

The Bisque-Phytomorph Image Analysis Project

Edgar Spalding (spalding@wisc.edu) and B.S. Manjunath (manj@ece.ucsb.edu)

Phenotypes due to induced mutations or natural genetic variation are rich sources of information about the genomic basis of plant form and function. Yet, the tools for studying phenotypes today are not nearly as sophisticated as those for studying the genetic side of the Genotype-to-Phenotype relation that needs to be mapped. Progress toward the overall goal will be accelerated when the two sides of the equation are more technologically balanced.

One problem to overcome is insufficient resolution to detect the influence phenotype. The other is insufficient throughput, making genome-level investigations infeasible. Image analysis has the potential to reduce both types of problems. In a broad sense, image analysis is the science of extracting information embedded in digital images. The images could be of molecular distributions in a cell, or texture distributions in a crop field, or at some biological level in between. The images may be 2D bitmaps or comprised of multiple layers of information, single snapshots or a time series, or combinations of these. Appropriately designed image processing algorithms can extract and quantify information suspected of being important by the biologist or of unknown importance. In some instances such algorithms can be completely automated. Depending on the resolution of the acquired images, the level of extracted information can be much more detailed and objective than a manual measurement. Enough progress has been made recently on understanding how to perform image-based plant growth and development experiments, on feature-extraction algorithm design, and incorporation of the enabling computer sciences to propose an activity to iPlant that will have broad impact on how some plant phenotypes can be quantified. Advancing this approach would help right the technological imbalance referred to above.

The Phytomorph project at the University of Wisconsin (<http://phytomorph.wisc.edu>) focuses on developing image-analysis tools for quantifying plant phenotypes in popular experimental systems, mostly Arabidopsis and maize. The emphasis has been on the early stages of seedling development. The Bisque project at the University of California, Santa Barbara (<http://www.bioimage.ucsb.edu/bisque>) presents and advances a sophisticated database environment for storing and manipulating biological images. At its core is a tree-structured database designed for images and metadata. The key feature of Bisque is that it can incorporate analysis-specific algorithms to create a workflow that takes images and the metadata and returns processed results required by the biologist. The following proposal prepared by Edgar Spalding (PI of Phytomorph) and B.S. Manjunath (PI of Bisque) and their colleagues describes a year-long activity that would integrate aspects of their two projects into a cyberinfrastructure-based tool taking the shape of a portal that enables the plant biologist to perform quantitative, image-based phenotype studies. The longer-term vision is to integrate the output of a Bisque-based image analysis pipeline with the outputs of activities closer to the genotype side so that plant biology research begins to build true multi-dimensional quantitative representations of the genotype-to-phenotype relationship.

Starting Points

As starting materials, the Phytomorph project will provide image analysis algorithms capable of measuring the size and morphology of Arabidopsis seeds (the same for maize kernels are in progress), and root growth. These methods already exist. Some are fully automated (e.g. Arabidopsis seeds, some root growth and shape algorithms). Some require bare minimum input (e.g. maize and rice coleoptile growth and shape, Arabidopsis root branching) and some require user supervision (e.g. Arabidopsis hypocotyl tracing). A tracking algorithm capable of kinematic analysis of cellular-level expansion rates in Arabidopsis roots is now matured. Thousands of experiments, each typically consisting of hundreds of images already exist in a database. Some of these experiments have already been processed from image acquisition all the way through the quantitative trait locus (QTL)-mapping stage. The Bisque project has immediately available a user environment in which the existing algorithms can be exposed to the user and through which the results will be presented.

Initial Progress

During the preceding six months we have had good progress to our initial goal to supporting several analysis algorithm through bisque. The initial work consisted of replicating the grid execution environment at UCSB used at Univ. of Wisconsin. Condor was installed on our local bioimage cluster of 32 machines. The condor-Bisque interface developed by the Spalding lab was ported and extended at UCSB. The condor interface allows bisque web-based resources to be utilized in the condor environment which prefers file-based interfaces. UCSB has extended bisque to allow multi-dimensional image construction based on groups (zip,tar) of 2D images sets. The initial set of algorithms was delivered to UCSB and included: A seed size analysis, a root-tip angle analysis and a multi-tip angle analysis. A sample MiniApp (workflow) is being developed for the multi-tip analysis. Work is continuing rapidly to finish the integration.

Next Steps

The next steps will be to finish incorporating these algorithms into Bisque, adapting both until a user-friendly tool with optimal performance is achieved. Creation of a relational database and a workflow that populates it as well as the xml-tagged Bisque database to provide users alternative methods of interacting with the data. In some cases, the tree structure of the Bisque database will be ideal and for other purposes a more rigidly defined relational database may be advantageous, especially when integrating with other types of 'omic' data. After the initial integration, in coordination with the Spalding group, we will assist integrating specific analysis visualizations.

After this initial analysis, visualization and workflow component integration, we will continue to identify algorithm of interest to the overall plant community, including image-based QTL analysis based on plant morphology. We will work closely with the Spalding and iPlant workgroups to determine what facilities are needed for a comprehensive image analysis solution.

Open ended support tasks

Over the next year, we expect the above activities to be replicated, in the sense that several open-ended tasks based on availability of new image-based analyses become available:

Support for new modules

The Bisque team will support and assist the creation of new modules by providing bisque access widgets, tools and programmer assistance as well as guidance.

Better support for identified workflows

Workflow application based on Bisque mini-apps will be created to guide users through the necessary steps. Essentially, the goal will be to make a biologist-friendly tool capable of converting a user request and image set into a computational job (for some cluster execution environment).

Bisque development

Our initial effort has identified several Bisque extensions and improvement that should be developed. The following section outlines specifically the tasks that UCSB will undertake during the following year in order to facilitate the incorporation of analysis, improve performance and allow better utilization of computational resources.

Abstracting module execution models (6 Man-months spread over 1 year)

Image datasets can often be more efficiently processed using large scale computational resources. Additional research and development directions will include incorporating the distributed computing environments (such as condor or hadoop) for module processing at the back end of Bisque. We will also be looking at the cloud computing paradigm that is becoming increasingly popular for large scale computing. This would allow researchers with or without local clouds access to practically infinite resources for image processing.

Extended module support (3 Man-months over 6 months)

Bisque module support is currently based on simple analyses with a few parameters. However, modern workflows require semi-automatic, interactive or with extended visualizations. The module system will be extended to support interactive and long-running analysis based on web-endpoints.

Programmer API (3 Man-months spread over 6 months)

Supporting the next generation of module execution modules (above) will require and improved programmer tools. APIs will be developed for Java, Matlab, C++ and python allowing module writers to interact with core bisque services simply and generically across different platforms.

Closer Integration of imaging technology (3 Man-months spread 6 months)

Close integration of imaging technologies including microscopes, cameras, and scanners shorten the time to analysis and permit better automation. A toolkit will be developed to permit local imaging infrastructure to integrated with remote bisque servers.

Integration of novel image types (3 Man-months spread 6 months)

Bisque current supports time-series and Z-stack images natively. New types of media include true video, and sparse images (collected on stage) would be added in the backend, web-frontend and API.

Bisque Core (12 Man months spread over 1 year)

Feature enhancements, improvements, and maintenance to the Bisque core including database infrastructure, large image support (tiled), fine-grained permissions and sharing, import/export of datasets, testing, module support, deployment options (cloud images) and user interface updates as needed.

Long term Bisque goals

While not part of this proposal we foresee activities that might be of interest to the iPlant community including extending Bisque facilities to emerging technologies including:

Extending to the Semantic Web (12 Man-months spread 1 year)

Support for incorporating and providing metadata compatible with semantic web tools would be added. Services would optionally export RDF data, and indexing (searching) would be extended to include aliased terms instead of keyword matching.

Support for social computing (12 Man-months over 1 year)

Internal support for provenance would be extended to support for collaborative editing, analysis and visualization. Project spaces, group notifications, and resource tracking would be developed or integrated.

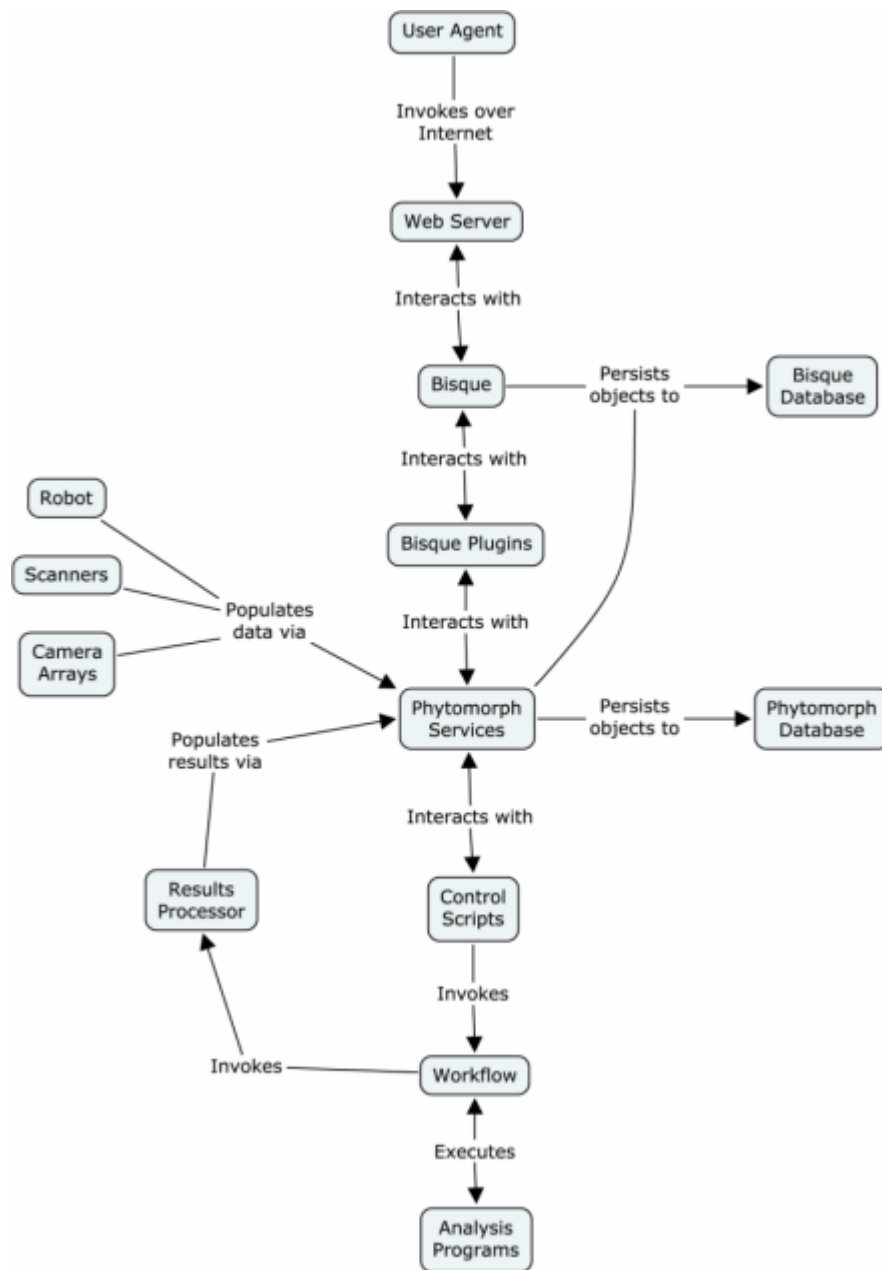
Extended visualization options (6 Man-months spread 1 year)

Better options for developing and integrating visualization will be required for some analysis types. This work will identify what sort of tools should be integrated and provide import/export options as well as integrating some initial visualization option directly.

The One Year Goal

After the progress during the initial six month period, we are confident that at the end of a year that a robust tool for efficient analysis of images of plant structures relevant for phenotype studies will be in operation. It will feature the algorithms that we know from experience are helpful to multiple labs that acquire compatible image data. Roots and shoots of monocot and dicot seedlings will be represented. As the community sees what can be accomplished by this way of quantifying phenotypes, we expect interest in algorithms capable of analyzing specific phenotypes at scales ranging from subcellular to field plots will drive path-breaking research. Any such algorithms could be incorporated into the tool.

The Figure below shows a preliminary flow chart for the proposed tool. In general, the Bisque team will focus on the databasing and workflow aspects while Phytomorph team will focus on algorithms, their integration, and connection to a condor grid. There will be substantial areas of overlap between the teams. For example, the Bisque team is also expert at image processing algorithms and the Phytomorph team will need to work at some times deeply with the Bisque source code and databases.



The One to Two Year Goal

The goals in the longer term will be to continue identify and integrate analysis and visualization workflows. The core tasks for UCSB will be the continued development and extension of the bisque platform for the purpose of providing a robust and capable infrastructure. Development and extensions will be guided by the needs of the iPlant community and workgroups. The figure below demonstrated the expected architecture to be developed during the 18 months.