Image Analysis
and quantifying PLA signals using CellProfiler

Sept. 2017

Carolina Wählby
Professor at the Centre for Image Analysis, Dept. of Information Technology, Uppsala University and SciLifeLab and Imaging Platform, Broad Institute of Harvard and MIT

Overview

Image analysis
• Why quantitative measurements?
• Why free and open source tools?
• Fundamental steps of image processing and analysis
  • Preprocessing; filtering
  • Object detection; segmentation
• Intro to CellProfiler

In situ PLA data analysis
• CellProfiler exercise and/or analyze your own data)
Why automated image analysis?

Does the cell count differ between these two images?
Is there a difference in the proportion of cells that express GFP?
Is there a difference in the localization of the GFP-tagged protein?

1. Detect cells
2. Segment cytoplasm and nuclei
3. Measure intensity of DAPI and GFP
Why automated image analysis?

Classify cells as having
- Cytoplasmic GFP
- Nuclear GFP
- No GFP expression

Does the cell count differ between these two images?
Is there a difference in the proportion of cells that express GFP?
Is there a difference in the localization of the GFP-tagged protein?

Bioimage informatics makes it possible to quantify image information in an unbiased way, and increases the scientific value of the experiment.
Bioimage informatics makes it possible to quantify image information in an unbiased way, and increases the scientific value of the experiment.

Why automated image analysis?

Does the cell count differ between these two images? - NO
Is there a difference in the proportion of cells that express GFP? - NO
Is there a difference in the localization of the GFP-tagged protein? - YES

Why quantitative measurements?

Which nucleus is brighter?

Same brightness...

Which nucleus is larger?

Same size...

What are the colors of the cells?

Same color...
Fundamental steps when extracting information from image data = ‘analysis pipeline’

- **Image acquisition**
- **Preprocessing, filtering**
- **Object detection, segmentation** (including 3D and tracking over time)
- **Making measurements, feature extraction**
- **Object classification, interpretation, recognition**

Knowledge about the application

Result: **Result**

Task: Count red and green dots per cell

Co-culture of mouse- and human cells, gap-fill sequencing of β-actin.

- **Mouse β-actin**: GCAAGCCGG (green)
- **Human β-actin**: GCAAGGCCGG (red)
- **DNA (blue)**

SNV
pixel = picture element; the numbers in the matrix that builds up the image

(voxel = volume element; builds up a 3D image volume)
'Modeling' of an object:
Define a dot as a group of pixels brighter than a fixed intensity value, e.g. >50.

Model; a dot is a group of touching pixels brighter than T.
A too low threshold; structures in the background are picked up.
A too high threshold; true dots are lost.

New model; a dot is a group of touching pixels of approximate 10-pixel radius, brighter than the local background intensity. (= tophat filtering)
A ‘minimum’ filter creates a new image where the output is replaced by the minimum pixel value (orange) within a pre-defined pixel neighborhood (3x3 pixels here).
A ‘minimum’ filter creates a new image where the output is replaced by the minimum pixel value (orange) within a pre-defined pixel neighborhood (3x3 pixels here).

**Image filtering: Top-hat filtering to enhance dots**

The filter size should always be larger than the objects of interest. These filters, and many others, are available in the ‘Morph’ module of CellProfiler.
Image filtering: **Min/max filtering** to enhance or suppress details

- Among the pixels within the filter in the input image, select the max (or min) value, and put in the output image.

Output of 7 x 7 max filter  original  Output of 7 x 7 min filter

=Dilation with square of size 7x7 pixels  

=Erosion with square of size 7x7 pixels

Once the dots are found, we need to define what a cell is.
Object detection: image segmentation

To divide the image into parts/regions/objects, which correspond to what we would like to analyze.

Can be based on

- Similarities: in intensity, shape, texture
- Discontinuities: edges
- High-level information: shape models

Approaches:

- Intensity thresholding
- Region based
  - Watershed segmentation
  - Region growing
  - Split and merge
- Edge based
- Active shape models
- Snakes
- Level sets
- Pixel classification...

Intensity thresholding

What is the best intensity threshold value for dividing the intensity histogram into foreground and background pixels?

Approach: Pick the method that provides the best results

- Otsu: Good for readily identifiable foreground / background
- Mixture of Gaussian (MoG): More accurate matching of intensity distribution → More robust than Otsu
- Fixed: assumes intensity distribution is constant throughout experiment
Segmentation by thresholding alone (output of CellProfiler IdentifyPrimary Objects, with no separation of clumps)

Segmentation by thresholding and watershed segmentation (output of CellProfiler IdentifyPrimary Objects, with separation of clumps based on intensity)

Carolina Wahlby, carolina@cb.uu.se

Watershed segmentation
-useful if objects touch

Note that if the objects vary a lot in intensity (many local maxima) watershed segmentation may lead to over-segmentation. This can be reduced by applying a smoothing (mean) filter prior to watershed segmentation.

Carolina Wahlby, carolina@cb.uu.se
Model: a nucleus is a group of touching pixels brighter than T, and with a round shape.
Model: a cell is a group of touching pixels brighter than T, linked to their ‘closest’ nucleus, measuring distance via pixels brighter than T (=seeded watershed segmentation, using nuclei as seeds).

Cell 1
- green dots: 23
- red dots: 3

Cell 2
- green dots: 4
- red dots: 32

Cell 3
- green dots: 65
- red dots: 0

Cell 4
- green dots: 69
- red dots: 5

Cell 5
- green dots: 8
- red dots: 74

Cell 2 green dots: 4
red dots: 32

Cell 3 green dots: 65
red dots: 0

Cell 4 green dots: 69
red dots: 5

Cell 5 green dots: 8
red dots: 74

Cell 1 green dots: 23
red dots: 3

CellProfiler

- Published in 2006 (Carpenter et al, Genome Biology)
- Cited in more than 3,500 scientific papers
- ~ 3 new published papers every day during 2015
- Launched >150,000 times/ year
- One of the Top 10 most-accessed papers of all time in Genome Biology
- Winner of Bio-IT World’s IT & Informatics Best Practices Award (2009)
- Interfaces with ImageJ/Fiji, ilastik, OMER, reads files via Bioformats, KNIME, DeNovo, GE IN Cell Analyzer, GenePattern (output as .gct), and more.
- Originally based on Matlab, in 2009 completely re-written in Python
- www.cellprofiler.org (forum, tutorials etc)
- object detection methods based on my PhD thesis

Developed by the Imaging Platform, Broad Institute of Harvard and MIT, Cambridge, MA, USA
Analysis Modules: setting up your pipeline

The analysis modules are built up based on a previous example pipeline or from scratch. Each step can be tested on a given or random image from the input modules, and the effects of the settings can be seen if the ‘eye icon’ is open.

A wide range of specific as well as general-purpose functions are available.

Once the pipeline is completed, all images are analyzed by clicking ‘Analyze images’

Measure ‘everything’; Counts, Sizes, Intensities, Textures, Correlations, Relationships ... Saved to a spreadsheet or database
Benefits of free and open-source software

- ‘Reproducible research’: with an open source solution, you can provide your analysis pipeline as part of the supplementary material of your published paper (along with sample data). E.g. use the macro recorder of Image J to document exactly what you do with each image.
- You can easily share pipelines and analysis approaches with collaborators.
- You can bring your methods with you when moving to a different lab, or working remotely.
- Educational value: anyone can go in and look at the source and learn.
- Commercial companies have trouble keeping up with the most recent developments: the latest things are more likely to be found in the open-source community.
- A user community for an open-source software will often be a more responsive and efficient source of help than what can be provided through an expensive service package for a commercial software (e.g. CellProfiler forum).
- Please remember to properly cite the software in your publications to facilitate future funding for the developers!!!
Support services

Education 10% (visibility)
Free consultancy
Simple projects

Support/education on:
- Experimental design
- Image acquisition
- Data management
- Image analysis development
- Statistical analysis
- Best practices
- Visualization

20% personal development / scientific meetings, infrastructure development / visibility (web)

Independent users

Approval = 20h free support per PI, after 20h, 800 SEK/hr fee, or education to continue on your own.

Typical projects
Collaboration

Joint research projects (co-authorship):
1. User fees
2. Joint application for external funding
   - Involve externally funded PhD students/PostDocs
3. (Staff’s own research time 20%)

Some concluding thoughts

• Quantification increases the value of an experiment
• Consider image resolution in relation to the phenotype of interest: lower resolution results in more objects per image and better statistics!
• It is not always necessary to find individual objects: always try the simplest approach first and use as benchmark for more advanced approaches

Carolina Wahlby, carolina@cb.uu.se