

Segmentation and Classification of Histological Structures in H & E Stained Images

S.B.Pawar^{1*}, V.S.Gaikwad²

¹Computer Science Engineering, JSPM Narhe, SPPU Pune, India

²Computer Science Engineering, JSPM Narhe, SPPU Pune, India

*Corresponding Author: swapnapawar19@gmail.com

Available online at: www.ijcseonline.org

Accepted: 13/Jul/2018, Published: 31/Jul/2018

Abstract-- Segmenting a broad class of histological structures is a prerequisite to determine the pathological basis of cancer, to clarify spatial interactions between histological structures in the tumor microenvironments, making precision medicine studies easy with deep molecular profiles, and provide an exploratory tool for pathologists. Histological structure determination helps elucidate spatial tumor biology. 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 to serve as an exploratory tool for pathologists. Breast cancer if detected early can be cured easily. Hence detection methods need to have more accurate diagnosis. Images obtained out of the scans done, processed to get the segments, which are then seen as clusters. These clusters are evaluated with classification techniques to reach the diagnosis result.

Keywords— histopathological image analysis, image segmentation, image statistics

I. INTRODUCTION

Histological structure determination helps elucidate spatial tumor biology and inform pathological basis of cancer. Accurate segmentation of histological structures can thus help build a spatial interaction map to serve as an exploratory tool for pathologists [1]. Segmentation can also facilitate precision medicine studies which perform micro dissection for deep 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 [2] have well-defined boundaries, but many others, invasive carcinoma and stroma for example, does not. Structural morphologies also vary significantly depending on tissue origins (e.g., breast vs. lung), tissue preparation and staining practices.

Historically, biomedical image analysis literature has focused on segmenting nuclei, since nuclei are building blocks for all higher level tissue structures [3]. 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 and classification methods.

The paper is organized as follows, Section I contains the introduction of need for the segmentation and classification of histological structures, Section II contain the related work

carried out in the histopathological domain to help diagnosis study, Section III contain the architecture and essential steps of methodology incorporated in the proposed system along with system design, Section IV describes results and discussion of system analysis in order to evaluate performance of the system, Section V concludes research work with future directions.

II. RELATED WORK

[1] 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.

[2] E. Bejnordi et al (2016) has represented automated detection of dcis in whole-slide h & e stained breast histopathology images. The author presents and evaluates a

fully automatic method for detection of ductal carcinoma in situ (DCIS) in digitized hematoxylin and eosin (H&E) stained histopathological slides of breast tissue. The proposed method applies multi-scale superpixel classification to detect epithelial regions in whole-slide images (WSIs). Subsequently, spatial clustering is utilized to delineate regions representing meaningful structures within the tissue such as ducts and lobules. A region-based classifier employing a large set of features including statistical and structural texture features and architectural features is then trained to discriminate between DCIS and benign/normal structures.

[3] F. Liu and L. Yang (2015) has represent a novel cell detection method using deep convolutional neural network and maximum-weight independent set. The author propose a novel algorithm for general cell detection problem: Firstly, a set of cell detection candidates is generated using different algorithms with varying parameters. Secondly, each candidate is assigned a score by a trained deep convolutional neural network (DCNN). Finally, a subset of best detection results is selected from all candidates to compose the final cell detection results. The subset selection task is formalized as a maximum-weight independent set problem, which is designed to find the heaviest subset of mutually non-adjacent nodes in a graph.

[4] J. Vicory et al (2015) has developed Appearance normalization of histology slides,” Computerized Medical Imaging and Graphics. The author presents a method for automatic color and intensity normalization of digitized histology slides stained with two different agents. In comparison to previous approaches, prior information on the stain vectors is used in the plane estimation process, resulting in improved stability of the estimates. Due to the prevalence of hematoxylin and eosin staining for histology slides, the proposed method has significant practical utility. In particular, it can be used as a first step to standardize appearance across slides and is effective at countering effects due to differing stain amounts and protocols and counteracting slide fading.

[5] 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 introduces a complete normalization scheme to address the problem of color variation in histopathology images jointly caused by inconsistent biopsy staining and nonstandard imaging condition. Method: Different from existing normalization methods that either address partial cause of color variation or lump them together, our method identifies causes of color variation based on a microscopic imaging model and addresses inconsistency in biopsy imaging and staining by an illuminant normalization module and a spectral normalization module, respectively. In evaluation, we use two public datasets that are representative of histopathology images commonly received in clinics to examine the

proposed method from the aspects of robustness to system settings, performance consistency against achromatic pixels, and normalization effectiveness in terms of histological information preservation.

[6] J. L. Fine (2014) has represented 21st century workflow: A proposal Digital pathology is rapidly developing, but early systems have been slow to gain traction outside of niche applications such as: Second-opinion telepathology, immunostain interpretation, and intraoperative telepathology. Pathologists have not yet developed a well-articulated plan for effectively utilizing digital imaging technology in their work. This paper outlines a proposal that is intended to begin meaningful progress toward achieving helpful computer-assisted pathology sign-out systems, such as pathologists’ computer-assisted diagnosis (pCAD). PCAD is presented as a hypothetical intelligent computer system that would integrate advanced image analysis and better utilization of existing digital pathology data from lab information systems. A detailed example of automated digital pathology is presented, as an automated breast cancer lymph node sign-out. This proposal provides stakeholders with a conceptual framework that can be used to facilitate development work, communication, and identification of new automation strategies.

[7] B.-R. Wei and R. M. Simpson (2014) has represented Digital pathology and image analysis augment bio specimen annotation and bio bank quality assurance harmonization. Standardization of bio repository best practices will enhance the quality of translational biomedical research utilizing patient-derived bio bank specimens. Harmonization of pathology quality assurance procedures for bio bank accessions has lagged behind other avenues of bio specimen research and bio bank development. Comprehension of the cellular content of bio repository specimens is important for discovery of tissue-specific clinically relevant biomarkers for diagnosis and treatment. While rapidly emerging technologies in molecular analyses and data mining create focus on appropriate measures for minimizing pre-analytic artifact-inducing variables, less attention gets paid to annotating the constituent makeup of bio specimens for more effective specimen selection by bio bank clients. Pathologist review of bio repository submissions, particularly tissues as part of quality assurance procedures, helps to ensure that the intended target cells are present and in sufficient quantity in accessioned specimens.

[8] 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 framework for unsupervised image segmentation, which is a critical step in a wide variety of image processing applications. We have found that most existing segmentation methods are not successful on histopathology images, which prompted us to investigate segmentation of a broader class of images, namely those without clear edges between the regions to be segmented. We model these images as

occlusions of random images, which we call textures, and show that local histograms are a useful tool for segmenting them.

[9] P. Isola et al (2014) has proposed Crisp boundary detection using point wise mutual information. The author propose a novel method for detecting such boundaries based on a simple underlying principle: pixels belonging to the same object exhibit higher statistical dependencies than pixels belonging to different objects. We show how to derive an affinity measure based on this principle using point wise mutual information, and we show that this measure is indeed a good predictor of whether or not two pixels reside on the same object.

III. METHODOLOGY

Methodology of system includes segmentation method and classification method. Image inputted to the system from the open available data sets gets preprocessed before passing to segmentation algorithm. Improving mechanism of clustering pixels into segments improves the segmentation of pixels into foreground and background.

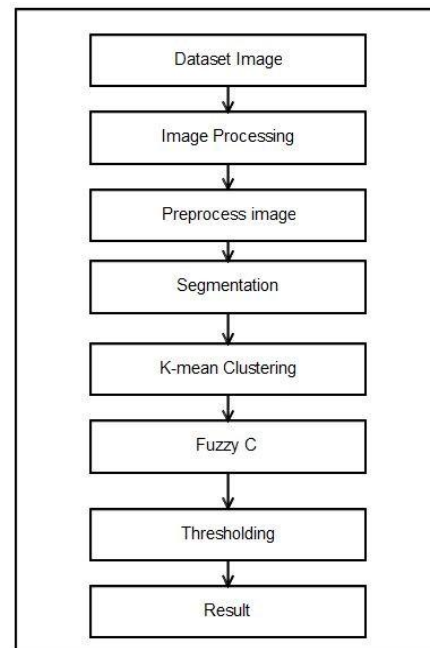
Image Preprocessing

1. **Image Filtering:** A software routine that changes the appearance of an image or part of an image by altering the shades and colors of the pixels in some manner. Filters are used to increase brightness and contrast as well as to add a wide variety of textures, tones and special effects to a picture
2. **Noise Removal from Image:** Images are often degraded by noises. Noise can occur and obtained during image capture, transmission, etc. Noise removal is an important task in image processing. In general the results of the noise removal have a strong influence on the quality of the image processing techniques. Digital images are prone to various types of noise. Noise can also be the result of damage to the film, or be introduced by the scanner itself.
3. **Grayscale Image:** Selected image is transformed into grayscale image. Grayscale is a range of shades of gray without apparent color. The darkest possible shade is black, which is the total absence of transmitted or reflected light. The lightest possible shade is white, the total transmission or reflection of light at all visible wavelength.

1. Segmentation

Proposed system segments the images using K-means algorithm to segment the image and form the clusters. K - means clustering algorithm is an unsupervised algorithm and it is used to segment the interest area from the background. But before applying K -means algorithm, first partial stretching enhancement is applied to the image to improve the quality of the image. Subtractive clustering method is data clustering method where it generates the centroid based on the potential value of the data points. So subtractive cluster is used to generate the initial centers and these centers

are used in k-means algorithm for the segmentation of image. Then finally medial filter is applied to the segmented image to remove any unwanted region from the image.



System Design

K-Means Clustering Algorithm

The cluster analysis procedure is analyzed to determine the properties of the data set and the target variable. It is typically used to determine how to measure similarity distance. Basically, it functions as follows:

Input: The number of k and a database containing n objects.

Output: A set of k-clusters that minimize the squared-error criterion.

Method:

1. Arbitrarily choose k objects as the initial cluster centres;
2. Repeat;
3. (Re) assign each object to the cluster to which the object is the most similar based on the mean value of the objects in the cluster;
4. Update the cluster mean, i.e. calculate the mean value of the object for each cluster; until no change.

2. Classification

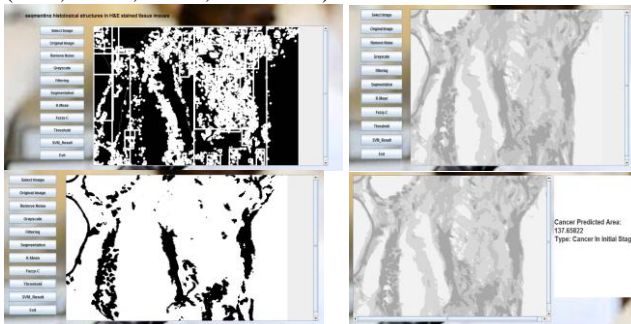
Segmented image is classified to check the status of disease. Classification step classifies it using effective sequential classifier approach support vector machine (SVM) to identify the tumor. Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each

feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

Support Vector Machine (SVM) is primarily a classifier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. For categorical variables a dummy variable is created with case values as either 0 or 1. Thus, a categorical dependent variable consisting of three levels, say (A, B, C), is represented by a set of three dummy variables: A: {1 0 0}, B: {0 1 0}, C: {0 0 1}

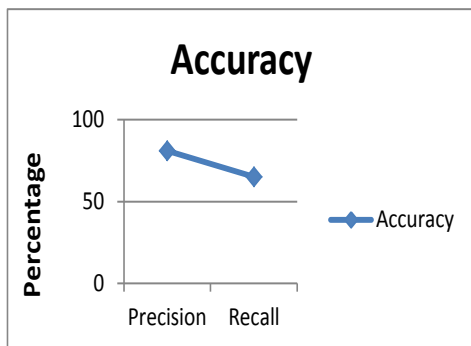
IV. RESULTS AND DISCUSSION

Dataset used to validate the segmentation and classification method is available open for medical research. Break His dataset provides microscopic images of breast tumor tissue collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X).

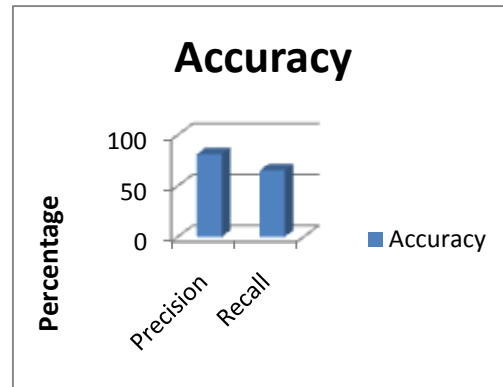


Precision and Recall:

- A dataset contains 100 images.
 - A search was conducted on that images and 80 records were retrieved.
 - Of the 80 records retrieved, 65 were relevant.
- Calculate the precision and recall scores for the search.



Line Chart



Result table:

Total no. of Images	Precision (%)	Recall (%)
100	81	65

V. CONCLUSION AND FUTURE SCOPE

In this system, segmentation method k-means is applied on data set breast cancer image stained with H & E stains. The highest accuracy was achieved when the K-means algorithm was applied but it was more time-consuming than EM. To reduce the time, further optimization should be carried out. This system uses publicly available data sets of images, to identify significant histological structures, and enable the understanding of their spatial relationships, and perhaps infer the status of the disease. The output of segmentation is passed to supervised classifier SVM (Support vector machine) to diagnose the disease. Traditional classification is improved to get better results by increasing the number of measures or attributes.

The future work for this topic is automating the color preprocessing step to be completely unsupervised using the recently proposed non-linear tissue-component separation method. This study attempts to identify segments with distinct spatial statistics, assigning them labels which falls under the challenging task of tissue recognition. Histological structures segmented by the algorithms proposed can be ranked from the most to least abnormal. This idea also raises the possibility of using spatial statistics in recognizing tissue origins. This work area needs initialization of open-source collaborative efforts among pathologists for annotating H&E images.

ACKNOWLEDGEMENTS

This work is supported in the area of histopathological structures analysis for the diagnosis of breast cancer. Author is thankful to ‘Department of Computer Engineering, JSPM, Narhe Pune’ and ‘Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, 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
- [12] S. Tamilarasan, P.K. Sharma, "A Survey on Dynamic Resource Allocation in MIMO Heterogeneous Cognitive Radio Networks based on Priority Scheduling", International Journal of Computer Sciences and Engineering, Vol.5, No.1, pp.53-59, 2017.

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.



V.S.Gaikwad pursued 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.

