Empirical gradient threshold technique for automated segmentation across image modalities and cell lines

J. CHALFOUN*, M. MAJURSKI*, A. PESKIN*, C. BREEN†, P. BAJCSY* & M. BRADY*

**Information Technology Laboratory, National Institute of Standards and Technology*

†*Princeton University, Princeton, NJ 08544, USA*

Summary

New microscopy technologies are enabling image acquisition of terabyte-sized data sets consisting of hundreds of thousands of images. In order to retrieve and analyze the biological information in these large data sets, segmentation is needed to detect the regions containing cells or cell colonies. Our work with hundreds of large images (each 21 000×21 000 pixels) requires a segmentation method that: (1) yields high segmentation accuracy, (2) is applicable to multiple cell lines with various densities of cells and cell colonies, and several imaging modalities, (3) can process large data sets in a timely manner, (4) has a low memory footprint and (5) has a small number of user-set parameters that do not require adjustment during the segmentation of large image sets. None of the currently available segmentation methods meet all these requirements. Segmentation based on image gradient thresholding is fast and has a low memory footprint. However, existing techniques that automate the selection of the gradient image threshold do not work across image modalities, multiple cell lines, and a wide range of foreground/background densities (requirement 2) and all failed the requirement for robust parameters that do not require re-adjustment with time (requirement 5). We present a novel and empirically derived image gradient threshold selection method for separating foreground and background pixels in an image that meets all the requirements listed above. We quantify the difference between our approach and existing ones in terms of accuracy, execution speed, memory usage and number of adjustable parameters on a reference data set. This reference data set consists of 501 validation images with manually determined segmentations and image sizes ranging from 0.36 Megapixels to 850 Megapixels. It includes four different cell lines and two image modalities: phase contrast and fluorescent. Our new technique, called Empirical Gradient Threshold (EGT), is derived from this reference data set with a 10-fold cross-validation method. EGT segments cells or colonies with resulting Dice accuracy index measurements above 0.92 for all cross-validation data sets. EGT results has also been visually verified on a much larger data set that includes bright field and Differential Interference Contrast (DIC) images, 16 cell lines and 61 time-sequence data sets, for a total of 17 479 images. This method is implemented as an open-source plugin to ImageJ as well as a standalone executable that can be downloaded from the following link: https://isg.nist.gov/.

Correspondence to: Joe Chalfoun, Information Technology Laboratory, National Institute of Standards and Technology, 100 Bureau Road, Gaithersburg, MD 20899. Tel: (+1) 301 975 3354; Fax: (+1) 301 975 6097; e-mail: joe.chalfoun@nist.gov

Background

Advances in microscopy image acquisition now allow the collection of large quantities of cell image data. Efficient processing of these terabyte-sized data sets demands novel algorithmic approaches to segmentation that enables both faster execution time and a high level of accuracy. To meet these demands, a segmentation method is needed that meets the following five criteria: (1) high accuracy as measured by the Dice index (0.9 or higher); (2) applicable to multiple cell lines with different cell densities and image modalities; (3) high throughput to process terabyte-sized data sets in a timely manner; (4) low memory footprint; (5) small number of user-set tuning parameters that are robust across an entire time-sequence of images.

Segmentation techniques can be classified into two categories based on their mathematical model: complex methods, such as level sets, and simple ones, such as thresholding. Complex methods are too computationally intensive for very large images. They include level set methods, approaches utilizing models based on partial differential equations, graph partitioning, watershed methods and neural networks (Egmont-Petersen *et al.*, 2002; Cremers *et al.*, 2006; Couprie *et al.*, 2009; Mobahi *et al.*, 2011). These computationally expensive techniques may satisfy the accuracy requirement, but do not meet the other criteria mentioned above: their execution time to process 1 TB of images takes longer than a day. Computationally simple segmentation methods include thresholding methods,

clustering methods and region-growing methods. Clustering methods range from simple and less accurate methods, such as k-means clustering (Dima *et al.*, 2011), to more complex fuzzy clustering methods that increase accuracy, but also increase computational cost (Despotović *et al.*, 2013). Region-growing segmentation methods are an example of methods that require user input in the form of seeds for growth that make this type of method difficult to automate (Kamdi & Krishna, 2012). Background reconstruction methods to improve segmentation accuracy require a lot of memory and have slow execution times (J Chalfoun *et al.*, 2013).

In general, pixel intensity gradients are higher for pixels at cell edges than for background pixels in an image. Based on this observation, we chose a segmentation approach based on thresholding the gradient image. Thresholding has a short execution time and a small memory footprint. However, finding the optimal threshold for each image remains a challenge. Sezgin and Bulent (Sezgin & Sankur, 2004) present an extensive survey of techniques for automating intensity threshold selection. Although many of these automated thresholding methods are implemented in ImageJ (Schindelin *et al.*, 2012) and Cellprofiler (Carpenter & Jones, 2006), an evaluation of these techniques against our reference data sets reveals that they fall short of meeting our criteria.

In this paper, we present the Empirical Gradient Threshold (EGT) method, a novel and empirically derived image gradient threshold selection method for separating foreground and background pixels in an image that meets all five requirements. EGT operates on the histogram of the gradient image and thus is a histogram shape-based thresholding method as classified by Sezgin and Bulent. The EGT method is derived from a reference data set using 10-fold cross validation where the data set is randomly split into 10 groups of similar size and nine groups are used for training and the remaining one for validation. The process is repeated 10 times to ensure that the empirical model is consistent across all groups. The reference data set consists of 501 validation images with manually determined segmentations and image sizes ranging from 0.36 Megapixels to 850 Megapixels. It includes seven different cell lines and two image modalities, phase contrast and fluorescent. We quantify the difference between our new approach and existing ones in terms of accuracy, execution speed, memory usage, and number of adjustable parameters on a reference data set. We show that EGT has the fastest execution time and the lowest memory utilization and it is the only method that segmented all data sets with a Dice value above 0.92. The results of this comparison are presented towards the end of the paper. Our method is also visually verified on a much larger data set that includes: bright field and Differential Interference Contrast (DIC) images, 16 cell lines, and 61 time-sequence data sets for a total of 17 479 images.

Figure 1 gives an overview of the work presented in this paper that led to creating the EGT segmentation. Second section describes the reference data sets, the selection process for

a gradient operator, and the empirically derived function that automatically computes a useful gradient threshold. Third section documents all experimental results. We describe how we measure segmentation performance and compare our technique with other segmentation approaches. We display the segmentation output on other image modalities and cell lines. Fourth section summarizes our results.

Methods

Biological motivation

The biological motivation of this work comes from four different applications displayed in Figure 2:

(1) Stem cell colonies: Pluripotent stem cells exist in a privileged developmental state with the potential to form any of the cell types of the adult body. Hence, there is great interest in understanding the relation between gene expression and cell state, in order to potentially engineer cell state for application to regenerative medicine. We used a cell line expressing GFP under the control of a critical pluripotency related transcription factor, OCT-4, to understand how normal stem cell cultures behave during routine feeding of cultures. These cells grow as isolated colonies, each colony comprising tens to thousands of cells as the culture progresses.
Because individual colony size is larger than the size of a single camera frame, colony tracking can only be done from movies of mosaics. In our case, we made a movie of 18×22 individual camera frames (total mosaic size $\sim$ 1GB) with a 10% overlap between frames in both the X and Y directions, over 162 time points (total movie size $\sim$ 350 GB). These large data set images were collected through time in the form of contiguous mosaics in phase contrast and GFP channels.

(2) NIH 3T3 cells: Despite numerous studies, the regulation of the extracellular matrix protein tenascin-C (TN-C) remains difficult to understand. By using live cell phase contrast and fluorescence microscopy, the dynamic regulation of TN-C promoter activity is examined in an NIH 3T3 cell line stably transfected with the TN-C gene ligated to the gene sequence for destabilized Green Fluorescent Protein (GFP). We found that individual cells vary substantially in their expression patterns over the cell cycle, but that on average TN-C promoter activity increases approximately 60% through the cell cycle. We also found that the increase in promoter activity is proportional to the activity earlier in the cell cycle. This work illustrates the application of live cell microscopy and automated image analysis of a promoter-driven GFP reporter cell line to identify subtle gene regulatory mechanisms that are difficult to uncover using population averaged measurements. The fully automated image
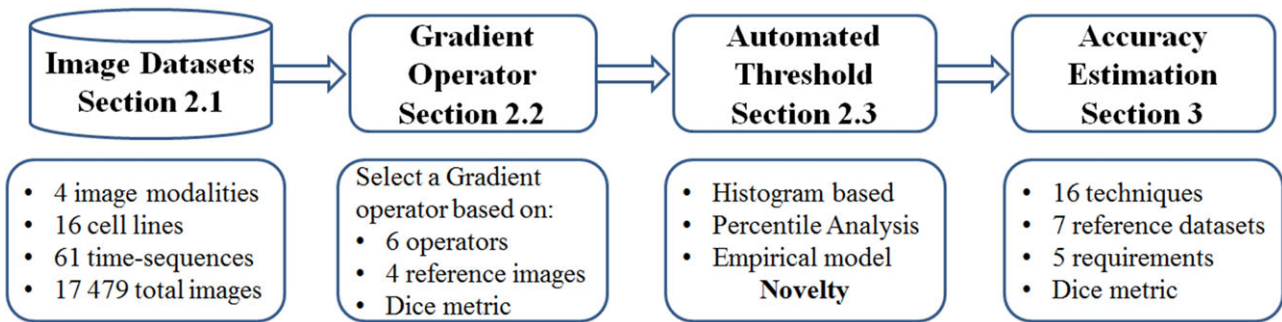
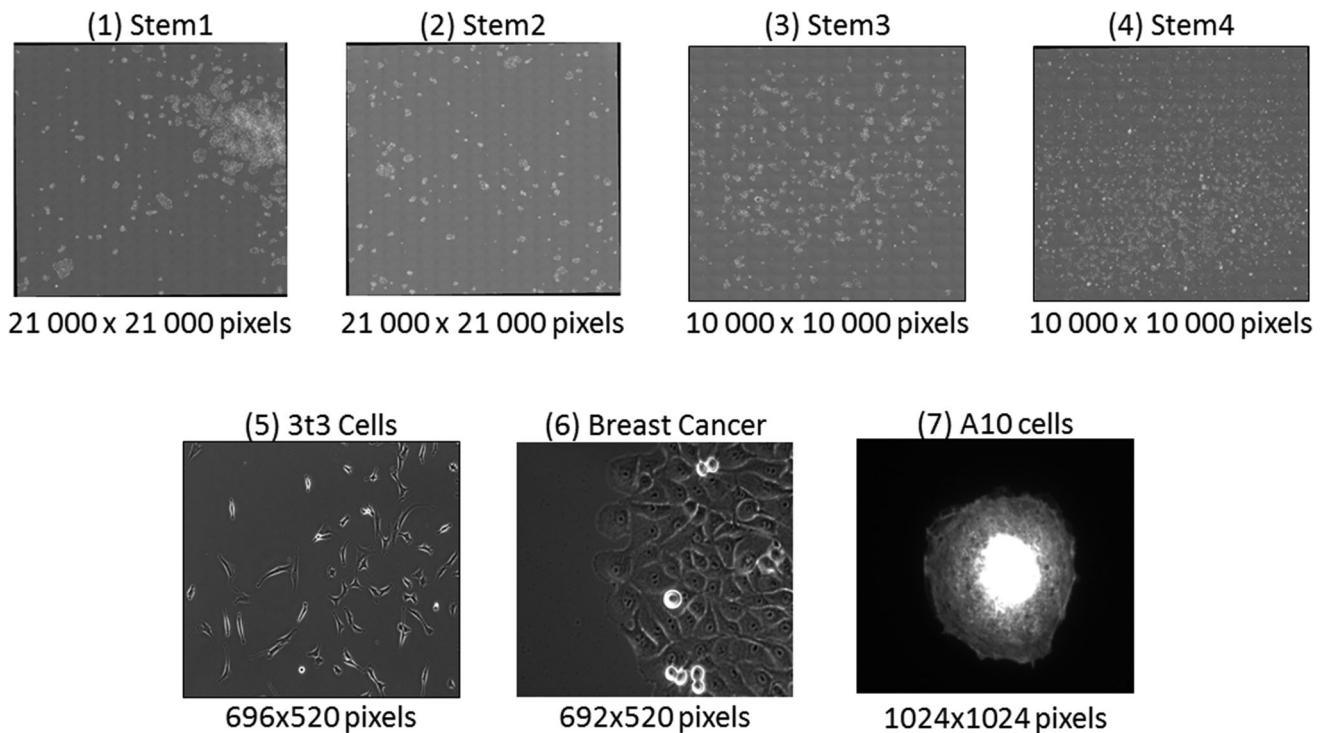**Fig. 1.** Overview of the work presented in this paper.

Fig. 2. Example images of the reference data sets. Six phase contrast images and one fluorescent images data sets. The first four images show stem cell colonies, the fifth NIH 3T3 cells; the sixth NIH breast epithelial sheets, and the seventh A10 rat cells.

segmentation and tracking are validated by comparison with data derived from manual segmentation and tracking of single cells. More detail about this work can be found in Halter *et al.* (2011a) and Chalfoun *et al.* (2013).

(3) MCF10A breast epithelial sheet cells: Many cell lines that are currently being studied for medical purposes, such as cancer cell lines, grow in confluent sheets. These cell sheets typically exhibit cell line specific biological properties such as the morphology of the sheet, protein expression, proliferation rate, and invasive/metastatic potential. However, cell sheets are comprised of cells of different phenotypes. For example, individual cells in a sheet can have diverse migration patterns, cell shapes,

can express different proteins, or differentiate differently. Identifying phenotypes of individual cells is highly desirable, as it will contribute to our understanding of biological phenomena of tumour metastasis, stem cell differentiation, or cell plasticity. Time-lapse microscopy now enables the observation of cell cultures over extended time periods and at high spatiotemporal resolution. Furthermore, it is now possible not only to label cells with fluorescent markers, but also to express fluorescently labelled protein, enabling spatiotemporal analysis of protein distribution in a cell sheet at a cellular level. More information about this project can be found in Weiger *et al.* (2013), Stuelten *et al.* (2010) and Chalfoun *et al.* (2014).

(4) A10 rat vascular smooth muscle cells: High resolution images of A10 cells are acquired in order to understand cell responses to the mechanical characteristics of extracellular matrix (ECM). The ECM represents the extracellular environment that affects cell behaviour. This is of great importance to elucidate the biological pathways of cancer and other pathologies and analyze and compare the behaviour of different proteins and organelles at subcellular level.

Reference image data sets

We have collected seven reference image data sets with four cell lines and two image modalities, six data sets in phase contrast and one in fluorescent. The first four data sets (Stem1, Stem2, Stem3 and Stem4) are images of stem cell colonies. These large data sets are terabyte-sized image sets and were acquired to study the temporal and spatial behaviour of stem cell colonies from the time of seeding (small colonies) to the time before differentiation (5 days later). The fifth data set is a time lapse sequence of NIH 3T3 cells; this experiment was performed to study the dynamic regulation of TN-C promoter activity (Halter *et al.*, 2011b). The sixth data set is a time sequence of a breast epithelial cell sheet (Stuelten *et al.*, 2010; Weiger *et al.*, 2013). The seventh data set contains fluorescent images of fixed A10 rat cells stained to analyze the subcellular protein expression for 129 cell images. Table 1 gives details about these 7 sets. Figure 2 shows an image example of each set. Higher resolution images can be downloaded from the following link: https://isg.nist.gov/.

A human expert manually segmented cells/colony edges in each frame using the pencil and brush tool in ImageJ (Schindelin *et al.*, 2012). A second expert inspected the manual segmentation to minimize human errors. For the large data sets, the manual segmentation is performed on a subset of each set. The total number of manually segmented images is 501. We use all the manually-segmented images in the reference data set to train the EGT model by a 10-fold cross-validation method.

Selection of a gradient operator

Many techniques are available to compute the gradient of an image with different kernel sizes. We began our analysis by comparing the most common gradient methods available to us in Matlab: (1) numerical gradient, (2) central difference, (3) intermediate difference, (4) Roberts, (5) Prewitt and (6) Sobel. More detailed information about each operator can be found in Gonzales and Woods (Gonzalez *et al.*, 2008).

To select an operator for this analysis, we randomly selected four images and their manual segmentations, one for each cell line. The gradient is computed for each image using six different operators and thresholded at every gradient percentile. The accuracy of all resulting masks is compared against the manual segmentation using the Dice index (Dice, 1945). The Dice index measures spatial overlap between two segmentations using the following formula: $Dice = 2 \times overlap/(area1 + area2)$ where $area1$ and $area2$ are the respective areas of the foreground masks. It ranges from 0 (no match) to 1 (perfect match). Figure 3 shows Dice index results for every percentile threshold for one of the large images using all 6 operators. This figure shows that a solution can be found by thresholding the image gradient to generate a segmented mask that is very close to a manual segmentation. Every operator gave a maximum Dice index value above 0.9. The uncertainty related to cell edge pixel locations has a margin of a couple of pixels due to the smooth transition between background and foreground intensities (Dima *et al.*, 2011) which minimizes the differences between all operators. However, after examining the results on all four test images, we chose the Sobel operator because it gave the highest average maximum Dice values across all images as shown in Table 2.

*Automatic selection of a gradient threshold value*

The gradient threshold selection is based on the assumption that there is a relationship between an optimal threshold value T and image histogram descriptors. First, we explored whether there exists a threshold that would yield a good segmentation result as assessed by the Dice index. We concluded that there is such a threshold and established a set of optimal threshold values by human inspection. Second, we observed a relationship between the histogram distribution shapes and these optimal threshold values. Third, we modelled the relationship mathematically by using empirical observations.

Existence of a solution. The gradient of every image in the reference data set is computed using the Sobel operator. The resulting gradient image is segmented by thresholding it at every gradient percentile value. The Dice index is computed between every segmented image and the corresponding manual segmentation. Figure 4 shows the maximum value of the Dice index (scaled between 0 and 100) computed for each image in the reference data set and the corresponding gradient percentile that generated that maximum Dice value. In all of our reference data set images, the best segmentation corresponded to percentiles between the 25th and 95th gradient percentiles. Figure 4 shows that an accurate segmentation solution exists using the percentile threshold across all reference images. The next step is to examine the histograms of these images.

Empirical observations. Figure 5 shows four examples of normalized histograms where the 95th, 75th, 55th and 35th percentiles gave the maximum Dice index, respectively. In an image where most pixels are background with low gradient values, higher percentages (> 75% for example) are needed

Table 1. Summary of reference image data sets

#	Data set	Image modality	Foreground type	Density of foreground objects	Image size (in pixels)	Image acquisition	Images manually segmented
1	Stem1	Phase contrast	Colony	Low, medium and high	21 000×21 000	157 images (45 min/image)	16, taken every 7.5 h
2	Stem2	Phase contrast	Colony	Low, medium and high	21 000×21 000	136 images (45 min/image)	14, taken every 7.5 h
3	Stem3	Phase contrast	Colony	Low, medium and high	10 000×10 000	477 images (15 min/image)	24, taken every 7.5 h
4	Stem4	Phase contrast	Colony	Low, medium and high	10 000×10 000	388 images (15 min/image)	19, taken every 7.5 h
5	NIH 3t3 Cells	Phase contrast	Cell	Low, medium and high	696×520	238 images (15 min/image)	All
6	Breast epithelial cell sheets	Phase contrast	Cell sheet	High	692×520	59 images (2 min/image)	All
7	A10 rat cells	Fluorescent	Cell	Low	1024×1024	131 images	All
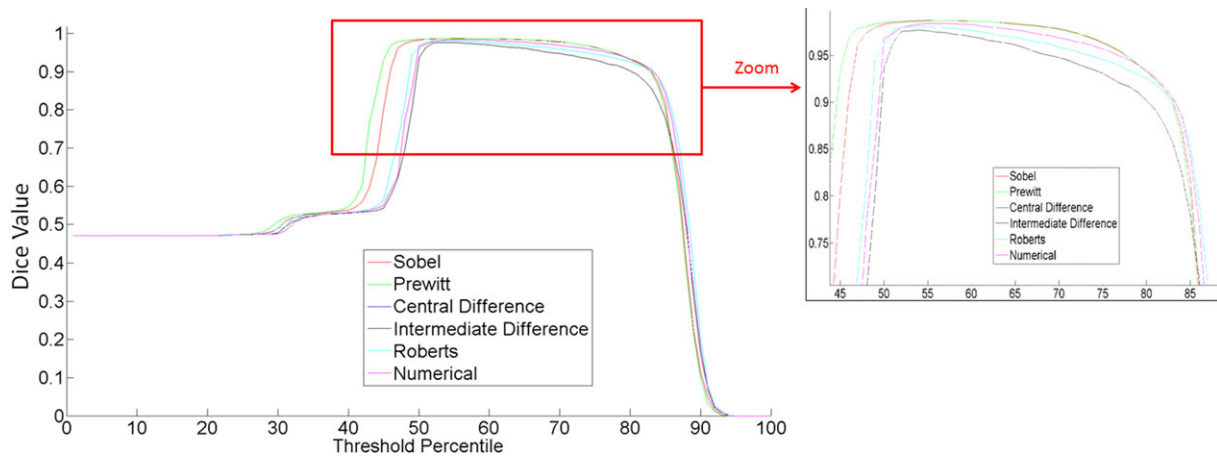
Fig. 3. The Dice index computed on every percentile threshold of one large image using all operators.

Table 2. Maximum Dice value reached for each operator and for each test image

Operator	Max Dice Image1	Max Dice Image2	Max Dice Image3	Max Dice Image4	Average Dice
Sobel	0.987	0.992	0.995	0.987	0.990
Prewitt	0.987	0.991	0.995	0.968	0.985
Central Difference	0.984	0.992	0.995	0.936	0.977
Intermediate Difference	0.977	0.992	0.991	0.920	0.970
Roberts	0.981	0.991	0.994	0.910	0.969
Numerical	0.984	0.992	0.995	0.936	0.977

to reach the correct percentile threshold for edge detection (Figure 5.1). In contrast, in an image where most pixels are foreground, lower percentages (> 35% for example) are needed to reach the correct percentile threshold for edge detection (Figure 5.3). The difference between the four plots in Figure 5 can be described by how much of the area X under the histogram curve lies to the right of the highest point of the histogram, the mode location.

The background of a biological image usually has low intensity variations in a small neighbourhood surrounding a pixel, which translates to low gradient magnitudes. Sharp changes in surrounding neighbour intensities around a pixel often correspond to noise in the acquired image. Gradient values for cell or colony edge pixels are usually between the highest gradient values and the lowest ones. Therefore, to measure the difference between the four curves in Figure 5, the area X under the histogram curve is computed between a lower bound (**lb**) and an upper bound (**ub**) for each image based on the location of the mode of the histogram, as outlined below.

Mathematical model. The previous section shows that there is a relationship between the histogram distribution and the gradient percentile values of which the threshold is computed. We will model this relationship with three equations relating (1) histogram H to area X under the histogram curve between
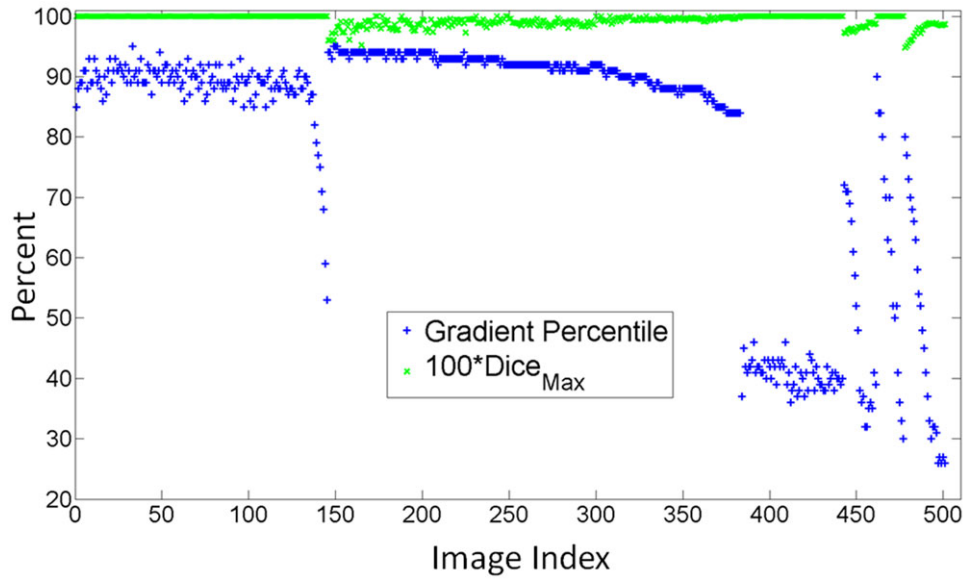
Fig. 4. Maximum Dice value (scaled between 0 and 100) and the corresponding gradient percentile threshold for every image in the reference data set.
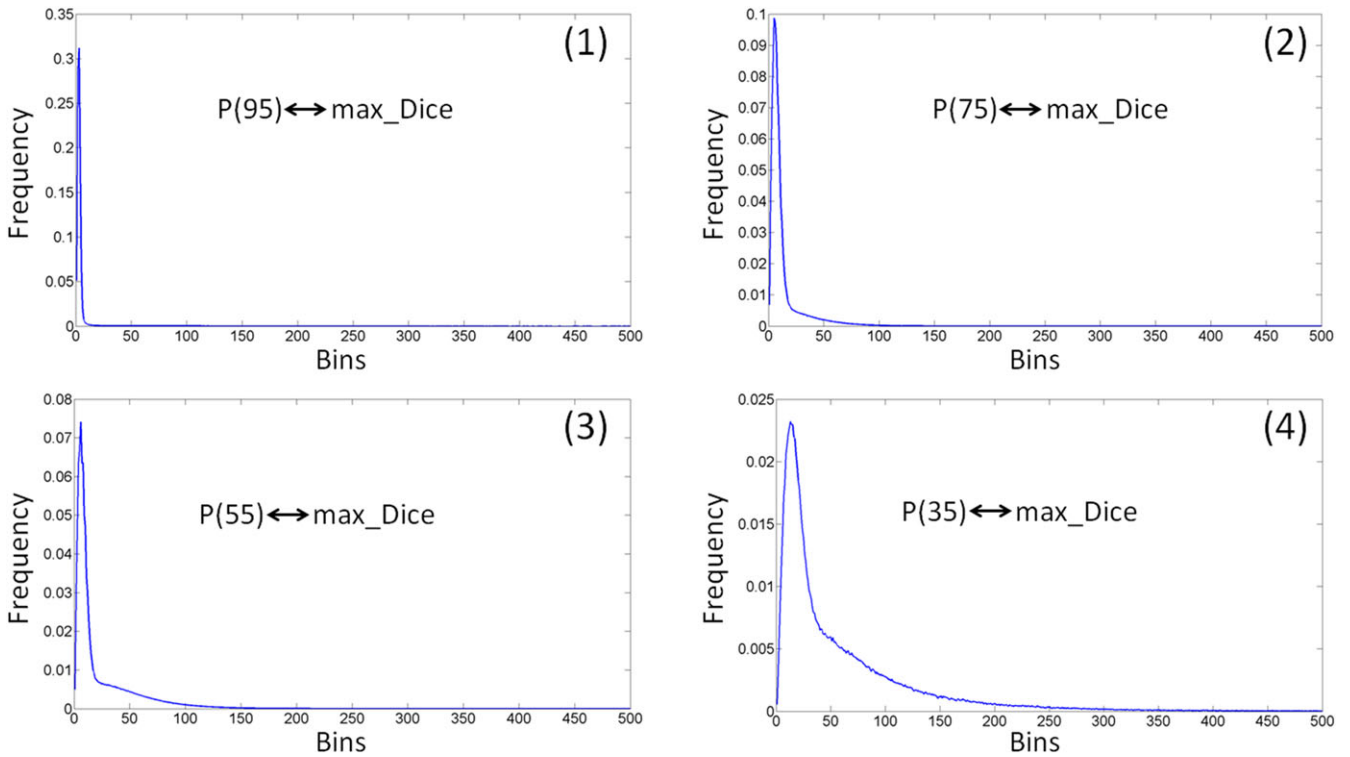

Fig. 5. Normalized histogram plots for images where (1) the 95th percentile, (2) the 75th percentile, (3) the 55th percentile and (4) the 35th percentile gave respectively the maximum Dice index. The plots are truncated at 500 instead of 1000 on the x axis to better highlight the difference.

a lower and upper bound, (2) area X to gradient percentile Y and (3) percentile Y to the optimal threshold value T :

$$
\begin{cases}
X = g(H) \\
Y = f(X) \quad, \\
T = p(Y)
\end{cases}
\tag{1}
$$

where

- H is the normalized histogram of the gradient image with respect to its cumulative sum ($sum(H) = 1$), represented by 1000 bins evenly spaced between the minimum and the maximum values found in the gradient image that are greater than 0.

- **X** is the area under the histogram between a lower and upper bound computed as a function of **H** .
- **Y** is the optimal gradient percentile value computed as a function of **X** .
- **T** is the gradient image intensity threshold value.
- **p** computes the threshold value **T** from the percentile value **Y** . **p**(**i**) is the threshold such that **i**% of image pixels have intensity gradients less than **p**(**i**) .

The percentiles are computed from the gradient image without the saturation values (where the gradient is equal to zero). Gradient magnitudes of zero correspond to neighbouring pixels in the image where the intensity is the same and thus do not correspond to edge pixels. Lower bounds are always greater than zero. Derivation of the lower and upper bounds are shown below.

The functions **f** and **g** and their respective arguments are determined empirically in the next sections. These functions compute the threshold value from the normalized histogram which constitutes the novelty of the EGT algorithm.

*Empirical derivation of function **g**.* The function **g** that computes the area **X** under the histogram curve is modelled as follows:

$$X = g(H) = \sum_{x=lb}^{ub} H(x) , \qquad (2)$$

where **lb** is a lower bound and **ub** is an upper bound that will be determined empirically from the mode location. The gradient magnitude mode value generally corresponds to pixels with low gradient variations (pixels that belong to the background or homogeneous pixels that do not belong to an edge). Since the mode is a statistical value of a histogram, we decided to empirically compute these bounds from an approximated mode location x_{mode} : **lb** = **n** * x_{mode} and **ub** = **m** * x_{mode} with **m** > **n** . We approximate the mode location x_{mode} using the average of the three highest estimated frequencies. The average mode location value is more accurate than the single maximum peak location and will minimize the uncertainty of computing the mode location in the presence of noise and artefacts in the background. The empirical derivation of the lower and upper bounds is made in such a way that enables a known fit (linear if possible) for function **f** . Therefore, we made an exhaustive search of these bounds looking for linearity of the function **f.**

Figure 6 displays the residual error of a linear fit to the function **f** colour coded between dark blue (lowest error value) to dark red (highest error value). The top portion is the residual error computed with regards to the exhaustive selection of a lower and an upper bounds as multiples of the mode location. The optimal solution corresponds to the global minimum of the lower and upper bounds exhaustive search. The lower portion displays the plots of 6 marked examples of area **X** vs. optimal percentile **Y** , where you can see the linearity. By analyzing the plots in Figure 6, we found that the optimal solution is the compute the area **X** between a lower bound equal to 3 × mode location on the **x** axis and an upper bound equal to 18 × mode location on the **x** axis. These empirically derived bounds ensured that most of the background or the very low varying pixel intensities and the high gradient magnitudes that correspond to image noise are removed. Furthermore, we impose an additional constraint on the upper bound to ensure that the area under the histogram is computed between the lower bound and at least the location x_{cs} that corresponds to a 95% drop in frequency value from the mode. Since the histogram is discretized, x_{cs} corresponds to the location where: $H(x_{cs} + 1) > 0.05 \times H(x_{mode})$ **AND** $H(x_{cs}) \leq 0.05 \times H(x_{mode})$.

Summary : **lb** = 3 × x_{mode} and **ub** = max (18 × x_{mode}, x_{cs})

Empirical derivation of function f. Figure 7 plots the percentile Y corresponding to the maximum Dice index values for all images when computed as a function of the area under the histogram X . This plot reveals a linear relationship between X and Y with a saturation of $Y = 25$ for $X \geq 50$. The function f derived empirically from the plot can be written as follows:

$$Y = f(X) = \begin{cases} 95 & X \leq s_1 \\ aX + b & s_1 < X < s_2 \\ 25 & s_2 \leq X \end{cases} , \qquad (3)$$

where s_1 and s_2 are derived from the plot with values equal to $s_1 = 3$, $s_2 = 50$.

To compute the linear relationship, we randomly arranged the reference data set into 10 groups of similar size. Nine of the groups are used for training and the remaining one as a validation set. A linear least squares fit is applied to the training set and the resulting linear equation is validated on the validation data set. This process is repeated 10 times. From the results shown in Table 3, we noticed only very small differences between all 10 iterations, showing that the selected model is very robust. We saw less than 1.3% variation in the slope parameter a of the linear function and less than 0.2% variation in the intercept parameter b. The linear function is computed as the average of all 10 values and is equal to $a = -1.3517$ and $b = 98.8726$.

When the percentile Y is computed, the image gradient threshold is then derived from the percentile by $T = p(Y)$ where $p(i)$ is the threshold such that $i\%$ of image pixels have intensity gradients less than $p(i)$.

The EGT algorithmic steps for segmenting an image are given below:

(1) Compute the gradient image G of the raw input image I using Sobel operator.
(2) Compute the histogram H of G with 1000 bins.
(3) Normalize the histogram with respect to its cumulative sum: $sum(H) = 1$.
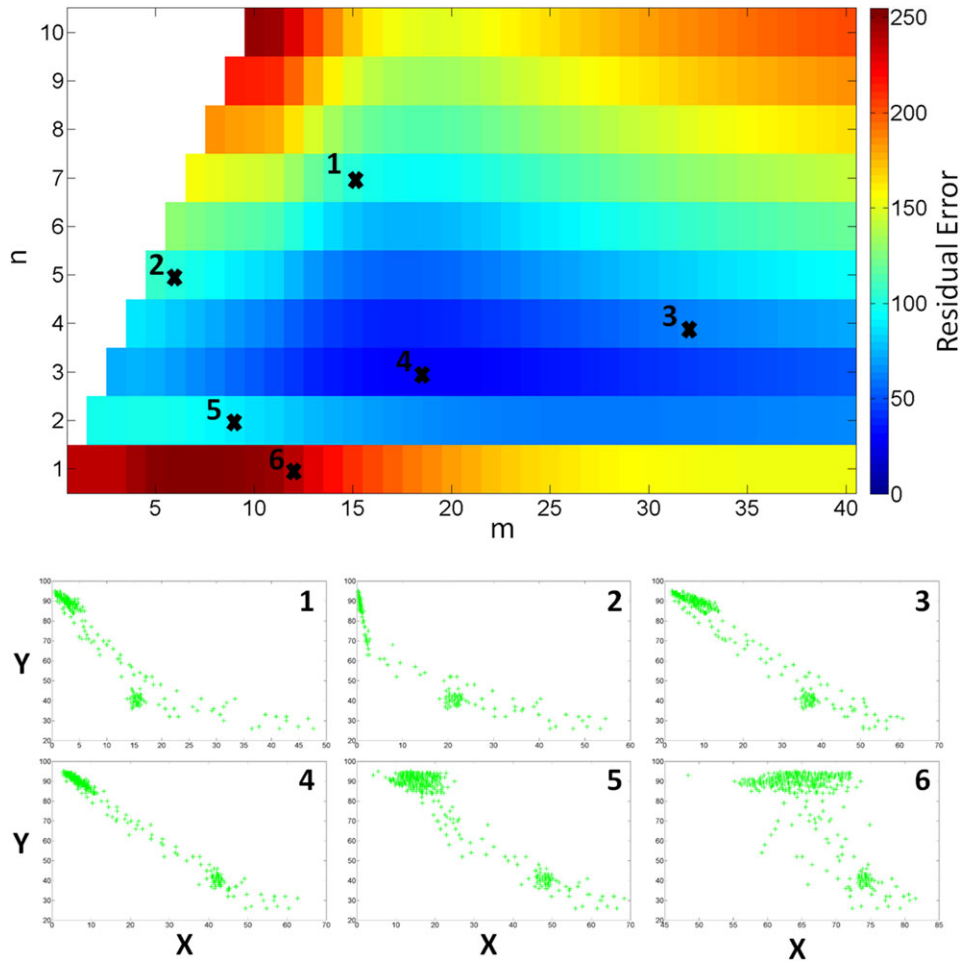
Fig. 6. Empirical derivation of the upper and lower bounds of function g. The top portion is the residual error of a linear fit between X and Y. The axis of this plot are the multiplicative factors (m, n) respectively of the mode location of which the lower bound $= n \times x_{mode}$ and the upper bound $= m \times x_{mode}$. The optimal solution corresponds to the global minimum of the lower and upper bounds exhaustive search. The lower portion displays the plots of the 6 marked examples.

(4) Average the top 3 histogram value locations to find an approximate mode location.

(5) Compute the area under the histogram X between the lower and upper bounds.

(6) Compute $Y = aX + b$.

(7) Compute the gradient threshold $T = p(Y)$ and segment the image.

(8) Fill holes in the resulting mask that are less than a user-input minimum hole size.

(9) Apply morphological erosion with a disk radius of 1 pixel to clean the noise around the edges.

(10) Filter small artefacts that are smaller than a user specified minimum cell size.

Figure 8 shows examples of segmentation results for the 7 reference data sets.

Handling special cases

Our assumptions for this analysis are (1) we can segment cells or colonies if edge pixel intensities are different from background intensities and (2) the background is locally uniform. However, edges of cell lines like Retinal Pigment Epithelial (RPE) cells are distinguishable by the human eye but are very close in intensity to the background as shown in Figure 9.1. The same type of edges are found when images are out of focus or have low Signal to Noise Ratio (SNR). To segment these types of cell images, we analyzed two data sets: (1) A large stem cell colony data set acquired with a lower exposure time and lower SNR than our four reference sets and (2) the RPE cell line images shown in Figure 9. Manual segmentation was performed on these two data sets and their respective plots corresponding to the maximum Dice indices are shown in Figure 10. The plots show that the relation between X and Y
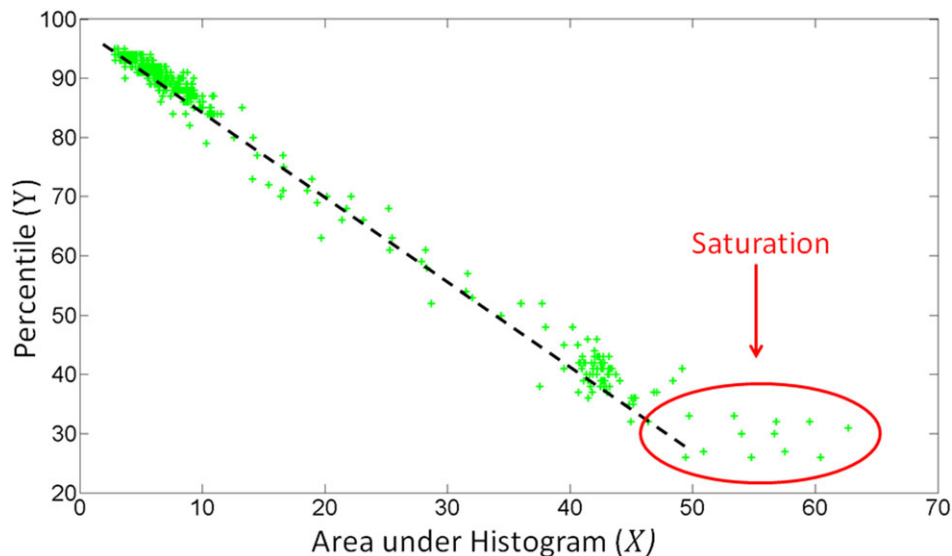
Fig. 7. Percentile Y as a function of the area under the histogram X . For each image of the reference data sets the percentile corresponding to the max Dice index is plotted. This plot shows a visibly linear relationship between X and Y .

Table 3. Results of the 10-fold cross validation for the linear function

Iteration	A	b	Mean Dice	Min Dice
1	-1.350	98.859	0.983	0.931
2	-1.356	98.940	0.982	0.938
3	-1.349	98.842	0.985	0.936
4	-1.351	98.853	0.985	0.945
5	-1.345	98.805	0.986	0.925
6	-1.345	98.788	0.986	0.957
7	-1.347	98.861	0.986	0.940
8	-1.361	98.963	0.984	0.943
9	-1.362	98.933	0.983	0.937
10	-1.349	98.880	0.986	0.962

remains linear for these special data sets, but the slope of the line drops by a constant factor compared with our previous data. Therefore, a user-defined parameter called "greedy" is introduced to control the percentile threshold for an entire time sequence data set for any data set that falls within these special cases. Figure 10 also shows that the greedy factor is consistent throughout a particular data set and hence the user needs to adjust this parameter only on one test image for the entire sequence. This parameter changes equation (1) to the following:

$$T = p\,(Y + \text{greedy})\,, \qquad (4)$$

with $-50 \leq \text{greedy} \leq 50$, $\text{greedy} \in \mathbb{N}$ and $0 \leq Y + \text{greedy} \leq 100$.

The greedy parameter lowers or raises the percentile threshold to capture the missed edge pixels that are in a low or high gradient region. Percentiles follow the intensity variations in the image better than just multiplying the current threshold by a factor, which is the case in the open source software Cellprofiler (Carpenter & Jones, 2006). Figure 9.2 shows the resulting segmentation after adjusting the greedy parameter. The RPE cells are imaged on a plastic plate which by manufacturing has scratched on it. Some scratches are very visible and thus will be picked up by the segmentation method and the result will appear as a line connecting cells.

Experimental results

Accuracy measurement and comparison with other techniques

Table 4 and Figure 11 present the accuracy summary comparison with the top 15 thresholding techniques compared against our method. Table 4 has a summary of the manual user input required (if any), the execution speed/image, and the memory consumption/image that shows that the EGT has the fastest execution speed and the lowest memory usage. Figure 11 shows the accuracy of segmenting each data set with the 16 methods as boxplots. The lower left corner of the figure shows the legend for all the plots. The green line in each plot indicates the accuracy threshold at Dice value of 0.9. A successfully segmented data set has the entire Dice boxplot above that line. These plots show that EGT successfully segmented all data sets with a Dice value above 0.92. Furthermore, EGT is the only method that segmented data sets 1 and 5 successfully. Three methods worked on data set 2, two methods for data set 3, three methods for data set 4, two methods for data set 6, and seven methods for data set 7. These results are summarized in the last column of Table 4 where the number of data sets successfully segmented by each method is presented, confirming that our new method is robust and accurate across multiple cell lines and image modalities.
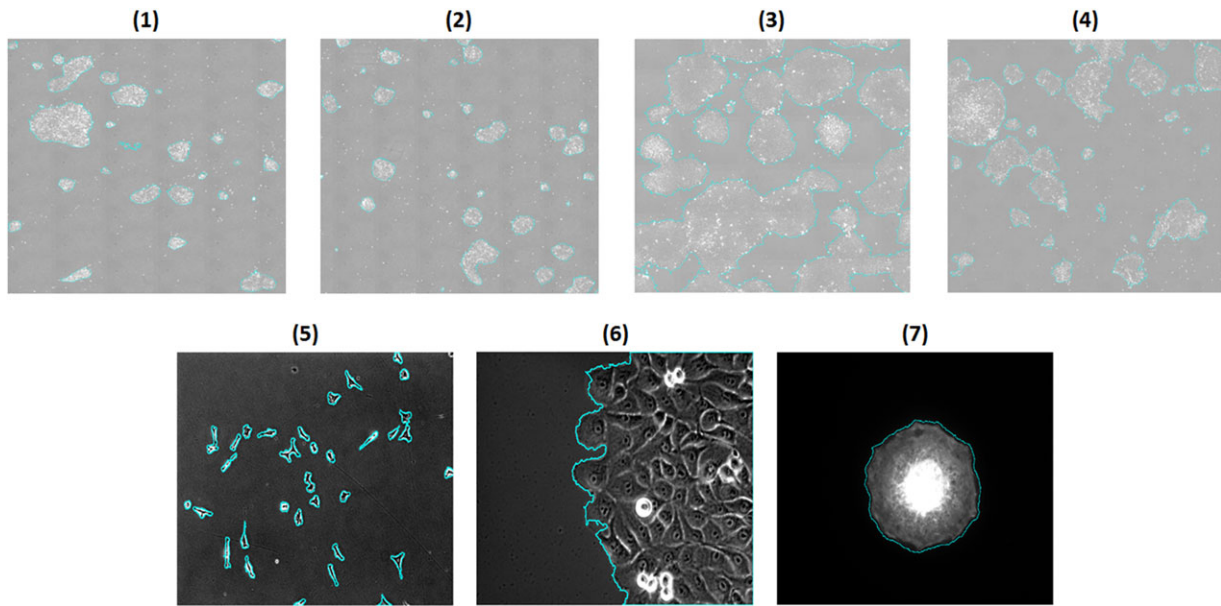
Fig. 8. Segmented images results with the contour overlaid on top of the original raw image. Large images (the first four) are zoomed in for better visualization. The cyan colour is only for edge highlighting.
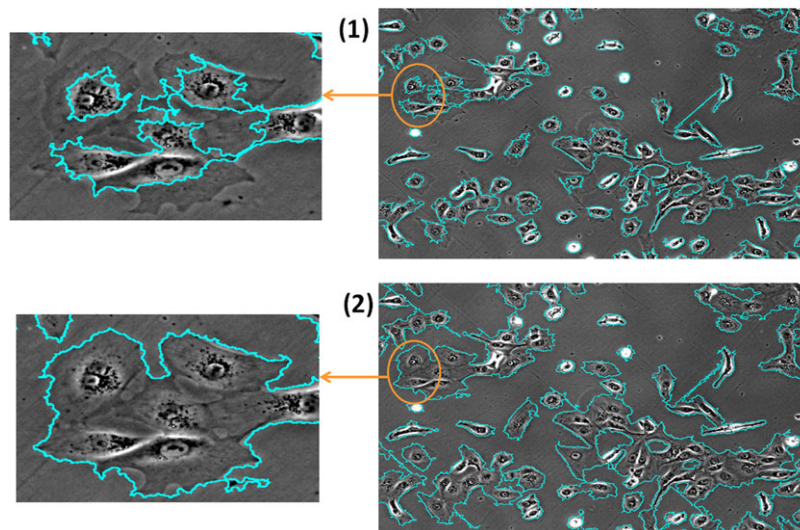

Fig. 9. Segmentation with (1) greedy parameter = 0 and (2) with greedy parameter = 20.

The free standalone executable, the plugin to ImageJ, the source code, and the data sets can be downloaded from the following link: https://isg.nist.gov/.

Application to different image modalities and cell lines

The EGT segmentation is empirically derived with the assumptions that edge pixel intensities are different from background intensities and the background is locally uniform. These assumptions are independent from a cell line and an image modality. As long as these two conditions are met, EGT should be appropriate to use on most cell images and cell lines. To

test this theory, we applied EGT on 4 image modalities, 16 cell lines, 61 time-sequence data sets for a total of 17 479 images. No quantification of the EGT performance is made on these data sets due to the lack of manual segmentation. However, the results have been inspected visually for accuracy. The segmentation results can be viewed and downloaded from https://isg.nist.gov/.

Figure 12 displays six example images of the EGT segmentation results. The cell lines and image modalities used are: (1) Bright field images of rat brain cells from NIH, (2) Fluorescent images of yeast cells downloaded from Duke University (Di Talia *et al.*, 2007; Wang *et al.*, 2010), (3) DIC images of iPS
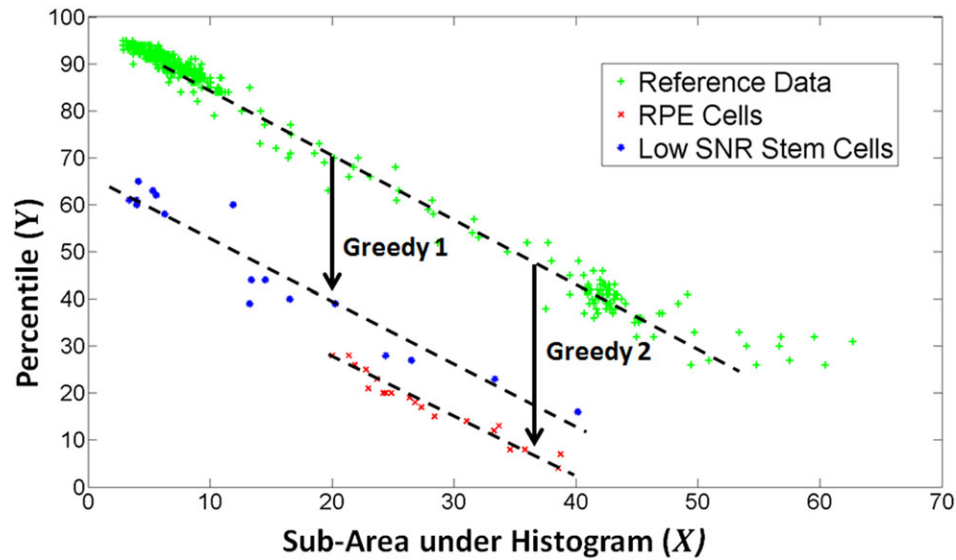
Fig. 10. Percentile *Y* as function of the area under the histogram *X* . For each image of the reference data sets in green and for the two special case data sets: low SNR stem cell colony images (blue) and RPE cells (red). The plot show that the slope of the line defined for the reference data set to compute the percentile, at which the threshold of the image gradient is defined, drops by a constant factor called "greedy" and the relation between *X* and *Y* remains a line for a particular special data set.

Table 4. Summary of relevant factors to large data sets: The execution speed (ES) in s and the memory usage (MU) in GB, the number of manual inputs (MI), and the number of data sets successfully segmented (DSS) by each method. A successful segmentation for a data set is when a minimum Dice value of 0.9 is obtained for each image in that data set. NA (not applicable) refers to a technique that could not be applied

Technique	ES Data (1-2)	ES Data (3-4)	ES Data5	ES Data6	ES Data7	MU Data (1-2)	MU Data (1-2)	MI	DSS
1. Huang (Schindelin *et al.*, 2012)	86.39	17.09	0.26	3.19	0.89	8.5	1.5	0	1
2. Li (Schindelin *et al.*, 2012)	52.67	11.88	0.09	0.08	0.32	8.5	1.5	0	1
3. Mean (Schindelin *et al.*, 2012)	92.65	17.22	0.09	0.08	0.32	8.5	1.5	0	2
4. MinError(I) (Schindelin *et al.*, 2012)	91.92	17.25	0.09	0.08	0.32	8.5	1.5	0	1
5. Shanbhag (Schindelin *et al.*, 2012)	95.01	18.86	2.57	46.39	8.78	8.5	1.5	0	0
6. Triangle (Schindelin *et al.*, 2012)	72.2	16.11	0.09	0.06	0.25	8.5	1.5	0	0
7. Background (Carpenter & Jones, 2006)	N/A	27.37	0.6	0.75	1.03	NA	21	0	0
8. Kapur (Carpenter & Jones, 2006)	N/A	32.22	0.65	0.77	1.08	NA	21	0	0
9. MoG (Carpenter & Jones, 2006)	N/A	31.64	0.73	0.79	1.06	NA	21	1	0
10. Otsu (Carpenter & Jones, 2006)	N/A	30	0.72	0.75	1.07	NA	21	0	1
11. RidlerCalvard (Carpenter & Jones, 2006)	N/A	32.25	0.71	0.75	1.06	NA	21	0	1
12. RobustBackground (Carpenter & Jones, 2006)	N/A	29.8	0.64	0.78	1.07	NA	21	0	1
13. Sobel (Gonzalez *et al.*, 2008)	59.91	10.37	0.06	0.07	0.21	11	2.5	1	0
14. LoG (Gonzalez *et al.*, 2008)	164.37	22.78	0.11	0.12	0.42	11	2.5	1	2
15. Canny (Gonzalez *et al.*, 2008)	186.79	43.73	0.2	0.2	0.83	12	2.8	1	2
16. EGT (new technique)	22.75	4.23	0.02	0.01	0.05	8.5	1.5	0	7

cells from the Lieber Institute, (4) phase images of bone cancer cells from the Broad Institute (Khan *et al.*, 2011), (5) Fluorescent images of E. coli cells from Duke University (Rosenfeld *et al.*, 2006; Wang *et al.*, 2010) and (6) Bright field image of hematopoietic progenitor cells (Buggenthin *et al.*, 2013).

Discussion

We quantified the difference between our new approach EGT and existing ones in terms of accuracy, execution speed,

memory usage and the number of adjustable parameters on a reference data set. EGT had the best results among all 16 tested methods, on 501 validation images with manually determined segmentation and image sizes ranging from 0.36 Megapixels to 850 Megapixels. Tests included seven different cell lines and two image modalities: phase contrast and fluorescent. EGT segmented 100% of the cells or colonies with a Dice index above 0.92. It was also visually verified on other image modalities like bright field and Differential interference contrast (DIC)
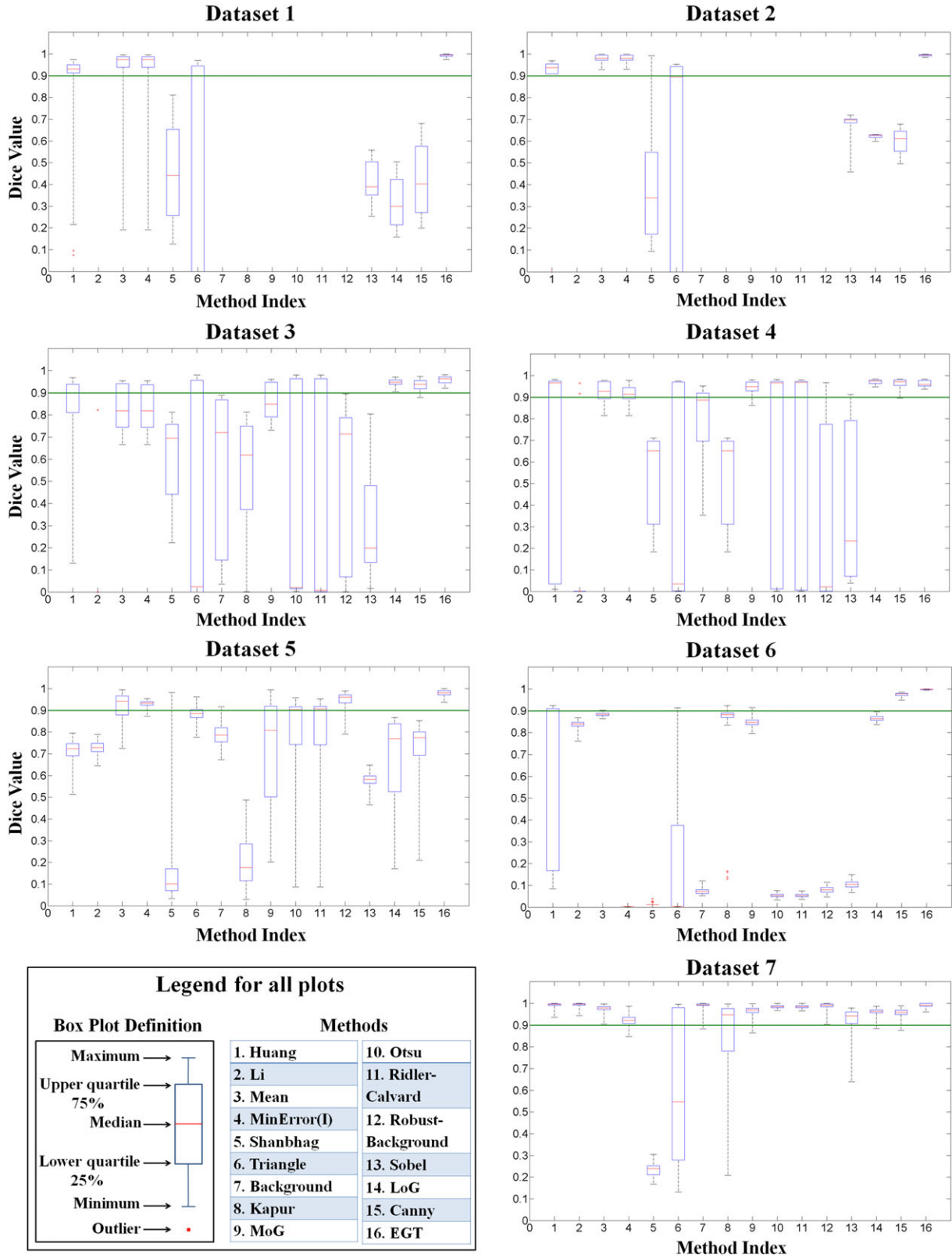
Fig. 11. Boxplots of all 16 methods applied to each of the 7 reference data sets. The green line in each plot indicates the accuracy threshold at Dice value of 0.9. A successfully segmented data set has the entire Dice boxplot above that line.
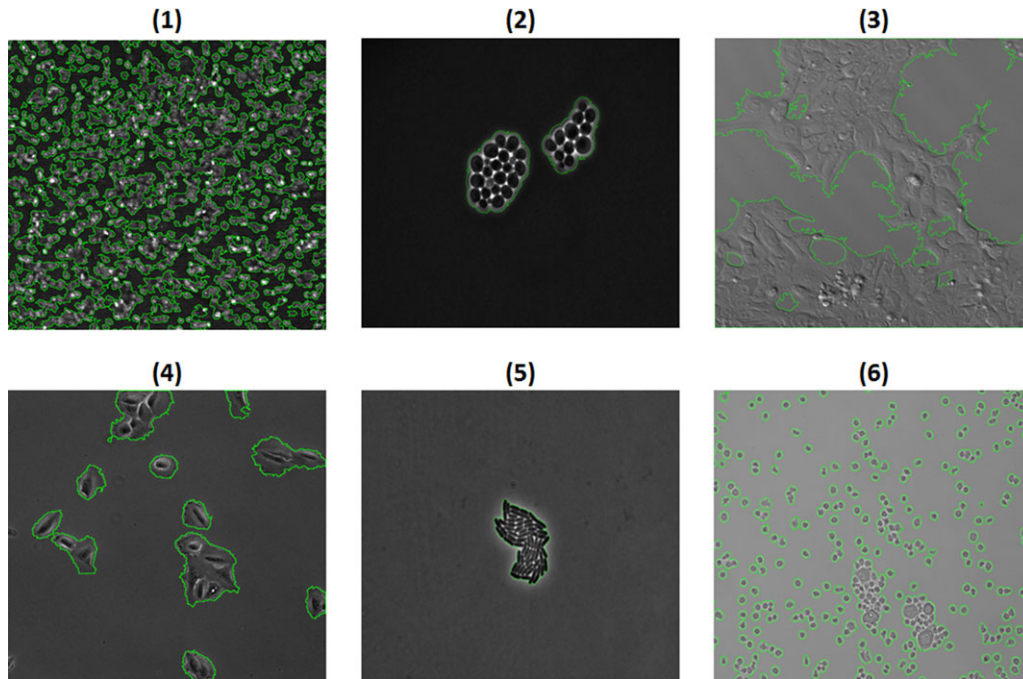
Fig. 12. Segmentation of different types of image modalities and cell lines with the contour overlaid on top of the original raw image. (1) Bright field images of rat brain cells, (2) fluorescent images of yeast cells, (3) DIC images of iPS cells, (4) phase images of bone cancer cells, (5) Fluorescent images of E. coli cells and (6) bright field image of hematopoietic progenitor cells.

images, with 16 cell lines, 61 time-sequence data sets and 17 479 total number of images.

One way how to improve the current segmentation results is to add a more sophisticated post-processing step that fills segment holes, for example, use of texture features per regions. Currently, the hole filling algorithm is based on the size of a hole (i.e. a region with background colour and size less than a user defined pixel threshold will get filled).

Another way how explore the EGT method robustness to the underlying image content is to apply it to medical images acquired using magnetic resonance imaging (MRI) or computer tomography (CT). We have conducted preliminary test with CT images of the American College of Radiology (ACR) CT accreditation phantom (Gammex 464). While qualitatively the EGT method successfully segmented the phantom objects on every image, systematic quantitative accuracy evaluations have to be performed in the future.

Conclusions

By developing this automated segmentation technique, we are able to automate segmentation to achieve higher reproducibility. This method has the potential to be applied in microscopy labs using phase contrast, DIC, bright field, and fluorescent imaging modalities in high-throughput environments. While we cannot make any statements about the use of the method in clinical practice at this point, we have

shown that the performance is satisfactory across multiple imaging modalities and cell types.

Working with terabyte-sized images requires segmentation techniques to be not only accurate, but also computationally fast and efficient, so that very large data sets can be processed in a timely manner with limited computational resources. We found that we had a need for a new fast and efficient segmentation method that was robust across all of the challenges of these data sets. The EGT is an empirical method derived from a very wide range of biological images, varying in image modality, cell density, and pixel intensity and gradient ranges. It satisfied all of the requirements and has shown to be highly accurate on all data sets we have used it on to date. We have released an open-source user interface for the community to test this technique on an even wider range of applications.

Brain Development for sharing the iPS cell images. Finally, we would like to acknowledge Steven Lund from the Statistics Engineering Division at NIST for his input on the percentiles part of the technique, David Nimorwicz from Software and Systems Division at NIST for creating the graphical user interface (GUI) for this work, and Walid Keyrouz from Software and Systems Division at NIST for his input on the manuscript writing.

Disclaimer

Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

References

Buggenthin, F., Marr, C., Schwarzfischer, M., Hoppe, P.S., Hilsenbeck, O., Schroeder, T. & Theis, F.J. (2013) An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. *BMC Bioinformatics* **14**.

Carpenter, A.E. & Jones, T.R. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**.

Chalfoun, J, Kociolek, M., Dima, A.A., Halter, M., Cardone, A., Peskin, A., Bajcsy, P. & Brady, M. (2013) Segmenting time-lapse phase contrast images of adjacent NIH 3T3 cells. *J. Microsc.* **249**, 41–52.

Chalfoun, J., Majurski, M., Dima, A., Stuelten, C., Peskin, A & Brady, M. (2014) FogBank: a single cell segmentation across multiple cell lines and image modalities. *BMC Bioinformatics* **15**.

Couprie, C., Grady, L., Najman, L. & Talbot, H. (2009) Power watersheds: a new image segmentation framework extending graph cuts, random walker and optimal spanning forest. *2009 IEEE 12th Int. Conf. Comput. Vis.* 731–738.

Cremers, D., Rousson, M. & Deriche, R. (2006) A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *Int. J. Comput. Vis.* **72**, 195–215.

Despotović, I., Vansteenkiste, E. & Philips, W. (2013) Spatially coherent fuzzy clustering for accurate and noise-robust image segmentation. *IEEE Signal Process. Lett.* **20**, 295–298.

Di Talia, S., Skotheim, J.M., Bean, J.M., Siggia, E.D. & Cross, F.R. (2007) The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448**, 947–951.

Dice, L. (1945) Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302.

Dima, A.A., John, E., James, F., *et al.* (2011) Comparison of segmentation algorithms for fluorescence microscopy images of cells. *Cytom. Part A J. Int. Soc. Anal. Cytol.* **79**, 545–559.

Egmont-Petersen, M., deRidder, D. & Handels, H. (2002) Image processing with neural networks—a review. *Pattern Recognit.* **35**, 2279–2301.

Gonzalez, R.C., Woods, R.E. & Eddins, S.L. (2008) Digital Image processing Pearson.

Halter, M., Sisan, D., Chalfoun, J., *et al.* (2011) Cell cycle dependent TN-C promoter activity determined by live cell imaging. *Cytom. Part A* **79**, 192–202.

Kamdi, S. & Krishna, R. (2012) Image segmentation and region growing algorithm. *Int. J. Comput. Technol.* **2**, 103–107.

Khan, I., Lupi, M., Campbell, L., *et al.* (2011) Interoperability of time series cytometric data: a cross platform approach for modeling tumor heterogeneity. *Cytometry. A* **79**, 214–226.

Mobahi, H., Rao, S.R., Yang, A.Y., Sastry, S.S. & Ma, Y. (2011) Segmentation of natural images by texture and boundary compression. *Int. J. Comput. Vis.* **95**, 86–98.

Rosenfeld, N., Perkins, T.J., Alon, U., Elowitz, M.B. & Swain, P.S. (2006) A fluctuation method to quantify in vivo fluorescence data. *Biophys. J.* **91**, 759–766.

Schindelin, J., Arganda-Carreras, I., Frise, E., *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.

Sezgin, M. & Sankur, B. (2004) Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **13**, 146–165.

Stuelten, C., Busch, J., Tang, B., *et al.* (2010) Transient tumor-fibroblast interactions increase tumor cell malignancy by a TGF-Beta mediated mechanism in a mouse xenograft model of breast cancer. *PLoS One* **5**, e9832.

Wang, Q., Niemi, J., Tan, C.-M., You, L. & West, M. (2010) Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry. A* **77**, 101–110.

Weiger, M., Vedham, V., Stuelten, C.H., Shou, K., Herrera, M., Sato, M., Losert, W. & Parent, C. (2013) Real-time motion analysis reveals cell directionality as an indicator of breast cancer progression. *PLoS One* **8**, e58859.