Fast and accurate automated cell boundary determination for fluorescence microscopy

Stephen Hugo Arce, Pei-Hsun Wu[&], and Yiider Tseng[†]

Department of Chemical Engineering, University of Florida and National Cancer Institute-Physical Science in Oncology Center, Gainesville, FL 32611

[†]Address correspondence to:

Yiider Tseng, Ph.D.

Chemical Engineering Building, Room 223
1006 Center Drive
University of Florida
Gainesville, FL 32611-6005
Tel: (352) 392-0862
Fax: (352) 392-9513
Email: ytseng@che.ufl.edu

[&] Current address:

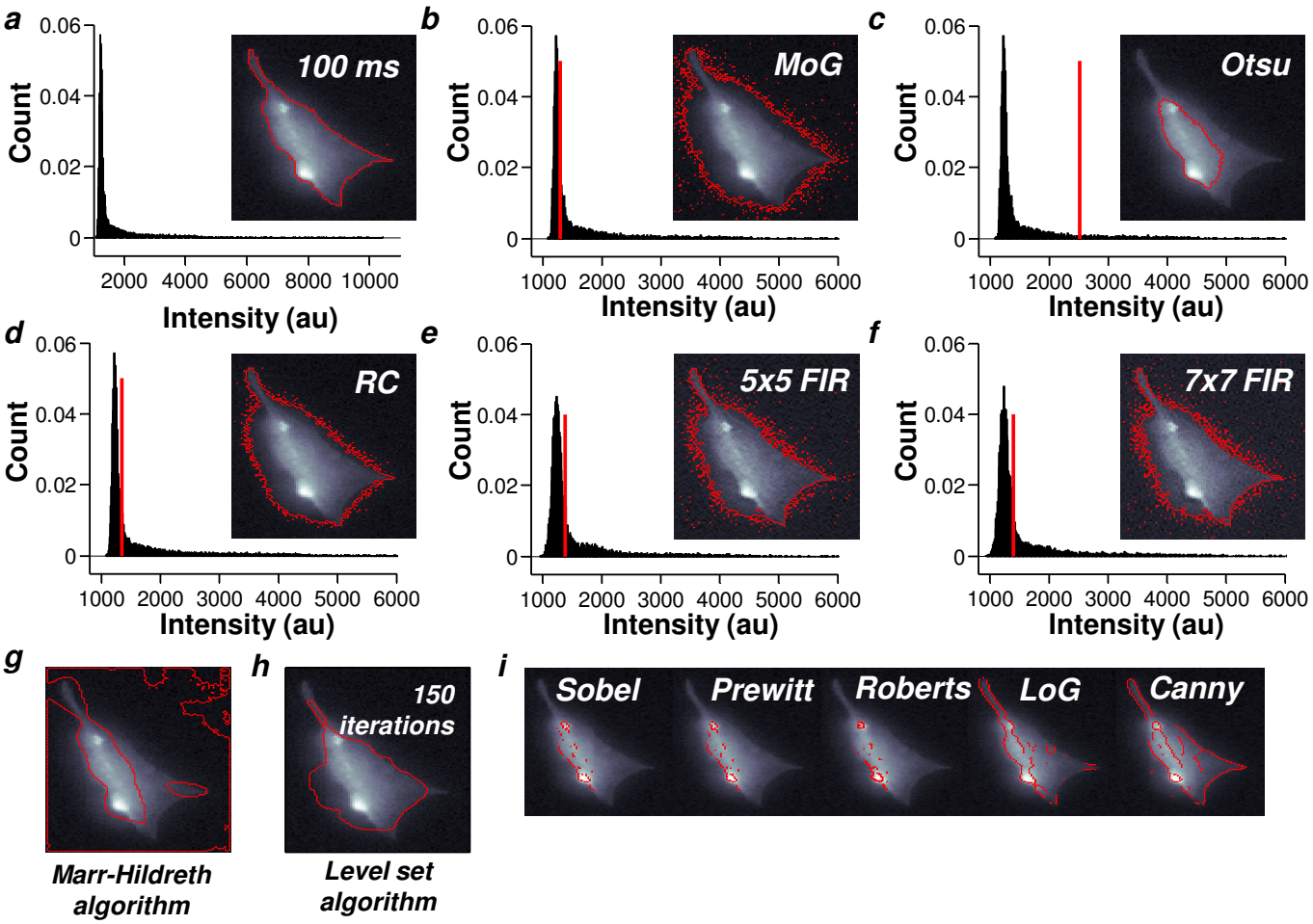
Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, Baltimore, MD 21218

Running Title: Fast and accurate automated cell boundary determination

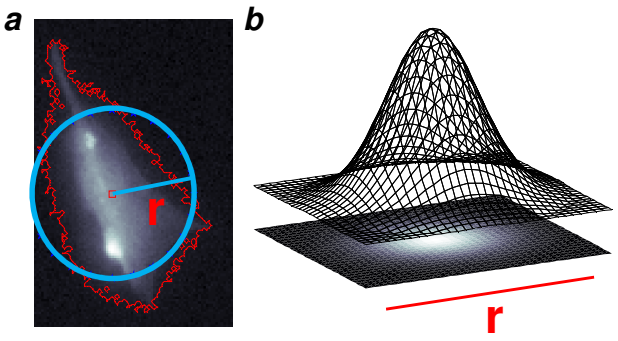
Category: Cell Biophysics

Key Words: cell boundary; fluorescence microscopy; digital image analysis; automated segmentation; threshold; video tracking; spatial filter

Abbreviations: Anisotropic Diffusion, AD; Differential Interference Contrast, DIC; Finite Impulse Response, FIR; Laplacian-of-a-Gaussian, LoG; Mixture of Gaussians, MoG; Red Fluorescence Protein, RFP; Ridler-Calvard, RC; standard deviation, STD; Variable Pressure Scanning Electron Microscopy, vpSEM.



Supplemental Figure 1 Arce et al.



Supplemental Figure 2 Arce et al.

SUPPLEMENTARY MATERIAL

Several segmentation strategies are currently available. Some use the highly regular and easy-to-segment nucleus fluorescence (commonly stained by Hoechst 33342) as a primary segmentation, and morphologically expand the initial nucleus boundary until the cell boundary is determined. This process requires a lot of extra user interaction: the extra staining step and microscopy necessary to get nuclei images, and then the precise tuning and visual inspection of cell-cell boundaries that are found. Also, the extra steps and morphological calculations take extra time to complete, adding to the workload of technicians and hurting high throughput capability. The boundaries generated by these routines do not necessarily segment cells effectively and must often be followed by steerable filtering to identify thin protrusions coming from cells. Hence, researchers may default to the simpler and easier to understand segmentation methods that employ a straightforward threshold to images.

ImageJ and CellProfiler (10, 11), two freely available imaging software packages, have a handful of built-in thresholding algorithms: the mixture of Gaussians (MoG) approach, Otsu's method, and the Ridler-Calvard algorithm, for example. The MoG method treats the histogram of pixel intensities as a mixture of two or more Gaussian distributions and uses the expectation-maximization algorithm to converge to a best fit. Then, either the parameters of the background distribution are used to calculate a threshold (done in this work), or an arbitrary factor that divides pixels based on their likelihood to be in one distribution or another is used (Supplementary Fig. S1 b). Otsu's method uses the histogram of image intensities and divides images into two classes of pixels by finding a threshold that minimizes the in-class variance of each pixel intensity distribution. (Supplementary Fig. S1 c). The Ridler-Calvard algorithm also

treats the histogram of the pixel intensities as two classes like Otsu's method, but instead of utilizing the variance, it finds the mean of each class and uses the average of the two means as a threshold once it converges (Supplementary Fig. S1 d). These methods work well in many cases, but for images of cellular fluorescence they break down against the inherent heterogeneity of cell signals and morphology.

Spatial filtering is a technique that works well to distinguish nuclei by adjusting the histogram of pixel intensities to facilitate the use of a simple thresholding method. Two finite impulse response (FIR) spatial high pass filters were transcribed directly from the work of Price (24) to try and aid segmentation of cell images: a 5x5 (Supplementary Fig. S1 e) and 7x7 (Supplementary Fig. S1 f). After the application of the FIR filters, the MoG method was used to segment each image. This approach slightly improved the boundaries around cells in comparison with straightforward thresholding, but some background noise is still included in the segmentation result nearby the brightest regions of the cell signal. The Marr-Hildreth (26) algorithm segments images by applying a Laplacian-of-a-Gaussian (LoG) filter to images and finding the zero crossing of the result. It can be implemented in two ways: a straightforward LoG filtering (Supplementary Fig. S1 i), or by applying a Gaussian spatial low pass filter (Supplementary Fig. S2) and then a Laplacian filter (12) (Supplementary Fig. S1 g).

A more state-of-the-art method to segment images is the level set algorithm. It iteratively modifies a 3-D surface using the gradient and Laplacian of raw images before segmenting the resulting surface by its zero crossing (Supplementary Fig. S1 h). The initial surface can be initialized in several ways, but the result shown was initialized using the RC threshold and creating a binary image of -1's and 1's. This produced the best boundary result after 150 iterations in comparison with other initializations and number of iterations (data not shown). The

graph cut method of segmenting images was too computationally expensive to run in Matlab. In particular, solving the eigenvalues and eigenvectors of a similarity graph causes the bottleneck for standard image sizes of about 120x120 pixels.

Finally, standard edge detection was performed on images of cell fluorescence using the built-in Matlab function 'edge'. The Sobel, Prewitt, Roberts and Canny edge detectors identified some parts of the cell boundary, but were insufficient to perform further measurements (Supplementary Fig. S1 i).

SUPPORTING MATERIAL FIGURE CAPTIONS

Supplementary Figure S1. A comparison between the boundary accuracy of several segmentation strategies.

a) A histogram of the intensity distribution in a typical fluorescent image of a NIH 3T3 fibroblast is shown. The image (inset) was captured using τ = 100 ms and 1000 bins were used to construct the histogram. The boundary in the inset image was generated using our segmentation strategy.

b) An intensity threshold calculated using the mixture of Gaussians (MoG) method is shown (vertical red line). This threshold segments the image by creating two classes of pixels: those higher than the threshold and those lower than it. A boundary has been drawn to illustrate the segmentation result by this method (inset).

c) An intensity threshold calculated using Otsu's method is shown (vertical red line). A boundary has been drawn to illustrate the segmentation result by this method (inset).

d) An intensity threshold calculated using the Ridler-Calvard (RC) method is shown (vertical red line). A boundary has been drawn to illustrate the segmentation result by this method (inset).

e) A 5 x 5 finite impulse response (FIR) spatial high pass filter was applied to the raw fluorescence image, and an intensity threshold was calculated using the mixture of Gaussians (MoG) method (vertical red line). A boundary has been drawn to illustrate the segmentation result by this approach (inset).

f) A 7 x 7 finite impulse response (FIR) spatial high pass filter was applied to the raw fluorescence image, and an intensity threshold was calculated using the mixture of Gaussians

(MoG) method (vertical red line). A boundary has been drawn to illustrate the segmentation result by this approach (inset).

g) The Marr-Hildreth method was applied to the raw fluorescence image. A spatial low pass filter result is filtered again by a Laplacian spatial high pass filter and the zero crossing of the result is used as the segmentation boundary. A boundary has been drawn to illustrate the segmentation result by this approach.

h) Implementation of the level set algorithm to segment the image was performed using a square initial boundary and run for 150 iterations. The penalization coefficient of the gradient term in the algorithm was set to 1, and the penalization coefficient of the curvature term was set to 0.1% of the squared maximum raw image intensity. More iteration would refine the boundary for the brighter region of the cell image, but the process has already neglected the cell protrusions in the image and without further user interaction to tune the coefficients of the algorithm, the segmentation result will remain inaccurate.

i) Standard edge detection was performed using the built-in Matlab function 'edge'. From left to right, the results of Sobel, Prewitt, Roberts, Laplacian-of-a-Gaussian (LoG) and Canny filters are shown.

Supplementary Figure S2. Finding the scale of cells in images and designing a spatial filter.

a) Spatial filtering requires careful consideration of the length scale during filter design. For a spatial low pass filter, the approximate length scale of the cell was estimated by first segmenting images using only a MoG threshold (in red). The segmented area was treated as if it was a circle (in blue) and the radius was calculated.

b) The approximated radius of the cell was used to determine the length scale of the spatial low pass filter. Low pass filters must contain elements that sum to 1 and should be radially symmetric; so, we employed a Gaussian filter. The sigma value for the Gaussian filter was set to $1/6^{th}$ of the radius value since about 96% of the Gaussian density lies within that range. A greater sigma value would flatten the filter and it would begin to approach the shape of a boxcar kernel at very high sigma. A lower sigma value would invalidate the calculation of the radius since contributions to the low pass result at the edge of the filter would be essentially zero. Finally, the elements of the filter in the corners were also set to zero by overlaying a flat disk of the same length scale on the Gaussian filter. This step avoided the small amount of bias for image elements located at the diagonals of the filter in the low pass result.