Benchmarking of image registration methods for differently stained histological slides

Jiří Borovec† Arrate Muñoz-Barrutia† Jan Kybic∗

∗ Department of Cybernetics, Faculty of Electrical Engineering, CTU in Prague, Czechia
† Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
‡ Universidad Carlos III de Madrid, Leganés, Spain

jiri.borovec@fel.cvut.cz http://cmp.fel.cvut.cz/~borovj3/

Introduction

Image registration is a common task for many biomedical analysis applications. The present work focuses on the benchmarking of registration methods on differently stained histological slides. This is a challenging task due to the differences in the appearance model, the repetitive texture of the details and the large image size, between other issues. Our benchmarking data is composed of 416 image pairs at two different scales — average image diagonal 2k and 4k pixels. We compare eleven fully automatic registration methods covering the widely used similarity measures. For each method, the best parameter configuration is found and subsequently applied to all the image pairs. The performance of the algorithms is evaluated from several perspectives — the registrations (in)accuracy on manually annotated landmarks, the method robustness and its computation time.

Materials

Whole slide microscopy images:
- 32 sets of consecutive sections — breast tumor, lung tumor, rat kidney
- stained by a different dye — Cytokeratin, CC10, posSPC, H&E, Ki67, CD31, CN31, EH, PH, Podocin. Negative

Evaluation measures:
- Target Registration Error (TRE) is mean Euclidean distance in pixels between the positions of the landmarks in the reference and the transformed images normalized by the size of the image diagonal.
- Robustness is the ratio of cases where registration improved the initial relative TRE.
- Execution time is measured on a computer using a single CPU/thread.

Numerical results

<table>
<thead>
<tr>
<th>Method</th>
<th>TRE [%]</th>
<th>Execution time [s]</th>
<th>Labels</th>
<th>Similarity</th>
<th>Features</th>
<th>Criterion</th>
<th>Optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenCV</td>
<td>0.83</td>
<td>0.28</td>
<td>PosSPC</td>
<td>Linear</td>
<td>OpenCV</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
<tr>
<td>M</td>
<td>2.32</td>
<td>0.23</td>
<td>H&amp;E</td>
<td>Linear</td>
<td>M</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
<tr>
<td>S</td>
<td>0.45</td>
<td>0.21</td>
<td>CC10</td>
<td>Linear</td>
<td>S</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
<tr>
<td>Ants</td>
<td>0.92</td>
<td>0.26</td>
<td>Ki67</td>
<td>Linear</td>
<td>Ants</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
<tr>
<td>Surf</td>
<td>2.01</td>
<td>0.24</td>
<td>CD31</td>
<td>Linear</td>
<td>Surf</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
<tr>
<td>BUnwarpJ</td>
<td>1.0</td>
<td>0.23</td>
<td>EH</td>
<td>Linear</td>
<td>BUnwarpJ</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
<tr>
<td>RNIftyReg</td>
<td>0.94</td>
<td>0.22</td>
<td>PH</td>
<td>Linear</td>
<td>RNIftyReg</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
<tr>
<td>Robust</td>
<td>0.5</td>
<td>0.21</td>
<td>Podocin</td>
<td>Linear</td>
<td>Robust</td>
<td>Surf &amp; MSER</td>
<td>RANSAC</td>
</tr>
</tbody>
</table>

The configuration of registration methods is presented in the following table.

Conclusion

- Selected registration methods cover the most common similarity criteria.
- Execution time of some methods is reasonably good (suitable for practical usage).
- Performance as measured show that the task is still not fully solved.
- In many cases the linear transformation performs better than the elastic.


References