Evaluation of Level Set-based Histology Image Segmentation Using Geometric Region Criteria

Adel Hafiane, Filiz Bunyak, and Kannappan Palaniappan

Department of Computer Science,
University of Missouri-Columbia, Columbia, MO 65211 USA

***Abstract*— There is a great deal of interest in developing automated histological grading of tissue biopsies. Current approaches involve sophisticated algorithms for image segmentation, tissue architecture characterization, global texture feature extraction, and high-dimensional clustering and classification algorithms. Although overall image classification accuracy is measured, there has been very little attention paid to the quantitative assessment of the image segmentation stage (glandular structure characterization stage) to provide feedback to the segmentation process. We describe a robust approach for tissue segmentation combining spatial clustering with multiphase vector level set active contours to extract nuclei, lumen and epithelial cytoplasm. Quantitative segmentation performance compared to manual ground truth is assessed using region-based geometric criteria.**

I. Introduction

The availability of high resolution multispectral multimodal imaging of tissue biopsies provides a new opportunity to develop improved tissue segmentation algorithms for developing computer-aided diagnostic classification of histological images in a clinical setting. Automated quantitative grading of prostate cancer tissue patches that is beginning to compare favorably with visual analysis by experts for assigning a Gleason grade to histogological imagery was demonstrated using a combination of low level image texture features and high level graph-based tissue architecture features based on manual [1] or semi-automated [2] image segmentation. A multiresolution approach using global texture features including first- and second-order statistics combined with a Gabor filter set was able to achieve over 90% overall accuracy in distinguishing between cancerous and benign tissue, and nearly 77% in distinguishing between two complex grades of cancer (Gleason grade 3 and 4 adenocarcinoma). In this paper, we focus on quantitatively evaluating our fully automatic robust image segmentation algorithm for histopathology imagery using fuzzy spatial clustering for class initialization, and tissue class refinement using vector-based level sets or geodesic to accurately extract, nuclei, lumen and epithelial cytoplasm regions [3]. Our main contribution is to introduce several new measures for quantitative evaluation of histopathology image segmentation results.

II. Multiphase Vector-based Active Contours

We briefly describe an extended version of the multi-class histology image segmentation algorithm introduced in [3]. The well known region-based Chan and Vese active contour algorithm for the two class case [4] was extended to a multiphase image segmentation framework in [5]. The multiphase approach enable efficient partitioning of the image into n classes using just $log(n)$ level sets without leaving any gaps or having overlaps between level sets. The Chan and Vese multiphase level set image segmentation approach involves minimization of a reduced or weak Mumford-Shah functional $F_n(\mathbf{c}, \Phi)$, otherwise referred to as a *minimal partition* Mumford-Shah functional [6]: $F_n(\mathbf{c}, \Phi) = \underbrace{\sum_{1 \le i \le n = 2^m} \lambda_i \int_\Omega (u_0 - c_i)^2 \chi_i \, d\mathbf{x}}_{\text{Energy Term}} + \underbrace{\sum_{1 \le i \le n = 2^m} \mu_i \int_\Omega |\nabla \chi_i|}_{\text{Length Term}}$ where, n is the total number of classes associated with m level set functions, u_0 is the gray-level image being segmented, Φ is a vector of level set functions, $\mathbf{c}$ is a vector of mean gray-level values (i.e., $c_i = mean(u_0)$ of class or phase i), χ_i is the characteristic function for each class i represented by the associated Heaviside functions $H(\phi_i)$, and $(\lambda_i,\ \mu_i)$ are constants associated with each energy and length term of the functional $F_n(\mathbf{c}, \Phi)$. In order to simplify computation of the length term in the above reduced Mumford-Shah energy function, we replace the measure of the characteristic functions by the sum of the length of the zero-level sets of ϕ_i, $\sum_{1 \le i \le m} \mu_i \int_\Omega |\nabla H(\phi_i)|$. Instead of an unweighted total length, this approximation weights some edges more than others, but is faster to compute and still leads to satisfactory segmentation results. Chan and Vese also extended their two-phase level set image segmentation algorithm for scalar valued images to vector-valued images such as color or multispectral images [7] where the image is partitioned into piecewise constant vectors in the spirit of spatially-based vector quantization. In this paper we combine the multiphase approach with the feature vector approach to handle both multiple image classes and vector-valued imagery such as segmenting color or multispectral images.

In the case of histopathology imaging derived from H&E stained cancer tissue biopsies four image classes have been shown to produce good feature sets for image classification-based cancer grading[1]. For the two level set case (i.e., $m = 2$) the image domain Ω is partitioned into *at most* four classes. Let $\mathbf{c} = \{\mathbf{c_{00}}, \mathbf{c_{01}}, \mathbf{c_{10}}, \mathbf{c_{11}}\}$ represent the set of average image feature vectors (ie three channel color) within each class or phase $\mathbf{c_{ij}}$, and $\Phi = (\phi_1, \phi_2)$ represent the

two level set functions. The multiphase vector-based energy functional $F_n(\mathbf{c}, \Phi)$ is defined as,

$$\begin{aligned} F_n(\mathbf{c}, \Phi) =& \lambda_1 \int_\Omega \parallel \mathbf{u_0} - \mathbf{c_{00}} \parallel^2 \ (1 - H(\phi_1))(1 - H(\phi_2)) d\mathbf{x} \\ &+ \lambda_2 \int_\Omega \parallel \mathbf{u_0} - \mathbf{c_{01}} \parallel^2 \ (1 - H(\phi_1))H(\phi_2) \ d\mathbf{x} \\ &+ \lambda_3 \int_\Omega \parallel \mathbf{u_0} - \mathbf{c_{10}} \parallel^2 \ H(\phi_1)(1 - H(\phi_2)) \ d\mathbf{x} \\ &+ \lambda_4 \int_\Omega \parallel \mathbf{u_0} - \mathbf{c_{11}} \parallel^2 \ H(\phi_1)H(\phi_2) \ d\mathbf{x} \\ &+ \mu_1 \int_\Omega |\nabla H(\phi_1)| \ d\mathbf{x} + \mu_2 \int_\Omega |\nabla H(\phi_2)| \ d\mathbf{x} \end{aligned} \quad (1)$$

The Euler-Lagrange equations are obtained by minimizing Eq. 1 and embedding $\mathbf{c}$ and Φ in a dynamical system as [5]

$$\begin{aligned} \frac{d\phi_1}{dt} &= \delta(\phi_1)\Big\{\mu_1 \ \mathrm{div}\big(\frac{\nabla\phi_1}{|\nabla\phi_1|}\big) \\ &- \big\{(\lambda_1 \parallel \mathbf{u_0} - \mathbf{c_{11}} \parallel^2 - \lambda_3 \parallel \mathbf{u_0} - \mathbf{c_{01}} \parallel^2)H(\phi_2) \\ &+ (\lambda_2 \parallel \mathbf{u_0} - \mathbf{c_{10}} \parallel^2 - \lambda_4 \parallel \mathbf{u_0} - \mathbf{c_{00}} \parallel^2)(1 - H(\phi_2))\big\}\Big\}, \\ \frac{d\phi_2}{dt} &= \delta(\phi_2)\Big\{\mu_2 \ \mathrm{div}\big(\frac{\nabla\phi_2}{|\nabla\phi_2|}\big) \\ &- \big\{(\lambda_1 \parallel \mathbf{u_0} - \mathbf{c_{11}} \parallel^2 - \lambda_2 \parallel \mathbf{u_0} - \mathbf{c_{10}} \parallel^2)H(\phi_1) \\ &+ (\lambda_3 \parallel \mathbf{u_0} - \mathbf{c_{01}} \parallel^2 - \lambda_4 \parallel \mathbf{u_0} - \mathbf{c_{00}} \parallel^2)(1 - H(\phi_1))\big\}\Big\} \end{aligned} \quad (2)$$

where, $\mathbf{c_{ij}}$ is the mean vector of all pixel-based vectors associated with each class or phase, and $\delta(\phi_k) = H'(\phi_k)$ is the Dirac delta function. For numerical stability of the delta function, Chan and Vese propose using a regularized Heaviside function. The motivation for using a multiphase, rather than a two-phase, level set framework is to accurately detect adjacent regions that meet at a junction (i.e., the triple junction in [5]). However, as the number of regions grows exponentially with the number of level set functions, its best to use a small set of level set functions (typically two or three or equivalently, four or eight regions, respectively).

III. Geodesic Level-sets Segmentation

In classical level set-based geodesic active contours [8], the level set function ϕ is evolved using the speed function,

$$\frac{\partial\phi}{\partial t} = g(\mathbf{u_0})(F_c + \mathcal{K}(\phi))|\nabla\phi| + \nabla\phi \cdot \nabla g(\mathbf{u_0}) \quad (3)$$

where F_c is a constant, $\mathcal{K}$ is the curvature term,

$$\mathcal{K} = div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) = \frac{\phi_{xx}\phi_y^2 - 2\phi_x\phi_y\phi_{xy} + \phi_{yy}\phi_x^2}{(\phi_x^2 + \phi_y^2)^{\frac{3}{2}}} \quad (4)$$

and $g(\mathbf{u_0})$ is the edge stopping function. Edge stopping can be any decreasing function of the image gradient. For histopathology image segmentation, we use Beltrami color edge stopping function defined as

$$g(\mathbf{u_0}) = exp(-abs(\mathbf{det}(\mathcal{E}))) \quad (5)$$

where $\mathcal{E}$ is

$$\mathcal{E} = \begin{bmatrix} 1 + \sum\limits_{i=R,G,B} \left(\frac{\partial \mathbf{u}_{0,i}}{\partial x}\right)^2 & \sum\limits_{i=R,G,B} \frac{\partial \mathbf{u}_{0,i}}{\partial x}\frac{\partial \mathbf{u}_{0,i}}{\partial y} \\ \sum\limits_{i=R,G,B} \frac{\partial \mathbf{u}_{0,i}}{\partial x}\frac{\partial \mathbf{u}_{0,i}}{\partial y} & 1 + \sum\limits_{i=R,G,B} \left(\frac{\partial \mathbf{u}_{0,i}}{\partial y}\right)^2 \end{bmatrix} \quad (6)$$

The constant velocity F_c pushes the curve inwards or outwards depending on its sign. The regularization term $\mathcal{K}$ ensures boundary smoothness. The external image dependent force $g(\mathbf{u_0})$ is used to stop the curve evolution at object boundaries. The term $\nabla g \cdot \nabla\phi$ is used to increase the basin of attraction for evolving the curve to the boundaries of the objects. The classical geodesic active contours is two phase and can segment an image into only two classes. In order to segment the three-class histopathology images we use two level sets, with one level set segmenting lumen regions, and the second level set segmenting nuclei regions.

IV. Evaluation of segmentation results

A typical segmentation evaluation criterion is based on the overlap between regions in the segmented image and segmented regions in the ground truth or reference image. Higher percentage overlap would indicate higher similarity relative to the ground truth image. However, overlap percentage alone is not a sufficient segmentation evaluation measure since under-segmentation, for example, can lead to large overlap percentages. A more versatile evaluation measure of segmentation results, as introduced in [9], takes into account both the geometric aspects of the segmented regions (localization) as well as the number of regions (over- and under-segmentation), where we use the following definitions:

- *Localization/overlap: the detected regions should be spatially coherent (eg. position, shape, size...) with those present in the reference image,*
- *Over-segmentation/fragmentation: where a region of the ground truth image is overlapped with two or more regions of the segmented image should be penalized,*
- *Under-segmentation/merges: Two or more regions of the ground truth image are overlapped with a single region of the segmented image should be penalized.*

Let C_i be the set of pixels in the image belonging to the class i, C_i^{Ref} and C_j^{Seg} are two classes from the reference image, I^{Ref}, and the segmentation result, I^{Seg}, respectively. The *matching index* M_I is defined as a weighted overlap ratio over all classes:

$$M_I = \sum_{j=1}^{NC_{Ref}} \frac{Card(C_{i^*}^{Seg} \cap C_j^{Ref})}{Card(C_{i^*}^{Seg} \cup C_j^{Ref})} \rho_j \quad (7)$$

where $i^* = \underset{i}{\mathrm{argmax}}(Card(C_i^{Seg} \cap C_j^{Ref}))$ is the class index of the class with the largest overlap (pixel-by-pixel basis) compared to each class in the reference segmentation, C_j^{Ref}, $Card(X)$ is the number of pixels in set X, NC_{Ref} and NC_{Seg} are the number of classes in the reference and segmented images respectively, and ρ_j is a size-based weighting

factor that controls the influence of each class so that smaller classes have less influence than the larger ones,

$$\rho_j = \frac{Card(C_j^{Ref})}{Card(I^{Ref})} \tag{8}$$

However, Eq. 7 does not take into account the region spatial homogeneity (ie. fragmentation or merges) of the class components/regions. Using the number of connected components in the reference and segmented images we define the following weight function, η, to penalize over- and under-segmentation errors,

$$\eta = \begin{cases} NR_{Ref}/NR_{Seg} & if \quad NR_{Seg} \geq NR_{Ref} \\ log(1 + NR_{Seg}/NR_{Ref}) & otherwise \end{cases} \tag{9}$$

where NR_{Ref} and NR_{Seg} are the number connected components in the reference and segmented images respectively. The weighting function decreases as $1/x$ for over-segmentation and is logarithmic in the number segmented regions for under-segmentation. The final evaluation criterion $\mathcal{H}$ is then given by the following equation:

$$\mathcal{H} = \frac{M_I + m \times \eta}{1 + m} \tag{10}$$

where m is a weighting coefficient which controls the importance of the over- or under-segmentation errors in the judgment. As shown heuristically in [9], $m = 0.2$, provides an adequate tradeoff between between the two parameters contributing to the evaluation measure.

V. Results and discussion

We compare three clustering-based image segmentation methods: K-means, Fuzzy C-means, and Spatial Constraint Fuzzy C-Means (SCFCM) and three types of level set methods: geodesic with SCFCM initialization, multiphase vector with random initialization and multiphase vector with SCFCM initialization (MVLS-SCFCM). Fig. 1 shows segmentation results of the MVLS-SCFCM method on a Gleason grade 3 tissue sample (Fig. 1(a)) compared to manual segmentation (Fig. 1(b)) with three classes.

For histopathology image segmentation, vector multiphase level sets are more reliable and robust compared to geodesic active contours. Geodesic active contours are more sensitive to initialization (should start either completely inside or outside of the regions of interest). They suffer from contour leaking on weak edges (ie. some nuclei edges), and early stopping on background edges (ie. cytoplasm texture). Due to these problems some significant nuclei regions are missed. The multiphase vector-based level sets (MVLS) are used with two different initializations: random circles and SCFCM segmented regions with the latter performing better especially in the lumen regions as quantified below.

Table I compares performance of different segmentation methods for each class in terms of area percentage Eq. 11, overlap percentage Eq. 12, and Jacard coefficient (normalized overlap percentage) Eq. 13. These additional measures comes to enforce the evaluation criteria described in Section IV. In table I Class 1, Class 2, and Class 3 represent nuclei, lumen and cytoplasm classes respectively.

$$\texttt{Area} = \frac{card(R_i)}{\texttt{Total number of points}} \times 100 \tag{11}$$

$$\texttt{Overlap} = \frac{Card(R_i^{Ref} \cap R_i^{Seg})}{\texttt{Total number of points}} \times 100 \tag{12}$$

$$\texttt{Jacard coefficient} = \frac{Card(R_i^{Ref} \cap R_i^{Seg})}{Card(R_i^{Ref} \cup R_i^{Seg})} \times 100 \tag{13}$$

Distribution of the different classes in the reference image is: nuclei 29%, lumen 11%, and cytoplasm 60%. For the nuclei class all tested algorithms provide areas similar to the area in the reference image. The major difference reside in lumen and cytoplasm classes due to the additional small lumen regions detected. Geodesic-SCFCM produces the best results for the area measure, followed by MVLS-SCFCM. The overlap measures between ground truth class and the corresponding one in the automatic segmentation are given in third major column of the table I. There are high overlap percentages for nuclei, but much less for lumen and cytoplasm. Level set methods provides better spatial precision particularly for lumen class. They tend to shrink lumen and nuclei regions, this reduces the overlap area with the reference image. Although the low percentage of overlapping in lumen area, level set methods provides better results in terms of accuracy and localization. This is shown in the next fourth major column of Table I. For the nuclei class all the methods presents good performances 64% in average. The lumen class produces low percentage 27% for K-means and FCM, 31% for SCFCM, Geodesic-SCFCM with 48%, MVLS-Random 40% and MVLS-SCFCM provides 43%. For cytoplasm class the level set algorithms give better results an average of 64%.

As more reliable measures, we compute the criteria $\mathcal{H}$ Eq. 10 and M_I Eq 7 (see section IV). These criteria are not symmetric, so the measure is applied in two different ways, automatic segmented image (Seg) versus groud truth (GT) and vice versa. Table II shows the segmentation quality compared to the reference image over all classes. Geodesic-SCFCM gives higher values for M_I measure producing 78% closer to the reference image. But when the over-under segmentation penalty is introduced in $\mathcal{H}$ criterion, MVLS-SCFCM outperforms all the tested methods.

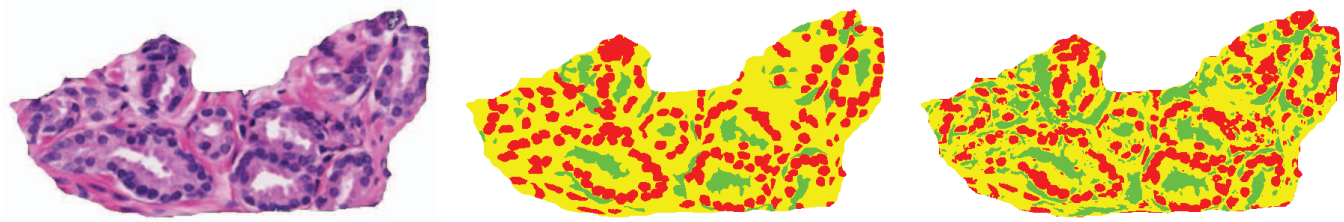
VI. Conclusion

In this paper we described a robust algorithm for fully automatic tissue segmentation of glandular structures in histopathology imagery. An accurate unsupervised initialization is provided using the spatial constraint fuzzy c-means developed previously by our group. The initial image clusters which may not be spatially contiguous with biological regions of interest are refined using an extended active contour

Method	Area %			Overlap %			Jacard Coefficient %		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
Ground truth	28.6	11.1	60.3	-	-	-	-	-	-
K-means	28.2	37.5	34.0	22.2	10.3	28.2	65.0	27.4	42.3
FCM	28.1	37.8	34.1	22.4	10.4	28.1	65.3	27.0	42.5
SCFCM	26.8	32.1	41.2	21.6	10.3	34.1	64.1	31.2	50.7
Geodesic-SCFCM	27.3	16.7	56.0	21.2	9.1	47.2	61.2	48.8	68.3
MVLS-random	28.9	23.6	47.6	22.3	9.9	40.5	63.3	40.2	60.1
MVLS-SCFCM	27.7	20.4	51.9	22.2	9.6	44.3	65.3	43.9	65.3

TABLE I

COMPARISON OF SEGMENTATION STATISTIC MEASURES OF THE NUCLEI (CLASS 1), THE LUMEN (CLASS 2) AND THE CYTOPLASM (CLASSE 3)



(a) Grade 3 (b) Gound truth (c) MVLS SCFCM

Fig. 1. Automatic segmentation of Gleason grade 3 histopathology image with nuclei shown in red, lumen in green, epithelial cytoplasm in yellow.

algorithms to handle complex biological structures in color imagery. We evaluate the segmentation accuracy according to the manual segmentation (ground truth). The proposed method outperforms the classical methods for all classes of regions, nuclei, lumen and cytoplasm. The quality of segmentation is important since it will be used for tissue classification in further future work.

Algorithm	M_I %	$\mathcal{H}$ %	M_I %	$\mathcal{H}$ %
	Seg $\rightarrow$ GT	Seg $\rightarrow$ GT	GT $\rightarrow$ Seg	GT $\rightarrow$ Seg
K-means	69.1	60.0	68.5	60.6
FCM	69.0	60.1	68.8	60.06
SCFCM	72.0	64.6	69.2	62.9
Geodesic-SCFCM	**78.9**	68.2	**78.35**	67.9
MVLS-random	75.8	67.8	71.4	64.9
MVLS-SCFCM	78.2	**70.8**	73.6	**68.1**

TABLE II

SEGMENTATION EVALUATION USING M_i AND $\mathcal{H}$ CRITERIA

REFERENCES

[1] S. Doyle, M. Hwang, K. Shah, A. Madabhushi, M. Feldman, and J. Tomaszeweski, "Automated grading of prostate cancer using architectural and textural image features," in *IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, April 2007, pp. 1284–1287.

[2] S. Naik, S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "Gland segmentation and computerized Gleason grading of prostate histology by integrating low-, high-level and domain specific information," in *Proc. 2nd MICCAI Workshop Microscopic Image Analysis with Appl. in Biology (MIAAB)*, Piscataway, NJ, USA, 2007.

[3] A. Hafiane, F. Bunyak, and K. Palaniappan, "Clustering initiated multiphase active contours and robust separation of nuclei groups for tissue segmentation," in *IEEE ICPR*, Genova, Italy, 2008.

[4] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Proc.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[5] L. Vese and T. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *Int. J. Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.

[6] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, pp. 577–685, 1989.

[7] T. Chan, B. Sandberg, and L. Vese, "Active contours without edges for vector-valued images," *J. of Visual Communication and Image Representation*, vol. 11, pp. 130–141, 2000.

[8] V.Caselles, R.Kimmel, and G.Sapiro, "Geodesic active contours," *Int. J. Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[9] A. Hafiane, S. Chabrier, C. Rosenberger, and H. Laurent, "A new supervised evaluation criterion for region based segmentation methods.," in *ACIVS*. 2007, vol. 4678 of *Lecture Notes in Computer Science*, pp. 439–448, Springer.