UCSB Bioimaging and OME

April 20 2006
Project

- Manjunath (UCSB) & Murphy (CMU)
- 5 Years bioimaging effort
- 2 ½ years in
- Biology + Image Processing + Catabases
Motivation

- **Diverse users**
  - UCSB Neurosciences institute
    - Retinal detachment: Fischer et al
    - Microtubule dynamics: Feinstein and Wilson
  - CMU Murphy Lab
  - Flybase
  - *Univ. of Utah Marc Lab* (internal development)

- **Dataset challenges**
  - Multiple types and collection techniques
  - Complex images (5D) and metadata
  - Acquisition and management of large sets
  - Security and Sharing of experimental data
Motivation

• Diverse users
  – UCSB Neurosciences institute
  – CMU Murphy Lab
  – Flybase
  – Univ. of Utah Marc Lab

• Dataset challenges
  – Multiple types and collection techniques
  – Complex images (5D) and metadata
  – Acquisition and management of large sets
  – Security and Sharing of experimental data
Motivation

- Diverse users
  - UCSB Neurosciences institute
  - CMU Murphy Lab
  - Flybase
  - Univ. of Utah Marc Lab

- Datasets challenges
  - Multiple types and collection techniques
  - Complex images (5D) and metadata
  - Acquisition and management of large sets
  - Security and Sharing of experimental data

- Analysis creates knowledge
  - Image analysis
  - Querying
  - Mining
Challenges

• Experimental metadata capture
• Dataset management
  – Personal collections
  – Multiple organizations
  – Privacy and security
• Analysis design
• Analysis integration
UCSB/CMU Timeline

- Kickoff Sep 2003
- Staffing Sep 2004
- 1\textsuperscript{st} Prototype delivered Jan 2005
  - (mysql, php, collections)
  - 2\textsuperscript{nd} prototype design begins
- Illya visits Feb 2005
- Internal Review Mar 2005
- Internal move to OME Oct 2005
- Review and BISQUE demo Mar 2006
• Current dataset collection
• BISQUE functionality
  – Data management
  – Ground truth acquisition
  – Analysis
• Architecture
Current collections

- UCSB Retinal Objects (confocal, EM)
- UCSB Microtubule Objects (light, AFM)

<table>
<thead>
<tr>
<th>Type</th>
<th>Current</th>
<th>Backlog</th>
<th>Rate/y</th>
<th>Expected 4Yrs</th>
<th>2D Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retinal EM</td>
<td>500</td>
<td>22000</td>
<td>500</td>
<td>23,000</td>
<td>23K</td>
</tr>
<tr>
<td>Retinal confocal P</td>
<td>3000</td>
<td>600</td>
<td>2400</td>
<td>10,000</td>
<td>10K</td>
</tr>
<tr>
<td>Retinal confocal Z</td>
<td>600</td>
<td>14000</td>
<td>12000</td>
<td>10,000</td>
<td>50K</td>
</tr>
<tr>
<td>Microtubule light</td>
<td>3000</td>
<td>2500</td>
<td>2500</td>
<td>13,000</td>
<td>400K</td>
</tr>
<tr>
<td>Microtubule AFM</td>
<td>0</td>
<td>500</td>
<td>1200</td>
<td>5000</td>
<td>30K</td>
</tr>
<tr>
<td>Flybase</td>
<td>0</td>
<td>125K</td>
<td>0</td>
<td>125K</td>
<td>125K</td>
</tr>
</tbody>
</table>

- Analysis feature sets can be several times raw data
- Total estimated size in TBs and growing for 1 Lab
- Complexity and analysis are the main issues

April 22nd, UCSB
Data management capabilities

- **Digital notebook**
  - Capture experimental/image parameters
  - Direct import process
  - Reconfigurable (confocal, 4 x microtubule, etc)

- **Web**
  - Web access and browsing
  - Organize images and metadata
  - Data sharing environment
  - Search by metadata or content
  - Integrated analysis

April 22nd, UCSB
Screenshots (Digital Notebook)
Data management capabilities

• Digital notebook
  – Capture experimental/image parameters
  – Direct Import process
  – Reconfigurable (confocal, 4 x microtubule, etc)

• Web
  – Web access and browsing
  – Organize images and metadata
  – Data sharing environment
  – Search by metadata or content
  – Integrated analysis
Screenshots (browsing)
Screenshots (search)
Screenshots (search)
Screenshots (search similar)
Screenshots (search similar)
Screenshots (personal collection)
Screenshots (5D Viewer)
Screenshot (5D with tracks)
Image analysis

- Integrated image analysis
  - Advanced image processing
  - Cell counter (ImageJ)
  - Quantify microtubule dynamics
  - Track identification

- Integration in progress
  - Segmentation and classification
  - Microtubule tracker
  - Modeling microtubule dynamics
  - Modeling uncertain data
  - Relevance feedback improving search
Screenshots (Image Analysis)
Image processing and infrastructure teams are developing universal “ground truth” collection tools able to retrieve data from data-base and feed user defined information back to the database. The main communication vehicle is XML interchange format.

At the current stage stand alone tools are being developed and tested that later on will be grouped in the universal application able to communicate directly to the data-base.
System description overview

- Current dataset collection
- Current functionality
  - Data management
  - Ground truth acquisition
  - Analysis
- Architecture
Hardware/software infrastructure

• **Hardware**
  – 16 node Cluster
    • dual Intel Xeon 3GHZ
    • Gigabit network switch
    • 2 TB Storage

• **Software**
  – Linux
  – OME (Apache, Postgresql)
  – Bisque
OME Base

- **Open Microscopy Environment**
  - “OME is an open source software project to develop a database-driven system for the quantitative analysis of biological images. OME is a collaborative effort among academic labs and a number of commercial entities”.
- Provides base for image/metadata storage and analysis integration
- **Boston: Sorger Lab**
- **Baltimore: Goldberg Lab**
- **Dundee: Swedlow Lab**
- **Madison: LOCI**
Bisque extensions

• Ongoing extensions with OME
  – Content based search
  – Analysis integration with OME
    • Segmentation
    • Cell Counting
    • Etc.
  – Front end dataset and analysis support
  – Schema additions for image types and analysis
  – Uncertainty modeling and queries
Calstate SB

- Scalable Distributed System for collaboration
- Cluster for analysis
- Distributed storage
  - Lustre filesystem
  - Shared Postgres metadata
  - Local image servers
- Bisque/OME over file system

April 22nd, UCSB
Conclusion

• Built 1st generation w/ >4000 images
  – Integrated several *useful* analyses
  – Being used internally
  – External Interest
• Immediate future
  – Continued development
    • analysis integration
    • Dataset integration
  – Local user feedback
  – Expand user base
  – Remote deployments
  – Integration with other projects
    • Flybase: Integration of schema/datasets in progress
    • BIRN: Mediation framework: when available
End

• Please visit:
  – http://bioimage.ucsb.edu/bisque
  – User: guest/bioimage

• Team:
  H. Alagud, A. Black, D. Fedorov, S. Jagadish, S. Jagannath, R. Kumar, T. Kuo, W. Smith