Software developments at CNB-CSIC

Biocomputing Unit
Natl. Center of Biotechnology (CSIC)
Madrid
Scipion: a workflow manager
Large community involvement

48 developers

<table>
<thead>
<tr>
<th>Repositories</th>
<th>165</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>50K</td>
</tr>
<tr>
<td>Commits</td>
<td>13K</td>
</tr>
<tr>
<td>Issues</td>
<td>1K</td>
</tr>
<tr>
<td>Discussions</td>
<td>0</td>
</tr>
<tr>
<td>Packages</td>
<td>4</td>
</tr>
<tr>
<td>Marketplace</td>
<td>1</td>
</tr>
<tr>
<td>Topics</td>
<td>7</td>
</tr>
<tr>
<td>Wikis</td>
<td>184</td>
</tr>
<tr>
<td>Users</td>
<td>48</td>
</tr>
</tbody>
</table>
1. Automatic acquisition
2. JAFIS and minimal hysteresis
3. Stream processing and monitoring
4. Automated sample evaluation
5. Automation in facilities workflows
6. Distributed processing
7. Giving the data back to the user
8. Remote access
9. User machines
Grid squares and holes detection and classification

- Setup threshold streaming
- View images and classifications
- Launch Ptolemy Neuronal Networks
- Filter and selection of final squares and holes
- Manage coordinates

- Scipion Ptolemy-plugin SerialEM-Protocol
  - Manage microscope magnification
  - Launch acquisitions
  - Manage coordinates

- Scipion Ptolemy-plugin Ptolemy-Protocol

- Images data

- CryoEM microscopy
Each beam shift accumulate a hysteresis in the microscope lens, causing aberrations (AFIS). Stopping the acquisition to calibrate the lens is the current solution.

**JAFIS** software, available for SerialEM, bring all the parameters involve in the aberrations issue for each hole (h).

Find a path that minimize the number of changes of sign in those parameters, is carried out by a **genetic algorithm**. Results:

- The genetic algorithm found a path with only **4 sign changes**:
  
  \[6, 11, 9, 3, 5, 2, 8, 0, 10, 4, 1, 7\]

- Following a default spiral path would result in **9 sign changes**:
  
  \[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\]
Stream processing and monitoring

- Micro 1 - 200 kV FEI TALOS Arctica
- Micro 2 - 300 kV JEOL CryoARM
- Processing server 1
- Processing server n
- NFS shares

Scipion
On-the-fly processing
Stream processing and monitoring
Stream processing and monitoring
Automated sample evaluation

Objectives

1) Constant feedback on the **quality of acquisition**
2) **Automated intelligent decisions** to filter bad quality images

1) **Quality of acquisition**: protocols to monitor the acquisition by extracting quality measurements such as gain, dose analysis or beam tilt analysis.

   - **Movie gain protocol**
   - **Poisson count protocol**
   - **Tilt analysis protocol**

These protocols can be added to your workflows for an extra level of quality checks.
2) **Automated intelligent decisions**: Consensus protocols combine estimations where at least two different algorithms agree from the same input data. Helps to make more robust estimations and discard bad quality results.

- **Alignment consensus**
  - Shifts trajectory correlation between two global alignments

- **CTF consensus**
  - Discarded micrographs

- **Particle picking consensus**
  - Picking consensus options

These protocols can be added to your workflows for more robustness and to act as quality filters.
Automation in facilities workflows

Topaz: 7486 particles

Relion: 2477 particles

Gautomatch: 3311 particles

+Deep Consensus
+Micrograph Cleaner
Automation in facilities workflows

On-the-fly processing (streaming) with reduced or non-human interactions workflows:

Stage 1:
1. From movies to micrographs
   - Alignment options (motioncorr, flexAlign, relion, etc.)
2. CTF estimation
   - Different options (gctf, cistem, xmipp, etc)
   - CTF consensus recommended

Stage 2:
1. Manual Checkpoint 1
   - Manually or time approved starts the picking
2. Particle picking strategy
   - Different picking strategies based in normal picking and consensus, using 2D references or training new models.
3. Trigger data
   - 2D Classification with eg: n=1,000, n=5,000, n=10,000 number of particles

Stage 3:
1. Manual Checkpoint 2
   - Manually or time approved starts centering particles for the initial volume
2. Initial volume estimation
   - Initial volume (eman2, cryosparc, relion, xmipp3, etc)

Checkpoints are used to manually or time specific start Stage 2 and Stage 3. This is done to allow the user to check the results of previous stages before advancing the next one.
Distributed processing

Micro 2 - 300 kV JEOL CryoARM

Processing server 1
Movie alignment

Processing server 2
All other steps

+GPU Queues
Giving the data back to the user
Giving the data back to the user
Remote access: VPN+VNC
User machines
User machines
Contents

1. Automatic acquisition
2. JAFIS and minimal hysteresis
3. Stream processing and monitoring
4. Automated sample evaluation
5. Automation in facilities workflows
6. Distributed processing
7. Giving the data back to the user
8. Remote access
9. User machines
Thanks