Electronic Supplementary Material (ESI) for Integrative Biology

Radler et al., Integr. Biol., 2011
DOI: 10.1039/c1ib00035g

Supporting Material

Technical Details and Algorithms

The different parts of the workflow are discussed in the order in which they are executed.

Calculation of a reference Background

We assume that only a small fraction of the pixels of each picture is part of a cell with the rest forming part of the background. We determine the background of the pixel at position x_i and y_j in slice k by calculating the median of all values of Pixel(x_i, y_j, k) from all movies of the experiment. The quality of this background movie increases with the number of movies that are grabbed during the experiment.

When too many bright pixels have been recorded, typically more than a third of the total, a reliable background cannot be calculated by the median method. In these cases we use backgrounds that have been directly measured. For all backgrounds, we do a smoothing with a Gaussian kernel.

Background Correction

The most prominent challenges in background correction are non-uniform illumination of the viewfield and background fluorescence of samples. The background fluorescence is routinely fluctuating between frames. In the background creation step, we have determined the most probable background fluorescence distribution. As this distribution depends on the illumination, we define an illumination factor f_{illu} as the ratio between the background for any pixel b and the mean background of the whole image <b>.

$$f_{illu} = \frac{< b >}{b}$$

Radler et al., Integr. Biol., 2011
DOI: 10.1039/c1ib00035g

The background values are normally distributed, so some values can be very low or even zero. Consequently, we have set an upper limit for this factor to two standard deviations of the background value distribution. The resulting pixel values $p_{corrected}$ are then given by

Segmentation

We detect fluorescently labeled cells using a density-based clustering algorithm combined with a threshold condition. This algorithm assumes that a fluorescent cell has higher fluorescence intensity values than the background on average. Bright pixels, with above-threshold brightness, are identified as part of a cell when more than *n* bright (above-threshold brightness) pixels in a distance not bigger than ε are collocated. The parameters *n* and ε can be adjusted according to the size and density of objects that are to be identified; the defaults are n=8 points and distance ε=2.0 pixels respectively. Neighboring objects that fulfil the density criterion are combined into a single object.

We use a density-based clustering algorithm combined with a threshold condition [16] The ideal threshold condition has to be defined in the tension zone between capturing small objects (low brightness value) and increasing the contrast to separate adjacent cells (high brightness value). Consequently, the initial segmentation often yields objects that do not correspond to the actual cells, in that they are either overly segregated or not detected at all (*over-segmentation*), or neighbouring cells are recognized as a single object (*under-segmentation*).

Electronic Supplementary Material (ESI) for Integrative Biology

Radler et al., Integr. Biol., 2011
DOI: 10.1039/c1ib00035g

Tracking

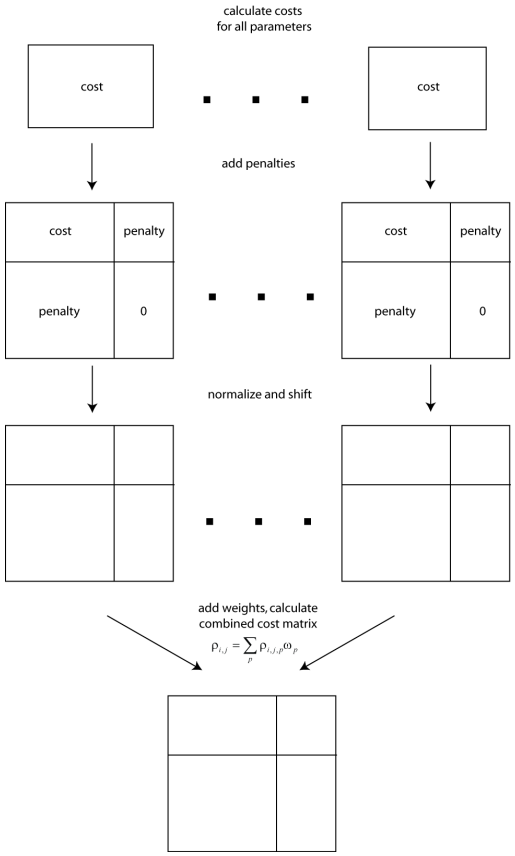

Figure A1
Preparation of the Multi Parameter cost matrix: Individual cost matrix are constructed as in Jaqaman et. al. (ref Jaqaman). Each cost matrix is then normalized to variance 1 and shifted to mean 0. Afterwards the total cost is calculated as weighted sum over all parameters.

Tracking is the art of identifying individual cells and connecting them through time.

As proposed by Jaqaman et. al., tracking can be treated as a linear assignment problem (LAP). A linear assignment problem is a mathematical formulation of all sorts of cost minimization or maximization problems. Here, we apply it to minimize the total connection costs for all cells between individual frames of a time-lapse, single-cell movie. The cost can either be derived from a single parameter, most commonly ‘position’, but it is also feasible to construct weighted mixtures of parameter values. To this end we construct cost matrices for all user-defined parameters. These contain sub-matrices of actual costs, the penalties, and a segment consisting of zeroes.

Radler et al., Integr. Biol., 2011
DOI: 10.1039/c1ib00035g

We combine these matrices by calculating individual mean and variance values for all cost sub-matrices. The complete matrices are then normalized to mean 0.0 and variance 1.0. The matrix elements $\rho_{i,j}$ of the resulting matrix are the sum of the products of the entries of the parameter matrices p with their corresponding weights ω_p:

$$\rho_{i,j} = \sum_p \rho_{i,j,p} * \omega_p$$

The resulting cost matrix can be evaluated with the regular LAP approach, which we solve by applying the Hungarian method [17].

Electronic Supplementary Material (ESI) for Integrative Biology

Radler et al., Integr. Biol., 2011
DOI: 10.1039/c1ib00035g

Event handling

The cost for connecting a cell between two adjacent frames is roughly continuous throughout the life cycle of a given cell. However, cells divide or die occasionally and in the particular case of a dense culture touch each other. Each of these *events* has a characteristic signature in terms of costs and context. We distinguish between two major classes of events: *life-cycle events* and *image analysis events* as introduced in Fig. "types of events". Here we list all event types in detail and describe their characteristic properties.

Image Analysis Events

Transient contact or *touch-and-go*: two or more cells are separate from each other in a given frame, then apparently fuse with each other, leading to segmentation as a single cellular shape. These may split into separate shapes in a later frame. Touch-and-go events are detected when several time series end in the same vicinity (defined by mean cell radius), and different time series start in the same area in the following frame. Both sets need to be completely disjunct. This approach may lead to the detection of too many events. To prevent such errors, a check is performed during the resolution of touch-and-go events (details below).

Asymmetrical contact or *eaten-cells*: one or more small cells appear to be engulfed by a bigger neighbor. This fusion of cell shapes is segmented into a single shape that is associated with the track of the bigger cell. These may split into separate shapes at a later time again. Asymmetrical contact events are detected, when at least one time series ends in the vicinity of a continuously tracked cell. This approach may likewise lead to the detection of too many events as well. For instance, it is possible that the smaller cell has undergone lysis. As is the case for touch-and-go events, a check is performed to prevent errors (details below).

Contour fusion or simply *fusion* exhibits the same motif as transient contact or asymmetrical contact but coupled with the impossibility to find a dividing line

Radler et al., Integr. Biol., 2011
DOI: 10.1039/c1ib00035g

between cells. Effectively, this is reported when the algorithm failed to automatically divide cell shapes after a transient contact or asymmetrical contact. The cause may either be that the event was erroneously detected (a cell has simply died) or that the cell resolving step failed.

Boundary losses or *borderline cells*: cells leave partially the view field.

Cell cycle events

Cell lysis: Cells exhibit an increase in size, while their fluorescence intensity declines before they disappear from view. This event is often accompanied by a apparent division into multiple daughter objects or simply the absence of the corresponding cell in the following frame. The disappearance of an object is the primary indicator for lysis.

Cell division: Cells divide into two approximately equal-sized children cells. Candidates for parent-children pair are selected using a neighborhood condition. Subsequently for each candidate a cell-division score is calculated according to

$$\begin{aligned} score = &(1-((\log(penalty_{cell-division})-c_{area:child-to-child})/\log(penalty_{cell-division}))) \\ &+(1-((\log(penalty_{cell-division})-c_{area:parent-to-children})/\log(penalty_{cell-division}))) \\ &+(1-((\log(penalty_{dist.})-c_{dist.:parent-to-children})/\log(penalty_{dist.}))) \end{aligned}$$

or set to positive infinity if one of the following condition applies:

- $\log(penalty_{cell-division}) < c_{area:child-to-child}$
- $\log(penalty_{cell-division}) < c_{area:parent-to-children}$
- $\log(penalty_{dist.}) < c_{dist.:parent-to-children}$

The kin relation is established if the score is less than positive infinity and in case of unambiguities the parent-children pair with the lowest score is selected.

Automatic filtering of valid cell traces

Radler et al., Integr. Biol., 2011
DOI: 10.1039/c1ib00035g

Contextual information allows the selection of cells that fulfill well-defined criteria, such as ‘contact-free’ and ‘only after first cell division’. The information from the event handling step enables the reliable, minimal filtering of defective cellular traces. This context-aware filtering can be adapted to all kinds of experimental data and be combined with classical quality criteria such as the length of a trace or the minimum fluorescence intensity.