

# A Quality Pre-Processor for Biological Cell Images

Adele P. Peskin<sup>1</sup>, Karen Kafadar<sup>2</sup>, and Alden Dima<sup>3</sup>

<sup>1</sup> NIST, Boulder, CO 80305

<sup>2</sup> Dept of Statistics and Physics, Indiana University, Bloomington, IN 47408-3725

<sup>3</sup> NIST, Gaithersburg, MD 20899

**Abstract.** We have developed a method to rapidly test the quality of a biological image, to identify appropriate segmentation methods that will render high quality segmentations for cells within that image. The key contribution is the development of a measure of the clarity of a biological cell within an image that can be quickly and directly used to select a segmentation method during a high content screening process. This method is based on the gradient of the pixel intensity field at cell edges and on the distribution of pixel intensities just inside cell edges. We have developed a technique to synthesize biological cell images with varying qualities to create standardized images for testing segmentation methods. Differences in quality indices reflect observed differences in resulting masks of the same cell imaged under a variety of conditions.

## 1 Introduction

High content screening (HCS) has become a critical method for large scale cell biology and is often used for drug discovery. HCS is the automation of cell biological investigation using automated microscopes and sample preparation and includes the acquisition and analysis of cellular images without human intervention. Quantitative fluorescent microscopy plays a key role in HCS as it does in cell biology in general. Typical HCS-based experiments can involve the analysis of more than a million cells [1][2][3][4][5].

Image segmentation is the most important part of analyzing biological image data [4]. Consequently, many segmentation methods have been published, including histogram-based, edge-detection-based, watershed, morphological, and stochastic techniques [4]. However, segmenting the same image with different methods can lead to different masks for a given cell, and hence different estimates of cell characteristics (e.g, area, perimeter, etc.). We refer to a mask as the set of pixels that are used to define an object within an image.

Accurate segmentation is one of the challenges facing high content screening [5]. Given the large number of images and the consequent impracticalities of human supervision, high content screening can be expected to pose additional

---

<sup>0</sup> This contribution of NIST, an agency of the U.S. government, is not subject to copyright.

challenges to ensuring reliable and accurate cell image segmentation. Indeed, the handling and analysis of large image sets has been identified as an impediment to the wider use of HCS [4]. Therefore it is important to obtain some predictive, objective and efficient measure of the segmentation quality resulting from a particular method, on a cell-by-cell basis, for different cells. Not all cells in the same image will lead to segmentations with the same accuracy. Certain clinically relevant cell lines are also known to be particularly difficult to segment, making the automated analysis of image data unreliable. For example, in one study, the dose response of morphological features varied significantly between well segmented and poorly segmented cells taken from high content screening data [6].

**Fig. 1.** One kernel at  $x_s$  (*dotted kernel*) or two kernels at  $x_i$  and  $x_j$  (*left and right*) lead to the same summed estimate at  $x_s$ . This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption. The last sentence of a figure caption should generally end without a period

The goal of this project is to provide a simple, yet reasonably faithful, measure of the quality of an individual cell in an image, to be used to facilitate the choice of a segmentation method and for the choice of focus setting for the cell image. To define a quality index for a cell image, we first examine a series of images and their ranges in quality (albeit subjective), revealing types of features associated with low/high quality. Section 2 describes the data images considered in this project. Examination of these images leads to the definition of the quality index in Section 3. We then illustrate in Section 4 our method of creating synthetic images with a given quality index. While our measure is shown to behave as expected (i.e., poor images associated with low quality indices), a field

study is needed to validate this measure for more general cases. We propose the investigation of this issue and others in the concluding Section 5.

## 2 Data description

We examined 16 sets of fixed cell images prepared by our biological collaborators at NIST. These images consist of A10 rat smooth vascular muscle cells and 3T3 mouse fibroblasts stained with Texas Red cell body stain [7]. For each set, we compared six different images for a total of 96 images. Each image comprises multiple cells. Five of the six images are repeated acquisitions under five sets of imaging conditions. The first three sets were acquired using three different exposure times with Chroma Technology’s Texas Red filter set (Excitation 555/28, #32295; dichroic beamsplitter #84000; Emission 630/60, #41834). This filter is matched to the cell body stain used and allows us to compare the effect of exposure time on acquired cell images. We used non-optimal filters to reduce intensity signal to noise and introduce blurring: Chroma Technology’s GFP filter set (Excitation 470/40, #51359; dichroic beamsplitter #84000; Emission 525/50, #42345).<sup>4</sup> The resulting images were blurred in a fashion similar to the Gaussian blurring operation found in many image processing systems. These five image sets were acquired using 2x2 binning in which the output of four CCD image elements are averaged to produce one image pixel. Binning is typically used to trade image resolution for acquisition speed and sensitivity [2].

The sixth image is a higher-resolution 1x1 binned image, using the full image range to define pixel intensities. We use this image to create a ground truth mask for each cell; the pixels are characterized (background, edge, cell interior) via expert manual segmentation to assure a “gold standard” image. The five conditions that generated the five images to be compared with the “gold standard” image are a combination of filter type (optimal or non-optimal) and illumination level (low, medium, high), and are given in Table 1. We visualize the overall quality of the image by color-coding the pixels. Figure 2 provides a visual rendition of the five conditions for cell number 4 from image set 1.

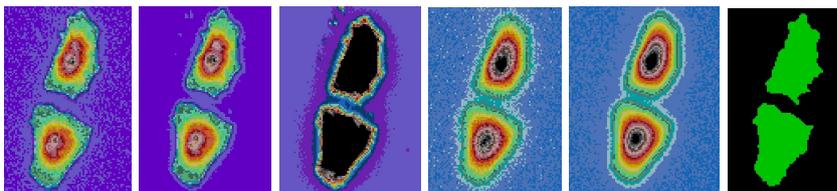
Each series of five color-coded images revealed differences in the implied masks representing a particular cell under the five imaging conditions. The apparent cell edges vary widely in clarity and sharpness across the five different images of the same cells. In particular, the images vary in terms of the number of pixel lengths (distance between pixels) needed to represent the thickness of the edge regions of the cells. We will attempt to quantify this feature in the next section.

---

<sup>4</sup> Certain trade names are identified in this report in order to specify the experimental conditions used in obtaining the reported data. Mention of these products in no way constitutes endorsement of them. Other manufacturers may have products of equal or superior specifications.

**Table 1.** The five sets of imaging conditions.

Image	Illumination Level	Exposure time(s)	Filter type
1	low	0.015	optimal filter (Texas Red filter set)
2	medium	0.08	optimal filter (Texas Red filter set)
3	high	0.3	optimal filter (Texas Red filter set)
4	low	1.0	non-optimal filter (GFP filter set)
5	medium	5.0	non-optimal filter (GFP filter set)
6	<i>(gold standard)</i>		

**Fig. 2.** Individual pixel intensities are color-coded over the range in each image, to show differences in edge sharpness. The ground truth mask is shown for comparison.

### 3 Quality Calculation

For each cell in an image, we look at pixel intensities within an isolated region containing the cell. We fit this distribution using a 3-component Gaussian mixture via the EM (Expectation-Maximization) algorithm. A 2-component mixture model, with a cell component and a background component, resulted in estimates for means, standard deviations, and component fractions significantly different from the actual data. Thus we reasoned that the edge pixels, whose intensities span a wide range of pixels, form their own third distribution. The 3-component model is illustrated in Figure 3 for the second and fifth cells of Figure 2 and better reflects the actual data. We provide the equations for fitting the 3-component model (described in general terms in [8] [9]) in [10]. As illumination is increased, the overall range of pixel intensities increases, giving a better separation of the background and edge peaks. At high illumination, however, the center of the cell is flooded with light, and the edge distribution is very thin and pushed back towards the background distribution. In the discussion below,  $x_p$  and  $s_p$  denote the estimated mean and standard deviation, respectively, of the intensities from pixels of type  $p$  from the EM algorithm, where  $p$  corresponds to  $B$  (background pixels),  $E$  (edge pixels), or  $C$  (cell interior pixels).

We examined 16 sets of data in detail. For each set of imaging conditions in Table 1, we find a consistent background distribution. Data for the background mean,  $x_B$ , and standard deviation,  $s_B$ , as well as for the edge pixel mean,  $x_E$ , and standard deviation,  $s_E$ , for the first image of each of the 16 sets, taken at low illumination with the optimal filter (set 1 in Table 1), are given in Table 2. The

**Table 2.** Background and Edge values for all low illumination, optimal filter images

Set	Cell number	Background Mean, $x_B$	Background SD, $s_B$	Edge Mean, $x_E$	Edge SD, $s_E$
1	1	115	4	141	19
2	5	119	4	193	42
3	2	117	4	178	36
4	1	115	4	188	50
5	2	116	4	238	85
6	1	119	4	195	40
7	1	119	4	202	38
8	2	116	3	131	12
9	1	119	5	152	24
10	1	113	4	156	33
11	1	115	4	141	20
12	1	116	3	158	31
13	1	113	3	152	25
14	1	117	5	190	59
15	1	114	4	178	48
16	1	117	4	171	31
Mean		116	3.94	173	37
SD		0.51	0.14	7.0	4.45

N.B. Pooled  $s_B = 3.98$  ( $\text{mean}(s_B) = 3.94$ ); pooled  $s_E = 40.87$  ( $\text{mean}(s_E) = 37.06$ )

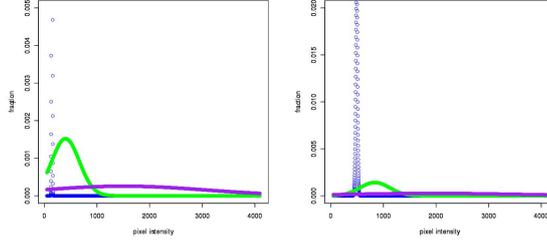
overall Gaussian shapes of the edge and cell distributions are similar in each of the 16 sets, although there is more variation in the edge distributions. The data for Table 2 refer to estimates from the 3-component model for the distribution of pixel intensities (background, cell, edge) for a single cell chosen from each image. We choose the first cell of each image for this table, unless a particular cell in the image was used consistently in another part of this analysis.

For each pixel intensity between  $x_B$  and  $x_E$  (i.e., between the mean of the background distribution and the mean of the edge distribution), we also calculate an average value for the magnitude of the gradient at that intensity. We look at the gridded data within a bounding box on the image that contains a single cell in an image. At each pixel location,  $p_{ij}$ , with pixel intensity,  $I_{ij}$ , we find a local derivative of that intensity and its magnitude, using a Sobel mask. In the  $i$  direction, a numerical estimate for the directional derivative at each  $p_{ij}$  is:

$$\partial I / \partial i |_{p_{ij}} = [2 \cdot I_{(i+1)j} - 2 \cdot I_{(i-1)j} + I_{(i+1)(j-1)} - I_{(i-1)(j-1)} + I_{(i+1)(j+1)} - I_{(i-1)(j+1)}] / 8. \quad (1)$$

In the  $j$  direction, we find the numerical estimate at each  $p_{ij}$ :

$$\partial I / \partial j |_{p_{ij}} = [2 \cdot I_{i(j+1)} - 2 \cdot I_{i(j-1)} + I_{(i-1)(j+1)} - I_{(i-1)(j-1)} + I_{(i+1)(j+1)} - I_{(i+1)(j-1)}] / 8. \quad (2)$$



**Fig. 3.** Normalized curves of the 3 components of pixel intensities, as an example, for the second and fifth cells of Figure 1, background(blue), edge(green), and cell(purple).

The total derivative gives the magnitude of the gradient of the intensity at  $p_{ij}$ , which we denote by  $g(i, j)$ :

$$g(i, j) \equiv dI(i, j) = \sqrt{(\partial I / \partial i)^2|_{p_{ij}} + (\partial I / \partial j)^2|_{p_{ij}}}. \quad (3)$$

We find a gradient magnitude value  $g(i, j)$  at each location  $p_{ij}$ . The average value of the gradient magnitude at those locations where the measured intensity level is a given value, say  $I^*$ , is given by:

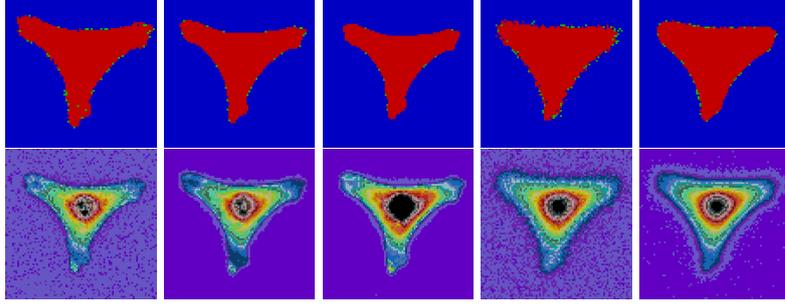
$$G(I^*) = \text{ave}\{g(i, j) : I(i, j) = I^*\}. \quad (4)$$

If  $I_{min}$  is the lowest intensity value found within the cell bounding box, and  $I_{max}$  is the largest intensity, then for each intensity  $I$  between  $I_{min}$  and  $I_{max}$ , we find the average gradient magnitude for that intensity. If, for example, a particular intensity value,  $I_v$ , occurs 3 times within the bounding box, at positions  $p_{ab}$ ,  $p_{cd}$ , and  $p_{ef}$ , the average magnitude of the gradient at intensity =  $I_v$  is:

$$G_v = (g_{ab} + g_{cd} + g_{ef})/3. \quad (5)$$

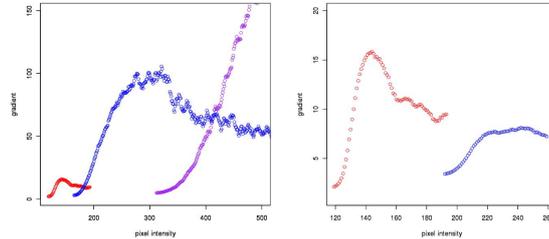
We will refer to this average gradient function as  $G(I)$ , where  $I$  is an intensity in the region of the edge. Recall that our 3-component model classified intensities as background, edge, and cell, so values of  $I$  near the edge likely lie between  $x_B$  and  $x_E + 2s_E$ . We denote by  $A$  the intensity for which  $G(I)$  is largest for  $x_B \leq I \leq x_E$ ; i.e.,  $A = \text{argmax}_{x_B \leq I \leq x_E} G(I)$ . In Figure 4, we show the locations of the pixels,  $p_{ij}$ , whose intensities are equal to Intensity A. Using our visualization tools, we see that these pixels lie within one or two pixel lengths of the apparent edge of the cell, which we see in accompanying images of more detailed contouring of the pixel intensities. Because the data are discretized, we can only approximate the position within 1-2 grid points in all directions.

The average gradient magnitude curve,  $G(I)$  vs. intensity  $I$ , reflects two cell characteristics. The initial slope of the curve directly indicates the sharpness of the edge: the steeper the slope, the sharper the edge. The shape of this curve also indicates a feature of the cell edges. For very sharp edges, the average



**Fig. 4.** Cell 2, set 1, imaging conditions 1-5: pixel intensities above Intensity A are red, Intensity A ( $\pm 1$  unit), green, and below A, blue. Green pixels fall close to the apparent cell edge. Below, each cell is color-coded in 40 equally-spaced pixel ranges.

gradient magnitude increases monotonically across the range from the mean of the background,  $x_B$ , to the mean of the edge distribution,  $x_E$ , as is seen for all high illumination images (type 3 in Table 1). For the low and medium illumination images, the gradient increases from  $x_B$ , reaches a maximum, and then falls. The graphs for the gradient magnitude for cell 5, image set 2, under low, medium, and high illumination with the optimal filter are shown in Figure 5a, and for low illumination for both filters in Figure 5b.



**Fig. 5.** Cell 5, set 2: a.) Averaged magnitudes of the pixel gradient for imaging conditions 1(red), 2(blue), 3(purple); b.) for imaging conditions 1(red), 4(blue).

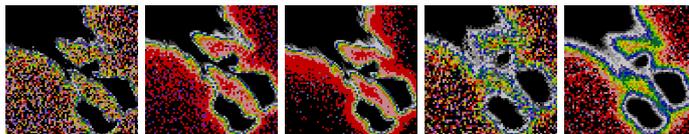
For the type 3, high illumination images, the gradient magnitude continually increases between  $x_B$  and  $x_E$ , so we set the maximum gradient value in this region, Intensity A, to be the magnitude of the gradient at  $x_E$ . The edge distribution is very close to the background distribution for these images, and we do not expect a rise and fall in the plot of the gradient magnitude within this small part of the curve (see the purple curve of Figure 5).

We assume that the quality of the image is a function of both the pixel intensities and their gradients at the edge of a cell. To assign a quality value to a cell,

we first analyze the background pixel distribution and edge pixel distribution, and compute the average gradient magnitude as a function of pixel intensity in the range between the means of these two distributions.

We measure the quality of an image in terms of an index that measures how rapidly the pixel intensities fall from the intensity at the maximum gradient at the edge (Intensity A) to the background mean value ( $x_B$ ). In particular, we calculate the expected fraction of this range that should lie on the image within two physical pixel lengths of intensity  $A = \operatorname{argmax}_{x_B \leq I \leq x_E} G(I)$ . We look at the quality indices  $QI$  of cells from several sets of images taken from two cell lines to illustrate the calculations of  $QI$ .

Consider the set of images from cell 5 in set 2, which yielded the gradient curves in Figure 5. The intensity range between  $x_B$  and Intensity A is divided into 10 equal-pixel-length ranges, and each range is displayed in Figure 6, enlarged pictures of the lower right corner of each of the five images in the region containing the cell pseudopodia. This example is chosen because the pseudopodia are connected to the rest of the cell over a very narrow region, of approximately the same size as the boundary edge region we use to define the quality measure.



**Fig. 6.** The mask for cell 5 of set 2, and enlarged pictures of the lower right hand portion of the cell containing the pseudopodia, under the five imaging conditions.

From the magnitude of the gradient, we calculate the expected fraction of the cell edge that lies within two pixel lengths of the maximum gradient at the edge. We use the following sequence of steps:

1. Find the 3-component pixel intensity distribution; denote means of the components by  $x_B$ ,  $x_E$ ,  $x_C$ .
2. Find the average gradient magnitude at each intensity between  $x_B$  and  $x_E$ .
3. Smooth the gradient in this region to fill in any gaps, and denote the resulting function by  $G(\text{Intensity})$ .
4. Find the intensity, Intensity A, at which the smoothed gradient magnitude is maximized.
5. Find the expected neighboring pixel to a pixel with Intensity A and denote this intensity as B; i.e., Intensity B = A -  $G(A)$ \*(1 pixel unit).
6. Find the expected neighboring pixel to a pixel with Intensity B; i.e., Intensity C = B -  $G(B)$ \*(1 pixel unit) = A -  $G(A)$ \*(1 pixel unit) -  $G(A-G(A))$ \*(1 pixel unit)\*(1 pixel unit).
7. Compute the quality index as  $QI = (A - C)/(A - x_B)$ .

We perform the above calculations for the five different images of cell 5, image set 2. The results of each of the steps are given in Table 3. Comparing these qualities with the images of these cells, the quality index appears to describe how well the very thinnest geometry of the cell, the place where the pseudopodia are attached, are apparent in the image. For the optimal filter images, increasing the illumination increases the range of pixel intensities in the image, separating the background and edge distributions from each other and from the pixel intensity distribution that represents the cell and its pseudopodia. The sharper the edge, the higher the quality index. The thin connection (2 to 3 pixel lengths) between the main part of the cell and the pseudopodia is clearly seen for the medium and high illumination figures, and pixel intensities between the two parts of the pseudopodia decrease to levels near  $x_B$ . For the images taken with the non-optimal filter, the connection between the main part of the cell and the pseudopodia is lost in the images, and the pixel intensities between the two parts of the pseudopodia do not approach  $x_B$ . It is clear from these figures that the decrease in pixel intensity between the inside of the cell and the background covers much more than two pixel lengths. In the medium illumination, non-optimal filter image, type 5, much less noise appears than in the corresponding low illumination image, type 4, but the edge of the cell covers, on average, about 4 to 5 pixel lengths, so no connection appears between the lower portion of the pseudopodia and the main portion of the cell.

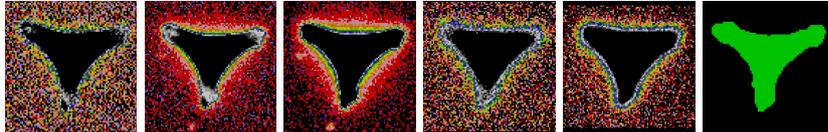
**Table 3.** Quality calculation for cell 5 of image set 2 for the five test images.

Type	$x_B$	A	G(A)	B	G(B)	C	Quality
1	119	146	16.06	130	7.91	123	0.852
2	165	319	109.13	210	45.23	165	1.000
3	321	799	480.50	319	5.38	314	1.015
4	192	222	7.87	215	7.02	208	0.467
5	524	692	39.92	653	33.67	620	0.429

## 4 Data Creation

To compare cell segmentation techniques with one another, we use the quality index we defined above to create sets of test images over a specified range of qualities. We assume that a given segmentation method will have some quality range over which the segmentation provides a given measure of accuracy. A quality index permits a quick assessment of the accuracy of the cell area from a particular segmentation method. To illustrate the creation of sets of synthetic cells with a specified range of qualities, we describe and show a set created from a ground truth mask of one of the cells in the 16 sets described above, cell 2 from set 10.

The five images of the cell of interest from that set have qualities equal to 0.842, 0.910, 1.137, 0.563, and 0.636. The images of those cells are shown in Figure 7. To test our method, we will construct a cell with pixel distributions and gradients that correspond with the second image, and compare its quality index to the index for that cell, 0.910. We now construct the new cell from the mask, such that its distribution has three components whose means and standard deviations agree with those of the original cell. Then we use the other characteristics of that cell, namely the intensity of maximum gradient at the edge and the cell quality, to compare with the new cell to ensure a match between the set values and the calculated values on the synthetic image. To begin, we read in the mask from Figure 7, and assign each pixel of the new image a “0” if it is part of the background of the mask, a “2” if it is part of the inside of the cell on the mask, and a “1” if it is part of the mask, but has a neighboring pixel that is not part of the mask. We build our cell edge by starting at pixels coded as “1”.



**Fig. 7.** The five different images of the cell whose mask is used for the cell synthesis and the corresponding mask.

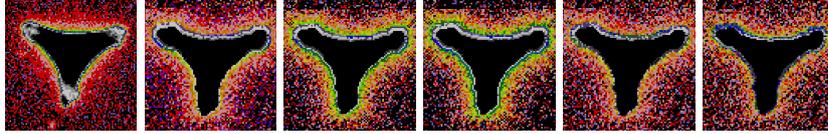
We assemble a pixel distribution for the new cell using the three components of the original cell. For cell 2 of image 10, the type 2 image,  $x_B$ ,  $x_E$ , and  $x_C$  respectively are 132, 225, and 826, and the  $s_B$ ,  $s_E$ , and  $s_C$  values are 5, 80, and 817. The fractional components are 0.800, 0.103, and 0.096. We calculate the expected number of pixels at each intensity by adding the contributions of each component at each intensity. If the total number of pixels in this region is  $T$ , then the expected number of pixels with intensity  $I$  is:

$$p_E = 0.800T; p_B = 0.103T; p_C = 0.096T; \quad (6)$$

$$\#pixels(I) = p_B\phi(I; x_B, s_B) + p_E\phi(I; x_E, s_E) + p_C\phi(I; x_C, s_C), \quad (7)$$

where  $\phi(I; \mu, \sigma)$  denotes the Gaussian density function with mean  $\mu$  and standard deviation  $\sigma$  at a particular value of the intensity,  $I$ . For each pixel intensity, we compute the difference between the raw data and the data computed from the 3-component distribution above to determine the standard deviation in the raw data. We compute the standard deviation over different pixel intensity ranges. Larger standard deviations are expected with more data in the range of the background pixel intensities.

We collect two sets of pixel intensities from the 3-component distribution. One set contains intensities at or below Intensity A, the intensity where the gradient is maximized. One set lies above Intensity A. For each pixel intensity in each set, we modify the actual pixel intensity by changing it to the Gaussian variate determined by  $N(x_p, s_p^2)$ , where  $p$  corresponds to  $B$  (background),  $E$  (edge), or  $C$  (cell interior) pixels. The pixels in the first set are placed outside of the mask according to their distances from the pixels marked with a 1. The pixels in the second set are placed inside of the mask according to their distances from the pixels marked with a 1. Figure 8b shows the resulting new cell alongside the original cell, 8a, from which it is made. We then test the new image by recalculating the 3-component analysis, pixel intensity gradient, and quality to ensure that it retained the same features. The means of the 3-component analysis are only slightly different, as would be expected by imposing random noise:  $x_B = 134$ ,  $x_E = 214$ , and  $x_C = 770$ . The intensity of maximum gradient at the edge occurs at Intensity A = 181, and the quality is 0.915. So we have created a cell with the same basic features as the original.



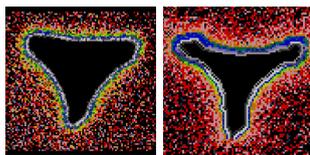
**Fig. 8.** a.)The original cell imaged with medium illumination, optimal filter. b.)Synthesized cell of the same type as (a), made as a test case. c,d.)Two cells identical except for the spread of the background pixels: standard deviations are 10 and 12 compared to original 5. e,f)Two cells identical except that the edge distribution has been shifted from a mean of 225 to a mean of 215 and 205.

Now we can change some of the characteristics of this synthetic cell and monitor the corresponding change in quality. First, we create a set of cells in the same way as the cell above, but change the standard deviation of the background peak. The original background peak has a standard deviation of 5, and we create two new cells with backgrounds of mean  $x_B = 132$  as before, but with standard deviations,  $s_B$  equal to 10 and 12 instead of 5 (see Figure 8c,d). The quality indices of these new cells change from 0.910 to 0.766 and 0.667, respectively.

Next we modify the edge distribution without changing the magnitude of the pixel intensity gradient. We shift the value of  $x_E$  from its original value of 225

to either 215 or 205, so that the modified edge distribution overlaps more closely with the background distribution. The results for these two examples are shown in Figure 8e,f. The image qualities for these two examples change from 0.910 to 0.878 and 0.825, respectively. As one would expect, the quality index decreases as the overlap between the background and edge distributions increases.

We can use these techniques to look at the effect of using a less optimal filter for cell imaging. We focus on the type 5 image of this same series, imaged with medium illumination and a non-optimal filter. If we try to reconstruct this cell, using the original cell's 3-component mean and standard deviation values, the same pixel gradient, and the same value of pixel intensity for the maximum gradient at the cell edge, we create a cell with basically the same quality as the original cell (0.59 compared with 0.58 for the original), but the shape of the cell looks very different, shown in Figure 9. Clearly the precise shape of the cell is influenced by the optical effects introduced with different filter sets, leading to different segmentation results, and hence the masks for the two cells in Figure 9 will be different. Thus this index appears to describe only the clarity of the cell edge, which will differ from the true mask as the quality decreases.



**Fig. 9.** Original cell imaged with medium illumination, non-optimal filter; cell synthesized with the same characteristics, starting from the ground truth mask.

## 5 Conclusions and future work

The features of a biological cell image are complex. The accuracy of a cell segmentation by any given method can depend on many different factors: the magnitude of the pixel gradient at the cell edges, the background pixel intensity distribution: the overlap between the background pixel intensity distribution and the edge pixel intensity distribution, and spreads in these distributions. It can also depend upon other geometric features of a cell, such as roundness of the cell and jaggedness of the edges of the cell.

The goal of this quality index is to provide an indication of the segmentation method that should be used to yield the highest accuracy in derived cell quantities (e.g., area, perimeter, etc.). Ideally, we would like to be able to distinguish between those segmentation methods that over- or under-estimate cell area from those that provide unbiased estimates of cell area, based on the quality of the edge of the cell. Because our quality index as defined here measures only edge

clarity, we may need to modify it to better capture cell image features. Our future work includes a more thorough investigation of the properties of this quality index, particularly its relation to accuracy of cell area. We plan to create sets of images with a very wide range of quality indices and cell geometries, to use as test images for a wide variety of segmentation methods. From them we hope to use this quality index (possibly modified) to choose segmentation techniques for very rapid analysis of large numbers of biological images.

## 6 Acknowledgements

Kafadar's support from Army Research Office Grant Number W911NF-05-1-0490 awarded to University of Colorado-Denver, and National Science Foundation Grant Number DMS 0802295 awarded to Indiana University is gratefully acknowledged. James Filliben of NIST's Statistical Engineering Division and John Elliott of NIST's Biochemical Science Division (BSD) designed the experiment that John performed to produce the cell images used in this study. Anne Plant and Michael Halter, of BSD, provided helpful discussion and input.

## References

1. Abraham, V. C., Taylor, D. L., Haskins, J. R.: High content screening applied to large-scale cell biology. *Trends in Biotechnology*. 22, 15–22 (2004)
2. Cox, G.C.: *Optical Imaging Techniques in Cell Biology*. CRC Press, Boca Raton, FL (2007)
3. Giuliano, K. A., Haskins, J. R., Taylor, D. L.: Advances in High Content Screening for Drug Discovery. *Assay and Drug Development Technologies*. 1(4), 565–577 (August 2003)
4. Zhou X., Wong, S.T.C.: High content cellular imaging for drug development. *IEEE Signal Processing Magazine*. 23(2), 170–174 (March 2006)
5. Zhou X., Wong, S.T.C.: Informatics Challenges of High-Throughput Microscopy. *IEEE Signal Processing Magazine*. 23(3), 63–72 (May 2006)
6. Hill, A.A., LaPan, P., Li, Y., Haney S.: Impact of image segmentation on high-content screening data quality for SK-BR-3 cells. *BioMed Central Bioinformatics*, 8, 340 (Sep 14 2007)
7. Elliott J.T., Woodward J.T., Langenbach K.J., Tona A., Jones P.L., Plant A.L.: Vascular smooth muscle cell response on thin films of collagen. *Matrix Biol*. 24(7), 489-502 (Oct 2005)
8. Hastie, T., Tibshirani, R.; Friedman, J.H.: *The Elements of Statistical Learning: Data Mining, Inference, And Prediction*. New York, Springer (2001)
9. Scott, D.W: Parametric Statistical Modeling by Minimum Integrated Square Error. *Technometrics*. 43, 274–285 (2001)
10. Peskin, A.P., Kafadar, K., Santos, A.M., Haemer, G.G.: Robust Volume Calculations of Tumors of Various Sizes. 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition (2009)