PLOS ONE

PombeX: Robust Cell Segmentation for Fission Yeast Transillumination Images

Jyh-Ying Peng[1,2,3]*, Yen-Jen Chen[1], Marc D. Green[4], Sarah A. Sabatinos[4], Susan L. Forsburg[4], Chun-Nan Hsu[5,6,7]*

1 Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, R.O.C., 2 Department of Education and Research, Taipei City Hospital, Taipei, Taiwan, R.O.C., 3 Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei, Taiwan, R.O.C., 4 Molecular and Computational Biology Program, University of Southern California, Los Angeles, United States of America, 5 Division of Biomedical Informatics, Department of Medicine, University of California, San Diego, La Jolla, United States of America, 6 Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C., 7 Information Sciences Institute, University of Southern California, Marina del Rey, California, United States of America

Abstract

Schizosaccharomyces pombe shares many genes and proteins with humans and is a good model for chromosome behavior and DNA dynamics, which can be analyzed by visualizing the behavior of fluorescently tagged proteins *in vivo*. Performing a genome-wide screen for changes in such proteins requires developing methods that automate analysis of a large amount of images, the first step of which requires robust segmentation of the cell. We developed a segmentation system, PombeX, that can segment cells from transmitted illumination images with focus gradient and varying contrast. Corrections for focus gradient are applied to the image to aid in accurate detection of cell membrane and cytoplasm pixels, which is used to generate initial contours for cells. Gradient vector flow snake evolution is used to obtain the final cell contours. Finally, a machine learning-based validation of cell contours removes most incorrect or spurious contours. Quantitative evaluations show overall good segmentation performance on a large set of images, regardless of differences in image quality, lighting condition, focus condition and phenotypic profile. Comparisons with recent related methods for yeast cells show that PombeX outperforms current methods, both in terms of segmentation accuracy and computational speed.

Introduction

Fission yeast (*S. pombe*) and humans share many orthologous genes required for the DNA damage response and maintenance of an intact genome [1]. We derived a set of yeast mutants from a collection of approximately 4,000 mutants [2], each of which contains two fluorescently tagged marker proteins: RPA-CFP and Rad52-YFP, in an attempt to identify mutations that create genome instability and change the distribution of these proteins. For each genotype (gene deletion) we acquired 3 to 5 sets of microscopic images with three channels: a transmitted illumination channel, a CFP channel (RPA-CFP/ssb1-CFP/rad11-CFP) and a YFP channel (Rad52-YFP/rad22-YFP). Details of this screen and the results will be presented elsewhere.

The challenge now is to analyze more than 20,000 image files (2 markers×3 images×~4,000 genes) to characterize representative phenotypes, and ultimately construct DNA repair pathways by correlating these quantitative phenotypes with known gene networks. The first task in the high content analysis of these images is to accurately segment cell boundaries, which may be obscured by inhomogeneous focus and contrast across a single image, due to imperfect alignment of substrate and the focal plane. Also images acquired in multiple sessions by different people may exhibit variant focus qualities, resulting in variable thickness and

intensity of the cell membrane in different images, and even different intensity gradients from background to cell interior. In this work, we present a robust cell segmentation system, PombeX, for *S. pombe* cells in transillumination images.

Methods

Our pombe cell segmentation method consists of the following steps: First we identify cell nucleus and background regions, which are then used to adjust focus gradient differences and enhance contrast between cell interior, membrane and background in the transmitted light image. Next a distance transform-based pixel classification method identifies approximate cell interiors to form an initial cell contour, and an approximate cell membrane edge map is generated to be used in the gradient vector flow (GVF) snake model. After obtaining the final GVF snake contour, machine learning classifiers are trained and employed to validate the cell contours. Figure 1 presents an overview of our method, in the following we present details of each step.

Nucleus and Background Region Detection

We first perform local shading correction of the transmitted illumination and fluorescent images by subtracting the original image by its average filtered image using a disk kernel of radius 40
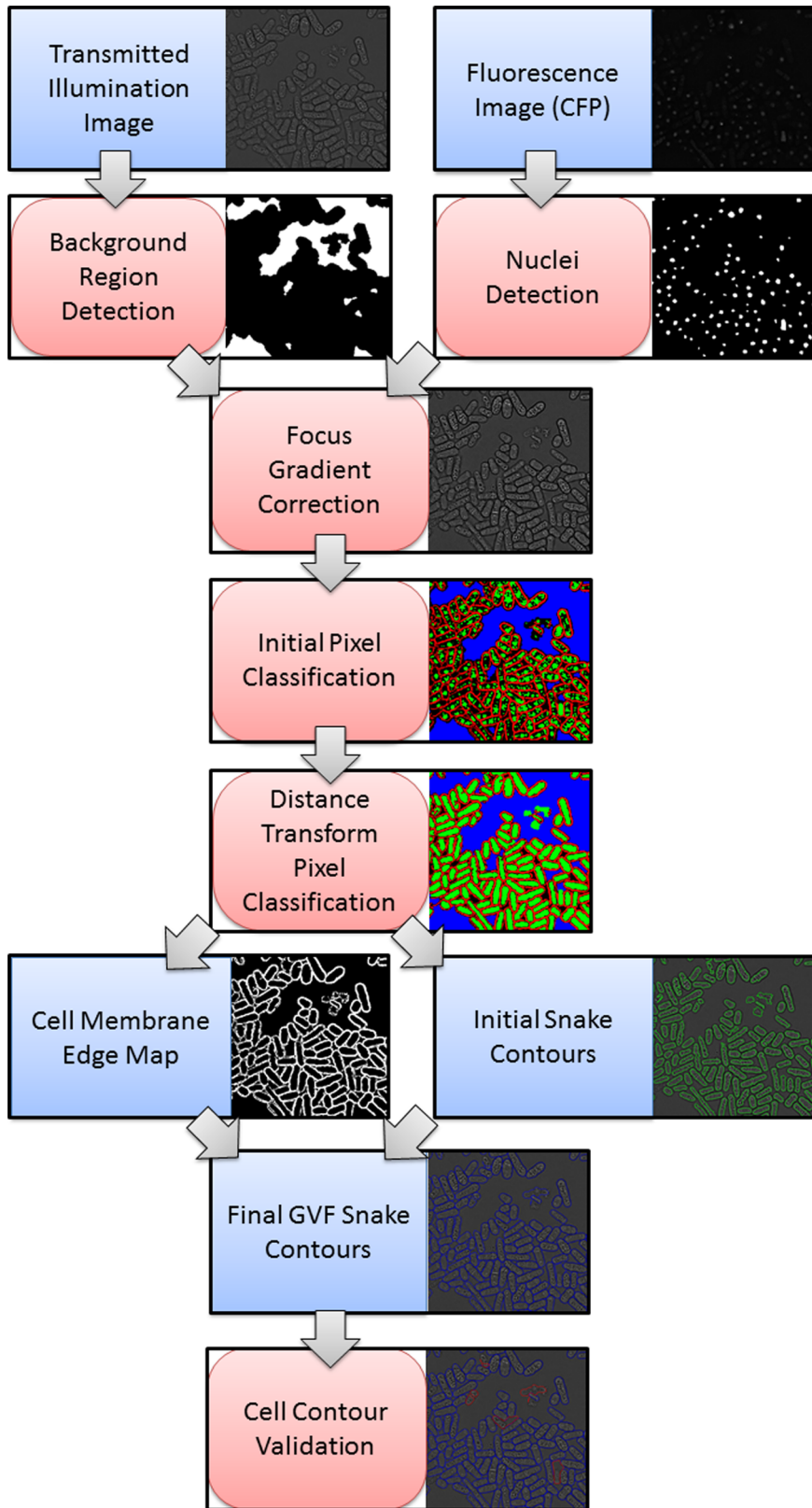
Figure 1. Overview of PombeX cell segmentation method.
doi:10.1371/journal.pone.0081434.g001

pixels. Then we detect the background in transmitted illumination images by Otsu's thresholding of its gradient image. The fluorescent signals are confined to the nucleus, so we further apply Otsu's thresholding on the fluorescent images to obtain the final nucleus boundaries. From these we can then calculate the average intensity of nucleus regions μ_N and background regions μ_B in the transmitted illumination image, which are then used to correct for the relative focus gradient across different images. If no nuclear fluorescent images are available, we set $\mu_N = \mu_B$ and skip the focus correction step.

Note that not all cells exhibit fluorescent signals in the nucleus, hence the intensity characteristics of the trans-illumination images are still needed to infer the position of non-fluorescent cells. Our system is able to correctly detect and segment most cells without requiring all cells to exhibit fluorescence.

Focus Gradient Correction

To derive our focus gradient correction procedure, we imaged pombe cells using changing focus positions, and investigated the intensity changes when we vary the focus position from above the coverslip and downward towards the specimen, going past optimal focus. We performed k-means clustering ($K = 3$) of image pixels using their intensity values at different focus positions as features. The top and middle plots in Figure 2 show that pixels are automatically clustered into cell interior, cell membrane and background regions. These three regions undergo different intensity changes through varying z-position (Figure 2, bottom plot). If we only look at the middle range (z-step 16–34) where the defocusing is not too severe, there seems to be a symmetry between the mean intensity of cell interior and cell membrane about mean background intensity. Thus we formulate a model where the average background intensity stays constant at μ_B, and the intensities of both the cell membrane and interior satisfies the relation

$$I^+ = 2\mu_B - I^-, \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \qquad (1)$$

where I^+ is the pixel intensity when positively defocused, and I^- is the intensity when negative defocused.

Thus to perform focus gradient correction, for each image we first calculate the average intensity of nucleus regions μ_N (representing cell interior intensity) and background regions μ_B in the transmitted illumination image, and then linearly adjust the intensity in the whole image so that μ_N–μ_B matches a predefined value (for the entire dataset), complementing the image if necessary. This procedure also enhances the overall contrast of the image. After this correction, most images will have cell interior regions consistently brighter than the background, and the cell membrane consistently darker. This facilitates the subsequent classification of pixels into cell interior, cell membrane and background, which serves as a basis in determining initial snake contours for each cell, and the gradient vector flow (GVF) edge map.

Some of the transmitted illumination images exhibit a focus gradient caused by the specimen plane not entirely in parallel to the focal plane, resulting in the intensity characteristics of cell interior and membrane changing across a single image. Consequently, the global focus correction presented in the last paragraph only works for part of the image. To analyze these images, we assume that the specimen plane is flat, and use a bilinear model to approximate the distance to the focal plane at each pixel position,

$$z = p_{00} + p_{10}x + p_{01}y, \qquad (2)$$

where z can be positive or negative depending on whether the specimen point is above or below the focal plane, and x, y is the specimen point position in pixels. To approximate z values, we make a second assumption that the intensity difference of nucleus to background is linearly related to z, which is justified when the defocusing is not too severe as shown in Figure 2.

Our adaptive focus gradient correction algorithm can be stated as follows:

1. If the image has focus gradient, then

    (a) Detect the location x, y of all nuclei and their intensity difference to background z.

    (b) Fit a bilinear model (eq. 2) to x, y, z.

    (c) For every pixel where $z < 0$, complement the pixel intensity i with respect to the average background intensity μ_B by $j = 2\mu_B - i$.

2. Globally and linearly adjust the intensity of the whole image so that the difference between average nucleus intensity (of the corrected image) and background intensity equal a predefined value.

Figure 3 shows how our method is applied and examples of corrected images. The original image contains a focus gradient effect where the average intensity of cell interior and membrane varies across the image and is higher than background intensity on one side of the image and lower on the other. After adaptive focus gradient correction, we can see that all image regions with sufficient contrast now consistently have the cell interior brighter than the background, and cell membrane darker, although the final position of detected cell membrane would vary somewhat. Since the ultimate goal is to detect fluorescence belonging to a single cell, it is tolerable to allow small fluctuations in detected membrane positions.

Cell Segmentation

We used active contour models [3] to obtain cell boundary contours. We first use adaptive thresholds [4] to classify pixels into four distinct sets: background, cell interior, cell membrane and ambiguous. Cell membrane pixels have intensity that are below an adaptive threshold calculated from all pixels darker than the background, whereas cell interior pixels have intensity greater than the average intensity of the nucleus region. Morphological close-opening and removable of isolated pixels were performed to smooth each region, then the remaining unclassified pixels X (ambiguous) were classified by a distance transform-based procedure (similar to watershed) that extends the background and cell interior regions to neighboring ambiguous regions:

1. Calculate distance transforms D_b of background set B, D_c of cell interior set C, and D_m of cell membrane set M.

2. $C \leftarrow C \cup \{x \in X | D_c(x) \leq D_b(x) \, \wedge \, D_c(x) \leq \frac{1}{2}D_m(x)\};$

    $B \leftarrow B \cup \{x \in X | D_b(x) < D_c(x) \, \wedge \, D_b(x)$
    $\leq \frac{1}{2}D_m(x) \wedge D_m(x) \leq D_c(x)\};$
    $$X \leftarrow X - (C \cup B).$$
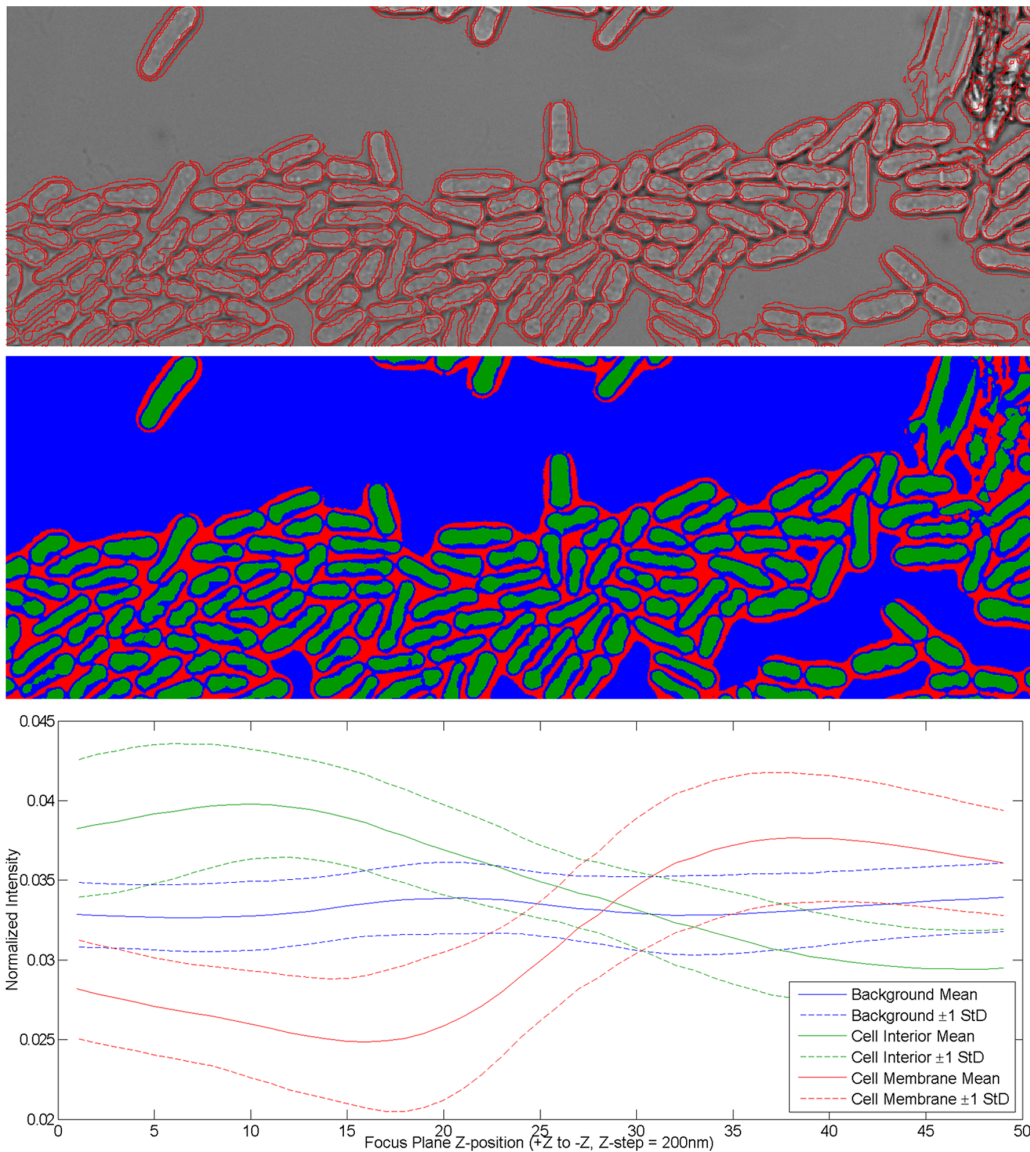
Figure 2. Analysis of pixel intensity variation of the same specimen imaged at different z-focus position. Top plot: original image at one z-position with red contours indicating k-means cluster boundaries. Middle plot: the same image with color-coded cluster membership for pixels. Bottom plot: Mean intensity of each pixel cluster through z-position.
doi:10.1371/journal.pone.0081434.g002

3. Repeat Steps 1 and 2 until X is unchanged.

Figure 4 shows some example results of this procedure. The initial snake contours are obtained from individual connected cell interior regions, while the cell membrane edge map is obtained from pixels classified as cell membrane.

Next we use the following snake model with energy functional E_{snake} for contour $\mathbf{x}(s) = [x(s), y(s)], s \in [0,1]$:

$$E_{snake} = \lambda_1 E_{int} + \lambda_2 E_{ext}, \qquad (3)$$

where λ_1, λ_2 controls relative weighting between internal and external energies. The conventional internal energy of the snake contour is

$$E_{int} = \frac{1}{2}\int_0^1 [\alpha|\mathbf{x}'(s)|^2 + \beta|\mathbf{x}''(s)|^2]ds, \qquad (4)$$

where the weighting parameters α and β control the snake's tension and rigidity, respectively. For the external image energy E_{ext}, we used the gradient vector flow (GVF) snake [5]. Given an edge map $f(x, y)$ derived from the image $I(x, y)$, the GVF field $\mathbf{v}(x, y) = [u(x, y), v(x, y)]$ minimizes the energy functional

$$E = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2|\mathbf{v} - \nabla f|^2 dxdy, \qquad (5)$$

where $\mathbf{v}$ denotes $-\nabla E_{ext}$, and f is the cell membrane edge map obtained previously. The parameter values used are $\lambda_1 = 1$, $\lambda_2 = 1$, $\alpha = 0.03$, $\beta = 0.2$ and $\mu = 0.8$. The computation of the gradient vector flow (GVF) field from the edge map, and the snake deformation iterations is as described in [5].

Cell Contour Validation

To further increase the robustness and efficiency of the system, we also employ machine learning classifiers to automatically
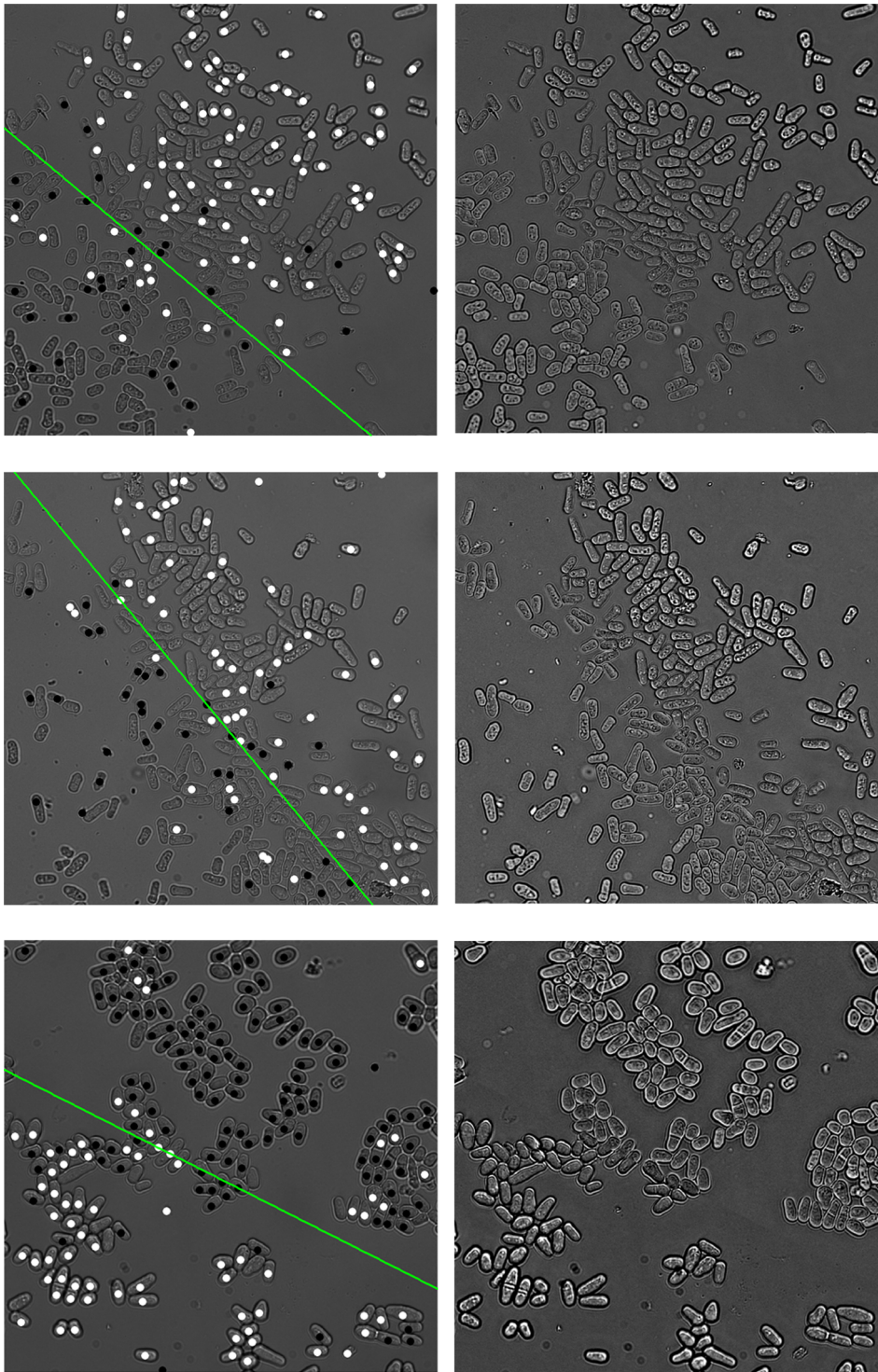
Figure 3. Examples of images before (left) and after (right) adaptive focus gradient correction and contrast enhancement. In the left panel the original images are overlaid with nucleus centroids (black and white spots) and the estimated $z = 0$ line (green line, where specimen plane intersects focal plane). For the nucleus centroids, black means its average intensity is lower than background, and white means it is brighter than background.

doi:10.1371/journal.pone.0081434.g003

validate the final snake contours. Specifically, we aim to detect errors such as merging of two or more cells, partial segmentation, and false detection of background or artifacts. Manually labeled training samples are generated for each type of error, and we used

support vector machines with radial basis function to learn and classify partial segmentation and merging errors. False detection of background and artifact are relatively rare and well distinguished from normal cells, thus a simple binary classification tree is trained
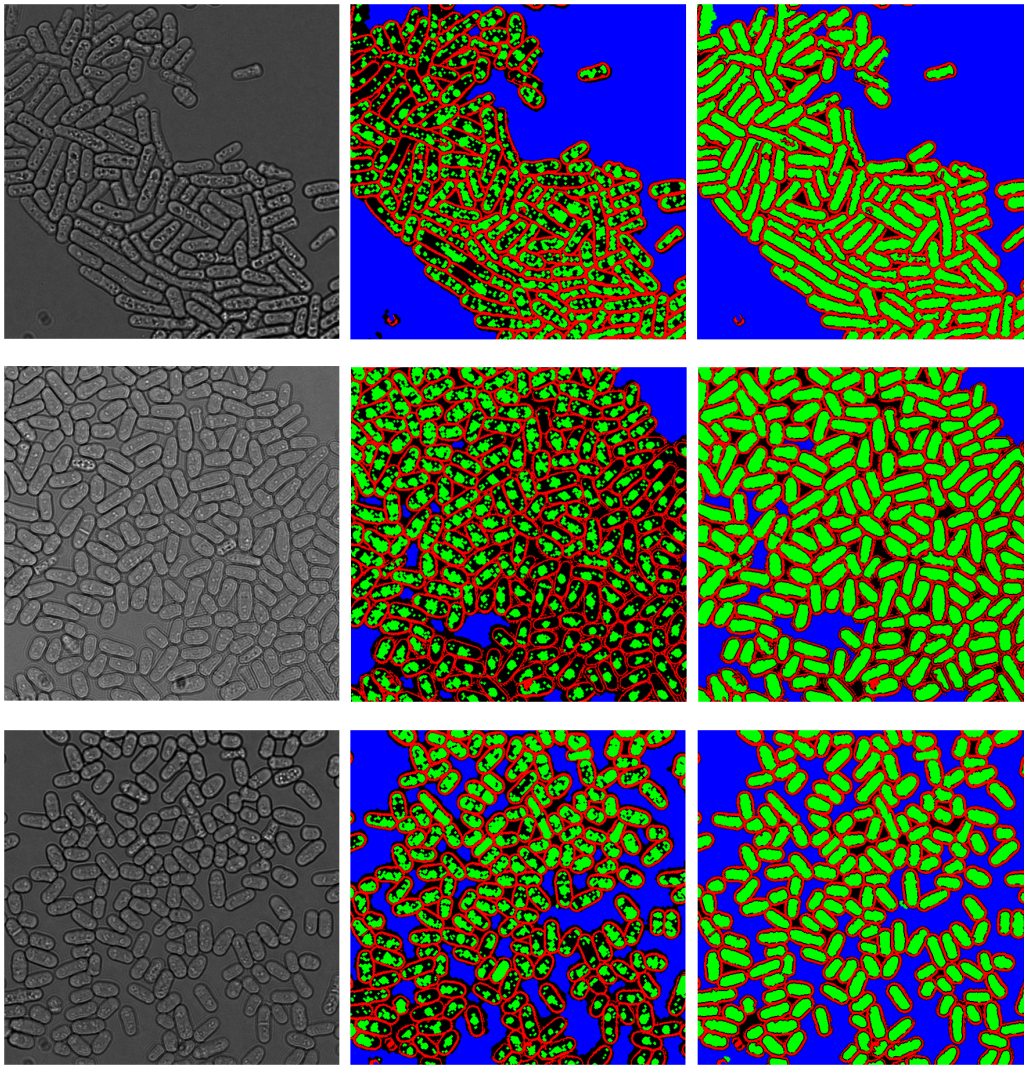
Figure 4. Example results of distance transform-based procedure. Original images (left), pixel classification before (middle) and after (right) distance transform-based procedure. Blue pixels indicate background, red is cell membrane and green is cell interior region.
doi:10.1371/journal.pone.0081434.g004

and used for this error type. Cells intersected by the image boundary share the same characteristics as a partially segmented cell, and are treated equivalently by our classifier.

Results

We applied PombeX to multi-channel images of *S. pombe*. The whole data set contains about 4,000 mutant genotypes, each with at least three sets of transmitted illumination (bright field), Rad52-YFP and RPA-CFP images. Our system is able to correctly segment a majority of cells in almost all images of sufficient quality. The performance is consistent over a wide variety of focal distance, field brightness, relative contrast and phenotypic characteristics. Moreover, some images contain dead cells with autofluorescence or no fluorescence, exhibited as enlarged nuclei or no nuclei, but our system is still able to correctly segment the transmitted illumination image given only a subset of nuclei positions. Figure 5 shows PombeX segmentation of a variety of phenotypes, including small and elongated cell shapes, cells in septation, and cells with various fluorescent signal profiles.

To further validate our method, we used heat treatment to induce phenotypic changes in mutant cells, and test the performance of PombeX on these classic phenotypes. In Figure 6, asynchronous wild type cells were taken from a late-exponential culture before treatment, causing a slightly smaller distribution of asynchronous wild type cell lengths. *mcm4ts* cells are an example of a classical cell division cycle (cdc) phenotype, which elongate during temperature treatment (*cdc21-M68* allele [6]). In contrast, *orp1ts* cells (*orp1-4* allele, the human ORC1 homologue [7]) do not replicate their DNA at 36°C, causing an accumulation of different cell sizes. The *rad4ts* (*cut5*) cells fail at both DNA replication and in the DNA damage checkpoint (*rad4-116* allele [8]). As a result they form small cells with a septum slicing through DNA in the centre of the cells, causing a cell untimely torn (cut) morphology. From the results we can see that PombeX can produce good segmentations for all of these phenotypes.

Finally, we performed quantitative evaluation and comparison with two recent automated segmentation algorithms for yeast cells, CellStat [9] and CellSerpent [10], using a set of hand produced gold standard segmentations of pombe cells, representing different image acquisition conditions and quality. CellStat first finds
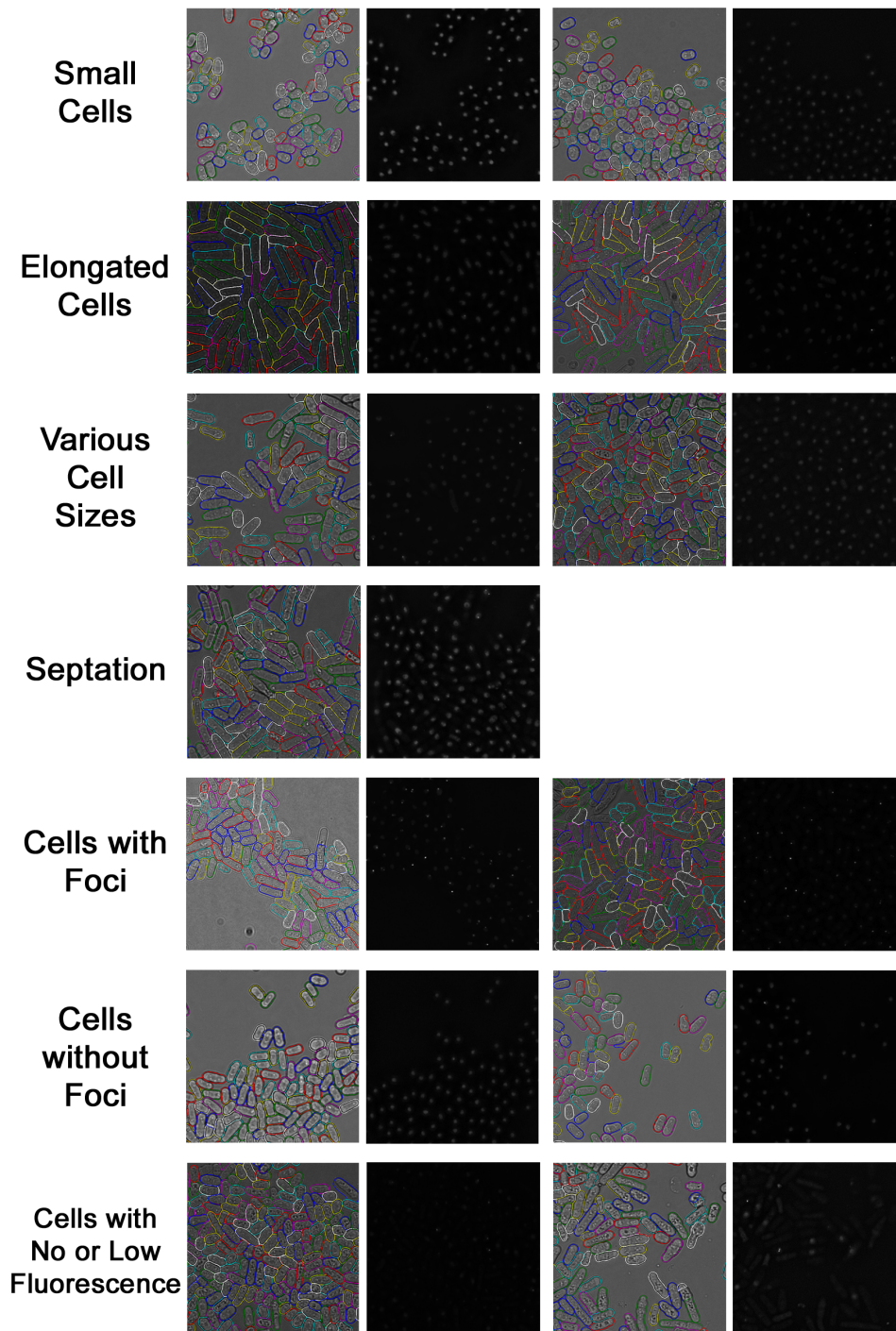
Figure 5. Representative segmentation results on various pombe mutants. For each image pair, trans-illumination images with randomly colored cell contours are shown on the left, and the corresponding RPA-CFP images on the right.
doi:10.1371/journal.pone.0081434.g005

candidate cell centers using a variation of the circular Hough transform, and then cell contours surrounding the candidate cell centers are extracted by directional derivatives and curve matching. CellSerpent used nonlinear degenerate elliptic smoothing to remove small features that are not cell membranes, generating an edge penalization image. Starting from local maxima in the edge penalization image, active contour models are used to find the final cell contours.

We evaluated the percentage of cells detected, the segmentation accuracy of the final snake contours, and the automatic error classifier. The whole set of 64 gold standard images contain a total of 16,170 pombe cells, averaging about 253 cells per image. The quantitative performances of CellStat, CellSerpent and PombeX are shown in Table 1. Figure 7 shows representative results from all methods. We tuned the parameters of CellStat and CellSerpent to optimize their performance on our data set. The parameter
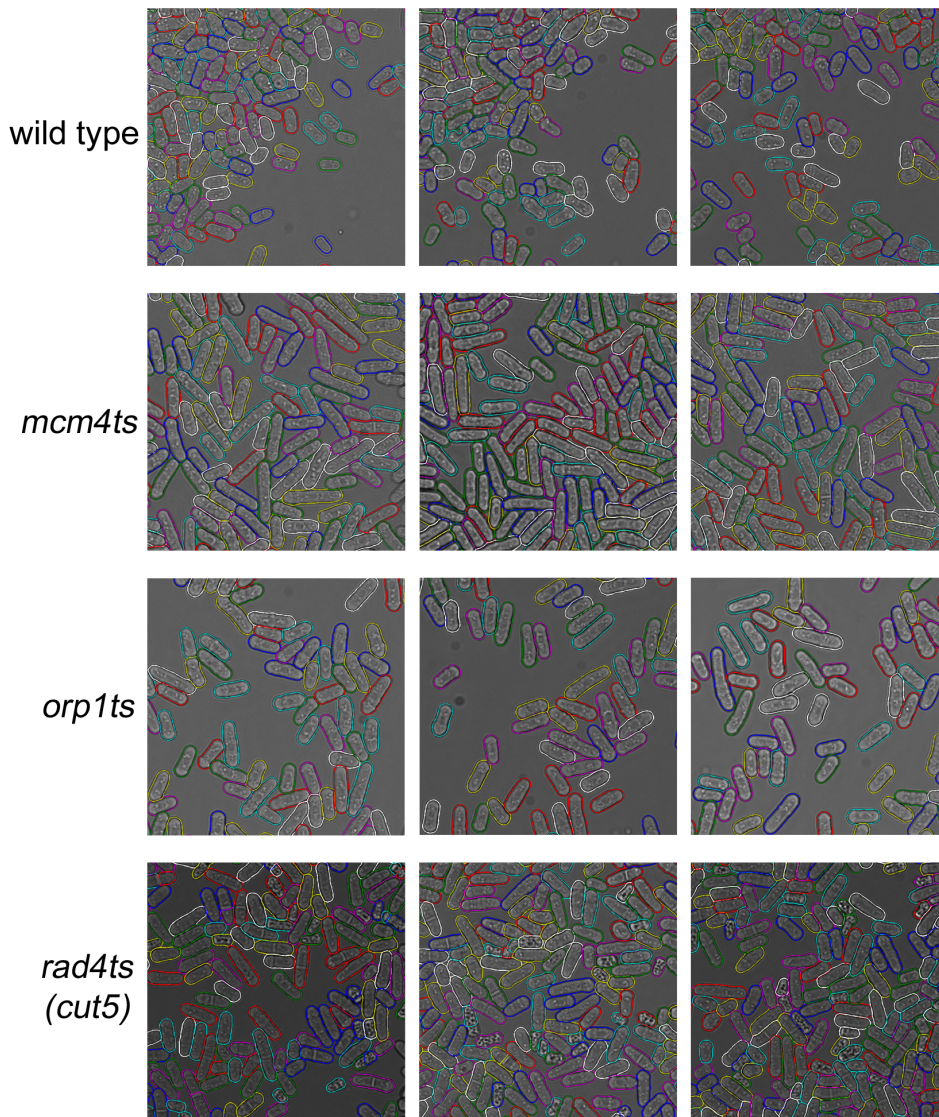
Figure 6. Segmentation results on well-known phenotypes. Cell contours are colored randomly.
doi:10.1371/journal.pone.0081434.g006

settings for CellStat are: Sigma Mult. Factor: 3 (default 7.5), Max Eccentricity: 15 (default 1.5), Max Area: 15000 (default 1500), and Max Radius to Centre: 100 (default 25). The parameter settings for CellSerpent are: Inflating force $\beta = 2.24$ (default $\beta = 2$). Since CellStat and CellSerpent did not use fluorescent signals, we also tested a version of our method that do not make use of the fluorescent channel images (listed as PombeX w/o FP in Table 1).

Our automatic cell contour validation system has an overall classification accuracy of 86.4%, with sensitivity 85.7% and specificity 85.4%. After erroneous cell contours are removed by our cell contour validation classifier, the remaining cell contours contain 97.8% true positives (precision, *i.e.* % relative to all detected cell contours). This shows that the final cell contours generated by our method after automatic validation are very reliable. For large scale high-throughput applications with huge amounts of data, in order to minimize the need for human intervention, the high reliability and robustness achieved by our system is valuable.

CellStat and CellSerpent were primarily designed for bright field images of *Saccharomyces cerevisiae* cells, which are round cells with good border contrast. This may in part account for their poor performance for correctly segmenting elongated pombe cells, since the final generated contours are biased toward circular shapes. CellSerpent was designed for images of crowded round budding yeast cells, and the elongated fission yeast cells in our images may be over-segmented into multiple round cells. On the other hand, CellStat assumes that each cell has a major part of its border touching the background, so cell misses and incorrect merging occurs for clustered or elongated cells.

Discussion

There are some recent works on cell image segmentation [11–13], which applied active contour models such as parametric snakes or level sets. Most of these methods are designed for phase-contrast images with higher contrast between cell membrane and background, and are designed with objective functions that
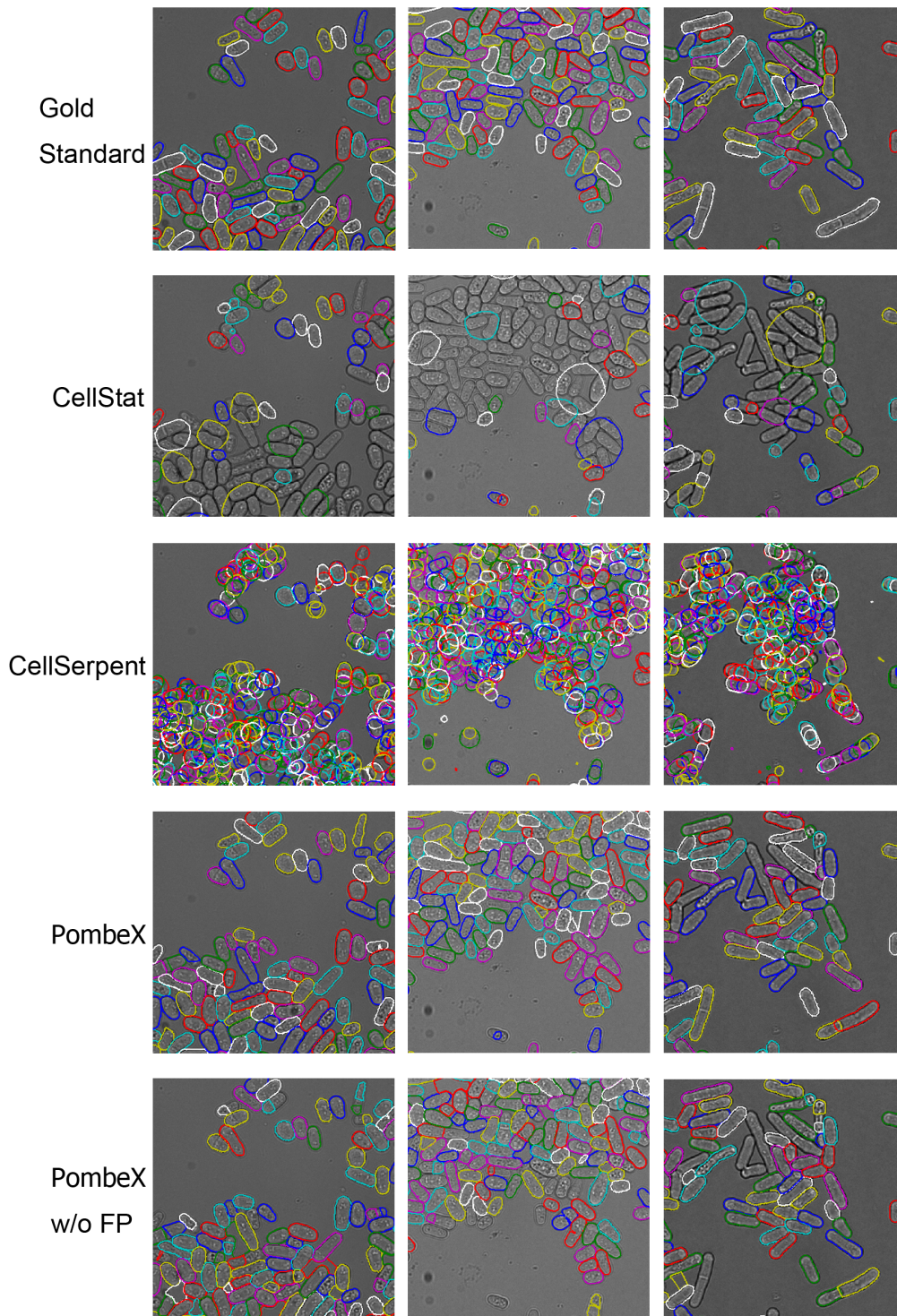
Figure 7. Segmentation method comparison. Contours of different cells are colored randomly.
doi:10.1371/journal.pone.0081434.g007

depend only on relative brightness. Suzuki *et al.* [14] segmented and quantitatively characterized budding and fission yeast cells, but their images are either stained with cell wall fluorescence (which is trivial to segment) or phase contrast images. They did not address the more complex problem of detecting and segmenting cells from ordinary bright field images. Tscherepanow *et al.* [15]

analyzed images very similar to the ones presented here (bright field images of Sf9 cells), but one of the main drawbacks of their method is its inability to deal with images with different focus properties, which are predominant in our dataset, and for which we have developed both global and locally adaptive methods of focus gradient correction.

Table 1. Quantitative performance results of yeast segmentation algorithms.

Method	%Cells Detected[1]	%Cells Segmented[2]	# of Contours Generated	PPV[3]	CPU time[4]
CellStat [9]	73.0%	10.2%	9164	18.8%	2382.5
CellSerpent [10]	99.8%	32.0%	104122	11.3%	1887.7
PombeX	97.0%	86.4%	18125	86.9%	135.7
PombeX w/o FP	94.5%	70.8%	19594	71.1%	125.0

[1]Relative to all cells.
[2]Defined as the percentage of cell contours with less than 10% pixel mismatch compared to gold standard cell contours, relative to all cells.
[3]Relative to all contours.
[4]Average CPU time per image, measured using a regular desktop PC (Windows 7, Matlab R2011a, Intel(R) Core(TM) i7-2600 CPU@3.40 GHz, 4 GB RAM).
doi:10.1371/journal.pone.0081434.t001

The adaptive focus gradient correction uses a bilinear model to estimate the intersection between the specimen plane and the objective focal plane, but this model may be too simplistic resulting in many cells not being focus corrected. It may be possible to improve this by explicitly using margin maximization methods, such as support vector machines, to find the exact boundaries between positively and negatively defocused regions. The adaptive focus correction algorithm is designed so that even when only a part of the image is complemented, C^0 continuity of the pixel intensities is still maintained because the complement is taken with the background intensity as center. But sometimes cells cut by the $z = 0$ boundary would still be severely distorted. This may require warping the boundary so that it only crosses the background pixels, though this solution is only feasible for images with cells not too densely packed.

The proposed segmentation system, PombeX, can be applied to a variety of experimental questions. For our screen, the next steps will be to characterize the fluorescent signals both within the nucleus and cytoplasm, as well as cell morphology. This will require identification and quantification of individual puncta at normal fluorescent intensity. Fluorescence that enters the cytoplasm can be used to automatically identify dead or dying cells. Additional information that may be culled from these data will include cell morphology to identify cell cycle stage. Finally, changes in cell dimension including average length or length × width dimensions can be quantified.

In conclusion we have developed a robust cell segmentation system for *S. pombe* cells that uses nucleus protein fluorescence to correct varying focus and contrast in the transmitted illumination image, combined with active contour segmentation and robust automatic contour validation. This system can be applied to similar bright-field microscopy images with or without corresponding fluorescence signal within the cell nucleus or cytoplasm, and can in principle be extended to deal with multiple cell types and image modalities.

Acknowledgments

Author Contributions

Conceived and designed the experiments: SLF. Performed the experiments: MDG SAS. Analyzed the data: JYP YJC CNH. Wrote the paper: JYP YJC CNH.

References

1. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, et al. (2002) The genome sequence of Schizosaccharomyces pombe. Nature 415: 871–880.
2. Kim DU, Hayles J, Kim D, Wood V, Park HO, et al. (2010) Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nature Biotechnology 28: 617–623.
3. Kass M, Witkin A, Terzopoulos D (1988) Snakes: Active contour models. International Journal of Computer Vision 1: 321–331.
4. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9: 62–66.
5. Xu C, Prince J (1998) Snakes, shapes, and gradient vector ow. IEEE Transactions on Image Processing 7: 359–369.
6. Nasmyth K, Nurse P (1981) Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast Schizosaccharomyces pombe. Molecular and General Genetics 182: 119–124.
7. Grallert B, Nurse P (1996) The ORC1 homolog orp1 in fission yeast plays a key role in regulating onset of S phase. Genes & Development 10: 2644–2654.
8. Saka Y, Yanagida M (1993) Fission yeast cut5+, required for S phase onset and M phase restraint, is identical to the radiation-damage repair gene rad4+. Cell 74: 383–393.
9. Kvarnström M, Logg K, Diez A, Bodvard K, Käll M (2008) Image analysis algorithms for cell contour recognition in budding yeast. Optics Express 16: 12943–12957.
10. Bredies K, Wolinski H (2012) An active-contour based algorithm for the automated segmentation of dense yeast populations on transmission microscopy images. Computing and Visualization in Science : 1–12–.
11. Ray N, Acton S, Ley K (2002) Tracking leukocytes in vivo with shape and size constrained active contours. Medical Imaging, IEEE Transactions on 21: 1222–1235.
12. Zimmer C, Labruyere E, Meas-Yedid V, Guillen N, Olivo-Marin JC (2002) Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cellbased drug testing. Medical Imaging, IEEE Transactions on 21: 1212–1221.
13. Mukherjee D, Ray N, Acton S (2004) Level set analysis for leukocyte detection and tracking. Image Processing, IEEE Transactions on 13: 562–572.
14. Suzuki G, Sawai H, Ohtani M, Nogami S, Sano-Kumagai F, et al. (2006) Evaluation of image processing programs for accurate measurement of budding and fission yeast morphology. Current Genetics 49: 237–247.
15. Tscherepanow M, Zöllner F, Hillebrand M, Kummert F (2008) Automatic segmentation of unstained living cells in bright-field microscope images: Advances in mass data analysis of images and signals in medicine, biotechnology, chemistry and food industry. In: Perner P, Salvetti O, editors, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, volume 5108. pp. 158–172. URL http://dx.doi.org/10.1007/978-3-540-70715-8-13.