Unsupervised Tissue Image Segmentation through Object-Oriented Texture

Akif Burak Tosun

Department of Computer
Engineering, Bilkent University,
Ankara, Turkey
e-mail: tosun@cs.bilkent.edu.tr

Cenk Sokmensuer

Department of Pathology,
Hacettepe University Medical
School, Ankara, Turkey
e-mail: csokmens@hacettepe.edu.tr

Cigdem Gunduz-Demir

Department of Computer
Engineering, Bilkent University,
Ankara, Turkey
e-mail: gunduz@cs.bilkent.edu.tr

Abstract— **This paper presents a new algorithm for the unsupervised segmentation of tissue images. It relies on using the spatial information of cytological tissue components. As opposed to the previous study, it does not only use this information in defining its homogeneity measures, but it also uses it in its region growing process. This algorithm has been implemented and tested. Its visual and quantitative results are compared with the previous study. The results show that the proposed segmentation algorithm is more robust in giving better accuracies with less number of segmented regions.**

Keywords- Quantitative medical image analysis; Image segmentation; Texture analysis.

I. INTRODUCTION

Cancer is one of the most important health problems that threat the human life [1]. The likelihood of curing cancer increases with its early diagnosis and correct grading, for which histopathological examination is routinely used. The number of computational studies on histopathological image analysis is increasing over the past few years. The main aim of these studies is to automate the diagnosis and grading process for reducing the subjectivity that can be observed in histopathological examination. These studies extract features from a histopathological tissue image and use the features in automated diagnosis and grading [2][3]. The images used in these studies are usually assumed to be homogeneous. However, this may not always be the case and tissue images may contain both normal and cancerous regions (Figure 1). Thus, before extracting features, heterogeneous images should be segmented into their medically uniform regions.

There are many approaches that have been proposed to segment heterogeneous images into their uniform regions; examples include region growing [4], graph-based [5], and stochastic [6] algorithms. These algorithms are proposed for segmenting generic images and do not employ a domain specific knowledge that can be necessary for interpreting some types of images such as histopathological tissue images. The interpretation of the tissue images requires the
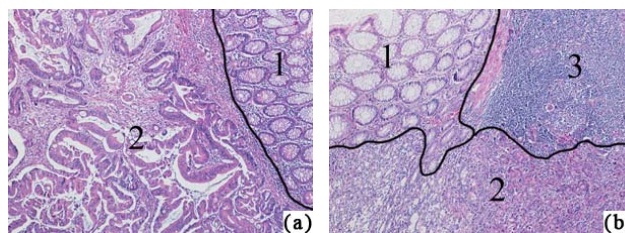

Figure 1: Two examples of colon tissues that are stained with the routinely used hematoxylin-and-eosin technique. The regions labeled with 1 include normal regions and the regions labeled with 2 include cancerous regions of different grades. The region marked with 3 in (b) can be included in either side without affecting the medical interpretation.

domain specific knowledge of a pathologist who mainly employs the distribution of cytological tissue components and abnormalities and irregularities observed in this distribution. Additionally, tissue images have some specific properties: they have similar color distributions in their heterogeneous regions and they may contain a large amount of noise and variations. The segmentation algorithms should also consider these domain specific properties.

In literature, there are few algorithms that work on heterogeneous tissue images. A common approach in these algorithms is to divide an image into grids and classify each grid according to its color and texture information, assuming that the grid is homogeneous [7][8]. These algorithms do not explicitly consider the domain specific knowledge of a pathologist in calculating the features. In our previous study [9], we proposed an algorithm, ObjSEG, which incorporates the knowledge of a pathologist into segmentation. It defines a set of primitive objects to represent cytological tissue components, computes its texture descriptors quantifying the spatial distribution of the objects, and uses these descriptors as homogeneity criteria in its region growing based segmentation. Although ObjSEG improves the results of its pixel-based counterparts, it has a problem of finding a common parameter set that works for all image instances, which reduces its robustness.
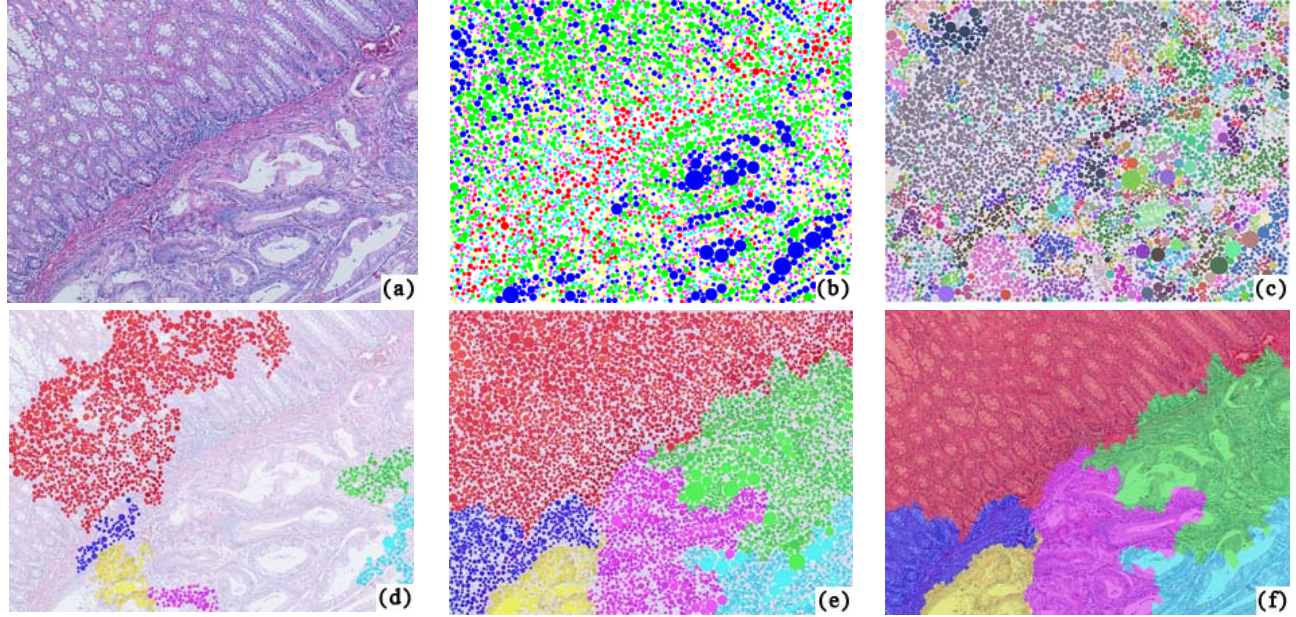
Figure 2: The illustration of the proposed region growing algorithm: (a) the original image, (b) circular objects that approximately represent the cytological tissue components, (c) all of the object groups, (d) seed groups after eliminating the small-sized object groups, (e) grown regions, and (f) final boundaries of the grown regions.

This paper extends the previous work to alleviate this problem. In this current work, we propose a new region growing algorithm, in which the growing process depends on object-to-object relationships, instead of pixel connectivity. It is different than the ObjSEG algorithm that grows the regions based on pixel connectivity. Our experiments show that the use of object-to-object relationships in region growing increases the segmentation performance. It also improves the robustness of the algorithm, enabling to select a common parameter set for all image instances that leads to good segmentation results.

II. METHODOLOGY

A tissue is not a random collection of its cytological components. The distribution of tissue components follows a pattern and this pattern changes with the existence of cancer. Pathologists diagnose and grade a tissue according to these changes. Thus, it is important to use spatial distribution of the tissue components in defining computational measures. The proposed algorithm makes use of the textural measures introduced by our previously proposed ObjSEG algorithm. ObjSEG quantifies the distribution of tissue components, approximately representing these components with circular objects and defining a set of texture descriptors on these objects. As it uses pixel connectivity in its region growing, it computes these descriptors for each pixel. On the other hand, this current work uses object-to-object relationships in the region growing process, and hence, it defines the texture descriptors for objects, instead of pixels. In the following subsections, we first give the details of object definition and the texture descriptors. Then, we explain the steps of the region growing algorithm that our current work introduces.

A. Object definition

Tissue components are approximately represented with circular objects using a heuristic algorithm given in [9]. This algorithm first clusters the image pixels into three by using the k-means method. Each cluster corresponds to one of the dominant colors in a tissue stained with hematoxylin-and-eosin. These colors are purple, pink, and white, and they mainly correspond to cell nuclei, stromal regions, and luminal regions, respectively. The algorithm then locates circular objects on the pixels of each cluster and groups them into two according to their sizes; one group is for larger objects and the other is for smaller ones. Figure 2(b) illustrates this object transformation for the original image shown in Figure 2(a); here six object types are represented with six different colors.

B. Homogeneity measures

The homogeneity measures used in this study rely on the following observation: in a homogeneous region, for each particular object, there should be another one with the same type and the same size and that object should be on the symmetrically opposite side of the particular object with respect to the centroid of this region. To quantify this observation for a single object, a window is located at its center and 12 descriptors are defined considering the objects falling in this window.

Let $O = \{O_i^j \mid j = 1, 2 \ldots 6; \, i \in N_j\}$ be the object set where O_i^j is the object with type j and id i and N_j is the number of objects with type j. Each object $O_i^j = \{X_i, Y_i, A_i\}$ is characterized with its centroids X_i and Y_i and its area A_i. And also let W be the window located at a given pixel; in our case, this pixel corresponds to the centroid of an object O_i^j

for which the texture descriptors are calculated. As the first set of descriptors, for each object type, the standard deviation Std^j of object areas are calculated as follows

$$Std^j = \sqrt{\left(\sum_{k \text{ s.t. } O_k^j \in W} (A_k - \mu^j)^2\right) / (n^j - 1)}$$

(1)

where μ^j is the average area of the objects that fall in window W and has type j and n^j is the number of these objects.

As the second set of descriptors, for each object type, the sum of the position vectors of the corresponding objects with respect to the window centroid (X_W, Y_W). Sum^j is defined as

$$Sum^j = \sqrt{\left(\sum_{k \text{ s.t. } O_k^j \in W} |X_k - X_W|\right)^2 + \left(\sum_{k \text{ s.t. } O_k^j \in W} |Y_k - Y_W|\right)^2}$$

(2)

C. Region growing algorithm

In its first step, seed groups are identified based on the similarity of adjacent objects. To determine the adjacent objects, a Voronoi diagram is constructed on the centroids of all objects and any two objects are labeled as adjacent if they share an edge on this Voronoi diagram. Then, similar objects are grouped together such that the Euclidean distance between any pair of the adjacent objects in this group is below a *similarity threshold*. Finally, the large-sized groups that contain more objects than an *object threshold* are considered as seeds. Figure 2(c) shows the objects of the same group with the same color. Figure 2(d) shows the seeds that are obtained by eliminating the small-sized groups.

In its second step, seeds are iteratively grown by appending the remaining objects to one of the seeds. For doing this, an individual remaining object is appended to an adjacent seed group if the distance between this object and the seed group is smaller than the *similarity threshold* that is relaxed by its 10 percent in every iteration. The descriptors of a seed group are calculated by averaging the descriptors of all objects that belong to this seed group. When all objects are assigned to a seed group, the algorithm employs the Voronoi diagram of the objects to find the final boundaries of the grown regions. The grown regions and their final boundaries are illustrated in Figure 2(e) and Figure 2(f), respectively.

III. EXPERIMENTS AND RESULTS

We conduct our experiments on 16 randomly chosen colon tissue images that contain both normal and cancerous regions. The tissues are stained with hematoxylin-and-eosin and their images are captured using a Nikon Coolscope Digital Microscope with 5× microscope objective lens. The

images are taken in the RGB color space and their resolution is 1920×2560.

The proposed algorithm has three parameters: window size (*winSize*), similarity threshold (*simThr*), and object threshold (*objThr*). To select the parameter set, we use leave-one-out cross validation; for each particular image, we determine the parameter set on all other images excluding this particular image and obtain its test segmentation result. For that, we consider all possible combinations of the following sets *winSize* = {32, 64, 96, 128}, *simThr* = {0.25, 0.50... 3.00, 3.50, 4.0}, and *objThr* = {10, 25, 50... 100, 150... 250} and select the one that leads to the best performance over all images except the excluded image. In defining the best performance, we consider both the accuracy and the number of segmented regions: we select the parameter set that leads to the best accuracy and that gives at most 10 segmented regions. Note that if only the accuracy was considered, we would select the parameter set that leads to very high accuracies but at the same time very high number of regions. Table I reports the average quantitative test results and the average number of segmented regions. The quantitative results are calculated comparing the segmented regions with the manual segmentation provided by our medical collaborator. The details of this calculation can be found in [9].

In order to understand the effectiveness of the proposed algorithm, we compare its results with our previously proposed ObjSEG algorithm [9], which uses a similar set of homogeneity criteria but a different region growing procedure in its segmentation. ObjSEG has also model parameters: small and large window sizes (*winS* and *winL*), area threshold (*area*), and merge threshold (*merge*). We select its parameter set also using leave-one-out cross validation, considering the following candidate sets: *winS* = {32, 64}, *winL* = {128, 256}, *area* = {5000, 7500... 20000, 25000... 50000}, and *merge* = {0.0, 1.0, 1.5, 2.0... 4.0}. The quantitative test results obtained by the ObjSEG algorithm are also given in Table I .

Figure 3 shows the visual results of the proposed algorithm (a1 – f1) and those of the ObjSEG algorithm (a2 – f2) on six example images. On these images, segmented regions obtained by each algorithm are shown in different colors. The images also include the boundaries of cancerous and normal regions that are manually drawn by our pathologist collaborator; regions that can be included in either side without affecting the medical interpretation are shaded in black. The quantitative and visual results show that the proposed region growing algorithm gives better segmentation performances. However, there are still errors in

TABLE I. QUANTITATIVE SEGMENTATION RESULTS

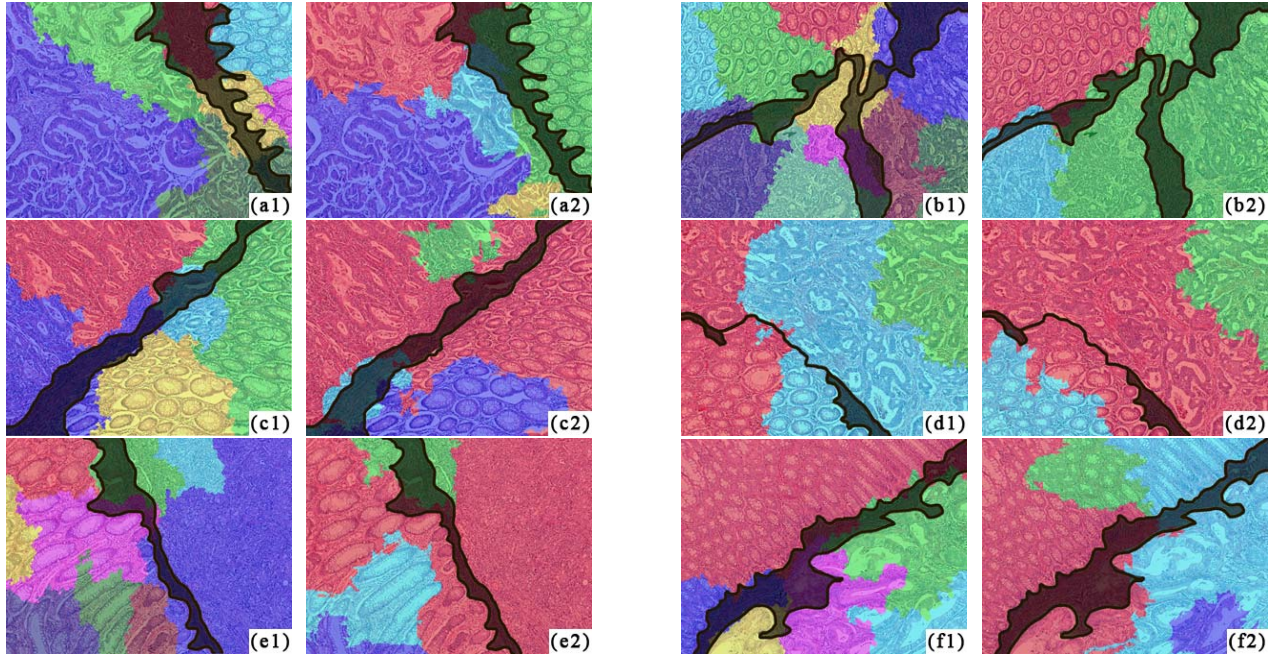|	Accuracy	Sensitivity	Specificity	Region no
Proposed algorithm	86.5 (±11.1)	86.0 (±26.5)	82.9 (±23.4)	5.9 (±2.2)
ObjSEG algorithm	82.8 (±12.5)	92.1 (±16.3)	66.1 (±32.8)	6.1 (±2.2)

Figure 3: The visual results of the proposed algorithm (a1 – f1) and the previous algorithm [9] (a2 – f2). Segmented regions are shown with different colors. The manual segmentations are also indicated in these images.

some images such as the one shown in Figure 3(d1). This is due to using a common parameter set; for such images, better results can be achieved with different sets. Nevertheless, the quantitative results demonstrate that the proposed algorithm yields better accuracy, sensitivity, and specificity values when a common parameter set is used for all images. Note that the quantitative results reported in [9] are the ones that are obtained by separately optimizing the merge threshold for each image. Moreover, the proposed algorithm leads to a reasonable number of segmented regions even though it does not have an explicit region merge step.

IV. CONCLUSION

This paper introduces a new region growing algorithm for the unsupervised segmentation of tissue images. This algorithm relies on using the similarity of objects that approximately represent cytological tissue components. Working with the images of colon tissues, our experiments show that the proposed region growing algorithm leads to better results compared to the previous algorithm that uses similar criteria but a different region growing procedure.

Our future work includes comparing the proposed algorithm with other state-of-the-art segmentation methods and defining different object-based textures for unsupervised segmentation algorithms.

REFERENCES

[1] WHO, "Cancer", World Health Organization, 2009.

[2] M.N. Gurcan, R. Ernst, A. Oto, S. Worrell, J.W. Hoffmeister, S.K. Rogers, "Measurement of colonic polyp size from virtual colonoscopic studies: comparison of manual and automated methods", Proc. of SPIE Medical Imaging, vol. 6144, pp. 1710-1721, 2006.

[3] C. Demir, S.H. Gultekin, B. Yener, "Learning the topological properties of brain tumors", IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 2 (3), pp. 262-270, 2005.

[4] Y. Deng, B.S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video", PAMI, vol. 23 (8), pp. 800-810, 2001.

[5] J. Shi, J. Malik, "Normalized cuts and image segmentation", PAMI, vol. 22 (8), pp. 888-905, 2000.

[6] L. Zhang, Q. Ji, "Image segmentation with a unified graphical model", PAMI, 99 (PrePrints), 2009.

[7] J. Smolle, "Computer recognition of skin structures using discriminant and cluster analysis", Skin. Res. Technol., vol. 6 (2), 58-63, 2000.

[8] Y. Wang, D. Crookes, O.S. Eldin, S. Wang, P. Hamilton, J. Diamond, "Assisted diagnosis of cervical intraepithelial neoplasia (CIN)", IEEE J. Sel. Top. Sign. Proces., vol. 3 (1), pp. 112-121, 2009.

[9] A.B. Tosun, M. Kandemir, C. Sokmensuer, C. Gunduz-Demir, "Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection", Pattern Recognit., vol. 42 (6), pp. 1104-1112, 2009.