

Simple Landscapes Analysis for Relevant Regions Detection in Breast Carcinoma Histopathological Images

Xiao Jian Tan

*School of Mechatronic Engineering
University Malaysia Perlis (UniMAP),
02600 Arau
Perlis, Malaysia.
xj_0506@hotmail.com*

Mohd Yusoff Mashor

*School of Mechatronic Engineering
University Malaysia Perlis (UniMAP),
02600 Arau
Perlis, Malaysia.
yusoff@unimap.edu.my*

Nazahah Mustafa

*School of Mechatronic Engineering
University Malaysia Perlis (UniMAP),
02600 Arau
Perlis, Malaysia.
nazahah@unimap.edu.my*

Wei Chern Ang

*Clinical Research Centre
Hospital Tuanku Fauziah,
01000 Kangar
Perlis, Malaysia.
wei.ang.1990@gmail.com*

Khairul Shakir Ab Rahman

*Department of Pathology
Hospital Tuanku Fauziah,
01000 Kangar
Perlis, Malaysia.
ksyakir@gmail.com*

Abstract— Breast carcinoma represents a huge global health problem among women in both developed and developing countries. It is estimated that over 508,000 women worldwide died in 2011 due to breast carcinoma. Nottingham Histological Grading (NHG) system is recognized as the gold standard to provide overall grade for breast carcinoma. One of the breast carcinoma criteria considered in the grading system is tubule formation. The assessment of tubule formation starts with visual inspection on breast histopathological image using 10x magnification. However, not all regions in the image provide meaningful information. Histopathological image with score 3 in tubule formation usually has a small tubule size. Thus, a visual inspection at a higher magnification is required. A continuous inspection at a higher magnification is time consuming. By eliminating the irrelevant regions in the histopathological image, histopathologist can focus on the relevant region for further examination. This study proposed a simple method to detect relevant region on the breast histopathological images using landscape analysis. The proposed method was tested using three groups of histopathological images: Group 1: relevant and irrelevant regions, Group 2: relevant regions only and Group 3: irrelevant regions only. The proposed method is found to be effective in eliminating irrelevant regions as the overall accuracy for Groups 1, 2 and 3 are 86.6%, 100.0% and 100.0%, respectively.

Keywords— *breast carcinoma; histopathological image; landscapes analysis; relevant region*

I. INTRODUCTION

Nottingham Histological Grading (NHG) system is recognized as the gold standard to provide overall grade for breast carcinoma [1]. Tubule formation is one of the three critical factors that is stated in the NHG system. The other two critical factors are mitotic count and nucleus pleomorphism [2, 3].

In recent years, pathology laboratories have undergone transformation where digital workflow has been introduced as standard practice [4]. The introduction of whole slide imaging (WSI) scanner allows a high throughput slide digitalization with relatively low cost [5]. The application of WSI scanner is fully automated. Slide digitalization is

recognized as a part of the standard practice in the pathology laboratory. The analogue histopathological slides obtained from surgical biopsy are converted to the digital slides using WSI scanner. Quantitative and qualitative analyses could be performed on the digital slides by implementing various image processing algorithms [5].

In the assessment of tubule formation, tumor regions that provide meaningful information which indicate the degree of differentiation in tumor cells are referred as relevant regions, whereas, the non-tumor regions and background are referred as irrelevant regions. Standard practice assessment of tubule formation starts with visual inspection at 10x magnification on a histopathological image. However, not all regions in the histopathological image provides meaningful information (i.e., relevant regions). Histopathological image with score 3 in tubule formation (obtained from NHG system) usually has a small tubule size. Histopathologist may require a visual inspection at a higher magnification (e.g., 20x to 40x magnification). A continuous visual inspection at a high magnification is time consuming [6]. A histopathological image could be formed by as high as 700,000 pixels. By eliminating the irrelevant regions in the histopathological image, histopathologist can focus on the relevant region for further examination. In Figure 1, images (a-d) and (e-h) show examples of relevant regions and irrelevant regions respectively found in the histopathological image.

Study to eliminate irrelevant regions from histopathological images of breast carcinoma for breast carcinoma grading using image processing technique is very few. [7-9] proposed pixel-wise labeling approaches which is suitable to be implemented in small size images. This is a good approach but not practical for a large size image. Implementing pixel-wise labeling approach on a large size image may slow down the overall computation time of the system. Therefore, this paper proposed a simple landscapes analysis that offers a fast and accurate detection of relevant region in breast histopathological images.

The organization of the paper is as follows: Section II provides details description on the proposed method, Section III provides a full description in experimental results and the conclusion is given in Section IV.

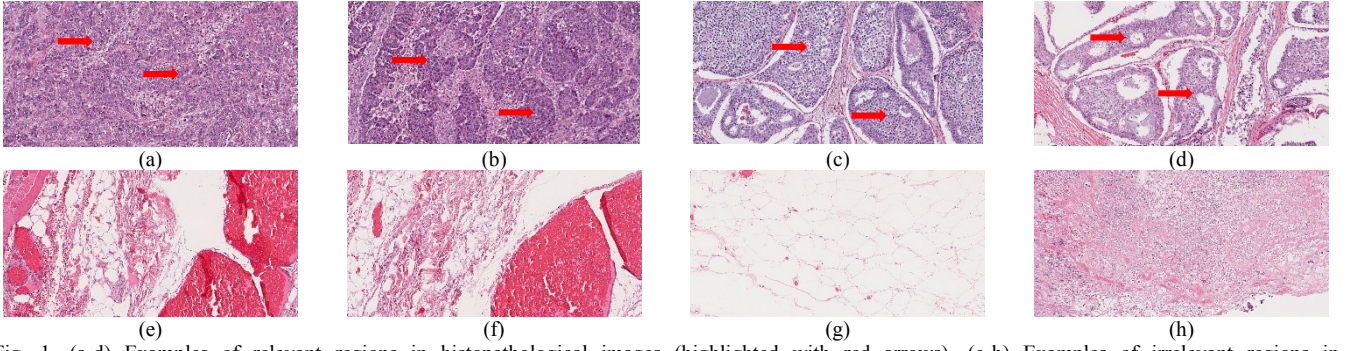


Fig. 1. (a-d) Examples of relevant regions in histopathological images (highlighted with red arrows), (e-h) Examples of irrelevant regions in histopathological images

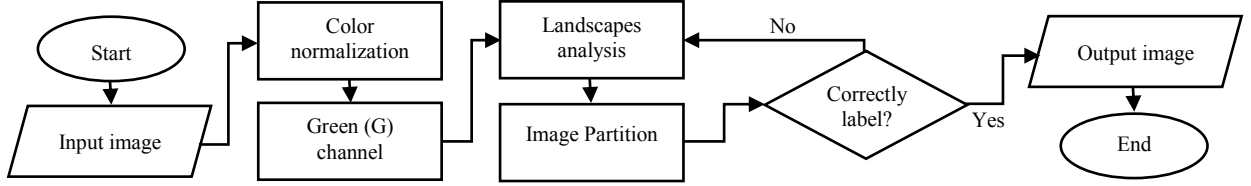


Fig. 2. Flow chart of the proposed method

II. METHOD

The overall flow chart of the proposed method is shown in Figure 2. The proposed method starts with a color normalization technique. This technique was implemented on the RGB input image to avoid color variation. Next, the Green (G) channel of the input RGB image was selected for the landscape analysis. The results obtained from the analysis were used to partition the histopathological image into relevant and irrelevant regions.

A. Color Normalization

Hematoxylin and Eosin (H&E) is the most common staining scheme that is used to discriminate histology structures in breast histopathological images [10]. However, color inconsistency may occur in the histopathological images due to the different manufacturers, different responses of the WSI scanners used, raw material, the method of application, the protocols across different pathology laboratory and the storage conditions prior to use [10-13]. Color inconsistency may hamper the implementation of an image processing algorithm across different histopathological images. To tackle this limitation, a simple color normalization technique namely histogram matching [14] was used. Histogram matching was used to match the intensity histogram of the input RGB image to a pre-selected reference image. Hence, the color inconsistency across different input images could be reduced.

B. Selection of Color Channel

Based on an empirical study, the G channel was found to be significant in structures discrimination specifically between the relevant and irrelevant regions in a histopathological image. In G channel, the intensities of the relevant region was found to be darker as compared to the irrelevant region. Therefore, the G channel was selected as input to the landscapes analysis.

C. Landscapes Analysis

Landscape can be defined as the visible features of an area of land, often considered in terms of their aesthetic

appeal [15]. In image processing, landscape analysis is often related to the analysis of the environmental features and related applications as in [16, 17]. This study assumed breast histopathological image as a landscape image [18]. The histopathological image used is a 2-dimensions matrix (dimensions of $M \times N$) with G intensity values ranging from 0 (i.e., black) to 255 (i.e., white). The visible features and/or patterns on the 'landscape' were then analyzed in vertical and horizontal directions. The landscape analysis in vertical direction ($landscape_n$) started by calculating the sum of the intensity values of each column. These values were normalized by the maximum value obtained from the respective column. For landscape analysis in horizontal direction ($landscape_m$), the sum of the intensity values of each row was calculated and normalized by the maximum value obtained from the respective row. Normalizations in (1) and (2) were used to ease data processing, however, it is not a must as the landscape features were invariant with respect to the scaling of data [18]. The equations of landscapes analysis in vertical and horizontal directions are given in (1) and (2), respectively.

$$landscape_n = \frac{\sum_{m=1}^M i_{nm}}{i_{nmax}} \quad (1)$$

where: i_{nm} = intensity of pixel i at location nm , where

$$n=1, \dots, N; m=1, \dots, M$$

i_{nmax} = the highest intensity in column n

$$landscape_m = \frac{\sum_{n=1}^N i_{nm}}{i_{mmax}} \quad (2)$$

where: i_{nm} = intensity of pixel i at location nm , where

$$n=1, \dots, N; m=1, \dots, M$$

i_{mmax} = the highest intensity in row m

The output values obtained from the $landscape_m$ and $landscape_n$ are referred as landscapes values. These values are always between 0 and 1. The columns or rows with mostly low intensity values tend to provide an output value approximate to 0, whereas, the columns or rows that with mostly high intensity values tend to provide an output value approximate to 1.

D. Image Partition

Based on the output of landscapes analysis, in each direction, two locations were selected: $Upper-limit_n$ and $Lower-limit_n$ for vertical direction; $Upper-limit_m$ and $Lower-limit_m$ for horizontal direction. In both directions, the locations of the first and last landscape values that are lower than k were selected as $Upper-limit$ and $Lower-limit$, respectively. k is a constant value between 0 to 1.

III. EXPERIMENTAL RESULTS

A. Dataset

A total of 50 histopathological images were selected. These images were divided into three groups as follows: Group 1, 30 images with relevant and irrelevant regions; Group 2, 10 images with only relevant region and Group 3, 10 images with only irrelevant region. These images were prepared under standard procedure and captured at 10x magnification by using Aperio CS2 WSI scanner. The captured images contained 8-bit RGB frames with a dimension of 614x1264 pixels. The images were presented in tiff file format.

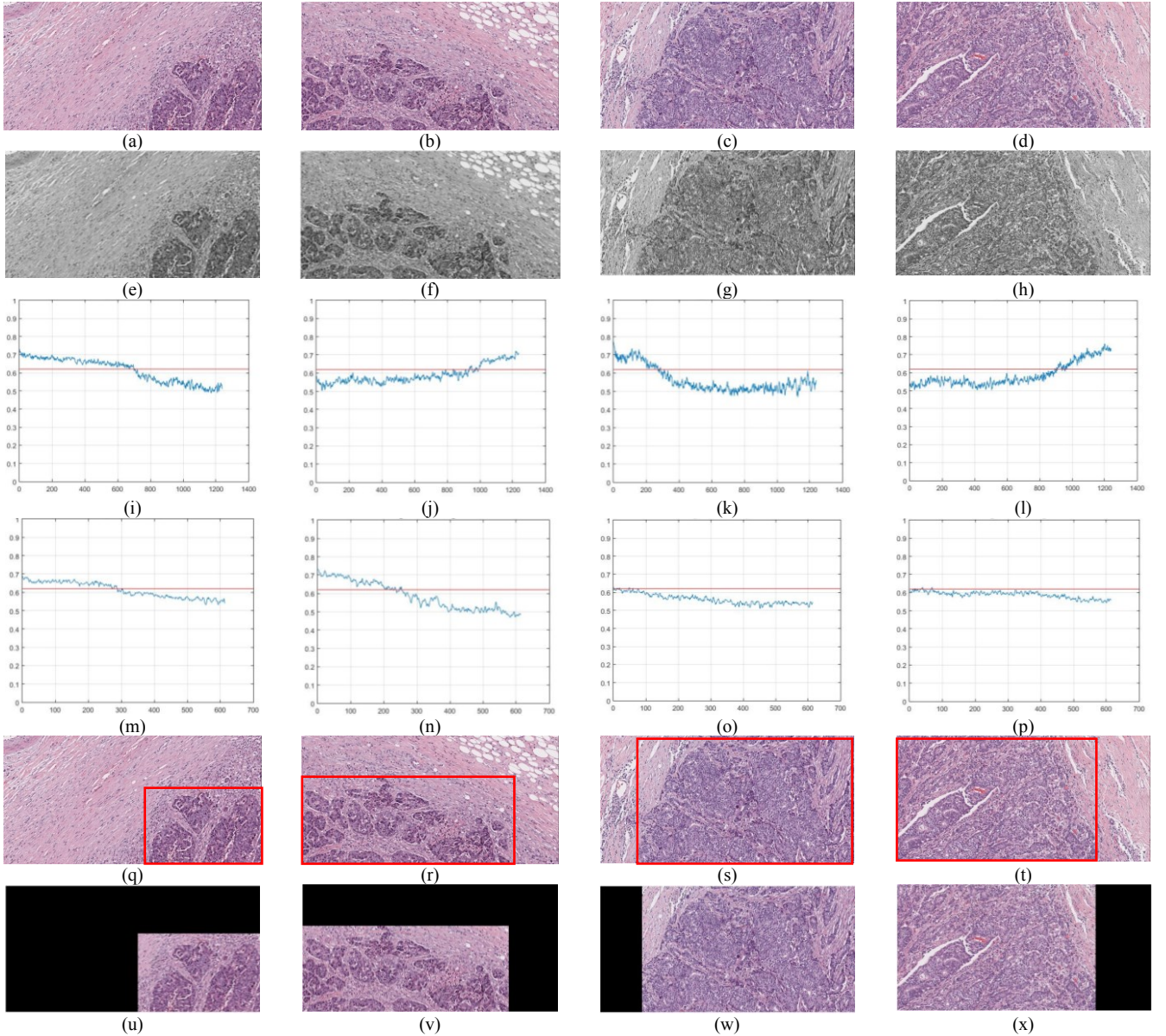


Fig. 3. Results of the proposed method implemented on Group 1 images (images with relevant and irrelevant regions). (a-d) original histopathological images, (e-h) G channel, (i-l) graph of $landscape_n$ against n , (m-p) graph of $landscape_m$ against m , (q-t) ground truth, highlighted in the red square, (u-x) results of proposed method, black region=irrelevant region

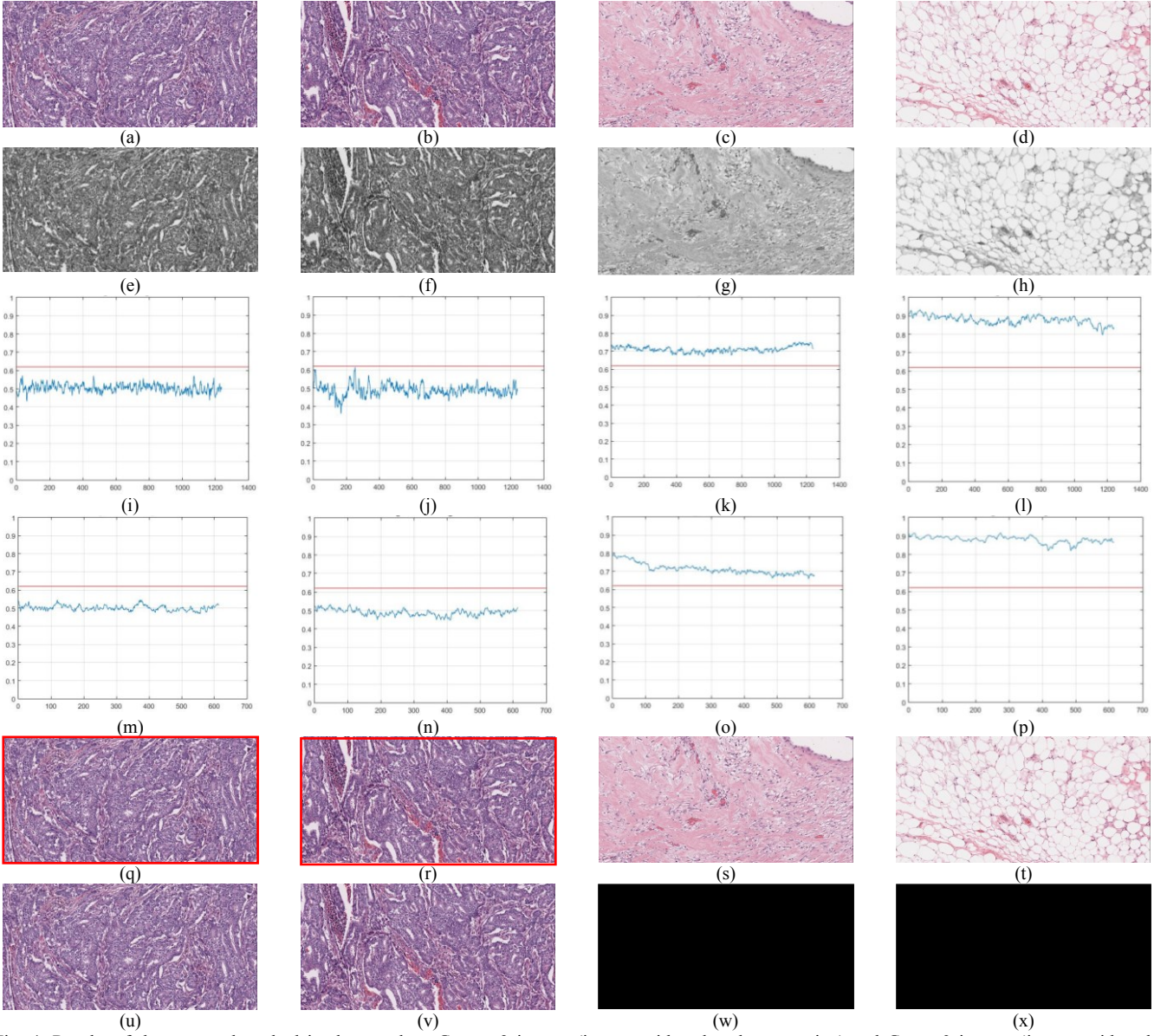


Fig. 4. Results of the proposed method implemented on Groups 2 images (images with only relevant region) and Group 3 images (images with only irrelevant region). (a-d) original histopathological images, (e-h) G channel, (i-l) graph of $landscape_n$ against n , (m-p) graph of $landscape_m$ against m , (q-t) ground truth, highlighted in the red square, (u-x) results of proposed method, black region=irrelevant region

B. Results and Discussion

Figures 3 and 4 show the results of proposed method implemented on histopathological images from Groups 1, 2 and 3: Figure 3 (a to d): original images with relevant and irrelevant regions, Group 1; Figure 4 (a and b): original images with relevant region only, Group 2; Figure 4 (c and d): original images with irrelevant region only, Group 3. In each figure, images (a to d) show the original histopathological images, images (e to h) show the G channel of the input images, images (i to l) show the graphs of $landscape_n$ against n , images (m to p) show the graphs of $landscape_m$ against m , images (q to t) show the ground truth where the relevant region is highlighted in the a red square, images (u to x) show the results of proposed method where the black region indicates the irrelevant region.

For the original images given in images (a to d) of Figures 3 and 4, the relevant regions appear in dark purple color. When converting into G channel, the relevant regions appear as regions with darker intensity (Figure 3 (e to h) and

Figure 4 (e and f)). Based on an empirical study conducted on the landscape analysis, the relevant region has a lower landscapes value whereas the irrelevant region has a higher landscapes value. A constant, $k=0.6200$ was selected such that the landscapes value lower than k is referred as relevant region and vice versa. This assumption is true and useful in detecting relevant regions of breast histopathological images.

Based on Figure 3 (q), the relevant region of the image is located at bottom right of the image. Stepping through the columns, the $landscape_n$ is decreasing from 0.7303 to 0.5103. The first location of the $landscape_n$ dropped below k is at the location, $n=643$. This location was selected as the $Upper-limit_n$. The last location of the $landscape_n$ lower than k is at location, $n=1240$. Thus, this location was selected as $Lower-limit_n$. The same steps were used to determine the $Upper-limit_m$ and $Lower-limit_m$.

For relevant region only (Group 2) in Figure 4 (a and b), the landscape values for both directions are always lower than k . These results are shown in Figure 4 (i, j, m and n). For irrelevant region only (Group 3) in Figure 4 (c and d),

the landscape values for both directions are always higher than k (Figure 4 (k, l, o and p)).

Table I shows the overall results of detection for the proposed method. Based on the results, the proposed method is found to be effective as the proposed method is able to correctly detect the relevant and irrelevant regions in 26 out of 30 images in Group 1. In Groups 2 and 3, the proposed method is able to correctly detect all the relevant and irrelevant regions in the datasets.

TABLE I. OVERALL RESULTS OF DETECTION FOR THE PROPOSED METHOD

Datasets	Correctly Label (CL)	Wrongly Label
Group 1: Relevant and irrelevant regions	26	4
Group 2: Relevant regions only	10	-
Group 3: Irrelevant regions only	10	-

To further evaluate performance of the proposed method, Acc was calculated. Acc is referred as the overall accuracy in detection of relevant region of the proposed method. The equation of Acc is given in (3).

$$Acc = \frac{CL}{T} * 100 \quad (3)$$

where: Acc = Accuracy

CL = Images that are correctly label

T = Number of images in the dataset

The overall result of detection of the proposed method is found to be promising as the overall Acc obtained for Groups 1, 2 and 3 are 86.6%, 100.0% and 100.0%, respectively.

IV. CONCLUSION

This study presents a simple landscapes analysis for relevant region detection in breast histopathological images. The intensity values of the G channel was used as input for landscapes analysis. The proposed method is found to be effective as the proposed method is able to partition the histopathological image into relevant and irrelevant regions. The irrelevant region is eliminated at the end of the algorithm. The overall Acc for Groups 1, 2 and 3 are 86.6%, 100.0% and 100.0%, respectively. As the proposed method did not involve complex mathematic equation, the overall computation time is low. Therefore, the proposed method is suitable to be used in large scale computation for considerable image size. This study could be further improved by increasing the Acc of the algorithm and testing by using a large scale dataset.

ACKNOWLEDGMENT

The authors would like to acknowledge the support from the Fundamental Research Grant Scheme (FRGC) under a grant number of FRGS/1/2016/SKK06/UNIMAP/02/3 from the Ministry of Higher Education Malaysia. The protocol of this study has been approved by the Medical Research and

Committee of National Medical Research Register (NMRR) Malaysia (NMRR-17-281-34236).

REFERENCES

- [1] H. J. G. B. and W.W.Richardson, "Histological grading and prognosis of breast cancer," vol. 22, no. 1, pp. 36–37, 1957. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [2] X. J. Tan, N. Mustafa, M.Y. Mashor, and K. S. Rahman, "Hyperchromatic nucleus segmentation on breast histopathological images for mitosis detection," Journal of Telecommunication, Electronic and Computer Engineering, 2017.
- [3] X. J. Tan, N. Mustafa, M.Y. Mashor, and K. S. Rahman, "Segmentation based classification for minimizing number of mitosis candidates on breast histopathological images," Journal of Telecommunication, Electronic and Computer Engineering, 2017.
- [4] N. Stathonikos, M. Veta, A. Huisman, and P. van Diest, "Going fully digital: Perspective of a Dutch academic pathology lab," *J. Pathol. Inform.*, vol. 4, no. 1, p. 15, 2013.
- [5] V. M., P. J.P.W., V. D. P.J., and V. M.A., "Breast cancer histopathology image analysis: A review," vol. 61, no. 5, 2014.
- [6] M. Peikari, M. J. Gangeh, J. Zubovits, G. Clarke, and A. L. Martel, "Triaging diagnostically relevant regions from pathology whole slides of breast cancer: A texture based approach," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 307–315, 2016.
- [7] N. Linder et al., "Identification of tumor epithelium and stroma in tissue microarrays using texture analysis," *Diagnostic Pathol.*, vol. 7, no. 1, p. 22, Jan. 2012.
- [8] A. M. Khan, H. El-daly, and N. Rajpoot, "Ranpec : Random projections with ensemble clustering for segmentation of tumor areas in breast histology images," *Med. Image Underst. Anal.*, pp. 1–7, 2012.
- [9] A. M. Khan, H. El-Daly, E. Simmons, and N. M. Rajpoot, "HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images," *J. Pathol. Informat.*, vol. 4, p. S1, Jan. 2013.
- [10] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.
- [11] K. Glatz-Krieger, U. Spornitz, A. Spatz, M. J. Mihatsch, and D. Glatz, "Factors to keep in mind when introducing virtual microscopy," *Virchows Arch.*, vol. 448, no. 3, pp. 248–255, 2006.
- [12] H. Journal, K. Sjukhuset, and K. Institutet, "Methodological aspects on immunohistochemistry in dermatology with special reference to neuronal markers," vol. 745, pp. 735–745, 1993.
- [13] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," *Proc. - 2009 IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI 2009*, pp. 1107–1110, 2009.
- [14] C. C. Vancea, V. C. Miclea, and S. Nedevschi, "Improving stereo reconstruction by sub-pixel correction using histogram matching," *IEEE Intell. Veh. Symp. Proc.*, vol. 2016–August, no. Iv, pp. 335–341, 2016.
- [15] A. Farina, "Chapter 1 introduction to landscape ecology," *Princ. methods Landsc. Ecol. Towar. a Sci. Landsc.*, vol. 2, p. 412, 2006.
- [16] E. A. Nilsen and M. Besterfield-sacre, "Landscape Analysis as a Tool in the Curricular Change Process," pp. 110–116, 2015.
- [17] N. Thanomsieng, N. Boonruam, P. Sirisawat, W. Nonsakhoo, and S. Saiyod, "Landscape analysis system using 3D stereoscopic for drone," *Proc. - 2017 IEEE 13th Int. Colloq. Signal Process. its Appl. CSPA 2017*, no. March, pp. 118–122, 2017.
- [18] W. Klonowski, R. Stepień, and P. Stepień, "Simple fractal method of assessment of histological images for application in medical diagnostics Simple fractal method of assessment of histological images for application in medical diagnostics," *Nonlinear Biomed. Phys.*, vol. 4, no. 1, p. 7, 2010.