OMERO.searcher: extensible, cross-platform content-based image search

Ivan E. Cao-Berg
Murphy Lab
Ray & Stephanie Lane Center for Computational Biology
School of Computer Science
Carnegie Mellon University
Content-based Image Search

- Also called Query-by-image-content (QBIC)
- Find images whose content, as reflected by image features, is similar to one or more query images
- Can use positive and/or negative examples
- Can be iterative (relevance feedback)
OMERO.searcher Workflow

(a) Flowchart:
- Image Importing via Importer
- Select Image(s)
- Content Search
- Select Additional Image(s)
- Satisfied?
  - Yes: End
  - No: Feature Calculation, Content DB Update

(b) Interface:
- # retrieved images 10
- Search by image content
- Negative and Positive classification buttons
- Search results with image thumbnails and details
OMERO.searcher: content-based image search for microscope images

Baek Hwan Cho, Ivan Cao-Berg, Jennifer Ann Bakal & Robert F Murphy

Affiliations  |  Corresponding author

Published online 28 June 2012

To the Editor:

Fluorescence microscopy is growing dramatically both in terms of technical capabilities and the volume of images generated. Online repositories have been created to provide public access to images and opportunities for joint research for many scientists¹. This has reintroduced challenges faced when sequence and structure databases were being established: developing fast and
OMERO.searcher ContentDB

- **ContentDB** key to content searching
- Contains features describing each image in a database and links to the image and its annotations
- Supports any feature set if Python code is provided for calculating it
- Portable
Components

OMERO.server
Python

Searcher Webapp
External Search Webapp
OMEROcli
Local Client

Content DB file(s)
OMERO.searcher Python libraries

- PySLIC
- ricerca
- PySLID
Python Libraries Updates

• ricerca: (1) updated definition of content databases to support changes implemented in Local Client
• PySLID: (1) deprecated older methods, (2) optimized existing methods and (3) improved handling of edge cases
• PySLIC: (1) minimal bug fixes
Searcher **Server** version update

- Searcher **Server** does queries using images from within an OMERO database for images in the same OMERO database
- Originally a patch to OMERO.web
- Now implemented as a Webapp
Searcher Webapp
Searcher Webapp
OMERO.searcher *External Search*

- *External Search* can use a user-supplied image(s) (e.g., from local computer) as a query to find images within an OMERO database.
- Originally implemented as a separate Django.
- Plan to make it a Webapp also.
- Tentative release at end of summer.
OMERO.searcher CLI

- PySLID and ricerca were designed so that users can write their own OMERO.scripts.
OMERO.searcher *Local Client*

- *Local Client* runs as a standalone program on your local computer, using your files as queries
- Can search any database for which you have access to a contentDB (both OMERO and non-OMERO)
- Automatically checks for updates to contentDBs and downloads
OMERO.searcher *Local Client*

- Runs from Python command line
- GUI version coming
- Tentative release at end of summer
OMERO.searcher Local Client

```
#!/bin/bash

PATTERN=CY-5
IID=46
OUTPUT_FILENAME=test_swedlow_HPA_{IID}_{HPA}_{PATTERN}.html

#INPUT ARGUMENTS
DATABASE=proteinatlas.org
FEATURESET=slf34
RESOLUTION=0.0581
NUMBER_OF_IMAGES=10
REFERENCE_CHANNEL_IMAGE=./images/swedlow/Bodis1_Plk1_B56a_01_05_R3D_D3D_DAPI_0039.bmp
PROTEIN_CHANNEL_OF_INTEREST=./images/swedlow/Bodis1_Plk1_B56a_01_05_R3D_D3D_{PATTERN}_0039.bmp

python omero.searcher.py -g $DATABASE -f $FEATURESET -e -5 -n $NUMBER_OF_IMAGES
   -r $RESOLUTION -o $OUTPUT_FILENAME $PROTEIN_CHANNEL_OF_INTEREST $REFERENCE_CHANNEL_IMAGE

# Wrote 49 lines
```

Get Help      ^G  WriteOut      ^Q  Read File      ^R  Prev Page      ^N
Exit          ^X  Justify       ^Y  Where Is      ^W  Next Page

Cut Text      ^K  To Spell     ^C  UnCut Text      ^U
Local Client
Local Client
Local Client

- Allow search across images of different resolutions
- Allow search of more than one contentDB at a time
Local Client

• Currently available contentDBs
  – Human Protein Atlas
  – PSLID RandTag
  – The Cell Library
  – more coming…
Future Directions

- Support for
  - Search across multiple servers from within the Webapps
  - BioFormats for reading images using *Local Client*
  - Trigger feature calculation using bridge
Image-Derived Generative Models

• Goal is to model the reality underlying microscope images
• Build models that capture
  – distributions of all cellular components
  – how these change from cell type to cell type
  – how specific aspects of these models are changed by perturbagens
May 17, 2013: Version 1.9.0 released!

**New:** Now allows synthesis of cell and nuclear shape instances for Hela cells using a diffeomorphic model.

The **CellOrganizer** project provides tools for

- learning generative models of cell organization directly from images
- storing and retrieving those models in XML files
- synthesizing cell images (or other representations) from one or more models

Model learning captures variation among cells in a collection of images. Images used for model learning and instances synthesized from models can be two- or three-dimensional static images or movies.

**CellOrganizer** can learn models of

- cell shape
- nuclear shape
- chromatin texture
- vesicular organelle size, shape and position
- microtubule distribution.

These models can be conditional upon each other. For example, for a given synthesized cell instance, organelle position is dependent upon the cell and nuclear shape of that instance.

Cell types for which generative models for at least some organelles have been built include human HeLa cells, mouse NIH 3T3 cells, and Arabidopsis protoplasts. Planned projects include mouse T lymphocytes and rat PC12 cells.

Support for **CellOrganizer** has been provided by grants GM075205 and GM090033 from the National Institute of General Medical Sciences, grants MCB1121919 and MCB1121793 from the U.S. National Science Foundation, by a Forschungspreis from the Alexander von Humboldt Foundation, and by the School of Life Sciences of the Freiburg Institute for Advanced Studies.
CellOrganizer

Statistical Model

- Nuclear shape
- Cell shape
- Object pos. probability
- Microtubule distribution
- Object appearance
- Object positions
- Object number
- Object distribution

Training

Synthesis

Cell Images

Synthetic Images

Zhao & Murphy, Cytometry 2007
Model Training

- Image Processing
- Parameter Estimation
- Model Building
PC12 cells undergoing differentiation
Image Synthesis

- Model loading
- Image generation
- Image linking
CellOrganizer

• v2.0 (expected release at the end of the summer)
• Open source (Matlab and C) [and python!]
• Contributions/collaborations welcome!
Future Directions

• Interface with OMERO for
  – Preprocessing images
  – Feature calculation
  – Storing models
  – Storing generated images

• Interfaces to cell simulation engines
  – Store and export models on demand