SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

An active-contour based algorithm for the automated segmentation of dense yeast populations on transmission microscopy images

Kristian Bredies Heimo Wolinski

SFB-Report No. 2011–028 December 2011

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Das Land Steiermark

Stadt GRAZ Wissenschaft

An active-contour based algorithm for the automated segmentation of dense yeast populations on transmission microscopy images

Kristian Bredies · **Heimo Wolinski**

Abstract An image-processing pipeline for the automated segmentation of yeast cells in microscopy images is proposed. The method is suitable for the non-invasive detection of individual cells in transmission data which can be acquired simultaneously with fluorescence data. It moreover takes the varying quality and highly heterogeneous characteristics of cells in transmission images into account, is capable to process images with dense yeast populations and can be used to extract quantitative cell-based data from transmission/fluorescence image pairs. Applicability and performance of the method is evaluated on a data set of 523 different yeast deletion mutant strains.

Keywords Cell segmentation · active contour algorithm · transmission microscopy · dense yeast populations

Kristian Bredies acknowledges support by the Austrian Science Fund (FWF) special research grant SFB F32 “Mathematical Optimization and Applications in Biomedical Sciences”.
This work was supported by grants from the Austrian Science Funds, FWF, project [F3005-B12 LIPOTOX] and the Federal Ministry of Science and Research [Project GOLD - Genomics of lipid-associated disorders, in the framework of the Austrian Genome program, GEN-AU], to Heimo Wolinski.

Kristian Bredies
Institute of Mathematics and Scientific Computing
University of Graz
Heinrichstraße 36/III
A-8010 Graz, Austria
E-mail: kristian.bredies@uni-graz.at

Heimo Wolinski
Institute of Molecular Biosciences
Department of Biochemistry
University of Graz
Humboldtstraße 50/II
A-8010 Graz, Austria
E-mail: heimo.wolinski@uni-graz.at

1 Introduction

In the post-genomic era, large-scale microscopic analyses of protein localization patterns or morphological traits provide powerful tools towards functional characterization of proteins. In this context, live cell imaging using fluorescent protein fusions has become a standard tool for studying alternations of protein localization or organelle morphology resulting from mutations or drugs.

In recent years, imaging-based screening platforms have been developed enabling the analysis of cellular events in a large number of experimental samples and with high throughput. These approaches complement biochemical and genetic methods and have been successfully applied for phenotypic characterization of subcellular components in different organisms including the yeast, Saccharomyces cerevisiae. The yeast is a widely used model system for studying fundamental cellular processes of eukaryotic cells. Due to its genetic tractability, its ease of genetic manipulation and the wealth of well-established biochemical, genomics and proteomics tools the yeast represents an excellent model for the analysis of protein function. In this respect, systematic analyses of yeast deletion mutant collections for abnormal morphology of fluorescently labelled subcellular structures in mutant strains provide an exceptional experimental platform for phenotypic screens and for large-scale analysis of protein function [23,21,17,15,14,13,10].

However, quantification of acquired large-scale image data is a challenge for image processing. For statistically relevant analysis of image-based yeast screening data not only the automated registration of fluorescently labelled subcellular structures but particularly the detection of individual cells is required. In this respect, methods have been established for labelling yeast cell boundaries or the entire cell volume using fluorescence dyes or fluorescent protein fusions. Due to the high contrast of fluorescence images, automated large-scale cell registration is facilitated, whereas various commercially and non-commercially methods and software tools are available to process such image data [3,8,18,9].

However, fluorescence-based approaches for cell detection may have some drawbacks such as fluorescence emission crosstalk between fluorescence dyes and the limited number of spectrally diverging fluorescence labels in multi-labelling experiments. Furthermore, the addition of external fluorescence dyes or co-expression of fusion proteins for cell labelling may affect the cellular physiology, which may lead to altered phenotypes of tagged proteins or organelles in living yeast cells.

Due to these reasons non-invasive methods for detection of yeast cells are appreciated. In this context, a few image processing methods and tools have been introduced for detection of yeast cells in conventional bright field (transmission) images or in differential interference contrast (DIC) images. The latter, however, is usually not applicable in high-resolution screening assays capable for simultaneous acquisition of fluorescence and transmission data, since DIC optics implemented in commercial microscopes and screening platforms usually impairs the detection of simultaneously acquired fluorescent objects significantly. This phenomenon is particularly limiting combined DIC/fluorescence imaging of yeast since its subcellular structures are typically at the limit of optical resolution.

On the other hand, methods for processing of yeast cells imaged using conventional bright field microscopy have been established. These methods have been shown to be suitable for specific yeast applications [12]. However, the use of these approaches for automated processing of large scale screening data may be limited by the quality of acquired images. For instance, variations in image contrast of acquired cell arrays, variations of the cell morphology of different yeast deletion mutant strains, optical dense cell content such as lipid droplets, subcellular structures showing high contrast such as vacuoles and very dense cell populations may interfere with reliable cell detection.

Here we present a new image processing method for automated registration of yeast cells acquired using bright field microscopy particularly in high-content screening assays. The method widely overcomes typical limitations of the registration of yeast cells in conventional transmission images.

A software tool is provided for automated processing both of transmission and of three-dimensional fluorescence data derived from different sources. In this respect, the tool is optimized for a microscope setup capable for simultaneous acquisition of transmission and of multi-dimensional fluorescence data as acquired using confocal microscopy or light-microscope based screening platforms. Notably, the software requires minimal user interaction and can be run in unsupervised batch-processing mode.

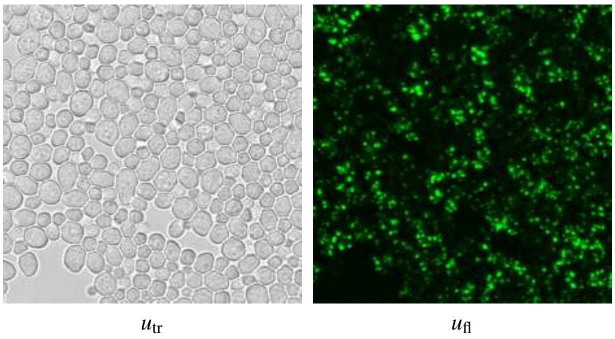


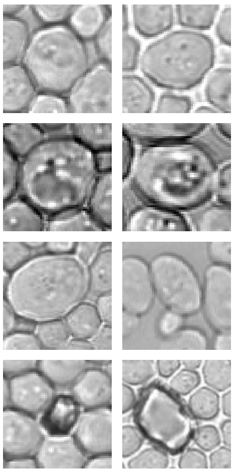
Fig. 1 A typical transmission (left) and fluorescence (right) microscopy image pair (yeast mutant strain YGL009C). The depicted fluorescence image is the maximum-projection over multiple layers (contrast-enhanced).

The paper is organized as follows. In Section 2, the segmentation problem and its mathematical formulation which leads to the active-contour algorithm is introduced. Furthermore, a special nonlinear denoising filter based on a degenerate partial differential equation and a slight extension of the well-known active contour method for detection of closed boundary curves is described. Section 3 shows the image processing steps used by the developed software tool for quantification of acquired yeast screening data. In Section 4, we briefly describe the features of the software tool. The software tool was applied to process a larger already published data-set of 523 test images with known characteristics [23]. The results are discussed in Section 5. Finally, in Section 6 we conclude the article with summarizing remarks and future outlooks.

2 Mathematical problem formulation

In the following, let u_{tr} and u_{fl} be a transmission and fluorescence microscopy image, respectively. Note that u_{tr} is always a two-dimensional image while u_{fl} can be three-dimensional, for instance, when taken with a confocal laser microscope. We are assuming that the quantities of interest are contained in u_{fl} whose information, however, is in general completely unrelated to the cell interior or the cell boundaries. Therefore, in order to obtain a quantitative cell-based evaluation, information about each single cell has to be extracted from u_{tr} only. This is usually the case when the lipid metabolism of yeast cells is studied. The transmission microscopy then provides a non-invasive way of obtaining information about the cell boundaries (see Figure 1 for an example image pair). Finally, it is assumed that one is interested in evaluating more than a few cells, resulting in u_{tr} showing a dense population of cells, i.e., not single objects on a homogeneous background.

The task of obtaining cell boundary information from a transmission microscopy image poses some challenges. One



– Cells can be separated by bright or dark edges, often both is appearing in the same image. Contrast is also varying.

– Organelles are often visible, some of them admit a cell-like structure.

– Extremely small and large cells can be present in the same image. Some cells have small buds attached.

– Some images show undesired objects like air bubbles, for instance.

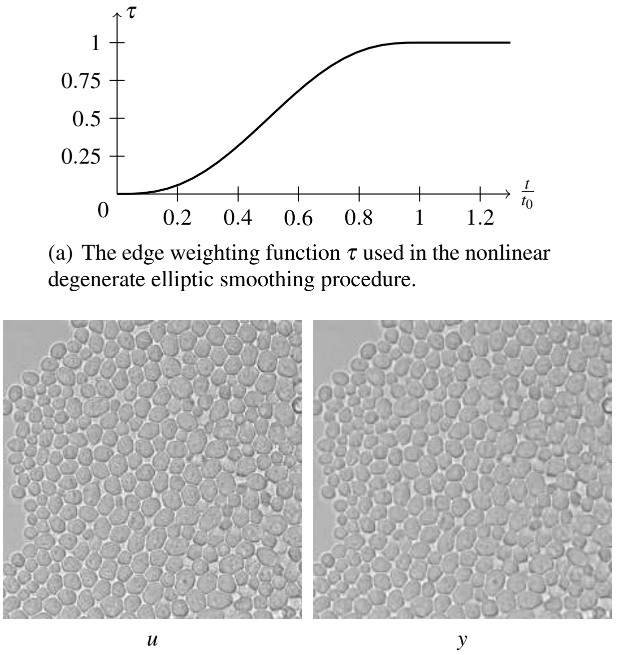
Fig. 2 Overview of the heterogeneous appearance of the yeast cells as well as detection error sources in transmission images.

of the main difficulties is that the cells appear on the microscopy image with high heterogeneity. Usually, there is interior structure present such as cell organelles and the depiction of the cell boundaries varies in intensity and contrast, see Figure 2. Hence, a method which is robust against these disturbances is needed. The algorithm we are proposing takes care of this requirement by applying, independently, two mathematical tools. The first is performing a suitable denoising step which is designed to preserve edges and thin structures like cell boundaries but to smooth out irrelevant details. This is done in order to provide some additional robustness for the subsequent active-contour algorithm which traces the boundary for a single cell. These two methods are now discussed in more detail.

2.1 Edge-preserving image denoising

Given a transmission image $u = u_{\mathrm{tr}}$, we would like to remove features which are below a certain size, as these are certainly not the cell boundaries. These features have much in common with noise, hence the task reduces to the denoising problem which is well-studied in mathematical image processing. Indeed, there is a variety of methods available [7, 2] and many of them give visually impressive results. Here, our focus is more on stability with respect to automated processing. We therefore employ a method based on partial differential equations. This method incorporates ideas from the Perona-Malik equation [16,6] as well as from anisotropic diffusion [22]. Specifically, it bases on the approach in [1], i.e., the solution of the *nonlinear degenerate elliptic equation*

$$\lambda y - \operatorname{div}\Big(\Big(I - \tau(|\nabla y_\sigma|)\frac{\nabla y_\sigma}{|\nabla y_\sigma|} \otimes \frac{\nabla y_\sigma}{|\nabla y_\sigma|}\Big)\nabla y\Big) = u \qquad (1)$$



(a) The edge weighting function τ used in the nonlinear degenerate elliptic smoothing procedure.

(b) The solution y of the nonlinear degenerate elliptic equation for a sample image u. Parameters: $\lambda = 10^{-3}$, $t_0 = 0.1$, image values are scaled to the range $[0,1]$.

Fig. 3 Illustrations of the data and parameters for the denoising procedure.

with respective homogeneous Neumann boundary conditions. Moreover, $\lambda > 0$ is a positive (small) parameter, $y_\sigma = y * G_\sigma$ denotes a smoothed version of y, (G_σ is, for instance, the Gauss function with variance $\sigma > 0$), $\tau : [0,\infty[\to [0,1]$ a monotone weighting function attaining 1 for some $t_0 > 0$, for instance

$$\tau(t) = \begin{cases} \left(\frac{t}{t_0}\right)^3 \left(\frac{t}{t_0}\left(6\frac{t}{t_0} - 15\right) + 10\right) & \text{for } 0 \leq \frac{t}{t_0} \leq 1, \\ 1 & \text{for } 1 < \frac{t}{t_0}, \end{cases} \qquad (2)$$

see Figure 3(a), and $a \otimes b$ denotes the tensor product (or dyadic product) of the two vectors a and b which is a rank-one matrix.

The diffusion tensor in (1) is designed such that in regions with edges, that is where $|\nabla y_\sigma|$ is large, it becomes degenerate, i.e., no diffusion in the direction of the edge, that is ∇y_σ, takes place. This leads to a preservation of the edge. There is only (anisotropic) diffusion in its tangential direction which has a regularizing effect on the shape of the edge and prevents irregular boundaries. Moreover, in flat regions where $|\nabla y_\sigma|$ is small, (1) induces nearly isotropic diffusion and leads, consequently, to the solution being smooth there. For more details on this model and a theoretical justification, we refer to [1].

In Figure 3(b), the effect of the nonlinear degenerate elliptic smoothing filter is depicted for a sample image. The solution of (1) is integrated into the image-processing pipeline as preprocessing step. The numerical implementation is based on the usual finite-difference discretization and a simple fixed-point iteration for the nonlinearity which fixes a y_σ, solves (1) (which is now linear) and hence, obtains a new y_σ. Although it might also be possible to employ more efficient algorithms like Newton's method, it turns out that this fixed-point iteration is sufficient for our purposes as one does not need a large number of iterations to obtain satisfactory results. Details on the interaction with the rest of the segmentation method can be found in Section 3.

2.2 Object boundary detection

The crucial step for obtaining information about the cells in u_{tr} is, of course, the detection of their boundaries. For each cell, we are interested in a function γ which parametrizes its boundary. This is also a well-studied task in mathematical image processing and usually solved by a so-called *active contour* or snake model. The standard snake model, introduced in [11], is usually modified in order to meet certain properties, for instance geometric criteria [4,5] or good detection of certain classes of objects [24]. Following this approach, we propose a functional based on the original snake model which also incorporates cell-detection specific requirements.

Recall the active contour model introduced in [11]: The sought parametrization γ is given by the solution of the minimization problem

$$\min_{\gamma:[0,1]\to\mathbf{R}^2} F(\gamma),$$
$$F(\gamma) = \int_0^1 g\big(\gamma(t)\big) + \alpha|\gamma'(t)|^2 + \beta|\gamma''(t)|^2 \,\mathrm{d}t \tag{3}$$

where g is a non-negative edge-penalization image representing edges with lower values and $\alpha,\beta > 0$ are regularization parameters. The model reflects two objectives: First, $\int_0^1 g\big(\gamma(t)\big)\,\mathrm{d}t$ penalizes to which amount γ lies not on an edge and $\int_0^1 \alpha|\gamma'(t)|^2 + \beta|\gamma''(t)|^2\,\mathrm{d}t$ penalizes the irregularity of the contour. The minimization of the sum therefore yields, under suitable choices of α and β, a smooth parametrization γ along an edge indicated by g. In practice, a numerical algorithm is employed which takes an initial γ^0 to a local minimum of F.

For cell-boundary detection within dense populations, it is necessary to modify (3): On the one hand, we like to detect closed curves, on the other hand, we have to approach the cell boundary by a initial γ^0 from the inside in order to register only a single cell at once. The former can be handled in a straightforward manner by considering periodic parametrizations $\gamma: S^1 \to \mathbf{R}^2$ with S^1 denoting the unit circle. For the latter, however, we have to introduce an additional term in F, since the penalization of only γ' and γ'' forces γ to a constant and hence, the numerical algorithm to find a contour of an object which is usually smaller than the initial curve. Approaching from the inside, however, means that the curve has to grow at some stage. We therefore propose the following functional to minimize:

$$\min_{\gamma:S^1\to\mathbf{R}^2} F(\gamma),$$
$$F(\gamma) = \int_{S^1} g\big(\gamma(t)\big) + \frac{\alpha}{2}|\gamma'(t)|^2\,\mathrm{d}t - \Big(\frac{\beta}{2}\int_{S^1} \gamma(t)\cdot\gamma'(t)^\perp\,\mathrm{d}t\Big)^p \tag{4}$$

where g is again the edge-penalization image, $\alpha,\beta > 0$ are regularization parameters, $\gamma'(t)^\perp = \big(-\gamma_2'(t), \gamma_1'(t)\big)$ and $p \in \,]0,1[$. Moreover, we are only considering feasible parametrizations γ which describe a non self-intersecting contour with positive volume and for which $(\gamma')^\perp$ points into the respective outward normal direction. While neglecting the second derivative of γ, the functional is designed in order to meet three objectives:

1. $\int_{S^1} g\big(\gamma(t)\big)\,\mathrm{d}t$ again accounts for how much γ is in an edge of the image (g is the edge-penalization image in which edges are represented by low values).
2. $\frac{\alpha}{2}\int_{S^1}|\gamma'(t)|^2\,\mathrm{d}t$ enforces again regularity of γ, ensuring that the contour results in a smooth curve.
3. $-\big(\frac{\beta}{2}\int_{S^1}\gamma(t)\cdot\gamma(t)^\perp\,\mathrm{d}t\big)^p$ measures the weighted negative volume of the area surrounded by the curve taken to the power p. It balances with the regularity term, such that the cells neither become too small or too large.

In particular, the third term, which is a negative volume term, forces an initial contour to grow when it spans a small volume. With increasing volume, it becomes less influential due to $p < 1$ and is therefore suitable for the approximation of a cell boundary from the inside.

Let us briefly discuss some numerical aspects of finding a (local) minimizer for (4). In the continuous setting, by differentiating F, one obtains the first-order necessary optimality conditions

$$F'(\gamma) = \nabla g \circ \gamma - \alpha\gamma'' - \beta\gamma'^\perp p\Big(\int_{S^1}\gamma(t)\cdot\gamma'(t)^\perp\,\mathrm{d}t\Big)^{p-1} = 0$$

whose solutions can, of course, be approximated numerically. In the implementation we chose, however, first to discretize the optimization problem (4) and then to solve by a gradient descent algorithm. The contour γ is discretized by periodic linear spline elements while g is represented by piecewise bilinear functions resulting from tensor-products of linear splines. Integration over g as well as ∇g can then be done in an exact manner and, consequently, the evaluation of $F'(\gamma)$. Our implementation performs a gradient descent algorithm, by successively evaluating $F'(\gamma)$ and searching a step-size s which ensures a sufficient decrease of F when

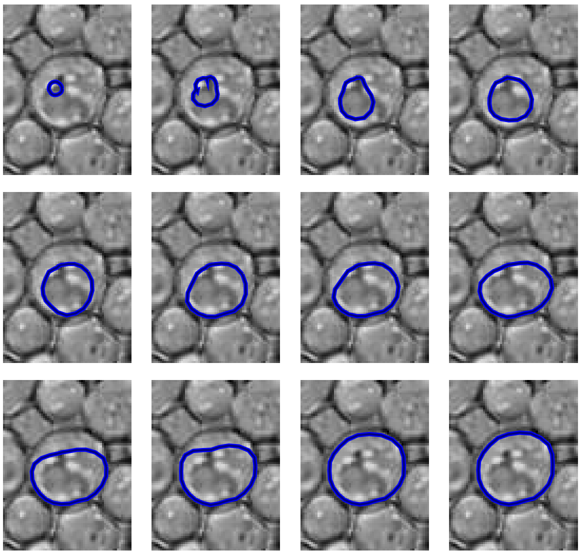


Fig. 4 The active contour method (4) for a single cell with complex interior structure. The initial contour (top left) is evolved by gradient descent steps (from left to right, top to bottom, some iterations are skipped) until a stationary point is reached (bottom right).

plugging in $\gamma - sF'(\gamma)$. This is repeated until the decrease of F is below a certain tolerance or a maximal number of iterations is reached. For a sample cell, the evolution of an initial curve is depicted in Figure 4. Typically, the algorithm stops after maximally 300 descent steps.

3 Description of the algorithm

The overall cell segmentation and data evaluation method consists of three independent parts: A preprocessing part aiming at performing image enhancements steps in order to improve robustness of the segmentation, the segmentation part where the majority of the work is done, and a postprocessing part computing cell-based characteristic values for the provided data. Each of the parts themselves incorporate several different techniques, which will be described in detail in the subsequent text. The pseudo-code algorithm stated in Algorithm 1 (Table 1) gives an overview over each step taken and the interdependencies of the given and computed data.

3.1 Preprocessing steps

*Background normalization (*NORMALIZEBACKGROUND*)*

Removal of inhomogeneities from the background is done by median filtering the input image with a mask roughly about the cell size. The outcome of this filtering detects the background inhomogeneities sufficiently well and subtraction from the original image yields the result.

Algorithm 1 Cell segmentation method

```
procedure CELLSEGMENTATION($u_{\mathrm{tr}}$, $u_{\mathrm{fl}}$)   ▷ $u_{\mathrm{tr}}$ transmission image
                                                   ▷ $u_{\mathrm{fl}}$ fluorescence image
    Begin preprocessing
        $u_{\mathrm{tr}} \leftarrow$ NORMALIZEBACKGROUND($u_{\mathrm{tr}}$)
                  ▷ remove low-frequency inhomogeneities from $u_{\mathrm{tr}}$
        $u_{\mathrm{bg}} \leftarrow$ DETECTBACKGROUND($u_{\mathrm{tr}}$)
                  ▷ detect cell-free regions
        $u_{\mathrm{tr}}^2 \leftarrow$ ELLIPTSOLVE($u_{\mathrm{tr}}$)
                  ▷ compute non-linear elliptic degenerate denoising
    End
    Begin segmentation
        $S \leftarrow$ FINDSEEDS($u_{\mathrm{tr}}^2, u_{\mathrm{bg}}$)
                  ▷ get starting points for active contour algorithm
        $g_1 \leftarrow$ EDGEIMAGE($u_{\mathrm{tr}}, u_{\mathrm{bg}}$)
        $g_2 \leftarrow$ EDGEIMAGE($u_{\mathrm{tr}}^2, u_{\mathrm{bg}}$)
                  ▷ compute edge penalization images
        for $s \in S$ do
            $(\gamma_s, F_s) \leftarrow$ ACTIVECONTOUR($s, g_1, g_2$)
                  ▷ for each seed, perform the active contour algorithm
                  ▷ $\gamma_s$ contour, $F_s$ quality measure
        end for
        $u_{\mathrm{label}} \leftarrow 0$                         ▷ label image
        for $s \in S$ such that $F_s$ is ascending do
            $u_{\mathrm{label}} \leftarrow$ ADDCELL($u_{\mathrm{label}}, \gamma_s$)
                  ▷ add cell region to $u_{\mathrm{label}}$, if it does not intersect
                  ▷ best ones are taken first
        end for
    End
    Begin postprocessing
        for each label $l$ in $u_{\mathrm{label}}$ do
            $v_l \leftarrow$ EVALUATECELL($l, u_{\mathrm{label}}, u_{\mathrm{fl}}$)
                  ▷ compute characteristic values for each cell
        end for
    End
    return $v$
end procedure
```

Table 1 The cell segmentation method in pseudo-code.

*Detection of cell-free regions (*DETECTBACKGROUND*)*

In order to disallow cell-free regions being detected by the active contour algorithm, these have to be identified. This is done by masking the parts of the image in which the gradient is below a certain threshold λ_{db1}. The resulting binary image is subject to a morphological erosion operation. Finally, the connected components of the resulting image are obtained and only the components whose area is above the threshold λ_{db2} are kept. Output is a binary image masking the cell-free regions.

*Nonlinear degenerate elliptic smoothing (*ELLIPTSOLVE*)*

In order to reduce the influence of noise and unwanted structures in a given image u, the edge-preserving smoothing method described in Subsection 2.1 is applied. The procedure ELLIPTSOLVE simply plugs the noisy input image u into the equation and obtains the denoised version y from its numerical solution.

3.2 Segmentation steps

*Starting points for the active contour algorithm (*FINDSEEDS*)*

The active-contour algorithm for the detection of a single cell boundary requires a starting point inside each cell. There has to be at least one of these points inside each cell and no starting points in the cell-free regions. This is done by finding the local maxima for a smoothed version of the edge-penalization image g given by

$$g(x) = \frac{1}{1 + \frac{|\nabla u(x)|^2}{\lambda^2}} \tag{5}$$

and bases on the observation that cell boundaries locally yield low values in g. Since g is given on a regular grid, this can be done by pointwise comparison of each value with the values in the neighboring points. A local maximum is by definition surrounded by values which are locally not greater and usually correspond to a point (ideally) in the center of a cell. The points obtained by this procedure are clustered such that a minimal distance is ensured and returned after checking that none of the points lie in the previously determined cell-free-region mask.

*Creating a robust edge penalization image (*EDGEIMAGE*)*

In addition to the starting points, the active contour algorithm requires information how to penalize areas which are not belonging to an edge. This is again done by applying (5) and performing some smoothing steps. Edge images are computed from both the original and the smoothed image. Finally, additionally penalization is added for the cell-free regions in order to avoid the active contour entering there.

*Trace the boundary of a single cell (*ACTIVECONTOUR*)*

For each "seed" s, an active contour is initiated around s and evolved according to the optimization problem (4) described in Subsection 2.2. The edge-penalization image $g = g_2$ is obtained from the smoothed image u_{tr}^2 via EDGEIMAGE. After meeting the stopping criterion, the procedure is repeated with the edge penalization g_1 from original image u_{tr} and returns the contour γ as well as the functional value $F(\gamma)$.

3.3 Postprocessing steps

*Compose a labelled image (*ADDCELL*)*

After the contour tracing is finished, the method creates a label image which assigns each pixel a positive natural number unique to each cell and 0 to the background. Starting with the contour associated with the lowest functional value for (4), the region occupied by the contour is determined and checked against the label image for intersection. If the area of intersection is too large, the current region is rejected, otherwise, it is incorporated into the label image with a new number. In the resulting label image, each cell can be identified through the corresponding pixels.

*Evaluate cell geometry and fluorescence data (*EVALUATECELL*)*

The last postprocessing step consists of computing some characteristic values for each cell and its associated fluorescence data and storing them in a list. These can be classified into values reflecting cell geometry and values associated with the fluorescence data. The values regarding cell geometry currently implemented are:

- center of mass,
- area and perimeter,
- width and height,
- spheroid volume and surface area.

The corresponding values extracted from the fluorescence data are:

- mass and perimeter,
- mean mass and perimeter,
- expectation and covariance,
- a scattering ratio.

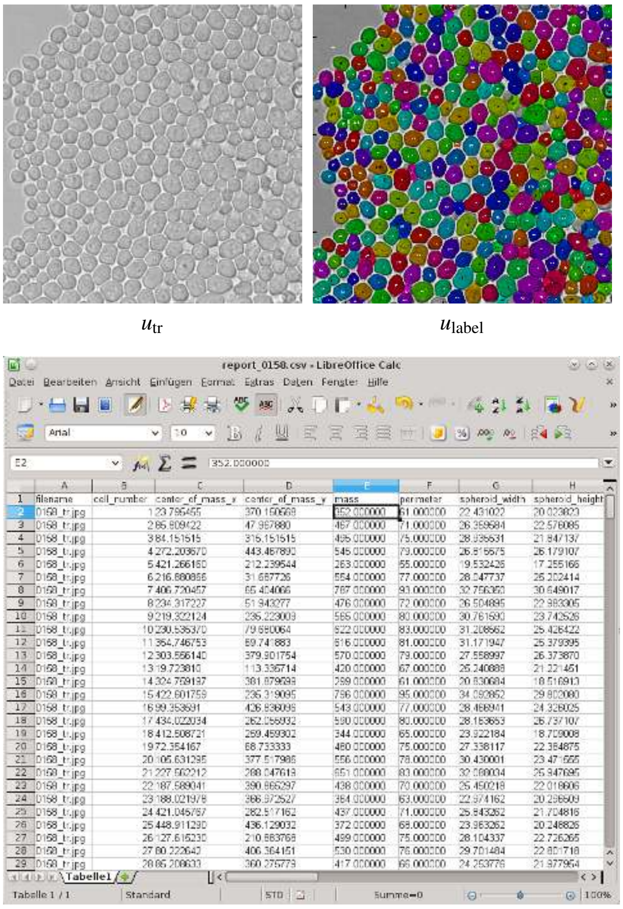
The corresponding output is a list of comma-separated values which can easily be imported into statistical evaluation software.

4 The software tool

The algorithm described in the previous section has been implemented in MATLAB [19]. The software is available for download at `http://microscopy.uni-graz.at/`. It comprises a graphical interface which allows for batch processing of large data sets as well as the configuration of the algorithm parameters. From the application point of view, the typical usage is specifying the data on which the cell segmentation should be performed (with default parameters) and starting the data processing pipeline which does not require further user interaction. For this task, no knowledge of the details of the algorithm is required. See Figure 5 for a sample output produced by the software.

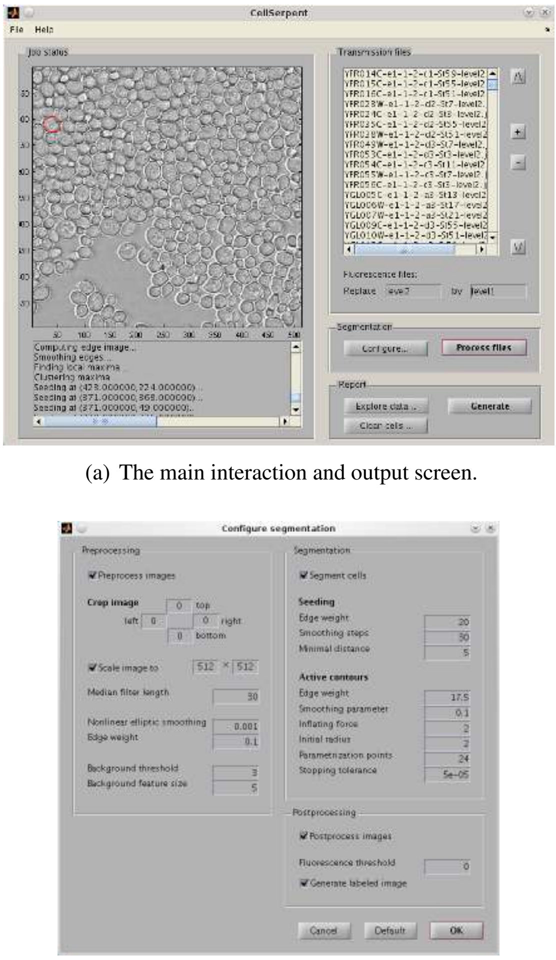
The parameters can, however, be changed on demand through a configuration dialog.

Figure 6(a) gives an impression of the software tool. The "transmission files" dialog can be used to specify the transmission image and their order for which the segmentation has to be carried out. Common image formats such as BMP, GIF, JPEG, PNG and TIF are supported as well as the MATLAB export generated by the visualization software AMIRA


u_{tr}
u_{label}
report_0158.csv - LibreOffice Calc
Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster Hilfe
Arial 10
E2
352.000000
Tabelle1
Tabelle 1 / 1
Standard
STD
Summe=0
100%

	A	B	C	D	E	F	G	H
1	filename	cell_number	center_of_mass_x	center_of_mass_y	mass	perimeter	spheroid_width	spheroid_height
2	0158_tr.jpg	1	23.796455	370.150968	352.000000	61.000000	22.431022	20.023823
3	0158_tr.jpg	2	85.809422	47.967880	467.000000	71.000000	26.359584	22.576085
4	0158_tr.jpg	3	84.151515	316.151515	495.000000	75.000000	28.936631	21.847137
5	0158_tr.jpg	4	272.203670	443.467890	545.000000	79.000000	26.815675	26.179107
6	0158_tr.jpg	5	421.266160	212.239544	263.000000	55.000000	19.532426	17.255166
7	0158_tr.jpg	6	216.880866	31.687726	554.000000	77.000000	28.047737	25.202414
8	0158_tr.jpg	7	406.720457	65.404066	767.000000	93.000000	32.756360	30.649017
9	0158_tr.jpg	8	234.317227	51.943277	476.000000	72.000000	26.504895	22.983305
10	0158_tr.jpg	9	219.322124	236.223009	565.000000	80.000000	30.761590	23.742526
11	0158_tr.jpg	10	230.535370	79.680064	622.000000	83.000000	31.208562	25.426422
12	0158_tr.jpg	11	364.746753	69.741883	616.000000	81.000000	31.171947	25.379395
13	0158_tr.jpg	12	303.556140	379.901754	570.000000	79.000000	27.558997	26.373870
14	0158_tr.jpg	13	19.723810	113.336714	420.000000	67.000000	25.240888	21.221451
15	0158_tr.jpg	14	324.759197	381.879599	299.000000	61.000000	20.830684	18.516913
16	0158_tr.jpg	15	422.601759	236.319095	796.000000	95.000000	34.092852	29.802080
17	0158_tr.jpg	16	99.353591	426.836096	543.000000	77.000000	28.466941	24.328025
18	0158_tr.jpg	17	434.022034	262.056932	590.000000	80.000000	28.153653	26.737107
19	0158_tr.jpg	18	412.508721	259.459302	344.000000	65.000000	23.922184	18.709008
20	0158_tr.jpg	19	72.354167	68.733333	480.000000	75.000000	27.338117	22.384875
21	0158_tr.jpg	20	105.631295	377.517986	556.000000	78.000000	30.430001	23.471565
22	0158_tr.jpg	21	227.562212	288.047619	651.000000	83.000000	32.089034	25.947695
23	0158_tr.jpg	22	187.589041	390.865297	438.000000	70.000000	25.450218	22.018606
24	0158_tr.jpg	23	188.021978	386.972527	364.000000	63.000000	22.974162	20.296509
25	0158_tr.jpg	24	421.045767	282.517162	437.000000	71.000000	25.843262	21.704816
26	0158_tr.jpg	25	448.911290	436.129032	372.000000	68.000000	23.963262	20.248826
27	0158_tr.jpg	26	127.615230	210.883768	499.000000	75.000000	28.104337	22.726265
28	0158_tr.jpg	27	80.222640	406.364151	530.000000	76.000000	29.701484	22.801718
29	0158_tr.jpg	28	85.208633	360.275779	417.000000	66.000000	24.253776	21.977954

Fig. 5 Top: A labelled transmission image according to the outcome of the segmentation algorithm. Each colored region and number corresponds to one segmented cell (yeast mutant strain YGR259C). Bottom: The characteristic values computed in the postprocessing step for each cell imported in a spreadsheet program.



```
Computing edge image...
Smoothing edges...
Finding local maxima...
Clustering maxima
Seeding at (423.000000,224.000000)...
Seeding at (371.000000,368.000000)...
Seeding at (371.000000,49.000000)..
```

```
YFR014C-e1-1-2-c1-St59-level2
YFR015C-e1-1-2-c1-St55-level2
YFR016C-e1-1-2-c1-St51-level2
YFR028W-e1-1-2-d2-St7-level2.
YFR034C-e1-1-2-d2-St3-level2.
YFR035C-e1-1-2-d2-St55-level2
YFR038W-e1-1-2-d2-St51-level2
YFR049W-e1-1-2-d3-St7-level2.
YFR053C-e1-1-2-d3-St3-level2.
YFR054C-e1-1-2-c3-St11-level2
YFR055W-e1-1-2-c3-St7-level2.
YFR058C-e1-1-2-c3-St3-level2.
YGL005C-e1-1-2-a3-St13-level2
YGL006W-e1-1-2-a3-St17-level2
YGL007W-e1-1-2-a3-St21-level2
YGL009C-e1-1-2-d3-St55-level2
YGL010W-e1-1-2-d3-St51-level2
```

Fluorescence files:
Replace level2 by level1

(a) The main interaction and output screen.

Preprocessing

☑ Preprocess images

Crop image: top 0, left 0, right 0, bottom 0

☑ Scale image to 512 × 512

Median filter length 30

Nonlinear elliptic smoothing 0.001
Edge weight 0.1

Background threshold 3
Background feature size 5

Segmentation

☑ Segment cells

Seeding
Edge weight 20
Smoothing steps 30
Minimal distance 5

Active contours
Edge weight 17.5
Smoothing parameter 0.1
Inflating force 2
Initial radius 2
Parametrization points 24
Stopping tolerance 5e-05

Postprocessing

☑ Postprocess images

Fluorescence threshold 0

☑ Generate labeled image

(b) The configuration dialog.

Fig. 6 The graphical user interface of the cell-segmentation tool.

[20]. For the evaluation of fluorescence data, a file name replacement rule can be set which is applied to find the fluorescence data associated with a transmission image (e.g., a for the transmission image `wt7trans.tif` and fluorescence image `wt7bin.mat`, the rule `trans.tif` → `bin.mat` has to be specified). The segmentation algorithm can be run and configured by "process files" and "configure" respectively. During the segmentation, status graphics and messages are displayed in "job status". Finally, the evaluation of the cell geometry and fluorescence data for all specified files can be done by "generate", which allows to choose from a detailed report listing each cell and a report which aggregates information for each image, both in CSV format.

The configuration dialog (Figure 6(b)) offers control over the algorithm described in Section 3. In detail, the following parameters can be adjusted:

- Preprocessing
  - *Preprocess images:* Disables preprocessing steps when unchecked,
  - *Crop image:* Allows to crop the transmission image, e.g., to get rid of labels,
  - *Scale image:* Optionally resizes the transmission image to the given dimensions,
  - *Median filter length:* Specifies the size of the median filter in NORMALIZEBACKGROUND,
  - *Nonlinear elliptic smoothing/edge weight:* The parameters λ and t_0, respectively, in (1) and (2),
  - *Background threshold/feature size:* Parameters $\lambda_{\mathrm{db}1}$ and $\lambda_{\mathrm{db}2}$ for DETECTBACKGROUND, respectively,
- Segmentation
  - *Segment cells:* Disables segmentation steps when unchecked,
  - *Seeding*
    - *Edge weight:* The value of the edge-penalization parameter λ in (5) in FINDSEEDS,
    - *Smoothing steps:* The number of linear smoothing steps performed in FINDSEEDS,
    - *Minimal distance:* Specifies the minimal distance under which seeds are no longer clustered,

  - *Active contour*
    - *Edge weight:* The value of λ in (5) used in ACTIVECONTOUR,
    - *Smoothing parameter:* Specifies the smoothness parameter α in (4),
    - *Inflating force:* Specifies the parameter β in (4),
    - *Initial radius:* The radius of the initial circle in ACTIVECONTOUR,
    - *Parametrization points:* The number of points for the discretization of the parametrization γ,
    - *Stopping tolerance:* ACTIVECONTOUR terminates if the functional decrease drops below this value,
- Postprocessing
  - *Postprocess image:* Skips postprocessing steps when unchecked,
  - *Fluorescence threshold:* When non-zero, performs a binary thresholding of the fluorescence image with the given threshold,
  - *Generate labeled image:* When checked, produces an image showing the region and number of each segmented cell.

There are moreover two helper tools available which facilitate the evaluation of the results. The first one, “Clean cells” can be called after the data processing has finished and may be used to remove specific cells from the segmented data. This is useful, for example, if a falsely detected cell should be erased or if only certain cells should be evaluated. To prevent data from being accidentally removed, a backup copy of the current data file will be created. In the process of cleaning, all cells will be renumbered such that no gaps will occur. With “Explore data”, the second helper tool, a graphical output of the cell geometry and fluorescence data can be created. For each file, the values of the selected feature will be sorted in descending order and plotted, on the y-axis, against its position in the rearranged data on the x-axis. The plot also provides some basic statistical characteristics such as mean, variance and median. It is possible to either plot a graph for each file individually or one graph for all files.

Finally, the configuration options as well as the file list can be saved and restored through the respective items in the “File” menu. Moreover, short instructions explaining the use of the program as well as some more details are available through the “Help” menu.

5 Experiments

In order to test the validity of the method, we performed two numerical experiments. For the first experiment, a transmission/fluorescence image pair was considered for which the transmission image of size 460 × 460 was also manually segmented (see Figure 7). Based on this segmentation, the characteristic cell-based values were computed and compared to the results of the automated segmentation procedure for which the implementation took approximately 312 seconds with MATLAB release 2010b on an Intel® Core$^{\text{TM}}$2 Quad CPU Q9550 @ 2.83GHz. In the manually segmented image, 324 cells could be identified while the automated segmentation method yielded 337 cells. Assigning corresponding cells, we found that the proposed method was able to identify 307 cells ($\approx 94\%$) correctly with 30 false positive cells ($\approx 9\%$). For the 307 cells, the characteristic values were compared in terms of the weighted L^2-error, see Table 2. In summary, the error is quite low for most of the values. However, for the cell mass, a rather large error could be observed. This is due to the inherent difficulties in the comparison: since the cell boundaries are usually ill-defined, the manual segmentation might also be inaccurate. In this respect, the automated algorithm tends to overestimate the actual cell mass. Nevertheless, the latter has only a mild impact on the fluorescence mass as no fluorescent material appears outside of cells, resulting in a smaller error for this characteristic value.

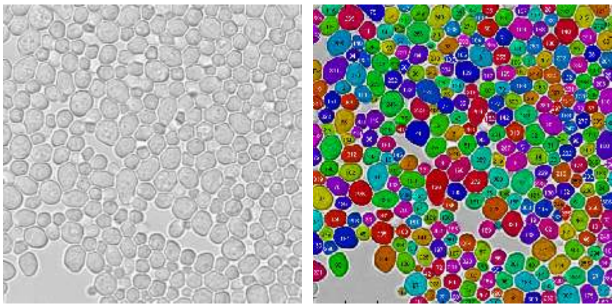
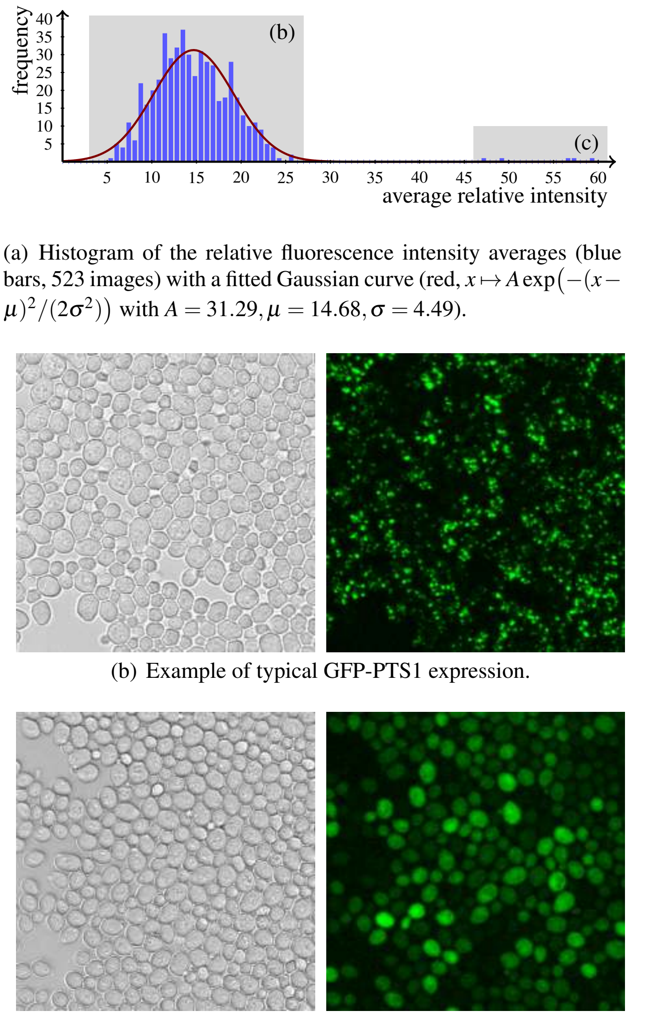


Fig. 7 A sample transmission image and its manual segmentation.

Characteristic value	Weighted L^2 error	
cell center of mass x	0.980722	pixels
cell center of mass y	1.373299	pixels
cell mass	83.750541	pixels
cell perimeter	8.044056	pixels
fluorescence mass	13.035571	pixels
fluorescence perimeter	8.079814	pixels
fluorescence mass/cell mass	0.014755	
fluorescence mass/perimeter	0.163840	

Table 2 The weighted L^2-error $E = \left(\frac{1}{N}\sum_{i=1}^{N}(v_i^{\text{manual}} - v_i^{\text{automatic}})^2\right)^{1/2}$ for a selection of characteristic values computed by the software.

In the second experiment, we performed the cell registration and quantitative analysis for 523 image pairs of size 460 × 460, each depicting a different yeast mutant colony. The fluorescence data describes the expression of the green fluorescent protein (GFP) fused to the peroxisomal targeting sequence 1 (PTS1), see [23] for details. It is known that this data set contains fluorescence images which correspond to yeast mutants with abnormal behavior. Our aim is to verify that these mutants can easily be found by evaluating the quantitative cell-based data produced by the algorithm.

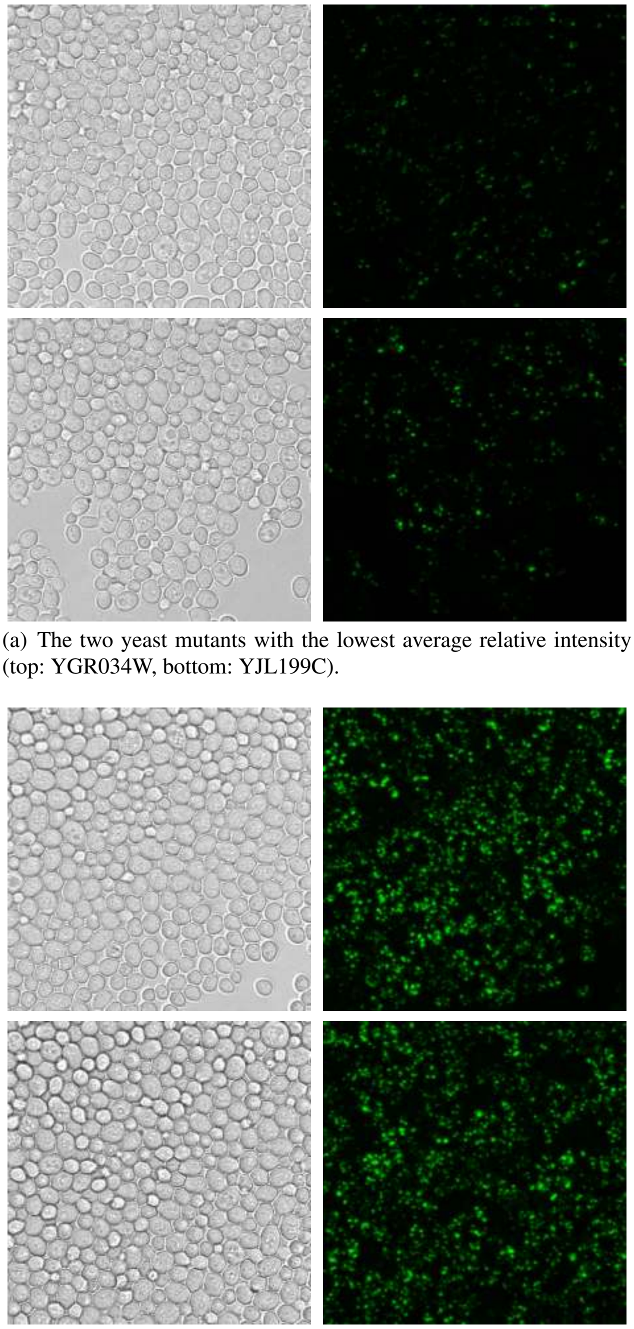


(a) Histogram of the relative fluorescence intensity averages (blue bars, 523 images) with a fitted Gaussian curve (red, $x \mapsto A\exp\big(-(x-\mu)^2/(2\sigma^2)\big)$ with $A = 31.29, \mu = 14.68, \sigma = 4.49$).

(b) Example of typical GFP-PTS1 expression.

(c) Example of abnormal GFP-PTS1 expression (outliers in the histogram).

Fig. 8 Quantitative evaluation of the average relative fluorescence intensity.



(a) The two yeast mutants with the lowest average relative intensity (top: YGR034W, bottom: YJL199C).

(b) The two yeast mutants with highest average relative intensity (top: YGR136W, bottom: YJL098W).

Fig. 9 The yeast mutant colonies associated with the extreme cases for the average relative intensity which are not outliers.

For this purpose, we chose to compute, for each image the average of the relative mass of fluorescence marker over all cells, i.e., the average, taken over all cells of an image, of the sum of the fluorescence intensity over the pixels of a cell divided by the pixel count of each cell. For qualitative evaluation and in order to identify outliers, these values have been put into a histogram, see Figure 8(a). The samples roughly follow a Gaussian distribution, a fact which was also observed in [23] where an image-based rather than a cell-based quantitative evaluation was performed. Furthermore, five outliers can be identified, corresponding to the yeast mutants YGL152C, YGL153W, YGR133W, YJL210W and YKL197C. By visual inspection, one sees that these correspond exactly to the fluorescence images which differ visually significantly from the rest (see Figures 8(b) and 8(c)).

Furthermore, the computed average relative intensity value can be used to distinguish those yeast mutants for which the average expression of GFP-PTS1 per cell is low and high, respectively, see Figure 9. Again, a visual inspection confirms the feasibility of this selection process. Overall, the cell-based quantitative evaluation confirms what might be deduced from visual inspection and image-based quantitative evaluation.

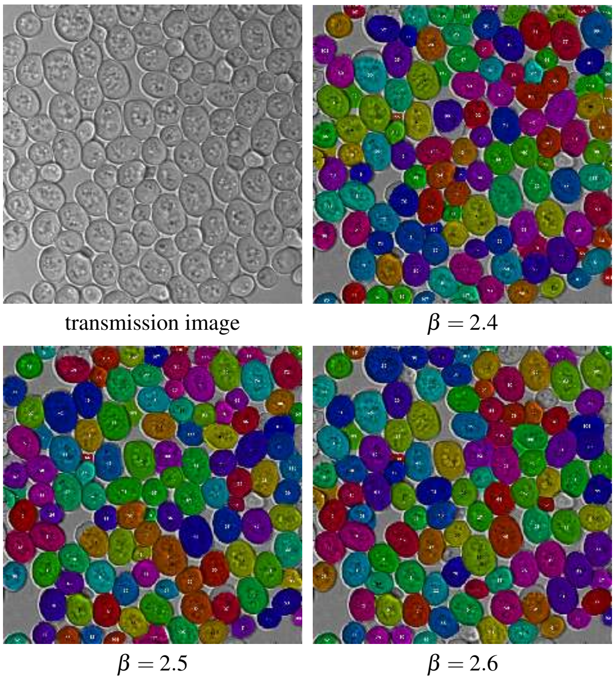


Fig. 10 The effect of varying the "inflating force" β on the outcome of the segmentation procedure. Top left: The test transmission image. Top right and bottom: The outcome of the segmentation procedure for different β.

Finally, we studied the behaviour of the algorithm for different parameters and image sizes. First of all, let us mention that the default parameters are in many cases already sufficient to obtain satisfactory results. The segmentation of the 523 test images was, for instance, performed with default parameters. In some cases, however, the parameters need to be adjusted in order to improve the outcome. Usually, there is only need to do this for a single image: once the parameters are fixed, they can be applied to a whole data set provided that the images are similar. For the segmentation of dense yeast colonies, the most relevant parameter appears to be the "inflating force" β in (4) which accounts for the cell size. To show the behaviour of this parameter, we ran the algorithm with different β on an image of size 384 × 384 showing large cells, see Figure 10 for the results. One can see that a rough tuning already yields satisfactory results for most cells. Choosing a smaller β leads to underestimation of the size for some large cells with complex interior structure while a larger β causes overestimation of the cell sizes and, consequently, the removal of smaller cells as these are ranked worse in terms of the objective functional F (also see the algorithm description in Section 3). The quality of the segmentation is, however, not affected for the majority of the cells, indicating that the procedure does not strongly depend on this parameter.

At last, we performed computations in order to see how image size affects the run time of the algorithm. The values can be found in Table 3. One can see that the method roughly scales linearly with respect to the number of pixels till 640 × 640. Beyond this size, time for data transfer which is often implicitly done in MATLAB gains more influence, resulting in a faster increase of the overall run time.

Image size (pixels)	Number of pixels	Run time (seconds)
128 × 128	16384	18.6
256 × 256	65536	67.4
384 × 384	147456	158.7
512 × 512	262144	315.1
640 × 640	409600	515.4
768 × 768	589824	834.0
896 × 896	802816	1453.5
1024 × 1024	1048576	2034.8

Table 3 Run times of the segmentation pipeline for different image sizes (MATLAB release 2010b, Intel® Core$^{\text{TM}}$2 Quad CPU Q9550 @ 2.83GHz).

6 Summary and conclusions

The segmentation of single cells in transmission images particularly of yeast cell populations is a challenging problem in mathematical image processing. The heterogeneity of the acquired image data, the varying characteristics of the cells and the density of cell populations demand high requirements on applied image processing algorithms.

In this paper, we proposed a variant of the active-contour model for effective detection of individual yeast cells in transmission images. Together with appropriate pre- and post-processing steps such as non-linear, edge adaptive smoothing, seeding and selection based on quality measures, a complete pipeline could be developed enabling registration of dense and heterogeneous yeast cell populations.

Furthermore, a MATLAB-based software tool has been developed in order to perform batch-processing and quantitative analysis both of transmission and 3D fluorescence data as acquired in a specific yeast screening approach. Notably, the software requires minimal user interaction to process the image data.

Experiments on real-life image sets confirmed the validity of the quantitative cell-based data obtained from the segmentation and evaluation process. We therefore expect that the proposed method and the provided software could be used to perform analysis on data admitting quantitative characteristics which are not easily deducible by visual inspection of image-based quantitative evaluation. Although our motivation is the segmentation of dense yeast cell colonies, it might moreover be possible that the method also works for other types of data sets. As the model bases on images which show, on a homogeneous background, a dense collection of round objects whose size varies in a certain range and which are separated by more or less sharp boundaries, it could be expected that data admitting these characteristics are also successfully segmented with similar quality. Exploring this could be a subject of further studies.

Currently, the method does not specifically differentiate between mother cells and their developing daughter cells, a feature appreciated for quantification of yeast cells during cellular growth. Moreover, graphical tools are limited to diagrams for comparative analysis of subcellular structures of different yeast mutants. In this respect, we currently implement 3D sharpness based methods to extract yeast buds and additional tools such as heat maps to further improve the statistical analysis of imaging-based yeast screening data.

Let us finally mention that in principle, the method could be extended to extract the 2D boundary surface of cells in 3D data sets, i.e., using active surfaces instead of active contours. However, this would require essential extensions in the model and a considerable effort in additional implementation for which efficiency issues will likely play a larger role than in the current approach. Again, this could be a direction of further research.

References

1. Bredies, K.: A variational weak weighted derivative: Sobolev spaces and degenerate elliptic equations. Tech. rep., University of Bremen (2008)
2. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Modeling and Simulation **4**(2), 490–530 (2005)
3. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P., Sabatini, D.M.: Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol **7**(10), R100 (2006)
4. Caselles, V., Catté, F., Coll, T., Dibos, F.: A geometric model for active contours in image processing. Numerische Mathematik **66**(1), 1–31 (1993)
5. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22**(1), 61–79 (1997)
6. Catté, F., Lions, P.L., Morel, J.M., Coll, T.: Image selective smoothing and edge detection by nonlinear diffusion. SIAM Journal on Numerical Analysis **29**(1), 182–193 (1992)
7. Chan, T.F., Shen, J.: Image Processing and Analysis - Variational, PDE, Wavelet, and Stochastic Methods. SIAM, Philadelphia (2005)
8. Hamilton, N.A., Pantelic, R.S., Hanson, K., Teasdale, R.D.: Fast automated cell phenotype image classification. BMC Bioinformatics **8**, 110 (2007)
9. Huh, S., Lee, D., Murphy, R.F.: Efficient framework for automated classification of subcellular patterns in budding yeast. Cytometry A **75**(11), 934–40 (2009)
10. Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., O'Shea, E.K.: Global analysis of protein localization in budding yeast. Nature **425**(6959), 686–91 (2003)
11. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision **1**(4), 321–331 (1988)
12. Kvarnstrom, M., Logg, K., Diez, A., Bodvard, K., Kall, M.: Image analysis algorithms for cell contour recognition in budding yeast. Opt Express **16**(17), 12,943–57 (2008)
13. Li, Z., Vizeacoumar, F.J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F.S., Min, R., Vandersluis, B., Bellay, J., Devit, M., Fleming, J.A., Stephens, A., Haase, J., Lin, Z.Y., Baryshnikova, A., Lu, H., Yan, Z., Jin, K., Barker, S., Datti, A., Giaever, G., Nislow, C., Bulawa, C., Myers, C.L., Costanzo, M., Gingras, A.C., Zhang, Z., Blomberg, A., Bloom, K., Andrews, B., Boone, C.: Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat Biotechnol **29**(4), 361–7 (2011)
14. Ohnuki, S., Oka, S., Nogami, S., Ohya, Y.: High-content, image-based screening for drug targets in yeast. PLoS One **5**(4), e10,177 (2010)
15. Ohya, Y., Sese, J., Yukawa, M., Sano, F., Nakatani, Y., Saito, T.L., Saka, A., Fukuda, T., Ishihara, S., Oka, S., Suzuki, G., Watanabe, M., Hirata, A., Ohtani, M., Sawai, H., Fraysse, N., Latge, J.P., Francois, J.M., Aebi, M., Tanaka, S., Muramatsu, S., Araki, H., Sonoike, K., Nogami, S., Morishita, S.: High-dimensional and large-scale phenotyping of yeast mutants. Proc Natl Acad Sci U S A **102**(52), 19,015–20 (2005)
16. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12**(7), 629–639 (1990)
17. Saito, T.L., Ohtani, M., Sawai, H., Sano, F., Saka, A., Watanabe, D., Yukawa, M., Ohya, Y., Morishita, S.: Scmd: Saccharomyces cerevisiae morphological database. Nucleic Acids Res **32**(Database issue), D319–22 (2004)
18. Saito, T.L., Sese, J., Nakatani, Y., Sano, F., Yukawa, M., Ohya, Y., Morishita, S.: Data mining tools for the saccharomyces cerevisiae morphological database. Nucleic Acids Res **33**(Web Server issue), W753–7 (2005)
19. The MathWorks Inc.: MATLAB, version 7.11.0 (R2010b) (2010)
20. Visage Imaging GmbH: Amira 5.3 Microscopy/ResolveRT (2011)
21. Vizeacoumar, F.J., van Dyk, N., Vizeacoumar, F.S., Cheung, V., Li, J., Sydorskyy, Y., Case, N., Li, Z., Datti, A., Nislow, C., Raught, B., Zhang, Z., Frey, B., Bloom, K., Boone, C., Andrews, B.J.: Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. J Cell Biol **188**(1), 69–81 (2010)
22. Weickert, J.: Anisotropic diffusion in image processing. Teubner, Stuttgart (1998)
23. Wolinski, H., Petrovic, U., Mattiazzi, M., Petschnigg, J., Heise, B., Natter, K., Kohlwein, S.D.: Imaging-based live cell yeast screen identifies novel factors involved in peroxisome assembly. J Proteome Res **8**(1), 20–7 (2009)
24. Xu, C., Prince, J.L.: Snakes, shapes and gradient vector flow. IEEE Transactions on Image Processing **7**(3), 359–369 (1998)