Current Practices, Better Options

Beata Turoňová

EMBL, Heidelberg
Optimization

- Reducing processing time
  - Computation
  - Manual processing
- Reducing (human-based) errors in processing
- Reducing amount of (low-quality) data
- Assuring robustness w.r.t. sample
Reducing Computation Time

- Good SW development
  - Fast, robust, and sustainable
  - Properly tested
  - Easy to parallelize if possible
  - Open-source

Subtomogram averaging

<table>
<thead>
<tr>
<th>Box Size</th>
<th>Matlab</th>
<th>C++ double precision</th>
<th>C++ single precision</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>1m 11s</td>
<td>0m 41s</td>
<td>0m 30s</td>
<td>0m 10s</td>
</tr>
<tr>
<td>72</td>
<td>8m 46s</td>
<td>6m 04s</td>
<td>3m 47s</td>
<td>0m 46s</td>
</tr>
<tr>
<td>144</td>
<td>85m 48s</td>
<td>65m 11s</td>
<td>38m 47s</td>
<td>6m 21s</td>
</tr>
<tr>
<td>288</td>
<td>805m 17s</td>
<td>558m 55s</td>
<td>423m 04s</td>
<td>47m 33s</td>
</tr>
</tbody>
</table>

- Times are for 1 iteration of 100 subtomograms and 100 rotations
- GPU is version **not** at all optimized
Reducing Errors in Processing

- Most caused by switching among different SW

Tilt-series Preprocessing
- Frame-Alignment
- CTF Estimation
- Dose-Filtering
- Alignment

MotionCorr (SerialEM)
CTFFind4, GCTF, ctfplotter
Matlab Script, eTomo
eTomo, protomo, autoTom

Subtomogram Averaging
Tom/AV3, PyTom, Dynamo, Relion

- Automate the transitions as much as possible
- Use SW providing a “complete” pipeline – emClarity, Warp
- Always check results after each step
Reducing Amount of Data

• In tomography less is often more
  • Prefer quality over quantity
  • Less data is easier to process, especially if some manual steps are required

• Starting with positions on a grid map
  • Improve your choice during the acquisition
  • Go back to a grid map after you processed a whole dataset

• Remove low-quality data in each step
  • Bad tilts, hard-to-align tomograms, bad particles etc.
  • Tools/scripts to facilitate analysis of the data
Robustness w.r.t. Sample

Immature HIV-1 CA-SP1 lattice

- Pixel size: 1.35Å
- Sample thickness: ~160nm
- Particles per tomogram: ~9 VLPs (~350 subtomograms per VLP)
- Symmetry: 6 fold
- Best reported resolution: 3.1Å

Human Nuclear Pore Complex (NPC)

- Pixel size: 3.35Å
- Sample thickness: ~450nm
- Particles per tomogram: 0-14 NPCs per tomogram
- Symmetry: 8 fold
- Best reported resolution: ~20Å
Robustness w.r.t. Sample

Immature HIV-1 CA-SP1 lattice  Human Nuclear Pore Complex (NPC)
# Robustness w.r.t. Sample

<table>
<thead>
<tr>
<th>HIV</th>
<th>NPC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tilt-Series Acquisition</strong></td>
<td>~40 min per TS +6 hours setup</td>
</tr>
<tr>
<td>~30 min per TS +5 hours setup</td>
<td><strong>Tilt-Series Preprocessing</strong></td>
</tr>
<tr>
<td>~1 h per TS</td>
<td>~1 h per TS</td>
</tr>
<tr>
<td>~20-60 min per TS Max 10 TS a day</td>
<td><strong>Tilt-Series Alignment</strong></td>
</tr>
<tr>
<td>~20-60 min per TS Max 10 TS a day</td>
<td>~90 min per reconstruction</td>
</tr>
<tr>
<td>~60 min per reconstruction</td>
<td><strong>Tomogram Reconstruction</strong></td>
</tr>
<tr>
<td>~90 min per reconstruction</td>
<td><strong>Particle Picking</strong></td>
</tr>
<tr>
<td>~15 min per tomogram</td>
<td>^15 min per tomogram Max 20 tomogram a day</td>
</tr>
<tr>
<td>Up to 1 week</td>
<td>Weeks</td>
</tr>
<tr>
<td>2-3 weeks</td>
<td>100 tomograms</td>
</tr>
<tr>
<td>8-10 weeks</td>
<td>Weeks</td>
</tr>
</tbody>
</table>
Robustness w.r.t. Sample

HIV

• Tilt-Series Alignment
  • Few fiducials but can be automatically tracked
  • Problem is the precision
  • Can be overcome with local alignments based on subtomogram positions (emClarity, Warp)

• Particle Picking
  • Manual picking using geometry is fast
  • Template matching works as well (emClarity, Warp?)

NPC

• Tilt-Series Alignment
  • Poor fiducial distribution
  • Problem is choosing “good” fiducials and track them (low SNR at high tilts)
  • Low SNR prevents (for now) successful use of local alignments approach

• Particle Picking
  • Manual picking is demanding and requires experience
  • Template matching does not work – easier to pick manually than clean the many false positive
  • Maybe deep-learning approaches might help here
SW with a “Complete” Pipeline

- **emClarity**
  - Very good results (HIV: 3.1Å)
  - Improvement of tomogram alignment based on subtomograms
  - Interesting classification method
  - Requires Imod and Chimera
  - Some parts still missing
  - Written in Matlab and uses GPU
  - Not flexible / modular
  - Hard-coded settings
  - So far rather user-unfriendly

- **Warp for SA**
  - Also very good results (HIV: 3.3Å)
  - Improvement of tomogram alignment based on subtomograms
  - Self-contained
  - Written in C#
  - GUI
  - Not explored: parameters, modularity
Summary

• Currently the processing is not robust w.r.t. samples
  • For difficult samples the automatization remains challenging

• Some samples can be processed efficiently and with little manual intervention
  • Optimize your sample as much as possible
  • Take your time during acquisition
    • Careful and experienced setup of positions
    • Prefer data quality over acquisition speed
  • Do not lose control over your data

• Share your data, parameters, and experience to help improve current SW and benchmark different approaches