeling framework can also be applied to questions at the tissue scale such as the behavior of electroporation and field-mediated DNA transport, physiological phenomena that have potential clinical use for the delivery of gene therapies. The CIBC is partnering with Dr. Scott Makeig from the University of California San Diego and Dr. Greg Worrell from the Mayo Clinic to study localization of epileptic sources (electric current dipoles) within the human brain. In the diagnosis of epilepsy, electroencephalography (EEG) and magnetoencephalography (MEG) inverse methods are used to support and complement the diagnostic hypotheses achieved by magnetic resonance imaging (MRI) techniques. The overall goal of research by Dr. Elliot McVeigh at the NIH Heart Lung and Blood Institute is to develop a computational model to predict the susceptibility to fatal ventricular arrhythmia. Dr. McVeigh and his colleagues have developed methods to measure electrical, mechanical, and geometrical properties of normal and abnormal myocardial tissue with unprecedented resolution, both in patient and experimental studies, and is studying the interplay between these properties and their relative roles in generating the substrate for ventricular arrhythmia. The CIBC is partnering with Dr. McVeigh in his research to assist in the display and analysis of data, especially diffusion-weighted MRI and electrical signals, and in model construction and simulation.

3. SOFTWARE PROCESS

Engineering high quality software is a difficult job. For software to be most useful, it must meet basic requirements for stability and be well documented. Scientific computing adds additional requirements: Developers usually need special expertise in areas of mathematics, numerical methods, or computer graphics. The computational demands of research code may also be very high and require specialized hardware, including parallel machines. The scientific application domain is often unfamiliar to the computer scientist, so an increased level of interaction with users (research scientists) is necessary to properly understand the requirements of a project. Finally, research code adds the additional burden of validation, which means continuous regression testing for correctness and stability.

Another important aspect of scientific computing to consider is the ultimate measure of the success of a project. The mission of a research computing lab like the CIBC is the advancement of science. In the "business" of research computing, success is measured by the impact of our software on the scientific community. To that end, we seek to maximize the quality and usefulness of our products for our collaborators and then to disseminate them as freely and widely as possible. Previous experience at the SCI Institute suggests that if we can produce software that is sufficiently useful, documented, and reliable, then the scientific computing community will participate in our maintenance and development effort. Similarly to how a successful project with our collaborators may eventually achieve independent funding, a successful software product may eventually be supported by its community of users.

The open source model of software development is a natural fit for our research-oriented business model. At the CIBC, we adhere to the definition of "open source" promoted by the Open Source Initiative, which, unlike some other academic definitions of open source, allows for-profit use and therefore the involvement of industry participants. This philosophy is in line with the priorities outlined in the NIH Roadmap initiative which encourage research collaborations between academia and industry. In order to meet the engineering demands imposed by open source scientific computing, we are implementing an agile software engineering process similar to those that have proved successful for other biomedical research computing initiatives. Specifically, we are modeling our development process after one introduced by the National Library of Medicine's Insight Toolkit (ITK) [2], which is currently in use by several other high-profile research groups and open source scientific computing projects, most notably, the National Alliance for Medical Image Computing [3] and the Visualization Toolkit [4]. This software process uses light-weight, rapid prototyping and development cycles that are made sustainable by continuous regression testing and supported by open-source distributed build and testing tools. The motivation behind this strategy is to minimize development overhead and maximize the quality and stability of our software.

Requirements gathering is an essential part of any software project and is made more difficult in our case by the highly technical nature of the research we are supporting. To address this problem, we deliberately foster strong working relationships with the research scientists in our collaborations. Our Center employs several technical managers who have special knowledge of relevant scientific and computational domains. Together with the CIBC faculty, these managers serve as liaisons between the scientific researchers and the core software development staff, coordinating development activities with specific research goals.

At the CIBC we use modular, generic coding principles. Our goal is to engineer reusable and extensible software modules that are interoperable with other open-source components. The SCI Institute has already developed a number of open-source code modules for scientific visualization and simulation. The CIBC is extending these modules for our work and building interfaces between our code and external open-source code libraries such as ITK for image processing. Our goal is not to duplicate the efforts of other open-source initiatives, but to be active users and contributors to those projects. For much of our work, the CIBC is using a platform called SCIRun

that was developed by the SCI Institute as a workspace for visually assembling computational work flows. SCIRun provides a graphical user interface layer around computational modules and an editor for connecting inputs and outputs of those modules into logical networks and advanced visualization windows[5].

A record of successful builds and regression testing over time will be an important guarantee to users of the stability of CIBC software. To support this level of rigor in testing, we are setting up a suite of dedicated testing machines and adopting a distributed build and testing environment. Distributed testing is important because many of our geographically diverse collaborators also contribute code and support for our tools and need access to our code base and our testing infrastructure. We use a variety of tools to support distributed development and testing. All of the code for our software projects is controlled by a revision control database that is accessible via the internet. We are migrating to the use of CMake for cross-platform configuration and building[6]. Along with CMake, we plan to use the Dart software quality system[7] to manage builds and testing on multiple platforms, at remote sites, and with various build configurations. The Dart tool compiles building and testing results into a web page that is accessible online, which allows local and remote developers to quickly see how their individual code edits affect the entire software project. Dart facilitates early bug detection and keeps everyone up-to-date on the status of the system.

The material disseminated from our Center includes not only software, but supporting documentation, scientific publications, and datasets. Dissemination is managed with a variety of tools and strategies. Our materials are freely available through a web interface and we maintain a user mailing list for informal support of our user community. We also make use of an online bug tracking database that allows users anywhere to enter bug reports. For our formal collaborators, we provide additional training and support through site visits, both by visitors to our NCRR Center and by visits to collaborator sites by Center personnel. The Center also plans to conduct workshops and tutorials for collaborators and other interested parties.

4. RESULTS

The CIBC is young, but already we are producing some exciting results. Using a combination of existing SCI Institute software modules, open-source libraries, and new development, we offer several applications for imaging, visualization, and bioelectric field simulations that are in use by our scientific collaborators. Our applications target users with various levels of computer expertise, from non-computer professionals such as biomedical scientists and clinicians, to more experienced computer scientists who need scripting and visual programming tools. CIBC software also supports a range of computer architectures and operating systems. Our software is designed to take advantage of large-scale multi-processor machines, but also runs on standard laptop and desktop configurations. This remainder of this section describes in more detail some of the software applications we have developed.

The Biomedical Problem Solving Environment, or BioPSE, is an application that integrates the SCIRun workspace and visualization windows with bioelectric field modeling components. BioPSE gives the user the ability to design and manage each step in a sequential computing process and create batch processes to execute repeated simulations. BioPSE is fully interoperable with ITK image processing filters and we are working to integrate a mesh-generation package to provide a unified application for simulation that takes biomedical image data through every step from model building to simulation results. Figure 1 shows the BioPSE interface, including

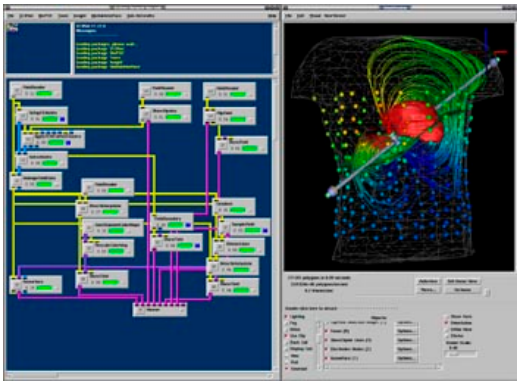

Fig. 1. The BioPSE problem solving environment.

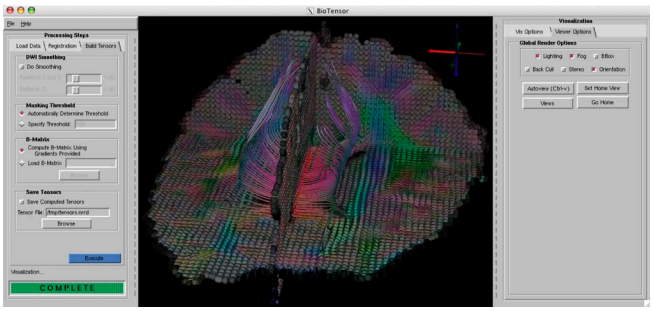

Fig. 2. The BioTensor tensor volume visualization application.

the visual module network editor and visualization of simulation results. BioPSE supports both current and voltage sources, as well as various internal boundary conditions (e.g. a subcutaneous patch electrode containing a region of constant potential on the front and an insulating backing). The geometry of BioPSE voltage sources can be arbitrarily complex, including wire electrodes, patch electrodes, can electrodes, and point electrodes. For current sources, the user can choose mix-and-match internal dipoles, surface point electrodes (e.g. for EIT), and complex user-defined charge distributions specified over user-defined surfaces (e.g. current densities defined over neural sheaths). An associated application, BioFEM, is built on top of BioPSE and SCIRun to provide a simplified interface to computing the electric field in a volume produced by a set of dipoles. BioFEM computes a solution to the bioelectric field forward problem and voltage values at electrode positions, which can be compared with values recorded via ECG or EEG.

BioTensor, shown in Figure 2 is a program that processes and visualizes diffusion tensor images, an important modality for many of our research collaborations. It can read diffusion weighted images (DWIs), perform correction for a common class of distortions in echo-planar imaging, estimate tensors from DWIs, and visualize the diffusion tensor field, which gives researchers an indication of the relative anisotropy of water diffusion in tissue. BioTensor's functionality is easily extended, and future versions will have expanded capabilities based on contributions from external collaborators and internal development.

BioImage is a program for volume visualization of medical image volumes. With BioImage, users can directly render a variety of native data formats (e.g. DICOM, Analyze, VFF, NRRD, PICT) and apply basic processing algorithms such as cropping, resampling, his-

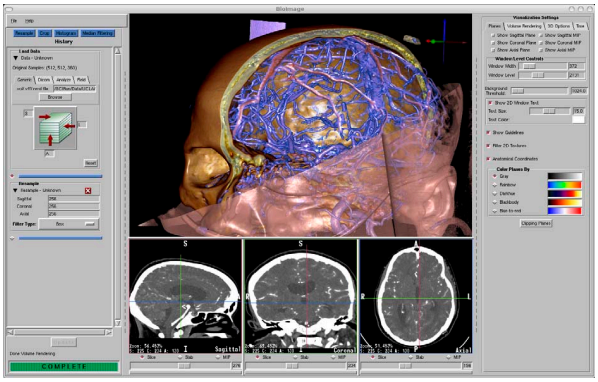

Fig. 3. The BioImage volume visualization application.

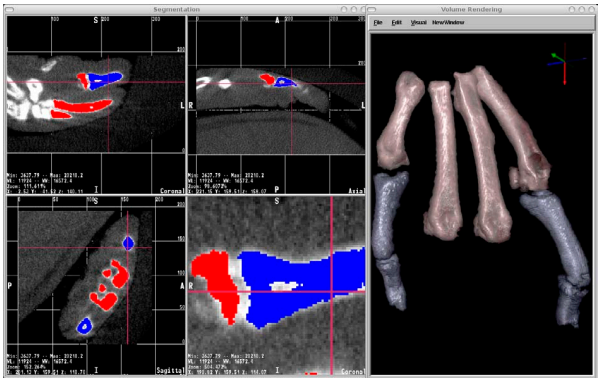

Fig. 4. The BioSegmenter volume segmentation and visualization application.

togram equalization, and median filtering. The BioImage visual interface allows users to interact with their data both in two-dimensional slice views, as well as in a full three-dimensional volume rendering view, which is powered by high-performance volume rendering modules. Users can move between the two- and three-dimensional views to control how different features of their data are displayed and gain both quantitative and qualitative insights into their data. BioImage supports interactive volume rendering of large datasets using a two-dimensional transfer function. With the novel approach of computing the gradient direction and magnitude on the GPU, graphics cards with limited memory (64MB) can still render large volumes (512^3 voxels) at interactive rates. The BioImage application is illustrated in Fig. 3.

The BioSegmenter application is an OpenGL-based viewer and editor of scientific voxel data that expands the functionality of BioImage. It can handle data types ranging from eight-bit to double precision floating point formats and features four-dimensional ($XYZ+time$) navigation and editing of scalar, vector, and RGBA data. Users can simultaneously view multiple orthogonal, two-dimensional slices of the data and create segmentation masks (labeled images) using manual drawing tools and automated segmentation algorithms. Any number of volumes can be displayed relative to each other with correct origin, orientations, and data spacing. Visualization tools include standard and custom color lookup tables, infinite zooming, paning, and surface and volume rendering. Multiple modalities acquired at different resolutions can be displayed using separate color maps with transparency for manual registration. Data may be cropped, resampled, and thresholded. BioSegmenter also includes powerful higher-dimensional image processing and segmentation algorithms from ITK filters including isotropic and anisotropic blurring, edge-detection, image statistics, flood filling, morphological operations, watershed segmentation, and level-set segmentation algorithms. We are also working to add automatic image registration methods from ITK to the BioSegmenter tool. Figure 4 illustrates the basic BioSegmenter interface.

5. CONCLUSIONS

The Center for Integrative Biomedical Computing (CIBC) is an NIH NCRR center with the mission to develop high performance image analysis, simulation, and visualization software in support of biomedical research. Development at the CIBC is collaborator-driven: software is designed in close cooperation with practicing scientists to address specific research needs. The CIBC believes in an open-source development process and participates in other biomedical open-source software initiatives where those efforts are complementary to our own. To that end, we are implementing an agile, open-source software process that supports distributed development, builds, and testing. Initial results at the CIBC include integrated software for bioelectric field simulation, high performance image and volume visualization, and image segmentation and analysis.

6. ACKNOWLEDGMENTS

The authors would like to acknowledge the rest of the CIBC team, including developers Darby Van Uitert and Michael Callahan, technical manager Jeroen Stinstra, and principal investigators Chris Johnson, Rob MacLeod, Dana Brooks, and Ross Whitaker.

This work was supported by NIH National Center for Research Resources grant P41 RR12553.

7. REFERENCES

[1] E. Zerhouni, "The NIH Roadmap," *Science*, vol. 302, pp. 63–72, October 2003.

[2] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*, Insight Consortium, http://www.itk.org/ItkSoftwareGuide.pdf, 2003.

[3] R. Kikinis, "National Alliance for Medical Image Computing," http://www.na-mic.org.

[4] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit, An Object Oriented Approach to 3D Graphics*, Prentice-Hall, 1998.

[5] S.G. Parker, D.M. Weinstein, and C.R. Johnson, "The SCIRun computational steering software system," in *Modern Software Tools in Scientific Computing*, E. Arge, A.M. Bruaset, and H.P. Langtangen, Eds., pp. 1–40. Birkhauser Press, Boston, 1997.

[6] K. Martin and B. Hoffman, *Mastering CMake, a Cross-Platform Build System*, Kitware Inc., 2004.

[7] GE Corporate Research, Development, and Insight Toolkit Project, "Dart Tests Reports and Dashboards," http://www.itk.org/Dart.