# A Quantitative Object-Level Metric for Segmentation Performance and Its Application to Cell Nuclei

Laura E. Boucheron[1,2], Neal R. Harvey[2], and B.S. Manjunath[1]

[1] University of California Santa Barbara, Electrical and Computer Engineering, Santa Barbara, CA 93106-9560

[2] Los Alamos National Laboratory, Space and Remote Sensing Sciences, P.O. Box 1663, Los Alamos, NM 87545

**Abstract.** We present an object-level metric for segmentation performance which was developed to quantify both over- and under-segmentation errors, as well as to penalize segmentations with larger deviations in object shape. This metric is applied to the problem of segmentation of cell nuclei in routinely stained H&E histopathology imagery. We show the correspondence between the metric terms and qualitative observations of segmentation quality, particularly the presence of over- and under-segmentation. The computation of this metric does not require the use of any point-to-point or region-to-region correspondences but rather simple computations using the object mask from both the segmentation and ground truth.

## 1 Introduction

The subject of objective and quantitative evaluation of segmentation performance has received less attention than has the development of various segmentation algorithms themselves. This has been noted by many researchers in the fields of computer vision and image analysis [1,2,3,4,5,6,7].

In our development of a quantitative metric, we avoid metrics of segmentation performance that rely on point-to-point or region-to-region correspondences (e.g., [2,8,9]). We also avoid empirical goodness metrics, as defined in [6], whereby properties of a "good" segmentation are defined a priori according to human perception of a "good" segmentation; in our application domain of cancer imagery it is difficult to define a single model which applies to all image objects. While there is much research in the use of multiple ground truths, often manually defined by multiple human experts, we stick to the case of one ground truth assumed to be the gold standard and quantify the segmentation performance at the level of image objects. This falls under the empirical discrepancy metrics as defined in [6].

We present here our research on the segmentation of cell nuclei in routine H&E stained histopathology imagery. In our use of the term "segmentation," we are referring to an object-level segmentation, i.e., a delineation of individual

G. Bebis et al. (Eds.): ISVC 2007, Part I, LNCS 4841, pp. 208–219, 2007.

nuclei, thus more standard metrics such as Receiver Operating Characteristics (ROC) curve analysis are not directly applicable. The pixel-level classification of nuclei pixels is described in our previous work [10].

In Section 2 we will first describe an object-level metric for segmentation accuracy (Section 2.1) as well as an analysis of the variation of our metric with segmentation quality (Section 2.2), a methodology for specification of object-level ground truth (Section 2.3), and a summary of our work on the segmentation metric (Section 2.4). We briefly discuss the application of our shape-based segmentation metric to non-elliptical objects in Section 3. We then present results for example nuclei segmentations in Section 4, including a standard watershed-based segmentation (Section 4.1), a combined shape-based and watershed-based segmentation (Section 4.2), and summarize our segmentation results (Section 4.3). Conclusions are presented in Section 5.

## 2 Segmentation Metric

The following metric was defined with the segmentation of cell nuclei, i.e., roughly circular or elliptical objects, in mind. For the segmentation of cell nuclei, we wish to penalize not only the size of regions missed and extraneous regions, but also the shape of those same regions. Additionally we include terms to penalize over- and under-segmentation. We introduce the quadrant sum as a method of quantifying deviation in shape from the ground truth by comparing the mass across two orthogonal axes through the object's center of mass. While this section will focus on elliptical objects, we will show the use of the quadrant sum for arbitrarily shaped objects in Section 3.

### 2.1 Definition

We define our segmentation metric as:

$$\begin{aligned} P = & \frac{1}{N_D}\sum_{i=1}^{N_D}\max\Bigg(0, \Bigg[1-\alpha_1\frac{SR-1}{\delta_{SR}}-\alpha_2\frac{1}{1.75}\left(\frac{PM}{GT}+\frac{2QS_{PM}}{GT}\right) \\ & -\alpha_3\frac{1}{1.75}\left(\frac{EP}{GT}+\frac{2QS_{EP}}{GT}\right)\Bigg]\Bigg)\cdot\left(1-\alpha_4\frac{N-N_D}{N}\right)-\alpha_5\frac{ER}{N\cdot\delta_{ER}} \end{aligned} \tag{1}$$

where

$$0 \le \alpha_i \le 1,\ i = 1, \ldots, 5 .$$

Taking each additive term Equation 1, we will define all variables. $N$ is the number of ground truth nuclei defined in the user markup and $N_D$ is the number of nuclei detected by the segmentation algorithm; thus the summation averages scores for individual nuclei. We penalize for each nucleus detected[1]:

[1] For the sake of clarity and brevity we have not included in Equation 1 the necessary clipping functions to assure that each term is less than 1. We will discuss the need for these clipping functions and explicitly display them in the discussions of individual terms to follow.

1. *The number of segmented regions:*

$$\mathrm{term}_1 = \alpha_1 \min\left(1,\ \frac{SR-1}{\delta_{SR}}\right) \tag{2}$$

We define $SR$ as the number of segmented regions overlapping the ground truth nucleus, and $\delta_{SR}$ as the upper limit for number of segmented regions. For a perfect segmentation there would be only one segmented region per ground truth region and $\delta_{SR} = 1$ would be an intuitive value for evaluation of very good segmentations; we leave this as a parameter, however, to allow for comparison of poorer segmentations with more tendency to oversegment. We use the minimum function to clip this term to a maximum value of 1. Overall, the weight $\alpha_1$ can be thought of as the penalty for an oversegmented nucleus, similar to the oversegmentation term of [4].

2. *The size and shape of the region of pixels missed:*

$$\mathrm{term}_2 = \alpha_2 \min\left(1, \frac{1}{1.75} \cdot \left(\frac{PM}{GT} + \min\left(1,\ \frac{2 \cdot QS_{PM}}{GT}\right)\right)\right) \tag{3}$$

We define $PM$ as the number of pixels missed: pixels belonging to the ground truth markup of the nucleus, but missed by the segmentation algorithm. $GT$ is the number of pixels in the ground truth markup, thus, $\frac{PM}{GT}$ quantifies the size of the region of missed pixels. This is similar to the percentage of misclassified pixels used in [5].

We also look at the spatial distribution of the missed pixels, since we wish to penalize certain spatial distributions more than others. For example, a distribution of missed pixels in an annulus about the centroid of the nucleus will affect the shape and other higher-level feature statistics far less than a distribution of missed pixels encompassing half of the nucleus. Note that this is a different approach than a simple pixel distance error as in [5] and tends towards an appreciation of accurate higher-level measurements as in [7]. We take the "quadrant sum" of the pixels missed, $QS_{PM}$ as follows:

$$QS_{PM} = \|r_1 + r_3 - r_2 - r_4\| + \|r_1 + r_2 - r_3 - r_4\| \tag{4}$$

where $r_i$ are the number of pixels in the $i = 1, 2, 3, 4$ quadrants:

$$\begin{aligned} r_1 &= \sum \left\| e^{j\theta_{PM}} \right\|, \quad \text{for } 0 < \theta_{PM} < \frac{\pi}{2} \\ r_2 &= \sum \left\| e^{j\theta_{PM}} \right\|, \quad \text{for } \frac{\pi}{2} < \theta_{PM} < \pi \\ r_3 &= \sum \left\| e^{j\theta_{PM}} \right\|, \quad \text{for } -\frac{\pi}{2} < \theta_{PM} < -\pi \\ r_4 &= \sum \left\| e^{j\theta_{PM}} \right\|, \quad \text{for } 0 < \theta_{PM} < -\frac{\pi}{2} \end{aligned} \tag{5}$$

where $\theta_{PM}$ is the angle of the polar representation of the pixels missed. Thus, $QS_{PM}$ is a measure of symmetry about the x- and y-axes of the region with the origin at the grouth truth centroid. Due to the discrete nature of the regions, it is possible that $QS_{PM}$ may slightly exceed $\frac{GT}{2}$; to compensate for this, we take the minimum of 1 and $\frac{2 \cdot QS_{PM}}{GT}$.

Overall, $\alpha_2$ can be thought of as the penalty regions of pixels missed, penalizing both size and shape. More details of the performance of this $QS$ term is explained in Figure 1 for circular and elliptical regions, including the motivation for our normalization factor of 1.75. While this is a simple and easy to compute metric, there is no reason why another shape metric could not be substituted, with appropriate attention to the inclusion of the size metric.

3. *The size and shape of the region of excess pixels:*

$$\text{term}_3 = \alpha_3 \min\left(1, \frac{1}{1.75} \cdot \left(\min\left(1,\ \frac{EP}{GT}\right) + \min\left(1,\ \frac{2 \cdot QS_{EP}}{GT}\right)\right)\right) \quad (6)$$

Similar to term 2, we define $EP$ as the number of excess pixels: pixels segmented as part of the nuclear region that do not correspond to the ground truth markup. We quantify the size of the region of extra pixels by $\frac{EP}{GT}$. We also take the "quadrant sum" of the excess pixels, $QS_{EP}$ and normalize by $\frac{GT}{2}$. Again, we take the minimum of 1 and $\frac{2 \cdot QS_{EP}}{GT}$ and normalize the sum of the two factors by 1.75. $\alpha_3$ is thus the penalty for size and shape of excess pixel regions, and is related to the degree of undersegmentation of the nucleus.

Averaging these three terms provides a measure of the segmentation performance on all detected nuclei. We also wish to weight this average by the general detection rate. Thus we scale the average of the previous three terms by:

4. *The fraction of nuclei detected:*

$$\text{term}_4 = 1 - \alpha_4 \frac{N - N_D}{N} \quad (7)$$

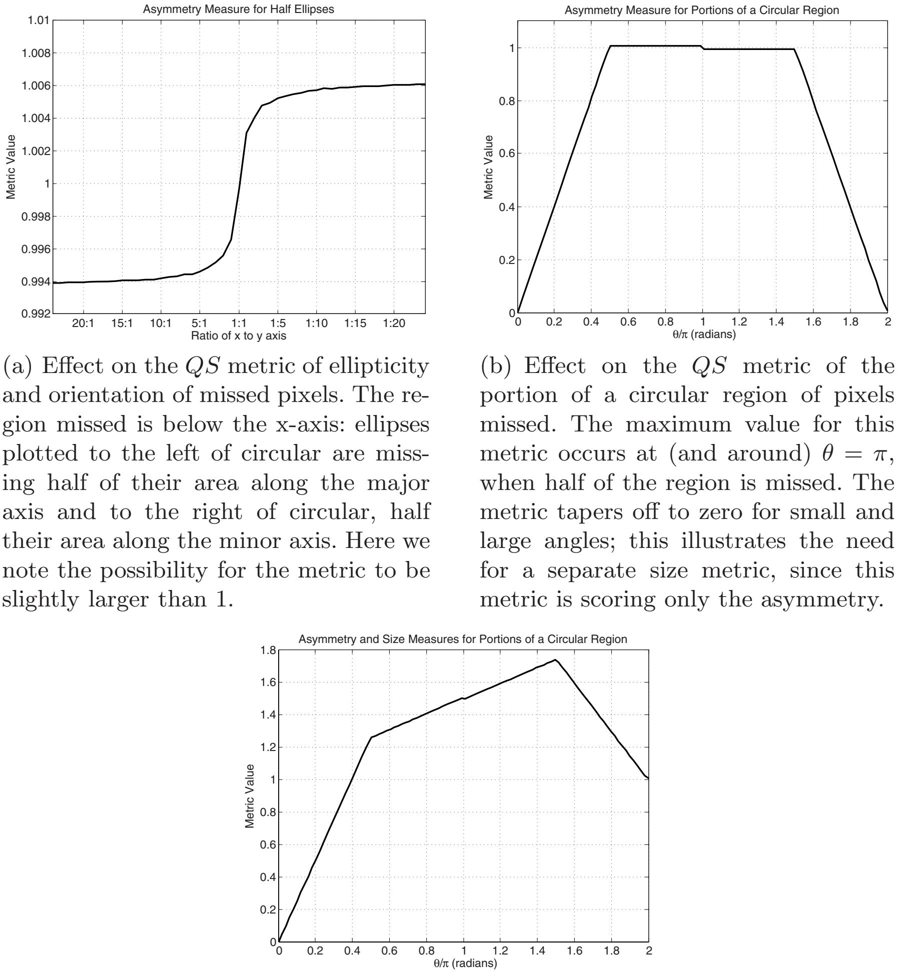
This term with $\alpha_4 = 1$ would simply be the detection rate. We leave this as a parameter since in the segmentation of nuclei, we are interested more in the accuracy of the nuclei that are segmented than in the actual detection rate. This harkens back to the theory of Ultimate Measurement Accuracy [7], wherein it is the accuracy of further image analyses that determine the accuracy of the underlying segmentation.

Finally we wish to penalize over the whole region of ground truth:

5. *The number of extra segmented regions:*

$$\text{term}_5 = \alpha_5 \min\left(1,\ \frac{ER}{N \cdot \delta_{ER}}\right) \quad (8)$$

This term looks at the excess segmented regions that have no correspondence to a ground truth nucleus. We define $ER$ as the number of excess segmented regions and $\delta_{ER}$ as the fraction of total ground truth nuclei that we will allow as excess regions. $\alpha_5$ is, therefore, the penalty for excess segmented regions, similar to the concept of noise in [4].



(a) Effect on the $QS$ metric of ellipticity and orientation of missed pixels. The region missed is below the x-axis: ellipses plotted to the left of circular are missing half of their area along the major axis and to the right of circular, half their area along the minor axis. Here we note the possibility for the metric to be slightly larger than 1.

(b) Effect on the $QS$ metric of the portion of a circular region of pixels missed. The maximum value for this metric occurs at (and around) $\theta = \pi$, when half of the region is missed. The metric tapers off to zero for small and large angles; this illustrates the need for a separate size metric, since this metric is scoring only the asymmetry.

(c) Additive effect of the $QS$ and size metrics. The combination of these two metrics yields the desired penalty. Note the maximum value of $\sim 1.75$.

**Fig. 1.** Values of the QS metric for pixels missed in discrete elliptical and circular regions. The QS metric in these plots has been normalized by $\frac{GT}{2}$, and the size metric by $GT$.

Overall, the choice of $\alpha_i$ reflects a weighting of the relative importance of the various penalties. Similarly, the choice of $\delta_{SR}$ and $\delta_{ER}$ reflects a choice in the latitude given to certain errors in segmentation. A reasonable choice for default parameters would be $\alpha = [0.5\ 0.5\ 0.5\ 1\ 0.5]$, $\delta_{SR} = 1$, and $\delta_{ER} = 1$, reflecting an equal penalty for under- and over-segmentation errors ($\alpha_1$, $\alpha_2$, and $\alpha_3$), a direct weighting by the detection rate ($\alpha_4$), equal importance given to the correct detection and segmentation of cell nuclei and the avoidance of erroneously detected and segmented nuclei ($\alpha_5$), one segmented region allowed per nucleus ($\delta_{SR}$), and weighting of the erroneously segmented regions proportional to the total number of cell nuclei ($\delta_{ER}$). It is important to note, however, that while the choice of these parameters will effect the absolute values of the metric terms, a direct comparison of segmentation performance for different algorithms may be achieved with any reasonable parameter choice.

### 2.2 Metric Variation Versus Segmentation Quality

We apply our segmentation metric (Equation 1) to the watershed transform of the complemented Euclidean distance transform (WSCDT) of a thresholded red channel for an example image. The threshold is varied over the entire range of values it can assume, [0,255], and we retain all pixels less than the threshold. The use of the red channel is due to the high contrast for nuclei in this channel.

We compute the WSCDT as follows:

1. Compute the negative of the Euclidean distance transform on the complemented binary image, setting the distance of all background pixels in the binary image to a depth of $-\infty$.
2. Compute the watershed transform on the resulting distance transform.

By varying the threshold we compute a variety of binary images. We compute the segmentation metric (Equation 1) of the WSCDT segmentation of these binary images to gain a sense of the expected variation in our metric for a range of segmentation possibilities. These possibilities include the two extremes whereby either all or none of the pixels has been classified as nuclei. We display the performance of the individual metric terms as well as the overall performance in Figure 2. It is important to note that we are plotting the *performance* of the individual terms rather than the terms themselves; thus we are plotting the subtraction of each term from a value of 1.

We see in Figure 2 that the performance is zero for both extremes of the threshold classification. Observation of individual terms shows expected trends, namely that:

- Term 1 (extra GT regions) decreases in performance as the threshold increases. This is due to the thresholded nuclei regions becoming larger with more complicated boundaries which results in the distance transform having multiple minima per connected component.
- Term 2 (pixels missed) increases in performance as more pixels are attributed to nuclei. The dip in performance at high thresholds is due to an assumption that the largest watershed region is the background which becomes invalid as nearly the entire image is classified as foreground.
- Term 3 (extra pixels) decreases in performance as nuclei tend to merge in the binary thresholded image.
- Term 4 (nuclei detected) increases in performance as more pixels are attributed to nuclei.

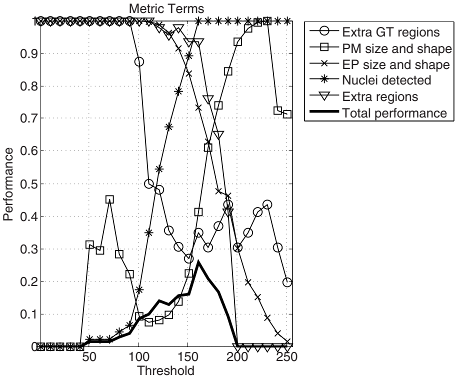


**Fig. 2.** Metric variation versus segmentation quality for an example image. The red channel was thresholded, retaining all pixels less than the threshold, and was then segmented with the WSCDT method. It should be noted that all terms plotted here are performance, i.e., one minus the penalty, where the penalties are the terms previously discussed in relation to the segmentation metric. The terms are denoted by a brief description in the legend, but they are also plotted in numerical order, i.e., blue circles are term 1, red squares are term 2, and so forth.

– Term 5 (extra regions) decreases in performance as more extraneous regions are thresholded as nuclei. The performance of this term returns to 1 for a threshold of 256, since there are no longer any extraneous regions; this is not apparent in Figure 2 since we have downsampled the plot for less clutter.

We note here that analysis of the individual metric terms is useful for quantifying segmentation, but we have integrated the terms into one metric to allow for single parameter to compare and/or optimize between different segmentations.

### 2.3 Ground Truth Image Markup Within a Truth Window

While it is easy to specify a pixel-level markup within a designated truth window, such a specification becomes more complicated with an object-level markup. In a pixel-level markup, an object that spans the truth window boundary can be marked up to the boundary without losing any important information for the overall classification. In an object-level markup, however, the actual extent and border of the object is of utmost importance. Moreover, if objects are marked within a rough concept of a truth window, the truth window may contain parts of objects that have not been delineated by the user. This will lead to erroneously low performance since the segmentation metric will assume that these regions were incorrectly segmented as image objects.

To help alleviate this problem, after the delineation of objects within the truth window is complete, the truth window is recomputed as the minimum bounding rectangle of the object markups. Using this new truth window, the user is asked to mark a minimum of one point for each unmarked object that is either completely or partially enclosed by the new truth window. This information is used in a connected-components analysis to determine if extra segmented regions are associated with an object that has not been delineated in the ground truth markup.

### 2.4 Summary

We have presented a general segmentation metric computed on an object level. This metric uses simple quantities that are easy to compute using the segmentation and ground truth masks, namely the regions of pixels missed by the segmentation and the regions of extra pixels not associated with a ground truth region. We have also shown the variation in this metric for a variety of segmentations using a simple watershed-based segmentation (WSCDT) and discussed the ground truth markup process for evaluation of the metric.

## 3 Application of the QS Shape Metric to Non-elliptical Objects

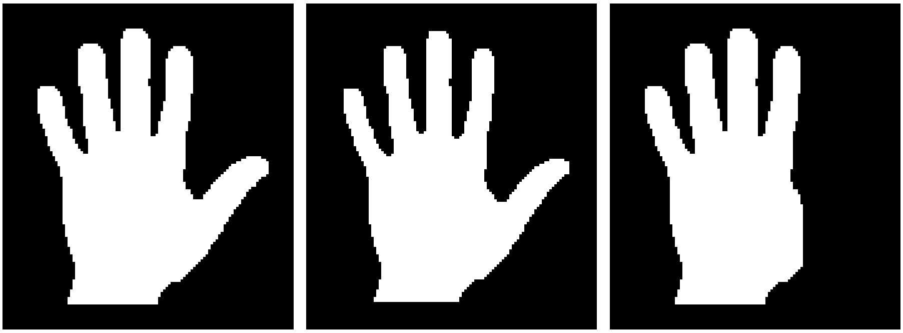
We would like to briefly discuss the applicability of the QS metric to non-elliptically shaped objects; we will be using the concepts of the PM QS metric, but the arguments are identical for the EP case. The use of the centroid of the ground truth object is what allows this metric to work for irregularly shaped objects. For a planar object with uniform density, the mass (number of pixels in our case) will be equal across any arbitrary line through the center of mass (equivalent to the centroid in the uniform density case). By defining orthogonal axes through the centroid, we can eliminate the chance of the arbitrary line corresponding to a reflectional symmetry of the region of pixels missed. We show an example of the application of the PM QS metric in Figure 3 for a hand silhouette. Further research into the use of our segmentation metric for arbitrarily-shaped objects is currently ongoing. In particular, the practical application of this metric may warrant a different normalization value than the theortical maximum of $\sim 1.75$ for the combined size and shape metrics for EP and PM.

## 4 Watershed-Based Segmentation

We investigate here two simple watershed-based segmentation methods for delineation of cell nuclei. We assign the default weights (discussed in Section 2.1) of $\alpha = [0.5\ 0.5\ 0.5\ 1\ 0.5]$, $\delta_{SR} = 1$, and $\delta_{ER} = 1$.

### 4.1 Watershed on the Complemented Distance Transform

We use the WSCDT method, as described in Section 2.2, on the pixel-level nuclei classifications from [10]. We present quantitative results in Table 1 and an example segmentation in Figure 4. The results in Table 1 represent the results averaged over the entire dataset of 58 images. The example semgnetaions in Figure 4 include the performance for the single example image.



(a) Original hand silhouette with $GT = 5270$ object pixels.

(b) Erosion by 1 pixel; total of $PM = 524$ pixels eroded (missed). $\frac{2 \cdot QS}{GT} = 0.188$, $\frac{PM}{GT} = 0.099$, $term_2 = 0.107$.

(c) Thumb removed; total of $PM = 524$ pixels missed. $\frac{2 \cdot QS}{GT} = 0.397$, $\frac{PM}{GT} = 0.099$, $term_2 = 0.227$.

**Fig. 3.** Application of the QS and size metrics to an example silhouette and "segmentations." Qualitatively, the segmentation in (b) retains more resemblance to the original silhouette in (a) than does the segmentation in (c), where the entire thumb is missed. A size metric alone would rank the two results in (b) and (c) as equally good segmentations, while the use of the QS metric penalizes the change in shape of (c). Note that in (b) the addition of the shape metric does not change the value of the original size-based metric by much (0.8%).

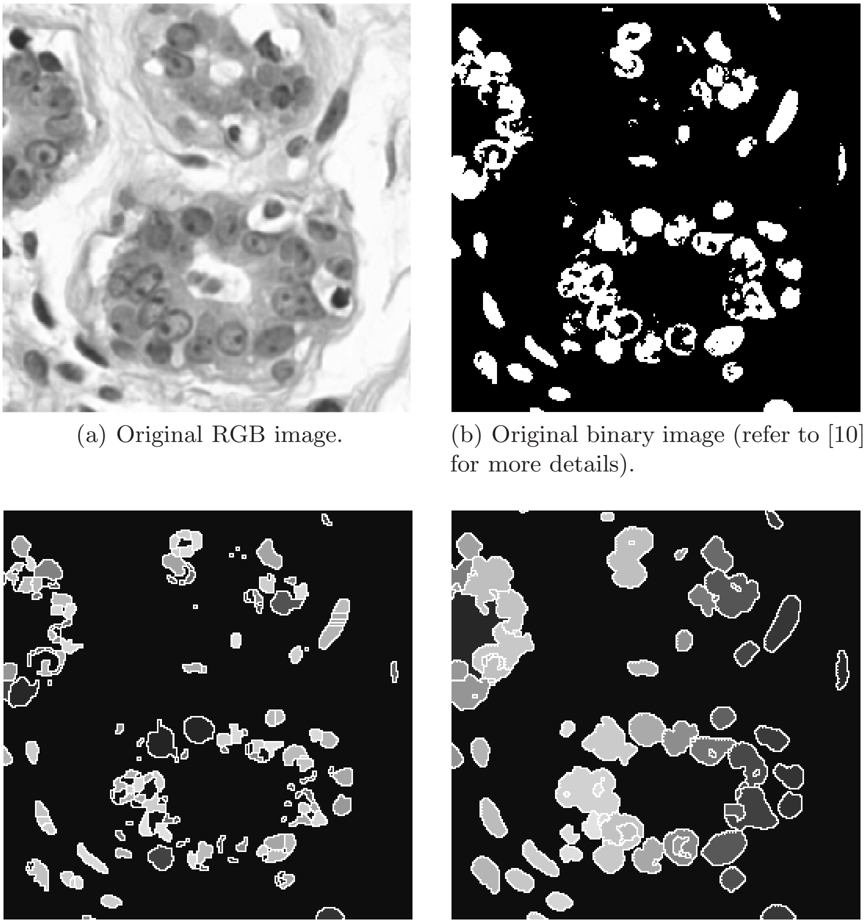
### 4.2 Marker-Based Watershed Segmentation

We use a prior assumption about the shape of cell nuclei, namely that they are roughly circular in shape and approximately the same diameter. Byun et al. [11] use an inverted Laplacian of Gaussian (LoG) filter for detection of nuclei in fluorescent confocal retinal imagery. For use in our brightfield imagery, we use a non-inverted LoG filter in the same "blobdetector" framework of [11].[2] We use the detection capabilities of this method as a seed for a subsequent watershed segmentation. This method (WSBlob) proceeds as follows:

– Detect nuclei using the red channel of the imagery and use these locations as foreground markers for the watershed transform. A filter size of 25 pixels in diameter (average nucleus diameter) and an inter-blob distance of 12 (half the filter diameter) was empirically chosen.
– Use the eroded complement of the binary nuclei classification as background markers.

Quantitative results for the WSBlob method are presented in Table 1 and an example segmentation in Figure 4. As for the WSCDT, the results in Table 1 are averaged over the entire dataset and example segmentations in Figure 4 include the performance for the example image.

[2] Code available at `http://www.bioimage.ucsb.edu/software.html`



(a) Original RGB image.

(b) Original binary image (refer to [10] for more details).

(c) WSCDT, $P = 0.2974$.

(d) WSBlob, $P = 0.3843$.

**Fig. 4.** Example watershed-based segmentations. Note the tendency of the WSCDT method to oversegment and the tendency of the WSBlob method to undersegment.

### 4.3 Summary

Referring to Table 1, we see terms 1 and 3 display the most difference between the two methods. In particular, WSCDT displays a worse performance for term 1 (extra GT regions) and better performance for term 3 (extra pixels), indicating a tendency for WSCDT to oversegment as compared to WSBlob which tends to undersegment. These observations are validated by observation of the example segmentations in Figure 4.

We have presented the application of our segmentation metric to the problem of segmentation of cell nuclei. We have shown the overall metric performance and the performance of individual terms for each segmentation method. We have also

**Table 1.** Nuclei segmentation term performance

| **Method** | **P** | **term$_1$** | **term$_2$** | **term$_3$** | **term$_4$** | **term$_5$** |
|---|---|---|---|---|---|---|
| **WSCDT** | 0.18 | 0.38 | 0.82 | 0.60 | 0.96 | 0.44 |
| **WSBlob** | 0.25 | 0.69 | 0.94 | 0.31 | 0.98 | 0.52 |

used the metric terms as a quantitative basis to compare the performance of the two methods, which corresponds well with the qualitative observations of general segmentation accuracy as illustrated by the example segmentations.

Further work should include observer studies to correlate the metric to the visual assessment of many individuals. Additionally, future and ongoing research includes the comparison of our metric to other applicable metrics, e.g., those presented in [2] and [9].

## 5 Conclusions

We have presented an object-level segmentation metric and its constituent terms and have shown that they correspond well with the qualitative observations of segmentation accuracy, including the general tendency of an algorithm to over- or under-segment an image. This metric also allows for a direct quantitative comparison between the outputs of different segmentation algorithms. While the metric defines a single performance, we have shown the usefulness of observing the performance of the individual metric terms.

We have also discussed a new method for specification of ground truth for this object-level segmentation problem. This involves not only the delineation of cell nuclei within an approximate truth window, but also the marking of non-delineated objects within the truth window. This allows us to focus our segmentation evaluation on only those objects that were delineated by the user.

In comparison to other work in segmentation evaluation, our metric does not require the computation of region or boundary correspondences which can be complicated. Instead we have introduced a metric based on simple subtrations of object masks and other object-level metrics (e.g., number of segmented regions). Additionally, we compute the segmentation performance on an object-by-object basis.

## Acknowledgments

The authors gratefully acknowledge Karen Glocer and Ed Rosten for the fruitful technical discussions on and suggestions for this work. Many thanks also to Dr. David Rimm in the Department of Pathology at Yale University for the histopathology data.

## References

1. Nattkemper, T.W.: Automatic segmentation of digital micrographs: A survey. In: Proc MEDINFO (2004)
2. Chalana, V., Kim, Y.: A methodology for evaluation of boundary detection algorithms on medical images. IEEE T. Med. Imaging 16, 642–652 (1997)
3. Udupa, J.K., LeBlanc, V.R., Zhuge, Y., Imielinska, C., Schmidt, H., Currie, L.M., Hirsch, B.E., Woodburn, J.: A framework for evaluating image segmentation algorithms. Comput. Med. Imag. Grap. 30, 75–87 (2006)
4. Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldgof, D.B., Bowyer, K., Eggart, D.W., Fitzgibbon, A., Fisher, R.B.: An experimental comparison of range image segmentation algorithms. IEEE T. Pattern Anal. 18, 673–689 (1996)
5. Yasnoff, W.A., Mui, J.K., Bacus, J.W.: Error measures for scene segmentation. Pattern Recogn. 9, 217–231 (1977)
6. Zhang, Y.J.: A survey on evaluation methods for image segmentation. Pattern Recogn. 29, 1335–1346 (1996)
7. Zhang, Y.J., Gerbrands, J.J.: Segmentation evaluation using ultimate measurement accuracy. In: Proc. SPIE, vol. 1657, pp. 449–460 (1992)
8. Lezoray, O., Cardot, H.: Cooperation of color pixel classification schemes and color watershed: A study for microscopic images. IEEE T. Image Process 11, 783–789 (2002)
9. Cohen, L., Vinet, P., Sander, P.T., Gagalowicz, A.: Hierarchical region based stereo matching. In: Proc. CVPR, pp. 416–421 (1989)
10. Boucheron, L.E., Bi, Z., Harvey, N.R., Manjunath, B.S., Rimm, D.L.: Utility of multispectral imaging for nuclear classification of clinical histopathology imagery. BMC Cell Biology 8(Suppl 1):S8 (2007), `http://www.biomedcentral.com/1471-2121/8/S1/S8`
11. Byun, J., Verardo, M.R., Sumengen, B., Lewis, G.P., Manjunath, B.S., Fisher, S.K.: Automated tool for nuclei detection in digital microscopic images: Application to retinal images. Mol. Vis. 12, 949–960 (2006)