Handling and Processing Big Data for Biomedical Discovery with MATLAB

Raphaël Thierry, PhD
Image processing and analysis – Software development
Facility for Advanced Microscopy and Imaging

23th June 2016
Key Takeaways – Benefits of MATLAB for Biomedical Research

1. Eliminate time-consuming programming tasks; Fast prototyping:
   *Ex: Achieved speed-up in programming can range from a factor 2 to 20*

2. Access to a wide range of state-of-the-art algorithms / libraries

3. Solve computationally- and data-intensive problems with available tools
   *Ex: More than 95% of projects involving big data are today treated with the help of Machine Learning approaches*
Facility for Advanced Imaging and Microscopy (FAIM)

Cutting-edge imaging software & tools and high-end microscopes for FMI

Main goals are:

- **Provide** high-end workstations for acquisition and analysis
- **Ensure** quality and efficiency by maintaining instruments and workstations
- **Development** of in-house image processing workflows
- **Train** scientists to take full advantage of hardware (cluster) and software
Innovation Challenges

- Prevalence of segmentation in many workflows.
- Manual intervention still a deeply anchored habit for biologists.
- Current automated methods are either too application specific or fail due to intrinsic algorithmic limitations.

Achievements

- Developments go toward speed-up (cluster computing), feature design & user interfaces
How did we get there and leverage MATLAB?

1. Preprocessing
   - 4D datasets: xycz
2. Training
   - Labelling
   - Training on cropped datasets
3. Pixel classification: Prediction
   - XY-plane, YZ-plane, XZ-plane
4. Segmentation & Quantification
   - Result
   - Prediction

Training on cropped datasets XY-plane, YZ-plane, XZ-plane.
Parallel Computing and MDCS with batch

1. Preprocessing
   - 4D datasets: xyzc

2. Training
   - Labelling
   - Training on cropped datasets

3. Pixel classification: Prediction
   - 4D : xyzc
   - 3D : xyz (c = dapi)

4. Segmentation & Quantification:
   - j = batch(myCluster,'fun')

MATLAB Desktop (Client)
Machine Learning in Behavioral Science

Neurobiology: Motor Circuit function

Single Mouse Tracking in full arena without camera calibration (Capelli, Group. S. Arber)

In accordance with the directives and ethics principles on animal testing in Switzerland
Machine Learning in Behavioral Science

Neurobiology: Neuronal circuits & computations

Tracking of Zebra Fish from dual view tomographic reconstruction (Huan Kuo, Group. R. Friedrich)
Machine Learning in Behavioral Science
Machine Learning in Cell Biology

Nuclear organization in development and genome stability
(Hauer & Gonzales, Group. S.Gasser)

Cell
Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in C. elegans Embryos

Graphical Abstract

MOVIES: 8 Z-stacks
$\Delta t = 80$ms
750 frames/ min
Machine Learning in Cell Biology

Cbx2 Targets PRC1 to Constitutive Heterochromatin in Mouse Zygotes in a Parent-of-Origin-Dependent Manner

Liu, Tardat, Group. Peters
Machine Learning in Quantitative Biology

Counting & quantification on Blastocyst
Julia Hacker, Group. A. Peters
Stack alignments: Nonrigid Transforms, Iterative Closest Point

<table>
<thead>
<tr>
<th>Image size</th>
<th>CPU time (s)</th>
<th>GPU time (s)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 x 1000</td>
<td>7.153</td>
<td>7.359</td>
<td>1.0</td>
</tr>
<tr>
<td>4000 x 4000</td>
<td>61.967</td>
<td>8.803</td>
<td>7.0</td>
</tr>
<tr>
<td>6000 x 6000</td>
<td>140.459</td>
<td>11.747</td>
<td>12.0</td>
</tr>
<tr>
<td>8000 x 8000</td>
<td>243.755</td>
<td>16.370</td>
<td>14.9</td>
</tr>
<tr>
<td>9000 x 9000</td>
<td>316.180</td>
<td>19.346</td>
<td>16.3</td>
</tr>
</tbody>
</table>

CPU i7-4770 4 cores @ 3.40GHz
GPU Tesla K40c

IEEE International Conference on Image Processing (ICIP 2014)

DEMIND DESTA INFRACOM ON EIGE PROCESSING

Raphaël Thierry, Moritz Kirschmann, Eric Hummel, Chris Hawes, Christel Genoud
Stack alignments: Multi-channel, Nonrigid Transforms, ICP

**before**

**after**

Extraction of the Nucleus membrane

Nucleus membrane
Concluding Remarks

- Machine Learning proved to be perfectly adapted for the analysis of Big Data in Biomedical Research

- We have found MATLAB to be a good solution to solve the multiplicity of issues encountered, such as:

  1. Rapid prototyping / proof of concept
  2. Handling diversity of problems as well as data formats
  3. Interaction/Interface with other softwares and languages
  4. Solutions for Speeding-up the process of Big Data
Next Steps

- Focus on Minimizing Memory Space Requirements and Execution Time

- Computation of Classifiers on-the-fly in MATLAB (while using Ilastik-GUI for Training and Feature selection on ND-Data)

- Deep Learning
Thank you for your attention