IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-6, Issue-2 E-ISSN: 2347-2693

A Survey on Advanced Methods for Segmentation of Structures in H & E Stained Images

S.B.Pawar[1*], V.Gaikwad[2]

[1*]Computer Science Engineering, Savitribai Phule Pune University, Pune, India.
[2] Computer Science Engineering, Savitribai Phule Pune University, Pune, India.

[*]*Corresponding Author: swapnapawar19@gmail.com*

Available online at: www.ijcseonline.org

Received: 24/Jan/2018, Revised: 02/Jan/2018, Accepted: 16/Feb/2018, Published: 28/Feb/2018

***Abstract*—** Segmenting a broad class of histological structures is a necessary to identify the presence of cancer, to clarify spatial relation between histological structures in the tumor environments, making precise medicine studies easy, and provide an exploratory tool for pathologists. Histological structure determination helps explain spatial tumor biology and adds an advantage for health care organizations. Role focuses on the segmentation of histological structures present in colored images with stains (H & E) of the breast tissue. Accurate segmentation of histological structures can help build a spatial interaction map identifying the relations between the pixels to serve as an exploratory tool for pathologists. Graph theory based methods proposed based on spatial color statistics and neighborhood of nuclei statistics as well as designed a new region-based score for evaluating segmentation algorithms. In the first method, pair wise pixel color statistics measures in an H&E optimized color space built to enhance the separation between hematoxylin and eosin stains. The first method is expected to be successful in segmenting structures with well-defined boundaries (e.g., adipose tissues, blood vessels).The second method is designed to segment large amorphous histological structures (e.g., tumor nests), the spatial statistics of inter-nuclei distances is considered. Working with expertly annotated breast H&E images, this paper demonstrated the ability of proposed algorithms to identify significant histological structures, and thus enable the understanding of their spatial relationships, and perhaps infer the status of the disease.

Keywords— histopathalogical image analysis, image segmentation, image statistics

I. INTRODUCTION

Histological structure determination helps to explain the spatial relations present between pixels in tumor environments and identify the basis of cancer. Accurate segmentation of histological structures can thus help to build a spatial relation map which will be used as tool to diagnose the presence of cancer. Segmentation can also help to study medicine studies precisely which perform micro dissection for molecular profiling. Accurately segmented image yields correct diagnosis on classification of image. Histological structure segmentation is very challenging because structures such as normal ducts and carcinoma in situ (outer tissues) have well-defined boundaries, but many others, invasive carcinoma (tissues deep down) and stroma for example, does not. Structural morphologies also vary significantly depending on origin of tissues, preparation of tissue and practices used for staining the images.

Historically, biomedical image analysis literature has focused on segmenting nuclei, since nuclei are building blocks for all higher level tissue structures [2].This strategy is unlikely to work in the case of breast carcinoma in situ, where the duct lumens may be completely filled by tumor cells. Accurate results to diagnose the cancer needs improvement in segmentation.

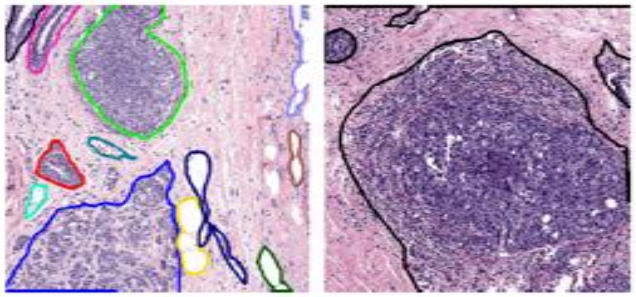

Fig.1.Broad class of histological structures found in breast tissue images [1]

As shown in Fig. 1, image at left side, hematoxylin and eosin stained tissue image, normal ducts present at top left corner have two layers of nuclei, epithelial i.e. is inner layer and

myoepithelial as the name suggests is outer layer, surrounding a cavity. The structure of the normal ducts is disturbed when ducts develop into tumor shown in green shown in Fig. 1, in which the epithelial nuclei multiply in close propinquity and capture the lumen cavity. The structure is further disbursed as the multiplying nuclei become invasive carcinoma (outlined in blue), destroying the duct confinement, freely penetrating through the breast stroma, and heading toward a blood vessel (outlined in teal), indicating an increased risk of cancerous tumor. Host response to infiltrating ductal carcinoma can be seen in Fig. 1 (right), where the tumor nest is pervaded with lymphocyte nuclei shown in small, dark purple color. The nuclei are small and found on one side of the cell wall surrounding large lipid droplets shown in white blobs for other histological structures such as adipose tissue. The system is targeting to segment such histological structures using graph based method to identify the cancer cells.

Proposed system focuses on improving the performance of segmentation and helps to improve pathological diagnosis of cancer. A computer-assisted work flow could tremendously reduce the amount of pathologist effort required to do this diagnostic work, which would be of great value to health care organizations.

II. LITERATURE SURVEY

Luong Nguyen (2017) has represented Special Statistics for segmenting histological structures in H & E stained tissue images. The author propose two segmentations of graph theory methods based on local spatial color and neighborhood of nuclei statistics as well as design a new region-based score for evaluating segmentation algorithms. In one of the methods, pair wise pixel color statistics is measured in an H&E optimized color space built to enhance the separation between hematoxylin and eosin stains.This is expected to be successful in segmenting structures with well-defined boundaries (e.g., adipose tissues, blood vessels).Another method is designed to segment large amorphous histological structures(e.g.tumor nests), where author relies on the spatial statistics of inter-nuclei distances[1].

F. Liu and L. Yang (2015) has represented a novel cell detection method using deep convolutional neural network and maximum-weight independent set.The author proposes a new algorithm for the general problem of cell detection: first, a set of candidates for the detection of cells is generated using different algorithms with variable parameters. Second, each candidate is assigned a score from a trained convolutional neuronal network (NCCD). Finally, a subset of the best detection results is selected among all the candidates to compose the final results of the cell detection. The subset selection activity is formalized as an independent set of maximum weight problem, which is designed to find the heaviest subset of mutually non-adjacent nodes in a graph[2].

J. Vicory et al (2015) has developed an appearance normalization of histology slides,Computerized Medical Imaging and Graphics.The author presents a method for the automatic normalization of color and intensity of digitized histological slides colored with two different agents. Compared to previous approaches, preliminary information on staining vectors is used in the process of estimating the plan, with the consequent improvement in the stability of the estimates. Due to the pervasiveness of hematoxylin and eosin staining for histological slides, the proposed method has a significant practical utility. In particular, it can be used as a first step to standardize the appearance between slides and is effective to counteract the effects due to different amounts and staining protocols and to counteract the fading of the slide[3].

X. Li and K. N. Plataniotis (2015) has demonstrated a complete color normalization approach to histopathology images using color cues computed from saturation weighted statistics. The author presents a complete normalization scheme to address the problem of color variation in images. Histopathology caused by inconsistent biopsy staining and nonstandard imaging. Method: Difference of the existing methods of normalization that approach the cause of the variation of color or of the limits, our method identifies the causes of the variation of color based on a model of microscopic images and addresses the inconsistency in the images and stains of biopsy using an illuminating normalization module and a spectral normalization module, respectively. In evaluation, we use public data sets that are representative of images, clinical tests, and clinical characteristics in terms of assessing the state of the system, the performance relationship with achromatic pixels, and the effectiveness of standardization in terms of preservation. of histological information[4].

J. L. Fine (2014) has represented pCAD approach in 21st century workflow: A proposal.Digital pathology is developing rapidly, but early systems have been slow to gain traction outside of niche applications such as: second opinion telepathology,immunosuppressed interpretation, and intraoperative telepathology. Pathologists have not yet developed a well-articulated plan to effectively use digital imaging technology in their work. This document describes a proposal that aims to initiate significant progress towards the achievement of systems of disconnection of computer-aided pathology, such as computer-aided diagnosis of pathologists (pCAD). pCAD is presented as a hypothetical intelligent computer system that would integrate advanced image analysis and better use of existing digital pathology data from laboratory information systems.

A detailed example of automated digital pathology is presented, such as the automatic disconnection of the mammary lymph node. This proposal provides interested parties with a conceptual framework that can be used to facilitate development work, communication and the identification of new automation strategies[5].

B.-R. Wei and R. M. Simpson (2014) has represented Digital pathology and image analysis augment biospecimen annotation and bio bank quality assurance harmonization.The standardization of best biorepository practices will improve the quality of translational biomedical research using biobanks derived from patients. The harmonization of disease quality assurance procedures for access to biobanks has lagged behind other biomedical research pathways and the development of biobanks. Understanding the cellular content of biorepository samples is important for the discovery of clinically relevant biomarkers for the diagnosis and treatment of tissues. While emerging technologies in molecular analysis and data mining focus on appropriate measures to minimize preanalytic variables that induce artifacts, less attention is paid to the composition of biospecimen annotations for a more effective selection of samples by clients.Both the preanalytical processing of the tissue and the composition of the sample influence the acquisition of macromolecules relevant to downstream assays. Pathological review of biorepository observations, particularly tissues as part of quality assurance procedures, helps ensure that the expected target cells are present and in sufficient quantities in the selected samples[6].

M. T. McCann et al (2014) has represented images as occlusions of textures: a framework for segmentation.The author proposes a new mathematical and algorithmic image for unsupervised image segmentation, which represents a fundamental step in a wide range of image processing applications. He discovered that most of the existing segmentation methods were not successful in the histopathological images, which led us to investigate the segmentation of a larger class of images, that is, without net margins between the regions to be segmented. The author models these images as occlusions of random images, which they call textures and show that local histograms are a useful tool for segmenting them[7].

P. Isola et al (2014) has proposed Crisp boundary detection using pointwise mutual information. The author proposes a new method to detect these limits based on a simple underlying principle: pixels belonging to the same object show higher statistical dependencies with respect to pixels belonging to different objects. The author exhibits the way to calculate an affinity measure based on this principle using wise mutual information, and the author confirms that measure is a good predictor of the presence or absence of two pixels in the same object[8].

N. M. Rajpoot et al (2013) has represented HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images.The segmentation of areas containing tumor cells in histopathological H & E images of the breast (and many other tissues) is a key task for computer-assisted evaluation and classification of histopathological slides. Good segmentation of tumor regions is also vital for automatic immunohistochemical color slide scoring to limit scoring or analysis to areas containing only tumor cells and to avoid potentially misleading results from the analysis of stromal regions. In addition, the detection of mitotic cells is essential for the calculation of key measures such as the mitotic index; a key criterion for classifying different types of cancer, including breast cancer. The author shows that tumor segmentation can allow the detection and quantification of mitotic cells from standard H & E slides with a high degree of precision without the need for special stains, which in turn makes the whole process more economic[9].

Vahadane et al (2016) has represented a structure-preserving color normalization and sparse stain separation for histological images . The staining and scanning of tissue samples for microscopic examination are full of undesirable color changes due to differences in raw material and dye production techniques, in laboratory color protocols and in the color responses of the digital scanners. When comparing tissue samples, color normalization and spot separation of tissue images can be useful for both pathologists and software. The techniques used for natural images do not use the structural properties of colored tissue samples and produce unwanted color distortions. The concentration of the stain can not be negative. The tissue samples are colored with only a few spots and most regions of the tissue are characterized by an effective spot at most. The author models these physical phenomena that define tissue structure first by decomposing images uncontrollably in stained density maps that are scattered and not negative. For a specific image and to combine their spot density maps with staining of a target image preferred by the pathologist, thus altering only its color while retaining its structure described by the maps[10].

III. METHODOLOGY

The system proposed in the paper focuses on segmenting histological structures in H&E stained images of breast tissues, in which hematoxylin stains nuclei to bluish-purple colors, and eosin stains cytoplasm and the stroma matrix to red-pink colors. Proposed system performs color preprocessing and preprocessed image is passed to segmentation Spatial image statistics present discriminative fingerprints for segmenting a broad class of histological structures. To test this, two graph-theoretic segmentation

methods are proposed, each of which relies on characterizing local spatial statistics.

H & E color preprocessing

This step is carried out to enhance the separation between the color appearances of the two stains.It is noted that the nuclei of the various cell types stain with different intensities of hematoxylin. An additional goal of this color representation is to homogenize the variations within the hematoxylin and eosin stain colors and thus, facilitate the extraction of histological structures which extend beyond subcellular components.An experts are asked to select a bag of dominantly stained hematoxylin and eosin pixels.Singular value decomposition is performed on this data to obtain an orthogonal projection matrix.From this construction,the hue values of hematoxylin and eosin stained pixels are expeted to be maximally separated in the complex color plane.

H & E color appearance normalization

Any inconsistencies in sectioning, staining, and imaging result in variation in color appearance of H&E images.It has been found the traditional methods are ineffective because the color distributions for some images in our dataset are skewed toward predominantly one stain, either hematoxylin or eosin.

To identify nuclei,stroma,shrinkage,others,H&E images converted into H&E hue, H&E saturation, and H&E brightness channels using a mixture of univariate von Mises. H&E hue space is angular and given the separation between hematoxylin-stained, eosin-stained, and white pixels in this space, the hue values modeled with a mixture of univariate von Mises distributions. To normalize color appearances,the statistics of the source and target images are matched. The statistics of a distribution can be characterized by an infinite set of moments. However, for analytical convenience,moments only up to the fourth order (mean, standard deviation, skewness, kurtosis) are computed.After normalizing the statistics in the H&E optimized color space, the resulting pixel values are converted into RGB space to obtain the normalized image, using the inverse of the rotation matrix.Normalized images will serve as inputs to two different segmentation strategies based on (1) spatial color statistics and (2) inter-nuclei distance distributions. spatial color statistics within a structure,such as ducts is stronger than across its boundaries. Since the H&E-hue is a angular variable, the joint distribution P(A,B) of hue values from two neighboring pixels lies on a torus.This joint density is modeled as a mixture of bivariate von Mises distributions.

The combination of hematoxylin-eosin pixel pairs across the Connective tisue duct boundary may have an equivalent or even higher probability than a pixel pair stained with hematoxylin inside the duct.This can beimproved by the use of mutual information (MI) to correct for relative abundance.

Joint Probability for pixel pair (i,j) with features fi and fj (fi and fj as hue angles) used to calculate PMI (pointwise mutual information). Graph formed using this pixels as node in graph and pixels connected with other pixels with weights (affinities).Affinity function uses PMI

Nuclei neighborhood statistics based segmentation

Local spatial statistics vary between As a first approximation, the spatial arrangement of the epithelial nuclei is characterized by the physical distances between pairs of nuclei various histological structures in breast tissue.

Nuclei segmentation in histopathalogical images is an extensively researched problem. However, the close proximity of epithelial cells and the prevalence of mitotic configures in breast cancer make it difficult to accurately detect nuclear boundaries. To avoid this issue, putative nuclei locations are identified in the form of super pixels in the hematoxylin stained tissue regions(H) and characterize neighborhood statistics using super pixels derived from eosin stained (E) and white (W) tissue regions. Tissue regions in the O class are ignored due to their small population and being diagnostically not relevant. posterior probability values calculated using the EM algorithm for the three classes. Delaunay triangulation is performed using their center coordinates.

The Delaunay triangulation preserves physical distances and allows building and partition graphs separately in each class. Additionally, the graph generated by the triangulation helps avoid the mistake of connecting a fibroblast nucleus with an epithelial nucleus when they are separated by a large area of stroma.Methods proposed can segment a broad class of histological structures such as fat clusters, blood vessels, lymphocyte aggregations, tumor nests while traditional method focuses on ductal carcinoma in situ. In addition, proposed methods are unsupervised at single scale, while traditional method involves supervised classification at multiple scales; Proposed methods include a rule for merging histological structures identified in different pixel classes.

Neighborhoods derived from the Delaunay triangulation can be richly characterized. However,in the system a simple property selected,namely pairwise distances between superpixels of the same class.In particular, for each class a separate graph is built in which each superpixel is a node, and neighboring superpixels of that class are connected by an edge if their distance is under a threshold. The distance threshold is to be at least the median of the distance distribution between neighboring superpixels.After building the graph over superpixels, greedy connected component analysis algorithm is used to cluster superpixels into segments.

IV. RESULTS AND DISCUSSION

In this paper, two graph-based image segmentation methods proposed based on local spatial color and nuclei neighborhood statistics. In colorStats, the spatial color statistics between neighboring pixels is analytically modeled

using bivariate von Mises mixture models in the H&E optimized color space. In inNucDist, histological structures segmented using nuclei neighborhood statistics in the form of pairwise distances between superpixels. Proposed method in the paper inNucDist, performs better than the state-of-the-art methods and the improvement is statistically significant. Segmentation methods adapted from the domain of natural images (JSEG, EGB, crisp-bound, gPb) performed much better than methods that have been specifically designed for H&E images (GraphRLM, GlandSeg). Notably, non-experts outperformed all the algorithms considered here, except for the method inNucDist on the region score. It is observed that non-experts tend to annotate more segments than experts. This is reasonable since experts with more training in histopathology are able to group together segments into relevant histological structures while non-experts focus on grouping pixels into coherent segments, without the histological knowledge

V. CONCLUSION AND FUTURE SCOPE

This paper aims to build a computer assisted diagnostic system for pathologists that parse WSIs and triages relevant histological structures for rapid diagnosis. Tumor diagnosis and classification is a difficult, labor-intensive task that requires the expertise of highly trained physicians. A computer-assisted workflow could tremendously reduce the amount of pathologist effort required to do this diagnostic work, which would be of great value to healthcare organizations. Two graph theory based segmentation methods discussed in this paper performs better segmentation as compared to traditional state-of-the-art methods. In colorStats method, the spatial color statistics between neighboring pixels is analytically modeled using bivariate von Mises mixture models in the H&E optimized color space. In inNucDist, histological structures segmented using nuclei neighborhood statistics in the form of pair wise distances between super pixels. Histological structures segmented by algorithms proposed in this paper can be potentially ranked from the most to least abnormal (cancer, atypia, inflammation, etc.). Proposed approach also raises the possibility of using spatial statistics in recognizing tissue origins. Finally this study initiates open-source collaborative efforts among pathologists for annotating H&E images.

ACKNOWLEDGEMENTS

This work is supported in the area of histopathalogical structures analysis for the diagnosis of breast cancer. Author is thankful to Faculty of Engineering and Technology (FET), Savitribai Phule Pune Univeristy, Pune for providing the facility to carry out the research work.

REFERENCES

[1] Luong Nguyen, "*Spatial statistics for segmenting histological structures in HE stained tissue images*", IEEE TMI, 2017.

[2] E. Bejnordi et al, "*Automated detection of dcis in whole-slide he stained breast histopathology images*", IEEE TMI, 2016.

[3] J. Vicory et al, "*Appearance normalization of histology slides, Computerized Medical Imaging and Graphics*",vol. 43, pp. 8998, 2015.

[4] X. Li and K. N. Plataniotis, "*A complete color normalization approach to histopathology images using color cues computed from saturation weighted statistics*"IEEE TBME, vol. 62, no. 7, pp.18621873, 2015.

[5] P. Isola et al, "*Crisp boundary detection using pointwise mutual information,*"ECCV 2014, 2014, pp. 799814.

[6] Vahadane et al, "*Structure-preserving color normalization and sparse stain separation for histological images,*"IEEE TMI, 2016.

[7] F. Liu and L. Yang, "*A novel cell detection method using deep convolutional neural network and maximum-weight independent set,*"MICCAI, pp. 349357, 2015.

[8] J. L. Fine, "*21st century workow: A proposal,"Journal of Pathology Informatics*, vol. 5, no. 1,p. 44, 2014.

[9] B.-R. Wei and R. M. Simpson, "*Digital pathology and image analysis augment biospecimen annotation and biobank quality assurance harmonization,*",Clinical biochemistry, vol. 47, no. 4, pp.274279, 2014

[10] M. T. McCann et al, "*Images as occlusions of textures: a framework for segmentation,*", Clinical biochemistry, vol. 47, no. 4, pp. 274279, 2014.

[11] N. M. Rajpoot, "*HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images,*",Journal of Pathology Informatics, vol.4,no.2,p.1,2013

Authors Profile

S.B.Pawar is a PG scholar in Computer Science and Engineering Department at Jayawant Shikshan Prasarak Mandal's Narhe Technical Campus,Narhe,Pune India. She has completed B.Tech. (CSE) from Walchand College of Engineering Sangli, India. Her research interests are Image Processing and Data Mining.

V.S.Gaikwad pursed Master of Computer Engineering from Walchand College of Engineering.He is currently pursuing Ph.D. and currently working as Senior Professor in Department of Computer Engineering JSPM Narhe. His research interests are Image Processing.