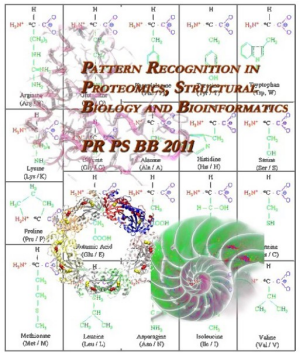



1st International Workshop on Pattern Recognition in
Proteomics, Structural Biology and Bioinformatics - PR PS BB 2011

Ravenna, September 13, 2011

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

Raffaele La Brocca[1], Filippo Menolascina[2], Diego di Bernardo[1,2], **Carlo Sansone**[1]

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

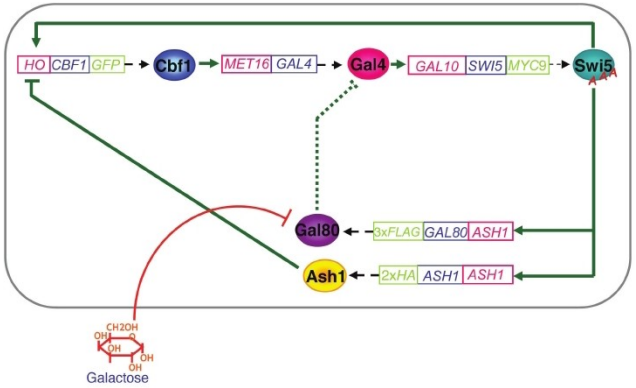
UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II

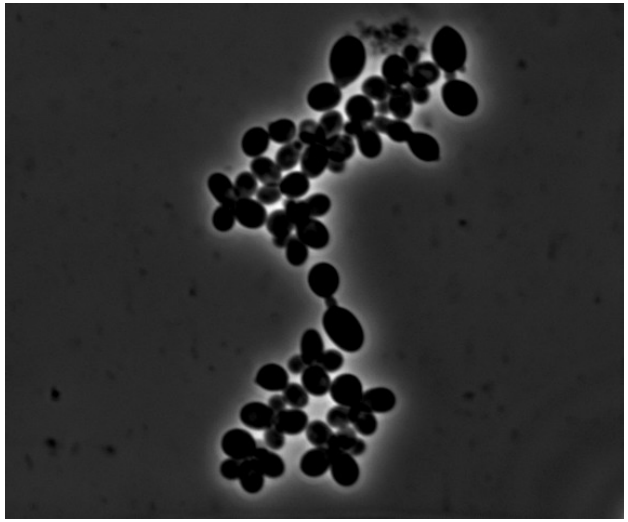
igem TELETHON INSTITUTE OF GENETICS AND MEDICINE

# Time-lapse microscopy

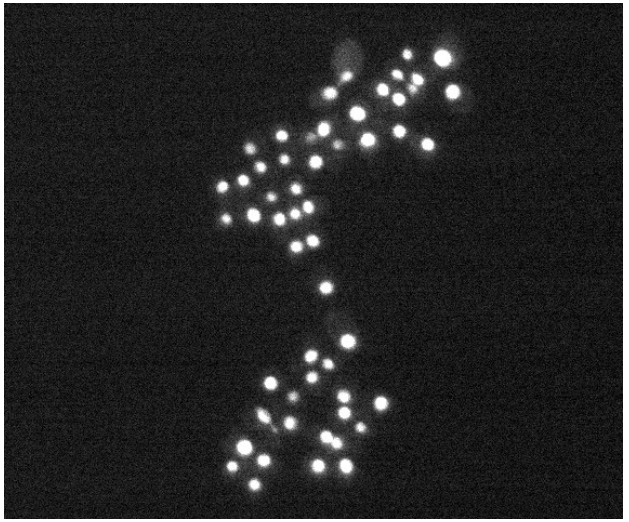
- Time lapse microscopy images are used by biologists to study gene circuit dynamics in single cells.
- Several applications in quantitative biology (e.g. Systems biology) require cells to be engineered to express fluorescent protein reporters allowing to follow the dynamics of a gene of interest.
- 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

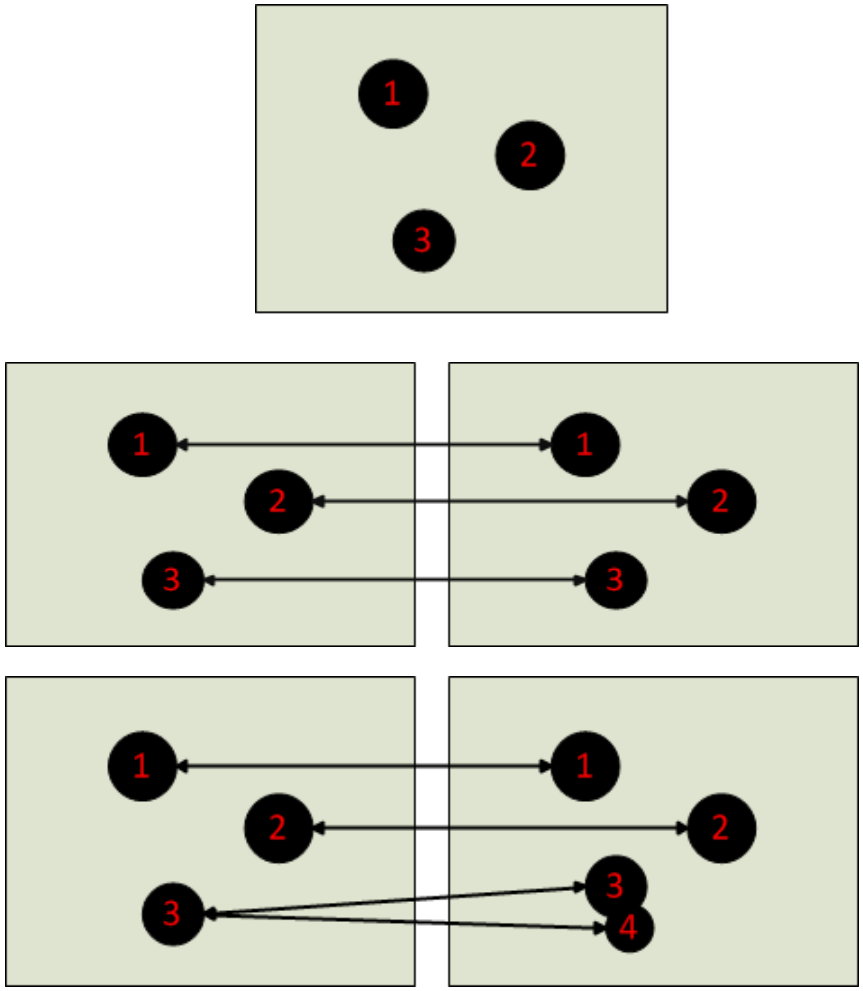
- Humans are good at cell identification, tracking and division detection in image sequences, but manual analysis is a tedious, time-consuming and error-prone task.
- Automatic cell segmentation and tracking are complex tasks whose success usually depends on strong assumptions.
- Many solutions had been 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

**Cell segmentation**

**Cell tracking**

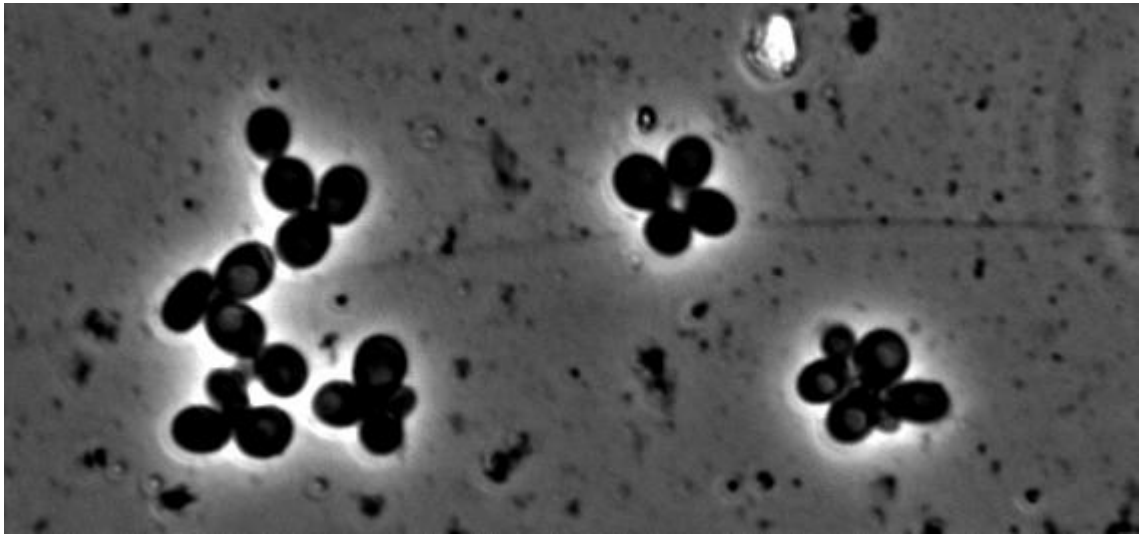
**Lineage analysis**

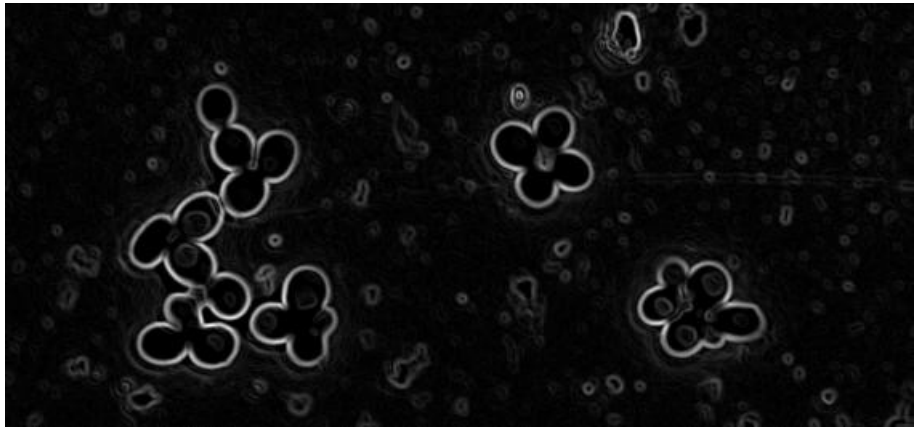



- To develop a solution to yeast cell tracking and cell division detection, which must be robust to experimental variability

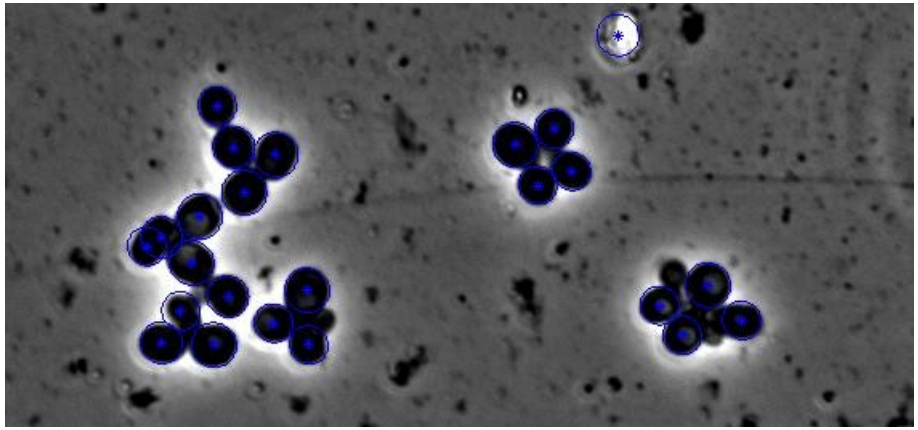
- The implemented solution must be used by biologist with little knowledge in the field of image processing

# Segmentation





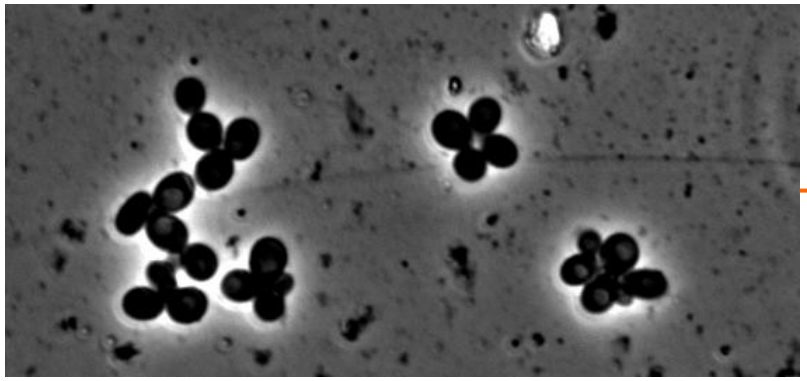
Edge points can be detected by the evaluation of the magnitude of thegradient calculated in each point of the image



Circular Hough-Transform

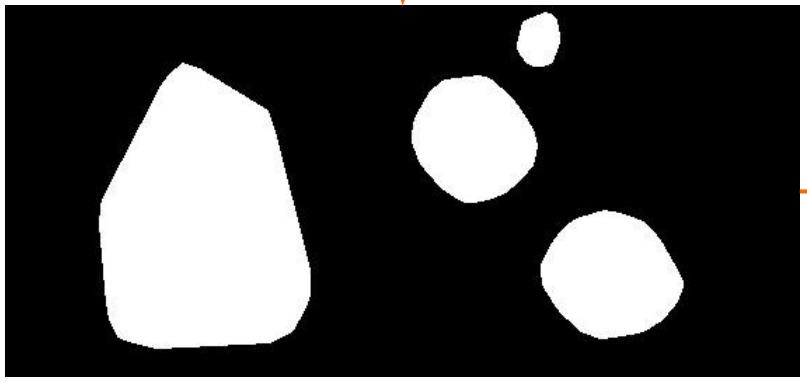
# Segmentation

**For computational reasons, CHT is applied only to a set of suitably chosen image sub-regions**

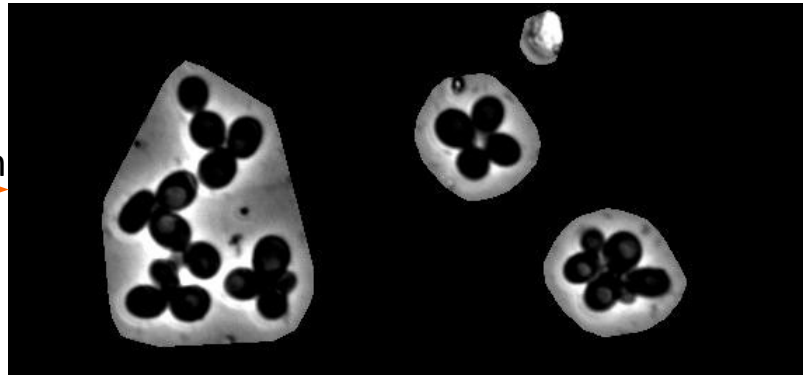


thresholding

morphological operations + convex hull of the connected components



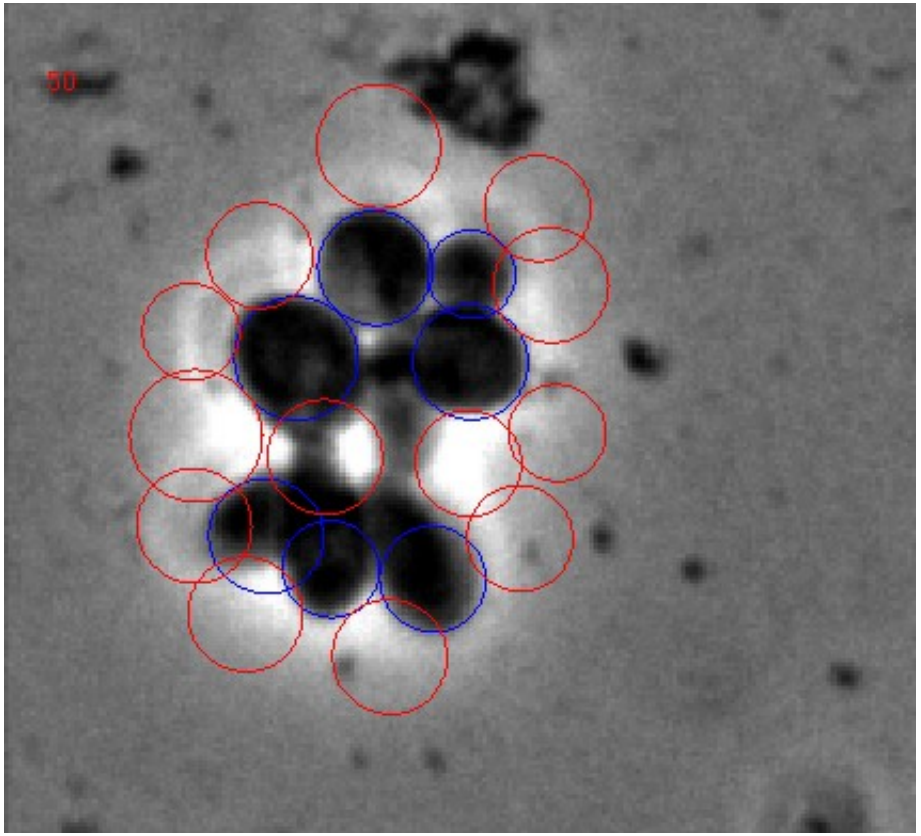
Region selection



# False positive detection

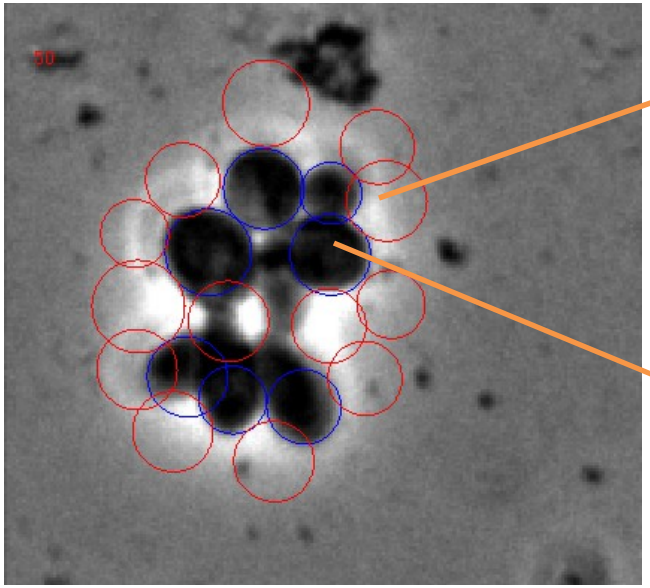
- The number of false positives is quite high

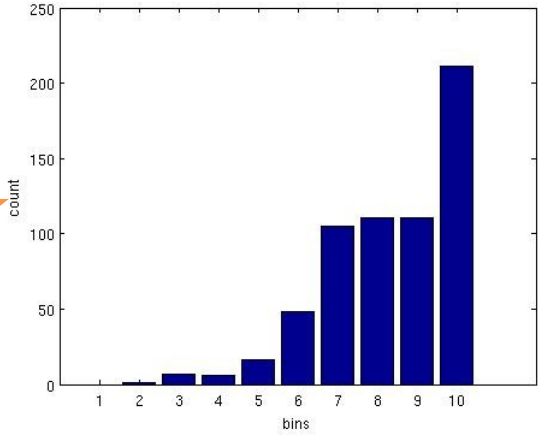
- Two proposed approaches:
    - Threshold-based
    - Machine learning-based

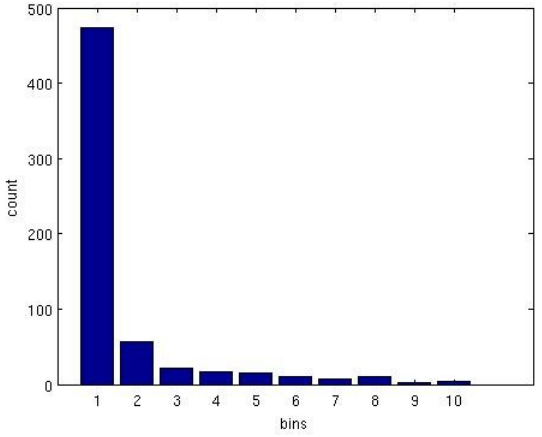


# False positive detection

**Fixed Threshold**





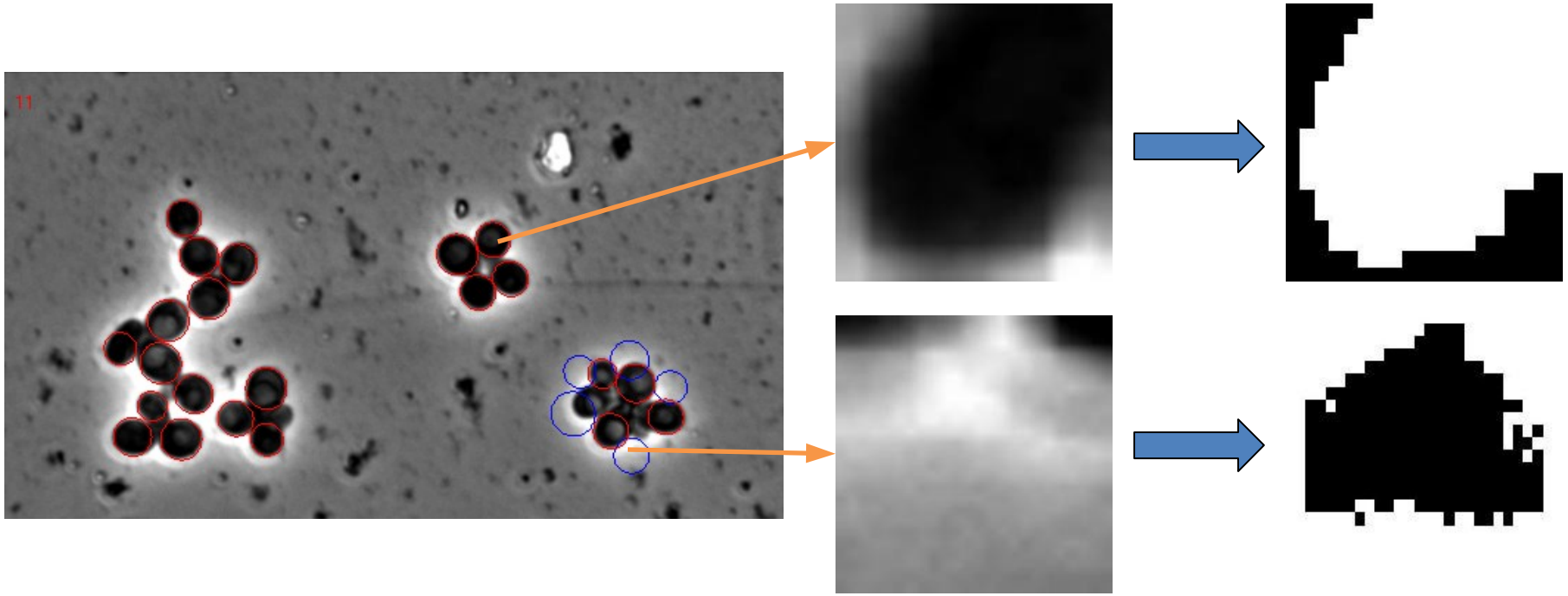

**A false positive occurs if the maximum of the histogram is greater than 3**

# False positive detection

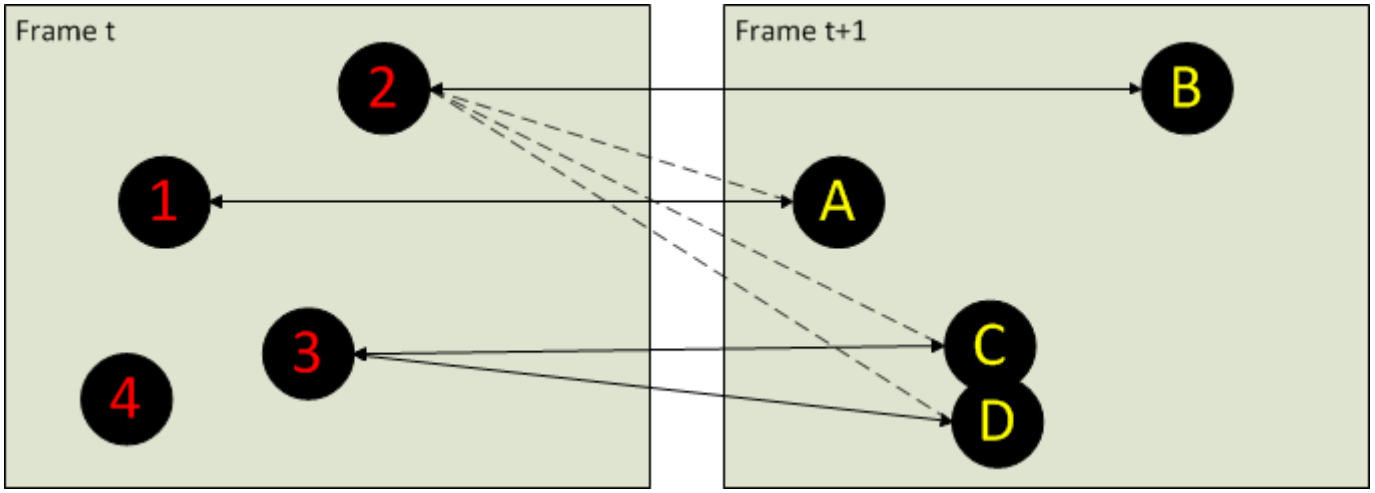
**A machine-learning based approach (by using Decision Trees)**

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)

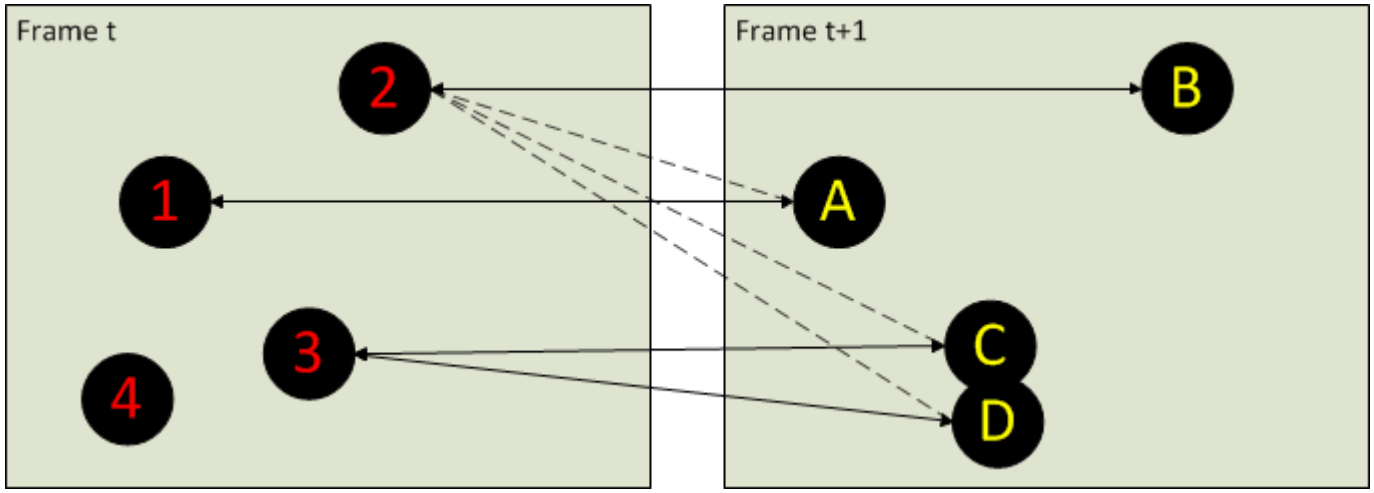


# Tracking and lineage analysis




- 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




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

$$p^t = \{p_1^t, \dots, p_n^t\}$$

$$p^{t+1} = \{p_1^{t+1}, \dots, p_m^{t+1}\}$$

$$min \sum_{j=1}^{n} \sum_{k=1}^{m} \phi_{j,k} c_{j,k}$$

$$s.p.$$

$$\sum_{j=1}^{n} c_{j,k} = 1 \qquad k = 1,..,m$$

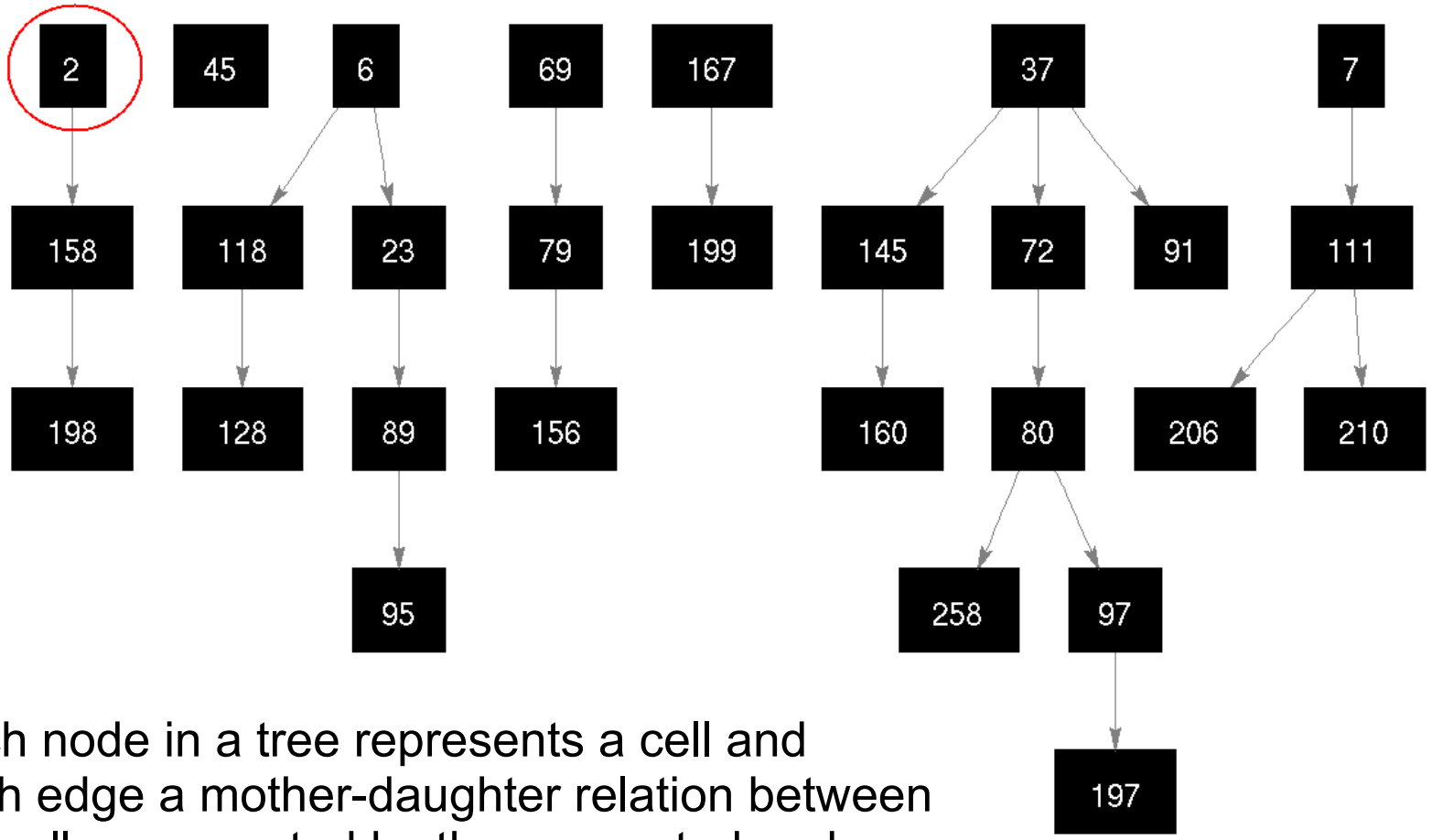
$$c_{j,k} \geq 0 \qquad j = 1,..,n \quad k = 1,..,m$$

$$C = \begin{pmatrix} c_{1,1} & \dots & c_{1,m} \\ \dots & c_{j,k} & \dots \\ c_{n,1} & \dots & c_{n,m} \end{pmatrix}$$

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

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.

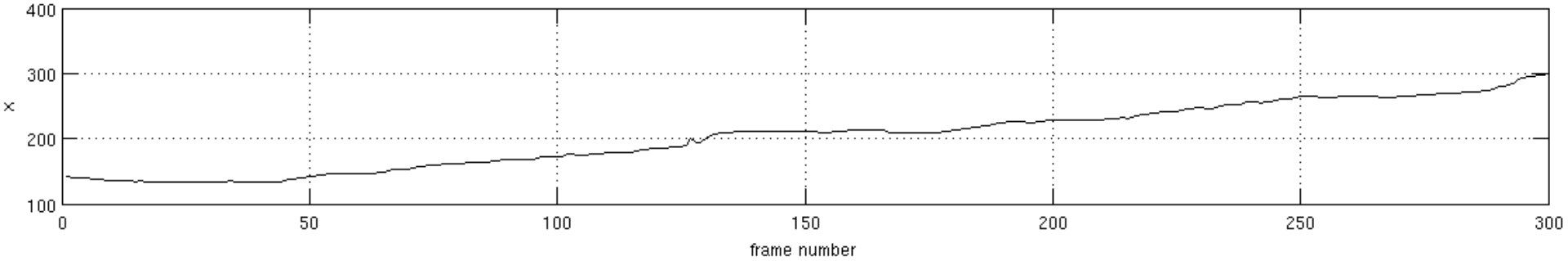
# Lineage trees

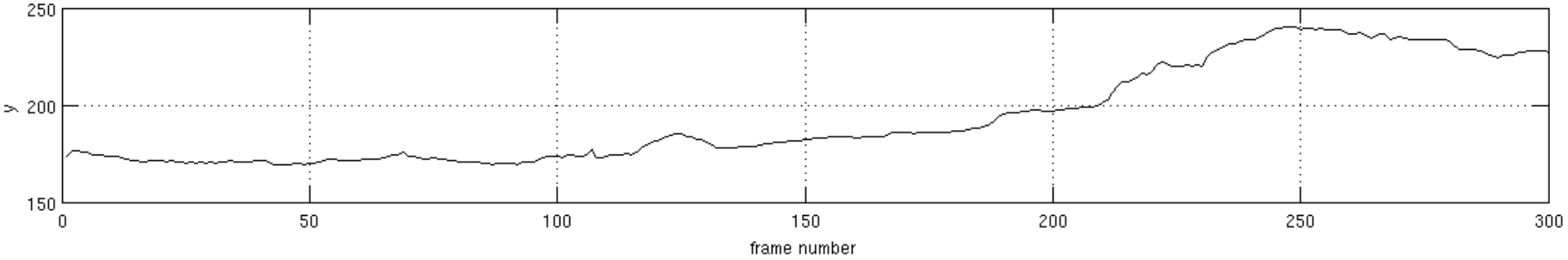


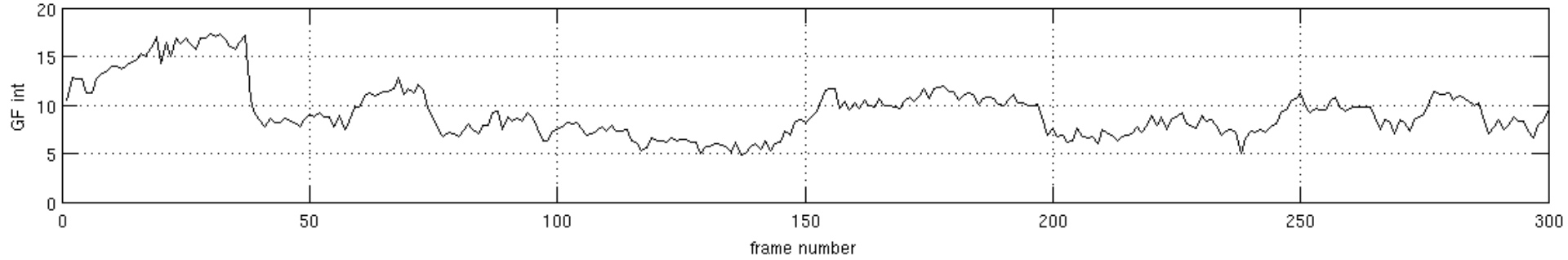

Each node in a tree represents a cell and each edge a mother-daughter relation between the cellsrepresented by the connected nodes.

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

# Tracking Analysis

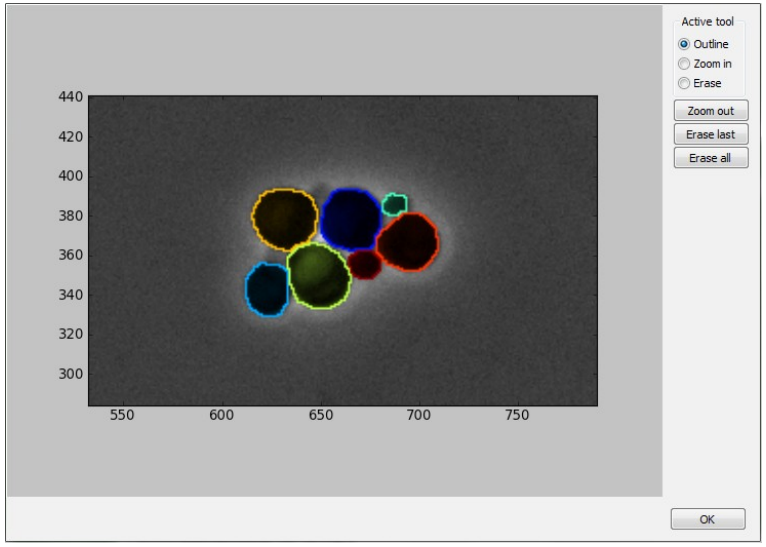


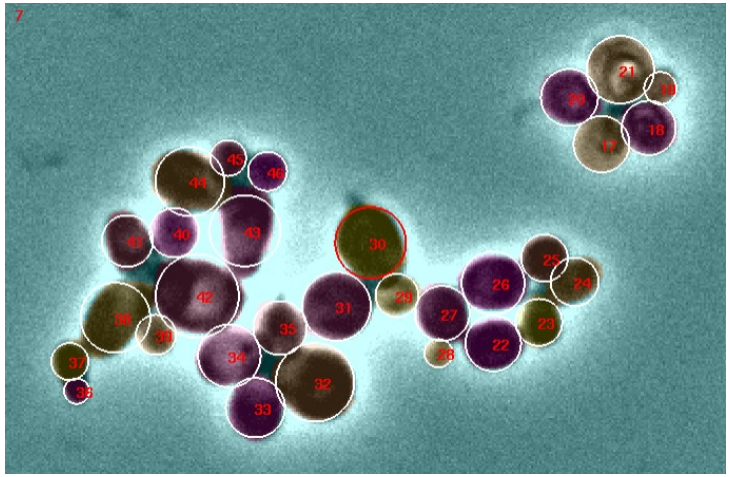
# Performance Evaluation

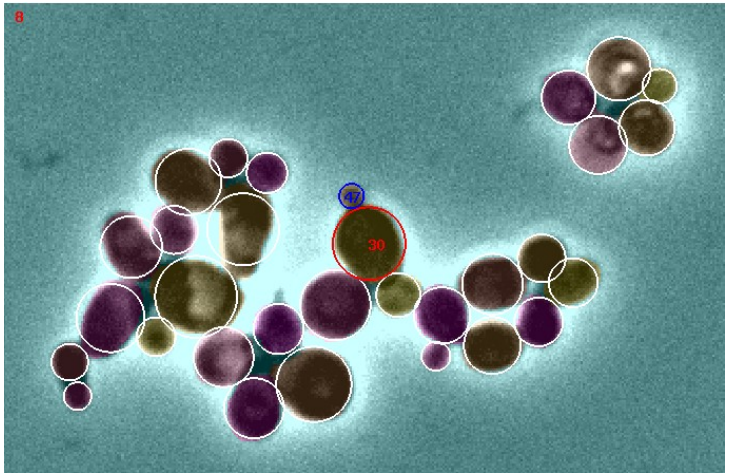
- Segmentation, tracking and lineage analysis
- We developed a tool for generating reference data



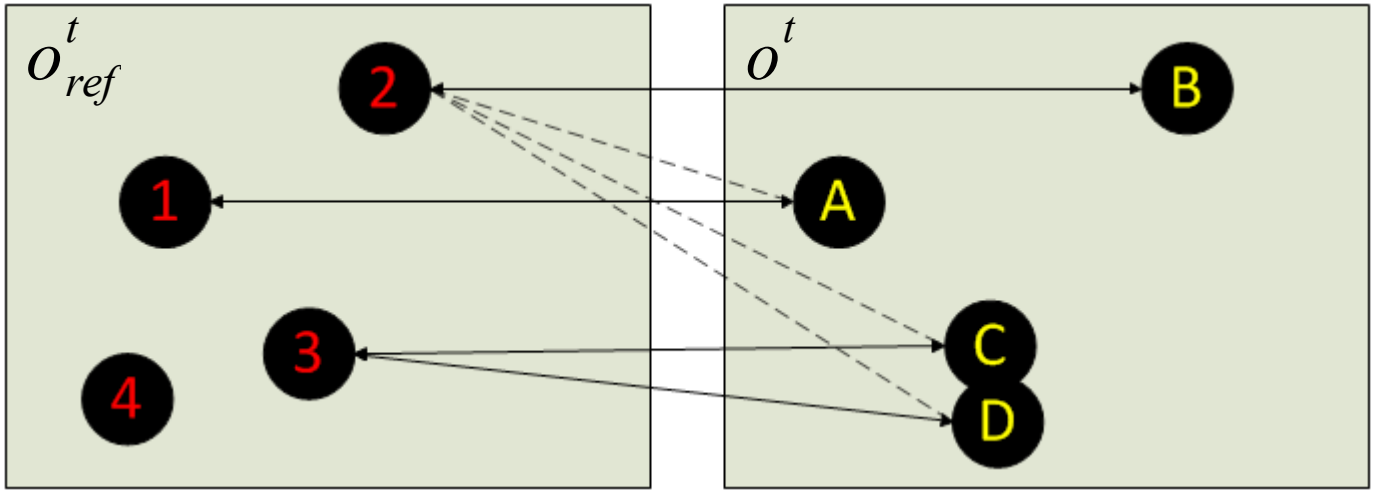

**CellProfiler for manual segmentation**

**GUI for manual tracking and lineage analysis**



# Performance Evaluation




$$o^t_{rif} \equiv \{ o^t_{rif,1}, \dots, o^t_{rif,n} \}$$

$$o^t \equiv \{ o^t_1, \dots, o^t_m \}$$

$$min \sum_{j=1}^{n} \sum_{k=1}^{m} \phi_{j,k} c_{j,k}$$

$$s.p.$$

$$\sum_{j=1}^{n} c_{j,k} = 1 \qquad k = 1,..,m$$

$$c_{j,k} \geq 0 \qquad j = 1,..,n \quad k = 1,..,m$$

$$\phi_{j,k} = \| p^t_{rif,j} - p_k \|^2$$

$c$ number of correspondences

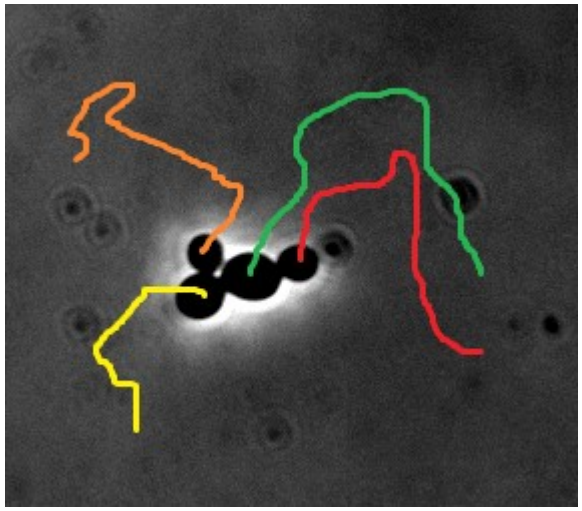
$$precision = \frac{c}{m}$$

$$recall = \frac{c}{n}$$

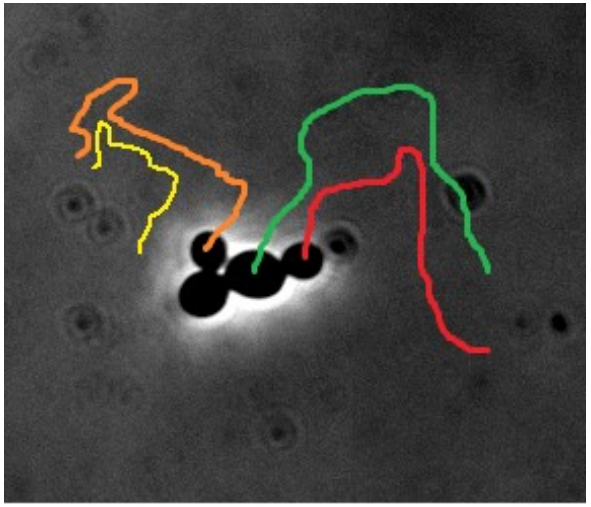
$$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

$$acc_{i,j} = \frac{A( r(o^t_j) \cap r(o^t_{rif,k}) )}{A( r(o^t_j) \cup r(o^t_{rif,j}) )}$$

# Performance evaluation



$tr_{rif}$



$tr$

$$tr_{rif} \equiv \{ tr_{rif,1}, \ldots, tr_{rif,n} \}$$

$$tr \equiv \{ tr_1, \ldots, tr_m \}$$

$$\phi_{j,k} = \| tr_{rif,j} - tr_k \| + \frac{|len(tr_{rif,j}) - len(tr_k)|}{ov(tr_{rif,j}, tr_k)} * 100$$

$$min \sum_{j=1}^{n} \sum_{k=1}^{m} \phi_{j,k} c_{j,k}$$

$$s.p.$$

$$\sum_{j=1}^{n} c_{j,k} = 1 \qquad k = 1,..,m$$

$$c_{j,k} \geq 0 \qquad j = 1,..,n \quad k = 1,..,m$$

# Experimental results (1/3)

| *no.* | *SDF* | *EDF* | *Min x* | *Max x* | *Med. x* | *Std. x* |
|---|---|---|---|---|---|---|
| 1 | 4 | 0 | 0.88 | 2.03 | 1.36 | 0.27 |
| 2 | 0 | 0 | 0.04 | 1.47 | 0.67 | 0.44 |
| 3 | 0 | 0 | 0.71 | 1.81 | 1.31 | 0.28 |
| *no.* | | | *Min y* | *Max y* | *Med. y* | *Std. y* |
| 1 | | | 0.20 | 1.37 | 0.73 | 0.34 |
| 2 | | | 0.45 | 1.52 | 1.08 | 0.27 |
| 3 | | | 0.33 | 1.98 | 1.21 | 0.46 |

# Experimental results (2/3)

**Table 1** – Confusion matrix of the trained classifier, evaluated on the training set.

| | True positives | False positives |
|---|---|---|
| **True positives** | 2879 | 76 |
| **False positives** | 100 | 3811 |

The misclassification rate evaluated with a leave-one-out cross validation was 0.1.

• 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

# Experimental results (3/3)

**Table 2** – Segmentation method performance indexes. Performance indexes evaluated with no false positive elimination are reported in brackets.

| *Precision* | *Recall* | *F* | Min acc. | Max acc. | Mean acc. |
|---|---|---|---|---|---|
| 0.92 (0.53) | 0.88 (0.92) | 0.90 (0.67) | 0.18 (0.12) | 0.95 (0.94) | 0.74 (0.73) |

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

| Tracking | | | Lineage | | |
|---|---|---|---|---|---|
| *Precision* | *Recall* | *F* | *Precision* | *Recall* | *F* |
| 0.87 (0.42) | 0.80 (0.95) | 0.83 (0.58) | 0.21 (0.27) | 0.43 (0.3) | 0.28 (0.33) |

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

| Method | Segmentation | | | Tracking | | |
|---|---|---|---|---|---|---|
| | *Precision* | *Recall* | *F* | *Precision* | *Recall* | *F* |
| *CellTracer* | 0.98 | 0.82 | 0.89 | 0.80 | 0.80 | 0.80 |
| This | 0.89 | 0.92 | 0.90 | 0.83 | 0.95 | 0.89 |

# 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.