Supplementary Material for “SCIP: A Single Cell Image Processor toolbox”

Leonardo Martins, Ramakanth Neeli-Venkata, Samuel M.D. Oliveira, Antti Häkkinen, Andre S. Ribeiro and José M. Fonseca

# Methods implemented in SCIP

# 1.1 Cell Segmentation

Two segmentation methods are available in SCIP. The first is based on ‘CellAging’ (Häkkinen *et al*., 2013) and ‘iCellFusion’ (Santinha, Martins, *et al*., 2015). Over segmentation is corrected using CART (Classification and Regression Trees Machine Learning algorithm) (Breiman *et al*., 1984), previously trained in various morphological images (Queimadelas *et al*., 2012). The second is based on applying a median filter to reduce noise while preserving cell edges and a multilevel Otsu threshold (Otsu, 1979). Under segmentation is corrected with by combining the Watershed Algorithm and the Distance Transform and finally by dividing the convex hulls of the objects. After the automatic segmentation is completed, it is possible to manually correct the results using the strategy developed in iCellFusion’ (Santinha, Martins, *et al.*, 2015)(Häkkinen *et al*., 2013). In this case, white borders correspond to manually added or modified cells, while blue borders represent automatically segmented cells (Figure 1).



Figure 1 - Example of an on-going process of manual segmentation correction; (A) using a free hand drawing function; (B) using a predefined polygonal shape; (C) end results of manual (white) and automatic segmentation (blue).

Overall, the cell segmentation algorithms provide information on the external borders of the cell, allowing the calculation of morphological data at a single-cell level, such as cell size and, consequently, cell growth, and enable the visualization of cell division events.

# 1.2 Image Alignment

After cell segmentation, the phase contrast images can be aligned to fluorescence images (typically called functional images), allowing the integration of morphological and functional features. SCIP automatic alignment uses the methodology proposed in CellAging (Häkkinen et al., 2013). In this, images captured by different cameras with distinct fields of view are automatically aligned by a 2-D affine geometric transformation (Sonka et al., 2014). If the images are captured by the same camera sensor, only translation and rotation are required to compensate for small drifts (Figure 2).



Figure 2 – Example of alignment between morphological segmentation and functional images.

# 1.3 Gaussian Segmentation algorithm for sub-cellular structures

SCIP allows the use of the Gaussian Segmentation algorithm modified for DAPI-stained nucleoids, which is described in (Santinha, Mora, *et al.*, 2015). This algorithm has already been used to study, e.g. how protein aggregates are excluded from the midcell position by the nucleoid and how that process functioned under sub-optimal temperatures (Oliveira *et al.*, 2016).

# 1.4 ‘TreshMorph’ Segmentation algorithm for sub-cellular structures

SCIP includes a new algorithm, ‘TreshMorph’, for automatic Morphological segmentation of sub-cellular structures. To locate the boundaries of the structures of interest inside each cell (e.g. nucleoids), the first step is the selection of a threshold to separate the objects from the cellular background. This threshold value is calculated separately for each cell (previously isolated by the cell segmentation algorithm).

Three thresholding methods are available. One is the Global Otsu's image threshold method (Otsu, 1979), which minimizes the intraclass variance of black and white pixels separating the structures with high fluorescence from the background. Another is the Multilevel image threshold which allows the selection of one threshold out of a specific number of thresholds using the Multilevel Otsu’s method (Otsu, 1979), allowing the algorithm to adapt to higher intensity variations inside each structure of interest. The last method is based on the mean and standard deviation of the pixels intensity inside each cell, in that it uses these values as the threshold to produce a separation between the structure of interest and the background (which is of value especially when their contrast ratios are high).

Next, the morphological ‘majority’ operation is applied to the obtained images (binary masks) of the structures based on the threshold, which sets a pixel to 1 if 5 or more pixels in its 3-by-3 neighborhood are 1s and to 0 otherwise. This operation removes small holes and smooths the structures.

The next step is the application of a ‘closing’ or ‘opening’ morphological operation to the resulting masks. The ‘closing’ operation performs a dilation, followed by an erosion, using a 3x3 structuring element. This slightly enlarges and smooths the boundaries of the structures, while removing small holes in the mask. The ‘opening’ operation performs an erosion followed by a dilation using the same 3x3 structuring element. This produces a small reduction and smoothing to the structures, while removing small objects in the mask. In the examples presented, the ‘closing’ operation resulted in higher statistical scores in images acquired with confocal microscopy, while the ‘opening’ operation had higher scores with images acquired with Highly Inclined and Laminated Optical (HILO) microscopy.

Finally, all objects smaller than X pixels can be removed (X is user selectable). In the presented examples, X=20 was found to be the most adequate for all structures.

Figure 3 shows examples of Nucleoid Segmentation using ‘TreshMorph’ with the Otsu's Global threshold. Figure 4 shows examples of FtsZ rings segmentation, using ‘TreshMorph’ with the multilevel Otsu's threshold. Figure 5 shows the segmentation of MinD proteins with ‘TreshMorph’ with a global threshold value based on the mean fluorescence intensity of each cell.

These structure segmentation algorithms (Gaussian and ‘TreshMorph’) provide morphological information on sub-cellular structures, such as the number of detected structures (in each cell), the location of the geometrical center of each structure, the distance of this center to mid-cell, as well as the major axis, minor axis and eccentricities of each structure. Temporal Features (e.g. mean and the variability of the period of oscillation) can also be extracted for structures with an oscillatory behaviour, such as the MinD proteins. These features can thus provide useful information on the spatial and temporal functions of each specific structure.



Figure 3 - Examples of segmentation of Nucleoids (in red): (Left) no Segmentation, (Right) Morphological Segmentation.



Figure 4 - Examples of segmentation of FtsZ proteins: (Left) no Segmentation, (Right) ‘TreshMorph’ Segmentation.



Figure 5 - Examples of segmentation of MinD proteins (in red): (Left) with no segmentation (Right) with ‘TreshMorph’ segmentation.

# 1.5 Detection of Inclusion Bodies from Phase-contrast images

Inclusion bodies can be detected by SCIP from segmented phase-contrast images (Figure 6A). To detect the presence of inclusion bodies, SCIP uses the GPL algorithm (Mora et al., 2011) (see Section ‘Gaussian Segmentation algorithm’), that creates seeds based on the image intensity gradient (yellow and red squares in Figure 6B). Inclusion bodies can then be detected using either Gaussian Segmentation (Figure 6C) or TreshMorph’ Segmentation (Figure 6D).

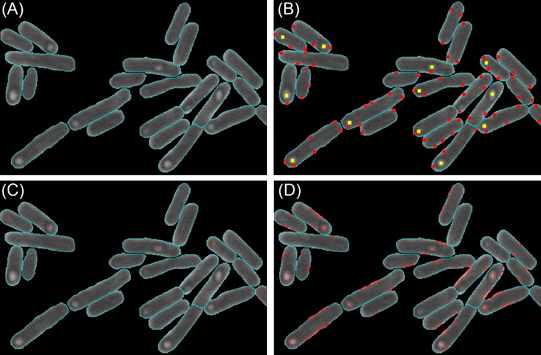


Figure 6 - Examples of segmentation of inclusion bodies. (A) No segmentation (B) Example of seed removal based on seeds in the border of the cell and seeds in the middle of the cell with no inclusion bodies (C) Gaussian Segmentation of Inclusion Bodies (D) ‘TreshMorph’ Segmentation of Inclusion Bodies, using a three level Otsu's threshold.

To remove false positive seed detections three methods are used. The first is based on the Euclidean distance between the seed center and the cell border closest pixel. If this distance is less than 5 pixels, that seed is removed (transforming yellow squares into red squares, as seen in Figure 6B). The second method removes seeds with a GPL ‘*path minimum amplitude*’ of zero. This value indicates that the seed is linked to the background independently of its position in the cell. Finally, we subtract the GPL ‘*path maximum amplitude*’ with the GPL ‘*path minimum amplitude*’ of the seed. If the subtraction is lower than a pre-defined threshold (for our images, 20 was found to be adequate), the seed is removed.

The algorithm for detection of inclusion bodies can be used to e.g. automatically remove cells with inclusion bodies from the analysis, which is commonly done (manually) in these studies (Neeli-Venkata *et al*., 2016).

# 1.6 Spot Detection algorithms

Spot detection has been implemented using all the methods used in ‘CellAging’ (H*ä*kkinen *et al.*, 2013) and ‘iCellFusion’ (Santinha, Martins, *et al.*, 2015). In all, three methods, differing in the filters used (Median, Low-Pass and Gaussian) were implemented. The parameters for each filter, as described in ‘CellAging’ (H*ä*kkinen *et al.*, 2013) and ‘iCellFusion’ (Santinha, Martins, *et al.*, 2015), are based on the intensity values inside the cell, in order to select the threshold values to detect fluorescent spots.

The algorithm removes spots exceeding the maximum accepted area and spots smaller than the minimum accepted area. This can be used to calculate the ‘integer-valued absolute number’ of RNA molecules (e.g. when tagged with MS2-GFP). This calculation is performed by subtracting the MS2-GFP “background” (mean pixel intensity of a cell, multiplied by the spot’s total number of pixels) from the spot’s total intensity and dividing it by a constant that represents the average intensity of one RNA (Golding and Cox, 2004)(H*ä*kkinen *et al.*, 2013)(Santinha, Martins, *et al.*, 2015).

# Single-cell and population-level co-localization tools

SCIP provides the following co-localization tools: the single-cell Pearson Correlation Coefficient (along the Major and the Minor Axes) and the Manders Coefficients (M1 and M2) of the first versus the second channel (Manders *et al.,* 1993). For the Manders Coefficients calculation, pixel intensities are normalized by subtracting the mean intensity inside the cell. Based on the fluorescent intensities of each channel, the pixel values inside one channel versus the corresponding values inside the other channel can also be provided.

These coefficients, when significant, can be used to study the function of one molecule, e.g. by inference from its spatial relationship with intracellular compartments or other molecules.

# Examples of Tool Usage and Performance Analysis of the Segmentation Methodologies

The supervised evaluation scores (Fawcett, 2014) of the segmentation (Nucleoids, FtsZ Rings, Min System) and detection (Inclusion bodies) are presented in Table 1 (namely Accuracy, Sensitivity, Specificity, Precision and F1-score) and are based on the manual evaluation performed by experts. Detection times are based on an Intel Core i5-3470 CPU @ 3.20 GHz with 16 GB RAM memory running a 64-bit Windows 7 operating system.

# 3.1 Nucleoid(s) and FtsZ Ring

We studied one time-series of 31 minutes (1 image per minute), which started with 116 cells and ended with 274 cells. Figure 7 shows the results of the segmentation of the nucleoids and the FtsZ ring using the ‘TreshMorph’ Algorithm with Global Otsu thresholding.

P:\Work Leonardo\Projects\Cell Division Analyser\Samuel Nuc - Ring Segmentation\TUT Soft - Working folder\Figures\RINGS_AND_NUCS_SEG.tif

Figure 7 - Examples of simultaneous visualization of Nucleoids (red) and FtsZ Rings (green): (A) with no segmentation (B) with ‘TreshMorph’ segmentation (red lines for nucleoid segmentation and green line for FtsZ Rings).

Table 1, presents the statistical metrics for the nucleoid segmentation algorithms, namely the Gaussian Segmentation with different ‘d’ parameter values and ‘TreshMorph’ (TM) with different threshold (T) values. Table 2 shows their Confusion Matrix Tables.

Table 1 - Statistical metrics of the nucleoid segmentation algorithms using various ‘d’ parameter values and threshold (T) values. Here ‘mean’ and ‘std’ represent the Mean and Standard Deviation of the pixel intensities inside each cell.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Sensitivity | Specificity | Precision | F1 Score | Detection Time (s) |
| Gaussian with d = 2 | 84.40 | 68.33 | 97.03 | 94.76 | 79.40 | 1461.8 |
| Gaussian with d = 3 | 85.67 | 70.74 | 97.41 | 95.55 | 81.30 | 1777.7 |
| Gaussian with d = 4 | 85.44 | 69.76 | 97.76 | 96.07 | 80.83 | 1737.0 |
| Gaussian with d = 5 | 84.81 | 67.98 | 98.04 | 96.47 | 79.76 | 1759.8 |
| Gaussian with d = 6 | 83.99 | 65.81 | 98.28 | 96.79 | 78.35 | 1860.3 |
| Gaussian with d = 7 | 83.36 | 64.15 | 98.46 | 97.03 | 77.24 | 1858.6 |
| Gaussian with d = 10 | 81.31 | 58.92 | 98.92 | 97.72 | 73.52 | 1831.8 |
| Gaussian with d = 15 | 79.17 | 53.67 | 99.22 | 98.19 | 69.40 | 1749.0 |
| Gaussian with d = 20 | 78.25 | 51.58 | 99.21 | 98.09 | 67.61 | 1735.2 |
| TM (T = Global Otsu) | 87.46 | 73.27 | 98.62 | 97.66 | 83.72 | 44.4 |
| TM (T = Multilevel Otsu - 2) | 74.06 | 41.14 | 99.94 | 99.83 | 58.27 | 58.73 |
| TM (T = mean) | 86.81 | 72.56 | 98.01 | 96.63 | 82.88 | 51.3 |
| TM (T = mean + 1/3 std) | 90.19 | 97.25 | 84.64 | 83.27 | 89.72 | 42.50 |
| TM (T = mean + 2/3 std) | 92.01 | 93.49 | 90.85 | 88.93 | 91.15 | 47.41 |
| TM (T = mean + 1 std) | 91.64 | 87.16 | 95.16 | 93.40 | 90.17 | 43.13 |
| TM (T = mean + 4/3 std) | 89.46 | 79.02 | 97.66 | 96.37 | 86.84 | 48.45 |
| TM (T = mean + 5/3 std) | 86.22 | 70.03 | 98.96 | 98.14 | 81.74 | 66.17 |

Table 2 - Confusion Matrix for nucleoid segmentation. Values are shown for the Gaussian Algorithm with different ‘d’ parameter values and the ‘TreshMorph’ Algorithm (TM) with different threshold (T) values. Here ‘mean’ and ‘std’ represent the Mean and Standard Deviation of the pixel intensities inside each cell.

|  |  |  |
| --- | --- | --- |
| **Gaussian with d = 2** | | |
| **Number of cells** | Condition Positive | Condition Negative |
| **Prediction Positive** | 2305718 | 127402 |
| **Prediction Negative** | 1068475 | 4165134 |
| **Gaussian with d = 3** | | |
| **Prediction Positive** | 2387020 | 111270 |
| **Prediction Negative** | 987173 | 4181266 |
| **Gaussian with d = 4** | | |
| **Prediction Positive** | 2353814 | 96244 |
| **Prediction Negative** | 1020379 | 4196292 |
| **Gaussian with d = 5** | | |
| **Prediction Positive** | 2293914 | 83996 |
| **Prediction Negative** | 1080279 | 4208540 |
| **Gaussian with d = 6** | | |
| **Prediction Positive** | 2220521 | 73761 |
| **Prediction Negative** | 1153672 | 4218775 |
| **Gaussian with d = 7** | | |
| **Prediction Positive** | 2164674 | 66213 |
| **Prediction Negative** | 1209519 | 4226323 |
| **Gaussian with d = 10** | | |
| **Prediction Positive** | 1988236 | 46411 |
| **Prediction Negative** | 1385957 | 4246125 |
| **Gaussian with d = 15** | | |
| **Prediction Positive** | 1810807 | 33425 |
| **Prediction Negative** | 1563386 | 4259111 |
| **Gaussian with d = 20** | | |
| **Prediction Positive** | 1740388 | 33923 |
| **Prediction Negative** | 1633805 | 4258613 |
| **TM (T = Global Otsu)** | | |
| **Prediction Positive** | 2472201 | 59243 |
| **Prediction Negative** | 901992 | 4233293 |
| **TM (T = Multilevel Otsu)** | | |
| **Prediction Positive** | 1388335 | 2227 |
| **Prediction Negative** | 1985858 | 4290309 |
| **TM (T = mean)** | | |
| **Prediction Positive** | 2448241 | 85397 |
| **Prediction Negative** | 925952 | 4207139 |
| **TM (T = mean + 1/3 std)** | | |
| **Prediction Positive** | 3281233 | 659250 |
| **Prediction Negative** | 92960 | 3633286 |
| **TM (T = mean + 2/3 std)** | | |
| **Prediction Positive** | 3154418 | 392811 |
| **Prediction Negative** | 219775 | 3899725 |
| **TM (T = mean + 1 std)** | | |
| **Prediction Positive** | 2940987 | 207916 |
| **Prediction Negative** | 433206 | 4084620 |
| **TM (T = mean + 4/3 std)** | | |
| **Prediction Positive** | 2666275 | 100529 |
| **Prediction Negative** | 707918 | 4192007 |
| **TM (T = mean + 5/3 std)** | | |
| **Prediction Positive** | 2363086 | 44854 |
| **Prediction Negative** | 1011107 | 4247682 |

Table 3 presents the statistical metrics for the FtsZ ring segmentation algorithms, namely the Gaussian Segmentation with different ‘d’ parameter values and ‘TreshMorph’ (TM) with different threshold (T) values. In Table 4 we present the Confusion Matrix Tables for each case.

Table 3 - Statistical metrics of the FtsZ rings segmentation algorithms. Results are shown for the Gaussian Algorithm with different ‘d’ parameter values and the ‘TreshMorph’ Algorithm (TM) with different threshold (T) values. Here ‘mean’ and ‘std’ represent the Mean and Standard Deviation of the pixel intensities inside each cell.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Sensitivity** | **Specificity** | **Precision** | **F1 Score** | **Detection Time (s)** |
| **Gaussian with d = 2** | 89.82 | 50.26 | 96.93 | 74.61 | 60.06 | 1500.8 |
| **Gaussian with d = 3** | 90.11 | 51.21 | 97.10 | 76.03 | 61.20 | 1644.8 |
| **Gaussian with d = 5** | 90.34 | 53.77 | 96.92 | 75.82 | 62.92 | 1805.3 |
| **Gaussian with d = 7** | 90.33 | 55.64 | 96.57 | 74.45 | 63.69 | 1740.4 |
| **Gaussian with d = 10** | 90.29 | 59.40 | 95.84 | 71.99 | 65.09 | 1676.3 |
| **Gaussian with d = 13** | 90.15 | 61.13 | 95.37 | 70.35 | 65.41 | 1864.4 |
| **Gaussian with d = 15** | 90.09 | 62.98 | 94.97 | 69.22 | 65.95 | 1616.6 |
| **Gaussian with d = 17** | 89.91 | 63.13 | 94.73 | 68.30 | 65.61 | 1742.1 |
| **Gaussian with d = 20** | 89.73 | 64.56 | 94.26 | 66.92 | 65.72 | 1672.3 |
| **TM (T = Global Otsu)** | 93.80 | 70.51 | 97.99 | 86.30 | 77.61 | 73.3 |
| **TM (T = Multilevel Otsu-2)** | 88.26 | 23.99 | 99.82 | 95.95 | 38.39 | 64.7 |
| **TM (T = Multilevel Otsu–1)** | 94.42 | 80.19 | 96.97 | 82.65 | 81.40 | 67.2 |
| **TM (T = mean - 1/6 std)** | 88.34 | 91.89 | 87.70 | 57.31 | 70.59 | 60.0 |
| **TM (T = mean)** | 90.59 | 89.04 | 90.87 | 63.69 | 74.26 | 57.7 |
| **TM (T = mean + 1/6 std)** | 92.02 | 85.41 | 93.21 | 69.35 | 76.54 | 53.4 |
| **TM (T = mean + 2/6 std)** | 92.80 | 80.78 | 94.96 | 74.22 | 77.36 | 52.6 |
| **TM (T = mean + 3/6 std)** | 93.05 | 75.59 | 96.19 | 78.08 | 76.82 | 55.1 |
| **TM (T = mean + 4/6 std)** | 92.96 | 70.05 | 97.09 | 81.17 | 75.20 | 57.4 |

Table 4 - Confusion Matrix for FtsZ ring segmentation. Values are shown for the Gaussian Algorithm with different ‘d’ parameter values and the ‘TreshMorph’ Algorithm (TM) with different threshold (T) values. Here ‘mean’ and ‘std’ represent the Mean and Standard Deviation of the pixel intensities inside each cell.

|  |  |  |
| --- | --- | --- |
| **Gaussian with d = 2** | | |
| **Number of cells** | Condition Positive | Condition Negative |
| **Prediction Positive** | 587173 | 199818 |
| **Prediction Negative** | 580990 | 6298748 |
| **Gaussian with d = 3** | | |
| **Prediction Positive** | 598252 | 188603 |
| **Prediction Negative** | 569911 | 6309963 |
| **Gaussian with d = 5** | | |
| **Prediction Positive** | 628096 | 200295 |
| **Prediction Negative** | 540067 | 6298271 |
| **Gaussian with d = 7** | | |
| **Prediction Positive** | 650016 | 223066 |
| **Prediction Negative** | 518147 | 6275500 |
| **Gaussian with d = 10** | | |
| **Prediction Positive** | 693895 | 270020 |
| **Prediction Negative** | 474268 | 6228546 |
| **Gaussian with d = 13** | | |
| **Prediction Positive** | 714073 | 300927 |
| **Prediction Negative** | 454090 | 6197639 |
| **Gaussian with d = 15** | | |
| **Prediction Positive** | 735695 | 327075 |
| **Prediction Negative** | 432468 | 6171491 |
| **Gaussian with d = 17** | | |
| **Prediction Positive** | 737474 | 342341 |
| **Prediction Negative** | 430689 | 6156225 |
| **Gaussian with d = 20** | | |
| **Prediction Positive** | 754130 | 372812 |
| **Prediction Negative** | 414033 | 6125754 |
| **TM (T = Global Otsu)** | | |
| **Prediction Positive** | 823613 | 130789 |
| **Prediction Negative** | 344550 | 6367777 |
| **TM (T = Multilevel Otsu-2)** | | |
| **Prediction Positive** | 280287 | 11838 |
| **Prediction Negative** | 887876 | 6486728 |
| **TM (T = Multilevel Otsu–1)** | | |
| **Prediction Positive** | 936735 | 196653 |
| **Prediction Negative** | 231428 | 6301913 |
| **TM (T = mean)** | | |
| **Prediction Positive** | 1040150 | 593052 |
| **Prediction Negative** | 128013 | 5905514 |
| **TM (T = mean - 1/6 std)** | | |
| **Prediction Positive** | 1073450 | 799430 |
| **Prediction Negative** | 94713 | 5699136 |
| **TM (T = mean + 1/6 std)** | | |
| **Prediction Positive** | 997672 | 440958 |
| **Prediction Negative** | 170491 | 6057608 |
| **TM (T = mean + 2/6 std)** | | |
| **Prediction Positive** | 943644 | 327716 |
| **Prediction Negative** | 224519 | 6170850 |
| **TM (T = mean + 3/6 std)** | | |
| **Prediction Positive** | 883040 | 247891 |
| **Prediction Negative** | 285123 | 6250675 |
| **TM (T = mean + 4/6 std)** | | |
| **Prediction Positive** | 818329 | 189836 |
| **Prediction Negative** | 349834 | 6308730 |

# 3.2 Min System

We studied one time-series of 121 minutes where the Min System was detected using Confocal Microscopy. For this example, we started with 7 cells and finished with 14, for a total of 1318 cells analyzed. Table 5 shows the statistical metrics for the MinD-GFP proteins segmentation algorithms, namely the Gaussian Segmentation with different ‘d’ parameter values and ‘TreshMorph’ (TM) with different threshold (T) values. Table 6 shows the Confusion Matrix Tables for each case.

Table 5 - Statistical metrics of MinD proteins segmentation algorithms. Results are shown for the Gaussian Algorithm with different ‘d’ parameter values and the ‘TreshMorph’ Algorithm (TM) with different threshold (T) values. Here ‘mean’ and ‘std’ represent the Mean and Standard Deviation of the pixel intensities inside each cell.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Sensitivity** | **Specificity** | **Precision** | **F1 Score** | **Detection Time (s)** |
| **Gaussian with d = 5** | 75.52 | 29.98 | 90.48 | 50.84 | 37.72 | 245.0 |
| **Gaussian with d = 10** | 76.35 | 41.99 | 87.62 | 52.70 | 46.74 | 280.2 |
| **Gaussian with d = 13** | 76.45 | 43.84 | 87.16 | 52.85 | 47.93 | 279.4 |
| **Gaussian with d = 14** | 76.44 | 46.21 | 86.36 | 52.66 | 49.23 | 281.2 |
| **Gaussian with d = 15** | 76.50 | 46.12 | 86.47 | 52.83 | 49.25 | 270.6 |
| **Gaussian with d = 16** | 76.49 | 45.86 | 86.56 | 52.83 | 49.10 | 276.0 |
| **Gaussian with d = 17** | 76.35 | 45.21 | 86.57 | 52.51 | 48.59 | 257.6 |
| **Gaussian with d = 19** | 76.30 | 45.11 | 86.54 | 52.38 | 48.48 | 273.9 |
| **Gaussian with d = 20** | 76.17 | 45.46 | 86.25 | 52.06 | 48.54 | 278.6 |
| **Gaussian with d = 25** | 75.78 | 44.50 | 86.05 | 51.15 | 47.59 | 266.9 |
| **TM (T = Global Otsu)** | 78.50 | 96.0 | 72.75 | 53.63 | 68.81 | 15.1 |
| **TM (Multilevel Otsu)** | 89.46 | 58.35 | 99.68 | 98.34 | 73.24 | 16.5 |
| **TM (T = mean – 2/6 std)** | 89.34 | 95.64 | 87.27 | 71.15 | 81.60 | 14.5 |
| **TM (T = mean – 1/6 std)** | 93.64 | 92.62 | 93.97 | 83.46 | 87.80 | 14.6 |
| **TM (T = mean)** | 95.28 | 85.13 | 98.69 | 95.54 | 90.01 | 14.4 |
| **TM (T = mean + 1/6 std)** | 92.20 | 69.84 | 99.54 | 98.03 | 81.57 | 14.9 |
| **TM (T = mean + 2/6 std)** | 88.96 | 55.92 | 99.81 | 98.99 | 71.46 | 14.8 |

Table 6 - Confusion Matrix for MinD protein segmentation. Values are shown for the Gaussian Algorithm with different ‘d’ parameter values and the ‘TreshMorph’ Algorithm (TM) with different threshold (T) values. Here ‘mean’ and ‘std’ represent the Mean and Standard Deviation of the pixel intensities inside each cell.

|  |  |  |
| --- | --- | --- |
| **Gaussian Algorithm d=5** | | |
| **Number of cells** | **Condition Positive** | **Condition Negative** |
| **Prediction Positive** | 124998 | 120856 |
| **Prediction Negative** | 291961 | 1149161 |
| **Gaussian Algorithm d=10** | | |
| **Prediction Positive** | 175091 | 157179 |
| **Prediction Negative** | 241868 | 1112838 |
| **Gaussian Algorithm d=13** | | |
| **Prediction Positive** | 182805 | 163069 |
| **Prediction Negative** | 234154 | 1106948 |
| **Gaussian Algorithm d=14** | | |
| **Prediction Positive** | 192697 | 173166 |
| **Prediction Negative** | 224262 | 1096851 |
| **Gaussian Algorithm d=15** | | |
| **Prediction Positive** | 192318 | 171713 |
| **Prediction Negative** | 224641 | 1098304 |
| **Gaussian Algorithm d=17** | | |
| **Prediction Positive** | 188530 | 170502 |
| **Prediction Negative** | 228429 | 1099515 |
| **Gaussian Algorithm d=19** | | |
| **Prediction Positive** | 188107 | 171000 |
| **Prediction Negative** | 228852 | 1099017 |
| **Gaussian Algorithm d=20** | | |
| **Prediction Positive** | 189581 | 174553 |
| **Prediction Negative** | 227378 | 1095464 |
| **Gaussian Algorithm d=25** | | |
| **Prediction Positive** | 185538 | 177185 |
| **Prediction Negative** | 231421 | 1092832 |
| **TM (T = Global Otsu)** | | |
| **Prediction Positive** | 400262 | 346036 |
| **Prediction Negative** | 16697 | 923981 |
| **TM (T = Multilevel Otsu-2)** | | |
| **Prediction Positive** | 243306 | 4103 |
| **Prediction Negative** | 173653 | 1265914 |
| **TM (T = mean)** | | |
| **Prediction Positive** | 351072 | 18457 |
| **Prediction Negative** | 65887 | 1251560 |
| **TM (T = mean – 2/6 std)** | | |
| **Prediction Positive** | 398782 | 161677 |
| **Prediction Negative** | 18177 | 1108340 |
| **TM (T = mean – 1/6 std)** | | |
| **Prediction Positive** | 386178 | 76520 |
| **Prediction Negative** | 30781 | 1193497 |
| **TM (T = mean + 1/6 std)** | | |
| **Prediction Positive** | 291190 | 5851 |
| **Prediction Negative** | 125769 | 1264166 |
| **TM (T = mean + 2/6 std)** | | |
| **Prediction Positive** | 233164 | 2378 |
| **Prediction Negative** | 183795 | 1267639 |

# 3.3 Inclusion bodies

To study the detection of inclusion bodies *E. coli* cells were exposed to osmotic stress as this type of stress leads to an increase in the amount of visible inclusion bodies (Oliveira *et al.*, 2016). 3 conditions were tested: no stress, medium and high stress (0, 125 and 300 mM of NaCl). We analyzed the phase-contrast images of cells under the osmotic stress for 60 minutes.

In this example, we calculated the time spent on the decision to remove the false seed. 405, 360 and 697 cells were analyzed for the conditions 0, 125 and 300 mM of NaCl, respectively. The performance scores are presented in Table 7. Table 8 shows the Confusion Matrix.

Table 7 - Statistical metrics of the inclusion body detection algorithm (Accuracy, Sensitivity, Specificity, Precision, F1 Score) for null, medium and high stress.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Sensitivity** | **Specificity** | **Precision** | **F1 Score** | **Removal Time (s)** |
| **No stress** | 98.75 % | 71.43 % | 99.88 % | 96.15 % | 81.97 % | 70.8 |
| **Medium stress** | 97.91 % | 88.94 % | 99.73 % | 98.53 % | 93.49 % | 120.3 |
| **High stress** | 97.81 % | 89.71 % | 99.39 % | 96.65 % | 93.05 % | 130.9 |

Table 8 – Confusion Matrix for the detection of Inclusion bodies based on the GPL seed placement and their respective deletion for null, medium and high stress.

|  |  |  |
| --- | --- | --- |
| **No Stress (0 mM of NaCl)** | | |
| **Number of cells** | **Condition Positive** | **Condition Negative** |
| **Prediction Positive** | 25 | 1 |
| **Prediction Negative** | 10 | 851 |
| **Medium Stress (125 mM of NaCl)** | | |
| **Prediction Positive** | 201 | 3 |
| **Prediction Negative** | 25 | 1110 |
| **High Stress (300 mM of NaCl)** | | |
| **Prediction Positive** | 375 | 13 |
| **Prediction Negative** | 43 | 2121 |

Finally, SCIP allows the manual addition of new seeds by clicking once inside the cell, allowing for fast corrections even for a large number of cells.

# 3.4 Structure detection of simultaneous channels

Figure 8 shows a single channel visualization of each structure of interest, while Figure 9 shows the three possible combinations of simultaneous visualization of 2 channels, with and without segmentation. In these images, MS2-GFP spots are shown in the green channel, Nucleoid(s) in the Blue channel and FtsZ Rings in the red channel.

P:\Work Leonardo\Projects\Cell Division Analyser\Samuel Nuc - Ring Segmentation\TUT Soft - Working folder\Figures\Figure_13.tif

Figure 8 - Single channel visualization of (A) Nucleoids (segmented in blue), (B) FtsZ Rings (segmented in red) and (C) MS2-GFP spots (segmented in green). Segmentation in both Nucleoids and FtsZ Rings was performed by the TreshMorph Algorithm while segmentation in MS2-GFP spots was done using a median filter.

P:\Work Leonardo\Projects\Cell Division Analyser\Samuel Nuc - Ring Segmentation\TUT Soft - Working folder\Figures\Figure_14.tif

Figure 9 - Examples of simultaneous visualization of two channels without (top) and with (bottom) segmentation: (A) Nucleoids (blue) and FtsZ Rings (red), (B) Nucleoids (blue) and MS2-GFP spots (green), (C) FtsZ Rings (red) and MS2-GFP spots (green), (D) Nucleoids (segmented in blue) and FtsZ Rings (segmented in red), (E) Nucleoids (segmented in blue) and MS2-GFP spots (segmented in green), (F) FtsZ Rings (segmented in red) and MS2-GFP spots (segmented in green). Segmentation in both Nucleoids and FtsZ Rings used the ‘TreshMorph’ Algorithm and segmentation in MS2-GFP spots used a median filter.

# 3.5 Single cell Co-localization of RNAp and Nucleoid

Here, RNAp-GFP molecules are shown in the green channel and HupA-mCherry-tagged nucleoids in the Red channel (776 cells analysed). As an example, we show the fluorescence along the normalized Major and Minor axis of each cell from two channels (Figure 10), from which we calculated the Pearson Correlation Coefficient (PCC) along each axis (Table 10). From the parameters provided by SCIP we plotted each pixel fluorescence intensity on the first channel versus its intensity on the second channel (Figure 11), from which it can calculate Manders Coefficients (Table 9).

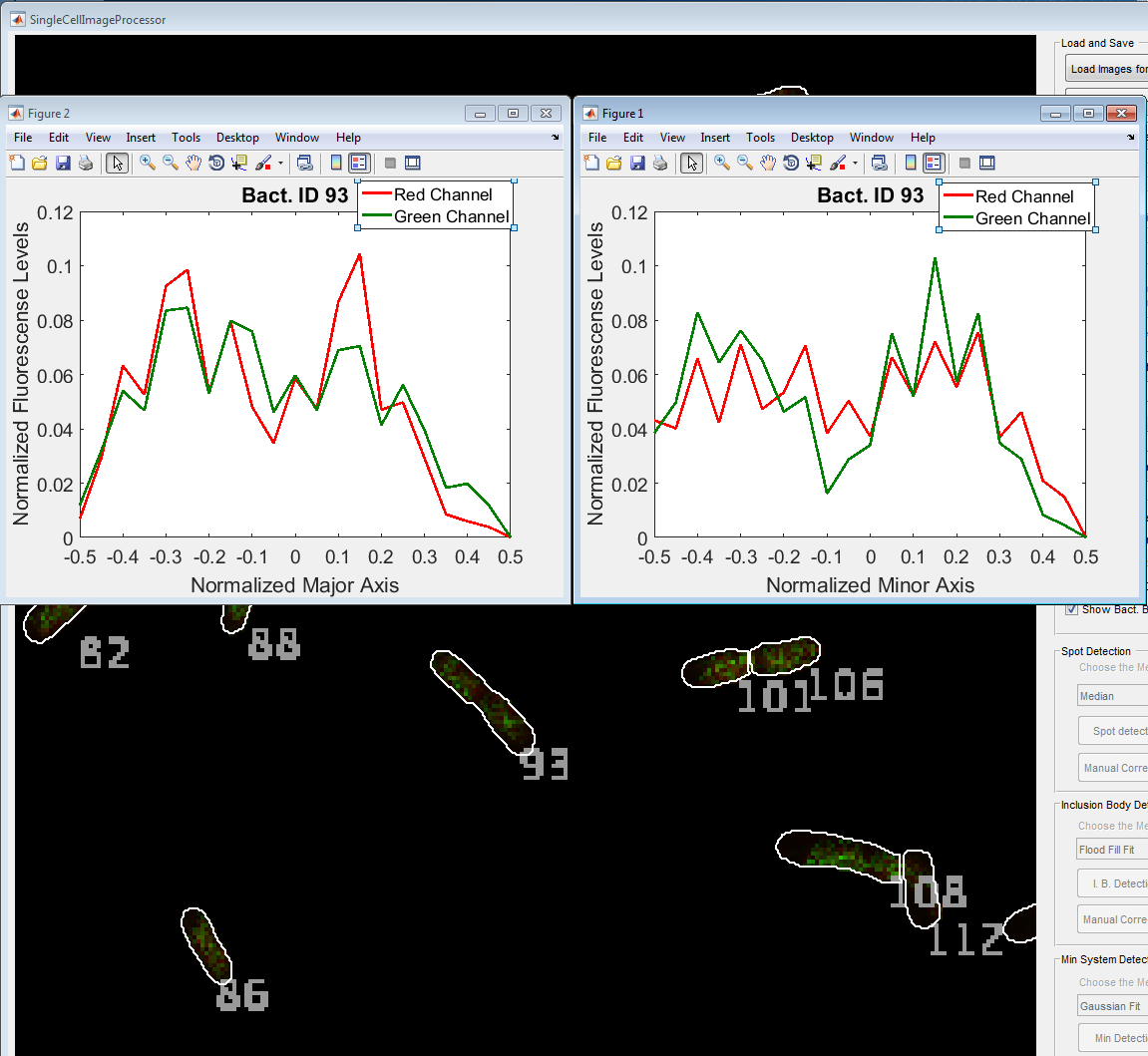


Figure 10 - Example of single-cell co-localization of Nucleoid and RNAp detection (cell with ID 93).



Figure 11 - RNAP fluorescence intensity vs. nucleoid fluorescence intensity values of Bacteria with ID 93.

Table 9 - Pearson Correlation Coefficient (PCC) between RNAp fluorescence and Nucleoid fluorescence in each cell, along the Major and Minor Axis of the specific cells. The Manders Coefficients were also calculated (M1 and M2 correspond to the Nucleoid and the RNAp, respectively).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Bacteria ID 86** | **Bacteria ID 93** | **Bacteria ID 101** | **Bacteria ID 106** |
| **Global PCC**  **(p-value)** | 0.4213  (5.5×10-25) | 0.3959  (5.1×10-38) | 0.5747  (1.5×10-47) | 0.4316  (6.6×10-26) |
| **Minor Axis PCC**  **(p-value)** | 0.951 (3.4×10-11) | 0.867 (3.6×10-7) | 0.956 (1.4×10-11) | 0.982 (2.5×10-13) |
| **Major Axis PCC**  **(p-value)** | 0.968 (1.3×10-9) | 0.928 (1.3×10-9) | 0.981 (4.9×10-15) | 0.933 (6.5×10-10) |
| **Manders Coefficients (M1)** | 0.447 | 0.445 | 0.5081 | 0.486 |
| **Manders Coefficients (M2)** | 0.667 | 0.654 | 0.7312 | 0.616 |

# Global Performance Analysis

A global performance analysis of all automatic algorithms in SCIP is performed based on the supervised evaluation scores (Fawcett, 2014), namely Accuracy, Sensitivity, Specificity, Precision and F1-score. The consolidated benchmark analysis of each algorithm that have better overall scores is presented in Table 10.

Table 10 - Benchmark results of the automatic detection algorithms. In Cell Tracking, we can only calculate Accuracy, because only True Positives and False Positives can be calculated. In Cell Segmentation/Detection, True Negative’s cannot be evaluated and consequently Accuracy and Specificity cannot be calculated.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Accuracy** | **Sensitivity** | **Specificity** | **Precision** | **F1 Score** |
| **Cell Border Segmentation** | - | 99.6% | - | 97.1% | 98.3% |
| **Cell Tracking** | 99.5% | - | - | - | - |
| **Spot Detection** | 98.2% | 99.7% | 97.2 | 96.0% | 97.8% |
| **Nucleoids Segmentation** | 92.0% | 93.5% | 90.9% | 88.9% | 91.1% |
| **FtsZ Rings Segmentation** | 94.4% | 80.2% | 97.0 | 82.7% | 81.4% |
| **Min System Segmentation** | 95.3% | 85.1% | 98.7% | 95.5% | 90.0% |
| **Inclusion Bodies Detection** | 98.0% | 88.5% | 99.6% | 97.3% | 92.7% |

All scores in Table 10 are between 99% and 85% (except for the detection of FtsZ Rings). This quality is usually sufficient to analyse, e.g., the effects of environmental shifts on the spatial and temporal behaviour of subcellular structures, as in (Oliveira *et al.*, 2016). The Sensitivity and Precision scores are lower in the case of FtsZ Rings segmentation (~80%) due to this processing involving large morphological changes of this cellular structure during the cell lifetime (Tsukanov *et al*., 2011), resulting in large fluctuations of the contrast ratio inside the cells. Due to this, it is expectable that a Machine Learning based algorithm (Neeli-Venkata *et al.*, 2018) that can adapt to each stage of the cell cycle (Tsukanov *et al*., 2011) could prove to be more efficient. Nevertheless, SCIP allows manual corrections when needed.

# Experiments

# 5.1 Bacterial strain, growth conditions, and induction

To collect microscopy data on the nucleoid, we used *E. coli* CM735-derived strain NK9386 expressing the endogenous gene *hupA* fused with fluorescent protein mCherry, under the control of its native promoter in the chromosome (Fisher *et al.*, 2013) (a kind gift from Nancy Kleckner, Harvard University, U.S.A).

To collect microscopy data on the Min system, we used *E. coli* FW1561, a derivative of W3110, expressing the endogenous gene *minD* fused with fluorescent protein superfolder GFP (sfGFP), under the control of the native promoter, incorporated into the chromosome (Wu *et al.*, 2015) (a kind gift from Cees Dekker, Delft University of Technology, Netherlands).

To collect microscopy data on the nucleoid, the Z-ring, and protein aggregates (MS2-GFP tagged RNA) simultaneously on individual cells, we used the strain FW1551 expressing the endogenous gene *hupA* fused with the fluorescent protein TagBFP, under the control of the native promoter, incorporated into the chromosome (Wu *et al.*, 2015) (a kind gift from Cees Dekker, Delft University of Technology, Netherlands). This strain was also transformed with medium copy plasmid expressing MS2-GFP reporter system, a single-copy BAC vector plasmid expressing the target gene *PtetA-mRFP1-96BS* with a 96 MS2-GFP binding site array and pEG12-*ftsz-mCherry*. All these plasmids have different origin of replications and different antibiotic resistances. We refer this strain as FW1551:mrfp1-96bs-FtsZ-mCherry.

In general, overnightcultures were grown in LB supplemented with appropriate antibiotics, when required, for 15 hours at 37 °C with shaking (250 rpm)*.* We subsequently made subcultures by diluting overnight cultures into fresh LB and the subcultures were left in the incubator at 37 °C with shaking (250 rpm) until the cells attain mid-logarithmic phase (OD600 ~ 0.3). Cells at this stage were induced with appropriate inducers (when needed) and, following the necessary amount of incubation time, taken to the microscopy chamber for imaging.

For the visualization of inclusion bodies, the cells were exposed to osmotic stress during time-lapse microscopy. Sodium Chloride (125 and 300 mM of NaCl) was added to the growth media and pumped into the thermal chamber (set to 37°C) for 1 hour. For population microscopy imaging, the cells were kept under osmotic stress for 60 minutes (osmotic stress-inducing media with 125 and 300 mM of NaCl).

To visualise 3 fluorescent proteins simultaneously in individual cells, first, the reporter plasmid activation was performed by adding 0.4 % of L-arabinose to the culture, and incubated at 37°C for 60 minutes. Following the reporter activation, the target plasmid was activated by adding 50ng aTc (Anhydro tetracycline) to these cultures and left in the incubator for 30 minutes. Cells were pelleted and then imaged.

# 5.2 Microscopy

For imaging, cells were placed on a microscope slide between a coverslip and 1% agarose gel pad prepared in a specific media. Phase contrast images were acquired for cell segmentation and, in time series, for lineage construction. Phase contrast images were captured by a CCD color camera (DS-Fi2, Nikon).

Fluorescent images were taken by a Nikon Eclipse (Ti-E, Nikon) inverted microscope with a C2+ point scanning confocal system and a 100x Apo TIRF objective (1.49 NA, oil). Blue fluorescence images were taken using a 461 nm laser (Melles-Griot) using a HQ514/30 filter. Green fluorescence images were taken using a 488 nm argon laser (Melles-Griot) using a HQ514/30 filter. Red fluorescence images were taken using a 543 nm He-Ne laser (Melles-Griot) using a HQ585/65 filter. We also used Highly Inclined and Laminated Optical sheet (HILO) microscopy to visualize the fluorescent labelled Min system. The green and red fluorescence signals were recorded using an EMCCD camera (iXon3 897, Andor Technology) with a 488nm laser, along with the HQ514/30 filter and the Texas Red filter (Nikon, Tokyo, Japan), respectively.

In general, for time-series, cells were visualized for 1 hour, with confocal images taken once every 1 minute and phase contrast images taken once every 5 minutes. The software for image acquisition was NIS-Elements (Nikon, Tokyo, Japan). Slides were kept in a temperature-controlled chamber (Bioptechs, FCS2) at stable temperature (37 °C, unless stated otherwise).

# References

Breiman, L. *et al*. (1984) Classification and Regression Trees Chapman and Hall/CRC.

Fawcett, T. (2014) An Introduction to ROC Analysis. Pattern Recognit. Lett., 27, 861–874.

Fisher, J.K. *et al*. (2013) Four-dimensional imaging of E. coli nucleoid organization and dynamics in living cells. Cell, **153**, 882–895.

Golding, I. and Cox, E.C. (2004) RNA dynamics in live Escherichia coli cells. Proc. Natl. Acad. Sci. U. S. A., **101**, 11310–11315.

Häkkinen, A. *et al*. (2013) CellAging: a tool to study segregation and partitioning in division in cell lineages of Escherichia coli. Bioinformatics, 29, 1708–9.

Manders, E.M.M. *et al*. (1993) Measurement of co-localization of objects in dual-colour confocal images. J. Microsc., **169**, 375–382.

Mora, A.D. *et al*. (2011) Automated drusen detection in retinal images using analytical modelling algorithms. Biomed. Eng. Online, **10**, 59.

Neeli-venkata, R. *et al*. (2016) Polar Localization of the Serine Chemoreceptor of Escherichia coli Is Nucleoid Exclusion-Dependent. Biophys. J., **111**, 2512–2522.

Neeli-Venkata, R. *et al*. (2018) The precision of the symmetry in Z-ring placement in Escherichia coli is hampered at critical temperatures. Phys. Biol., **15**, 056002

Oliveira, S.M.D., *et al*. (2016) Increased cytoplasm viscosity hampers aggregate polar segregation in Escherichia coli. Mol. Microbiol., **99**, 686–699.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, **9**, 62-66.

Queimadelas, C. *et al*. (2012) Segmentation and tracking of Escherichia coli expressing tsr-venus proteins from combined DIC/Fluorescence images. In, MEDSIP 2012 - 5th International Conference on Advances in Medical Signal and Information Processes, Liverpool, UK. Liverpool, UK, pp. 1–2.

Santinha, J., Mora, A., *et al*. (2015) Detection and Segmentation of nucleoids based on Gradient Path Labelling. In, 2nd International Conference on Biomedical and Pharmaceutical Engineering (ICBPE 2015). May 12-13, 2015, Warsaw, Poland.

Santinha, J., Martins, L., *et al*. (2015) iCellFusion: Tool for fusion and analysis of live-cell images from time-lapse multimodal microscopy. In, Karâa,W.B.A. and Dey,N. (eds), Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes. IGI Global, pp. 71–99.

Sonka, M. *et al*. (2014) Image processing, analysis, and machine vision. Cengage Learning.

Tsukanov, R. *et al*. (2011) Timing of Z-ring localization in Escherichia coli. Phys. Biol., 8, 66003.

Wu, F. *et al*. (2015) Multi-color imaging of the bacterial nucleoid and division proteins with blue, orange, and near-infrared fluorescent proteins. Front. Microbiol., 6, 607.