Segmentation, tracking and lineage analysis of yeast cells in bright field microscopy images

Raffaele La Brocca¹, Filippo Menolascina², Diego di Bernardo¹,², Carlo Sansone¹

¹Dipartimento di Informatica e Sistemistica, Università degli Studi di Napoli Federico II
²Systems and Synthetic Biology Laboratory, Telethon Institute of Genetics and Medicine
Time-lapse microscopy

- Time lapse microscopy images are used by biologists to study gene circuit dynamics in single cells.
- Several applications in quantitative biology (e.g. Systems biology) require cells to be engineered to express fluorescent protein reporters allowing to follow the dynamics of a gene of interest.
- Microscopy images can be used to obtain quantitative measures of the protein concentration levels over time in each cell through image processing routine.
  - Bright field images are used to track cell movements over time and construct lineage trees reporting mother-daughter relationships between cells
  - Fluorescent field images are used to evaluate the expression level dynamics in every tracked cell.
Time-lapse microscopy

Bright-field image

fluorescent-field image
Cell segmentation and tracking

- Humans are good at cell identification, tracking and division detection in image sequences, but manual analysis is a tedious, time-consuming and error-prone task.
- Automatic cell segmentation and tracking are complex tasks whose success usually depends on strong assumptions.
- Many solutions had been developed in this field
  - watershed and active contours methods
    - need consistent effort to adapt to the specific characteristics of the experiments of interest.
- Existing software, such as CellTracer and CellProfiler, have been found to be heavily dependent on parameters' choice and to possibly perform poorly on new data unless a long search for the optimal parameters' set is carried out
Our aim

- To develop a solution to yeast cell tracking and cell division detection, which must be robust to experimental variability
- The implemented solution must be used by biologists with little knowledge in the field of image processing
Segmentation

Edge points can be detected by the evaluation of the magnitude of the gradient calculated in each point of the image.
Segmentation

For computational reasons, CHT is applied only to a set of suitably chosen image sub-regions.
False positive detection

• The number of false positives is quite high

• Two proposed approaches:
  • Threshold-based
  • Machine learning-based
A false positive occurs if the maximum of the histogram is greater than 3
False positive detection

A machine-learning based approach (by using Decision Trees)

Used features:
- the mean intensity value of the extracted subregion
- the proportion of the pixels in the convex hull containing the subregion that are also in the subregion (solidity)
- the displacement from the centroid specified by the object to the center of the subregion, divided by the radius specified by the object
- the proportion of the pixels in the region that are also in the subregion
- the values of the histogram with ten bins of the region represented by the object (intensity features)
Tracking and lineage analysis

- Tracking can be performed by finding the correspondences between the objects detected in two consecutive frames by considering a minimum cost configuration.
- This association cost increases as long as the displacement between the centroids of the corresponding objects.
- The minimum cost configuration can be determined by setting up and solving a linear programming problem (LPP).
Tracking and lineage analysis

The equality constraints impose that each object detected in frame $t+1$ have to correspond to one and only one object detected in frame $t$. Each object detected in frame $t$, indeed, can correspond to one, many or no object detected in frame $t+1$.

Let $p^t = \{p_1^t, \ldots, p_n^t\}$ and $p^{t+1} = \{p_1^{t+1}, \ldots, p_m^{t+1}\}$.

$$
C = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0
\end{pmatrix}
$$

$$
\phi_{j,k} = \|p_j^t - p_k^{t+1}\|^2
$$

Subject to:

$$
\sum_{j=1}^n c_{j,k} = 1 \quad k = 1, \ldots, m
$$

$$
\sum_{j=1}^n c_{j,k} \geq 0 \quad j = 1, \ldots, n \quad k = 1, \ldots, m
$$

The equality constraints impose that each object detected in frame $t+1$ have to correspond to one and only one object detected in frame $t$. Each object detected in frame $t$, indeed, can correspond to one, many or no object detected in frame $t+1$. 
Lineage trees

Each node in a tree represents a cell and each edge a mother-daughter relation between the cells represented by the connected nodes.

By using the software we developed, the user can visualize the trajectory performed by the corresponding cell by clicking on a node.
Tracking Analysis
Performance Evaluation

- Segmentation, tracking and lineage analysis
- We developed a tool for generating reference data

CellProfiler for manual segmentation

GUI for manual tracking and lineage analysis
Performance Evaluation

\[ o_{rif}^t \equiv \{ o_{rif,1}^t, \ldots, o_{rif,n}^t \} \]
\[ o^t \equiv \{ o_1^t, \ldots, o_m^t \} \]
\[ \min \sum_{j=1}^{n} \sum_{k=1}^{m} \phi_{j,k} c_{j,k} \]
\[ s.p. \]
\[ \sum_{j=1}^{n} c_{j,k} = 1 \quad k = 1, \ldots, m \]
\[ c_{j,k} \geq 0 \quad j = 1, \ldots, n \quad k = 1, \ldots, m \]
\[ \phi_{j,k} = \| p_{rif,j}^t - p_k \|^2 \]

\[ c \quad \text{number of correspondences} \]

\[ \text{precision} = \frac{c}{m} \]
\[ \text{recall} = \frac{c}{n} \]
\[ F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]
\[ \text{acc}_{i,j} = \frac{A \left( r(o_j^t) \cap r(o_{rif,k}^t) \right)}{A \left( r(o_j^t) \cup r(o_{rif,k}^t) \right)} \]
Performance evaluation

\[ \text{tr}_{\text{rif}} \equiv \{ \text{tr}_{\text{rif},1}, \ldots, \text{tr}_{\text{rif},n} \} \]
\[ \text{tr} \equiv \{ \text{tr}_{1}, \ldots, \text{tr}_{m} \} \]
\[ \Phi_{j,k} = \| \text{tr}_{\text{rif},j} - \text{tr}_{k} \| + \frac{\text{len}(\text{tr}_{\text{rif},j}) - \text{len}(\text{tr}_{k})}{\text{ov}(\text{tr}_{\text{rif},j}, \text{tr}_{k})} \times 100 \]
\[ \min_{j=1}^{n} \sum_{k=1}^{m} \Phi_{j,k} \sum_{j=1}^{n} c_{j,k} = 1 \quad k = 1, \ldots, m \]
\[ c_{j,k} \geq 0 \quad j = 1, \ldots, n \quad k = 1, \ldots, m \]
## Experimental results (1/3)

<table>
<thead>
<tr>
<th>no.</th>
<th>SDF</th>
<th>EDF</th>
<th>Min x</th>
<th>Max x</th>
<th>Med. x</th>
<th>Std. x</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0.88</td>
<td>2.03</td>
<td>1.36</td>
<td>0.27</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
<td>1.47</td>
<td>0.67</td>
<td>0.44</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.71</td>
<td>1.81</td>
<td>1.31</td>
<td>0.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>no.</th>
<th>Min y</th>
<th>Max y</th>
<th>Med. y</th>
<th>Std. y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
<td>1.37</td>
<td>0.73</td>
<td>0.34</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
<td>1.52</td>
<td>1.08</td>
<td>0.27</td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>1.98</td>
<td>1.21</td>
<td>0.46</td>
</tr>
</tbody>
</table>
The misclassification rate evaluated with a leave-one-out cross validation was 0.1.

<table>
<thead>
<tr>
<th>True positives</th>
<th>False positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>2879</td>
</tr>
<tr>
<td>False positives</td>
<td>100</td>
</tr>
</tbody>
</table>

The method has been then tested with reference to data coming from two image sets, parts of two independent experiments.
- The first image set is a 50 frames sequence from one of our experiments, where a high cellular replication rate is observed.
- The second one is a 50 frames sequence extracted from the sample set available in *CellTracer* website.
Table 2 – Segmentation method performance indexes. Performance indexes evaluated with no false positive elimination are reported in brackets.

<table>
<thead>
<tr>
<th>Precision</th>
<th>Recall</th>
<th>F</th>
<th>Min acc.</th>
<th>Max acc.</th>
<th>Mean acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.92 (0.53)</td>
<td>0.88 (0.92)</td>
<td>0.90 (0.67)</td>
<td>0.18 (0.12)</td>
<td>0.95 (0.94)</td>
<td>0.74 (0.73)</td>
</tr>
</tbody>
</table>

Table 3 – Tracking method performance indexes. Performance indexes evaluated with no false positive elimination are reported in brackets.

<table>
<thead>
<tr>
<th>Tracking</th>
<th>Lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>0.87 (0.42)</td>
<td>0.80 (0.95)</td>
</tr>
</tbody>
</table>

Table 4 – A comparative of the performance indexes evaluated for CellTracer and our solution.

<table>
<thead>
<tr>
<th>Method</th>
<th>Segmentation</th>
<th>Tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>CellTracer</td>
<td>0.98</td>
<td>0.82</td>
</tr>
<tr>
<td>This</td>
<td>0.89</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Conclusions and future works

- In this paper a robust method for yeast cell segmentation, tracking and lineage analysis is presented.
- A reliable performance evaluation method is also introduced.
- The results of the comparative analysis we carried out confirms the competitive performance of our approach, making it a good choice for biologists looking for simple and out-of-the-box solutions.

These results encourage further improvements in segmentation accuracy and mother-daughter relationships detections.