

Breast Cancer Histopathology Image Cluster Analysis

Bhagirathi Halalli¹, Aziz Makandar²

Department of Computer Science^{1,2}, Akkamahadevi Women's University, Vijayapura^{1,2}

Email: bhagyaigali@gmail.com¹, azismakandar@gmail.com²

Abstract- Pathology examination remains the major evolution in cancer detection and grading. It constitutes the highest standard in many cancer diagnoses and also plays critical role in prognostic assessments. Recently, computerized techniques have been evolving in diagnosing digital pathology for emerging applications related to nuclei detection, segmentation and classification. Nuclei segmentation remains the challenging task because of cell morphology and architectural distribution. In many of the cancer research, these approaches are playing a key role in reducing human intervention, providing accurate clinical report and considering the second opinions. Pathological studies have been conducted for detection and grading many cancers like cervix, lung, brain, prostate and breast cancer. This paper, focused on nuclei segmentation in the pathological image (H&E staining) conducted for breast cancer. The proposed method helps to segment the nuclei from breast biopsy histopathology image. In this paper, we compared and contrasted the two nuclei segmentation methods such as K-means clustering and L*a*b* Color Space segmentation algorithm. The segmentation result proved that L*a*b* color space algorithm is suitable for nuclei segmentation in H&E staining images.

Index Terms- Image processing, Breast Cancer, Histopathology, Cluster Analysis.

1. INTRODUCTION

Breast cancer can be detected at an earliest through mammography [1] but grading can be done only through the histopathology images and it is the standard clinical practice for the diagnosis and prognosis of breast cancer. Pathology is biopsy microscopic study of the cell morphology changes. The tissue sample collected by biopsy and prepared for viewing under microscope by fixing it. For the purpose of visualizing under micro-scope, the components of the tissues dyed with different stains. Then, the tissue components are analyzed [2][3]. This plays an important role in therapeutic decision making. Hence, diagnosis of pathology image remains the utmost standard in identifying the cancer similarly for breast cancer. In digital pathology, the automated detection and grading the breast cancer is the challenging one. As a consequence, the use of Computer Aided Diagnosis (CAD) in digital pathology, enhances the efficiency and accuracy of the pathologist's decision and it benefits to patients as well as doctors [4][5].

H&E staining method is extensively used method in pathology diagnosis. It has been practiced by pathologist for over hundred years and still it is widely used for observing the changes of cell structure and morphological structure analysis under microscopes [6]. In Hematoxylin stains, nuclei is in dark blue color and other structures like cytoplasm, stroma etc are in pink color as seen the figure 1.

In last decade many researches have been conducted on digital pathology including brain, breast, cervix, liver, lung and prostate cancer detection and grading. Among all the studies, segmentation of nuclei and

classification is recurring task and difficult to perform on histopathology images comparing cytopathology images [7]. In cytopathology nuclei are well separated and simple tissue structure but accuracy is less. In histopathology, structure of nuclei is complex and irregular in visual perception [8-10]. Hence recent researches more focused on nuclei detection in H&E staining images.

This paper aims to present an automated segmentation of nuclei in H&E staining images of breast cancer by using K-means clustering algorithm and L*a*b color space segmentation algorithm.

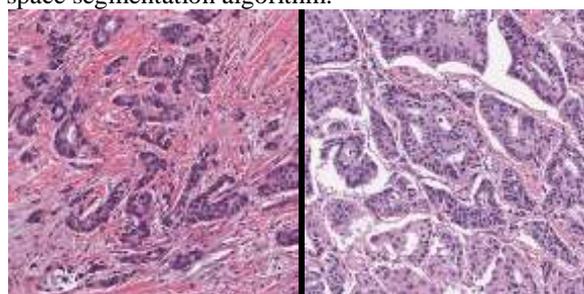


Fig 1. Original histopathology images

2. DATASET

In this paper we have used the dataset used by Andrew Janowczyk [11]. The dataset of original H&E image consist of 143 images scanned at 40x. Each image is of 2000X2000 pixels. The samples collected from 137 patients and 143 images, more than one image associated with few patients. The dataset consist of ground truth of nuclei.

3. METHODOLOGY

Segmentation of nuclei is an important task in digital pathology because of the reasons identified are, there is strong belief that the structure of nuclei is related to outcome and morphology of the nuclei is the key component for grading the cancer [12]. A current research on nuclei segmentation noticed that detection of nuclei and accurately finding the borders or separating the overlapping nuclei is the challenging task [13]. Hence In this paper, we segmented nuclei from histopathology image by using K-means clustering and Color based $L^*a^*b^*$ segmentation methods.

3.1. K-Means Clustering

K-means algorithm is classical clustering algorithm based on distance [14]. It adopts distance as the evaluation parameter of similarity, the shorter distance between two objects is the more similar. K-means clustering algorithm aims to partition number of partitions in cluster space. Each partition or observation in K-means clustering belongs to the particular cluster with nearest mean value is the prototype cluster. We take an advantage of this feature to partition the strained image based on color information of the each cluster [15]. In this paper, K-means algorithm works iteratively to partitions the image into three clusters. The K-means algorithm works in following steps.

Algorithm: K-means clustering

Step 1. Read the histopathology image from dataset.

Step 2. Convert the RGB color image into LAB color space.

Step 3. Pick the Kth cluster randomly for center to distance measure.

Step 4. Label the cluster to each pixel in the histopathology image to minimize the distance between these pixels and center of the assigned cluster.

Step 5. Recomputed the center of the cluster by averaging the pixels in the new cluster.

Step 6. Repeat the step 3 to 5 until the resultant cluster converge.

Step 7. Segment the nuclei into separate image.

The algorithm converges the difference and similarity typically based on intensity value of each pixel, texture information and location or weighted combination of these factors. Its robustness depends on the initialization of primary cluster.

3.2. Color Space

As observed in a^*b^* color space the H&E staining images are having maximum four colors including background. As noticed the different colors easily in histopathology image, $L^*a^*b^*$ color space algorithm helps to separate the color space clusters. It is also called as CIELAB color space. It is derived from the CIE XYZ tristimulus values. And it consists luminosity (L^* -Brightness layer), chromaticity (a^* -color in red-green axis) and Chromaticity (b^* -color falls in blue-yellow axis). In this approach major task is to select the region of interest based on average color in a^*b^* color space [16] [17]. The algorithm works in following steps.

Algorithm: $L^*a^*b^*$ Color Space

Step 1: Read H&E image from dataset

Step 2: Calculate Sample Colors in $L^*a^*b^*$ Color Space for Each Region

Step 3: Classify Each Pixel in H&E Using the Nearest Neighbor Rule

Step 4: Display Results of Nearest Neighbor Classification as different clusters recursively

Step 5: Segment the nuclei on Separate Image.

4. EXPERIMENTAL RESULTS:

Segmentation of nuclei in H&E staining image is as given in the experimental results for qualitative analysis. The nuclei are segmented by K-means cluster algorithm at third cluster as shown in the figure 2. The nuclei segmented in color based $L^*a^*b^*$ algorithm the nuclei identified in a^*b^* space with four iterations. The images represented in the results are selected based on different appearance of the nuclei in H&E image as large and small nuclei.

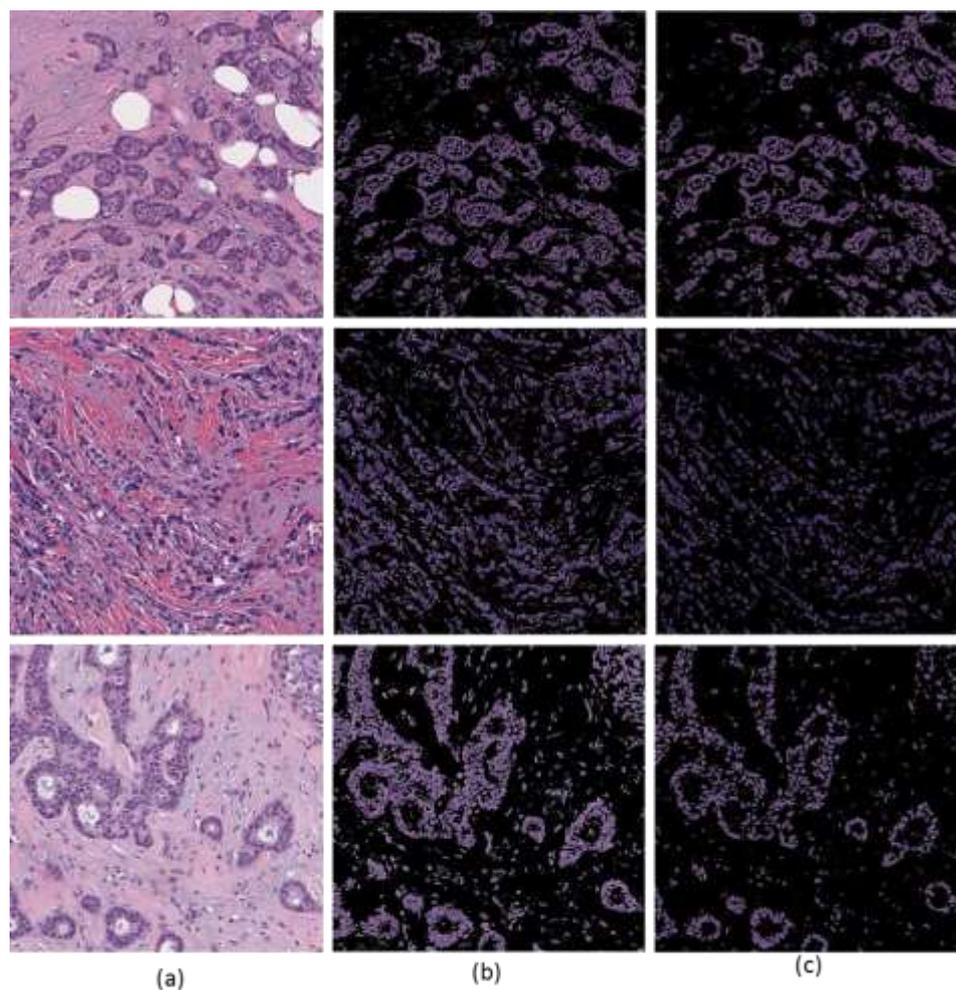


Fig 1. Experimental results (a) Original Image (b) nuclei segmented by lab color space algorithm (c) nuclei segmented from K-means clustering algorithm

The results of segmentation as observed in the above figure suggest that nuclei segmentation is better achieved by color space $L^*a^*b^*$ segmentation algorithm comparing to the traditional k-means algorithm.

Table 1. PSNR values of both the segmentation algorithm

Images	k-means Cluster	Color space $L^*a^*b^*$
img1	36.2171	37.2555
img2	36.6622	37.2211
img3	36.415	37.2555
img4	36.4319	36.4215
img5	36.4959	36.5802
img6	36.4209	36.6441

img7	36.4602	36.7511
img8	36.5335	36.3887
img9	36.5956	36.6478
img10	36.1334	36.5102
img11	36.7373	36.8123
img12	36.3267	36.7213
img13	36.5842	37.1847
img14	36.4656	37.3177
img15	36.3224	36.5245
AVG	36.45346	36.81574667

The quality of the segmentation output analyzed by using Peak Signal to Noise Ratio (PSNR). The PSNR is the proportion between maximum attenuation of power and corrupting noise that influences likeness of image.

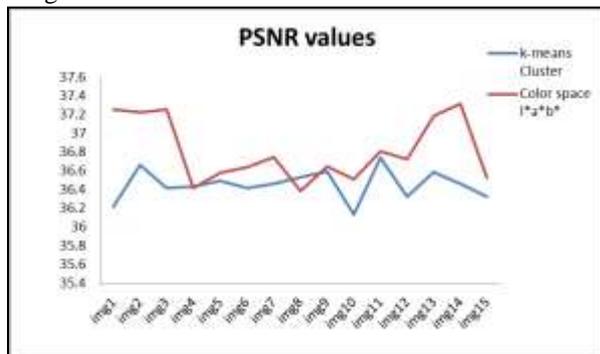


Fig 2. PSNR values plotted

5. CONCLUSION

Since from the last decade, many numbers of researches have been proposed in the field of pathology imaging, mainly focusing on nuclei segmentation and classification. Segmentation of nuclei is the challenging one because of its irregular shape and structure. Moreover, it is not possible to numerically compare the experimental results because of lack of benchmark dataset; different studies have been using their own clinical datasets. We have segmented nuclei by using K-means clustering algorithm and color space $L^*a^*b^*$ algorithm. The K-means algorithms results the nuclei in cluster 3. The PSNR value observed that the color space $L^*a^*b^*$ algorithm have higher than the traditional K-means algorithm. This paper suggests that computer image processing techniques shows the significant changes in pathological image analysis. In future segmentation can be improved by morphological transformations.

REFERENCES

[1] Mammography: Medline plus <https://medlineplus.gov/mammography.html> accessed on 09-06-2017.

[2] Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2008, May). Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on (pp. 284-287). IEEE.

[3] Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., & Madabhushi, A. (2016). Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE transactions on medical imaging, 35(1), 119-130.

[4] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. IEEE reviews in biomedical engineering, 2, 147-171.

[5] Yamaguchi, M. (2017). Computer-Aided Differentiation for Pathology Images. In Image-Based Computer-Assisted Radiation Therapy (pp. 67-84). Springer Singapore.

[6] Veta, Mitko, et al. "Breast cancer histopathology image analysis: A review." IEEE Transactions on Biomedical Engineering 61.5 (2014): 1400-1411.

[7] Irshad, H., Veillard, A., Roux, L., & Racoceanu, D. (2014). Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. IEEE reviews in biomedical engineering, 7, 97-114.

[8] He, L., Long, L. R., Antani, S., & Thoma, G. R. (2012). Histology image analysis for carcinoma detection and grading. Computer methods and programs in biomedicine, 107(3), 538-556.

[9] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. IEEE reviews in biomedical engineering, 2, 147-171.

[10] Loukas, C. G., Wilson, G. D., Vojnovic, B., & Linney, A. (2003). An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections. Cytometry part A, 55(1), 30-42.

[11] Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of pathology informatics, 7.

[12] Veta, M., Huisman, A., Viergever, M. A., van Diest, P. J., & Pluim, J. P. (2011, March). Marker-controlled watershed segmentation of nuclei in H&E stained breast cancer biopsy images. In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on (pp. 618-621). IEEE.

[13] Veta, M., van Diest, P. J., Kornegoor, R., Huisman, A., Viergever, M. A., & Pluim, J. P. (2013). Automatic nuclei segmentation in H&E stained breast cancer histopathology images. PloS one, 8(7), e70221.

[14] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100-108.

[15] Chitade, A. Z., & Katiyar, S. K. (2010). Colour based image segmentation using k-means clustering. International Journal of Engineering Science and Technology, 2(10), 5319-5325.

[16] Recky, M., & Leberl, F. (2010, August). Windows detection using k-means in cie-lab color

- space. In Pattern Recognition (ICPR), 2010 20th International Conference on (pp. 356-359). IEEE.
- [17] Heckbert, P. (1982). Color image quantization for frame buffer display (Vol. 16, No. 3, pp. 297-307). ACM.