

Fuzzy C-Means Clustering with Histogram based Cluster Selection for Skin Lesion Segmentation using Non-Dermoscopic Images

Salam Shuleenda Devi*, Ngangbam Herojit Singh, Rabul Hussain Laskar

Department of Electronics and Communication Engineering, National Institute of Technology Meghalaya (India)

Received 14 June 2019 | Accepted 28 August 2019 | Published 20 January 2020



ABSTRACT

Purpose – Pre-screening of skin lesion for malignancy is highly demanded as melanoma being a life-threatening skin cancer due to unpaired DNA damage. In this paper, lesion segmentation based on Fuzzy C-Means clustering using non-dermoscopic images has been proposed.

Design/methodology/approach – The proposed methodology consists of automatic cluster selection for FCM using the histogram property. The system used the local maxima along with Euclidean distance to detect the binomial distribution property of the image histogram, to segment the melanoma from normal skin. As the Value channel of HSV color image provides better and distinct histogram distribution based on the entropy, it has been used for segmentation purpose.

Findings – The proposed system can effectively segment the lesion region from the normal skin. The system provides a segmentation accuracy of 95.69 % and the comparative analysis has been performed with various segmentation methods. From the analysis, it has been observed that the proposed system can effectively segment the lesion region from normal skin automatically.

Originality/Value – This paper suggests a new approach for skin lesion segmentation based on FCM with automatic cluster selection. Here, different color channel has also been analyzed using entropy to select the better channel for segmentation. In future, the classification of melanoma from benign naevi can be performed.

KEYWORDS

Medical Image Segmentation, Melanoma, Histogram, Fuzzy C-means, Local Maxima, Non-Dermoscopic Image, Cluster Size.

DOI: 10.9781/ijimai.2020.01.001

I. INTRODUCTION

MELANOMA is one of the threatening skin cancer diagnostics in the human body. It is mainly developed due to unpaired DNA damage by ultraviolet radiation from sunlight. Also triggers mutations which cause the rapid multiplication of skin cells and form malignant tumors. The death rate due to melanoma skin cancer is rapidly increasing as compared to other cancer except lung cancer. In United States, 1-78-560 cases are reported as melanoma in 2018. Out of those, 87-290 cases are noninvasive and 91-270 cases are invasive. Melanoma is curable in early stage, but if it is not, it becomes fatal and hard to treat. About 9-320 people died due to melanoma cancer in the US (men: 5990, women: 3330).

There are two possible ways to lower the death rate from melanoma cancer, i.e. to diminish the number of new tumors by identifying and eliminating or preventing the development of melanoma. Many researchers have developed many systems to identify and prevent the situations but it does not hold great.

Classification of melanoma from benign melanocytic naevi is not an easy task. Hence, the importance of digital image analysis techniques

is increasing [1]. The analysis can be done using two different images, i.e. dermoscopic (microscopic) and non-dermoscopic (clinical images). Non dermoscopic images are easily available and accessible as compared to dermoscopic. These images are captured by using conventional user grade cameras which is easily accessible by naked eye [2]. Non-dermoscopic images are preferred over dermoscopic images as they have the benefit of being easily accessible. Whereas dermoscopic images are captures by using a special instrument called dermatoscope which is not easily available. In US, dermoscopic images are used by 50% of the dermatologists for cancer examination. Non-dermoscopic image has also some issues such as non-uniform illumination and less contrast results in tough lesion segmentation.

II. LITERATURE REVIEW

Image segmentation is much used in the medical field for computer aid detection of different types of cancer as the breast one [3] or the skin one, targeted in this research. Various computerized techniques for skin lesions segmentation are developed by researchers. Different segmentation approaches such as region merging, active contours and thresholding have been used for this purpose [4] [5]. Friedman *et al.* [6], proposed a diagnostic technique using the criteria such as Asymmetry, Border irregularity, Color variation and a Diameter greater than 6

* Corresponding author.

E-mail address: shuleenda26@gmail.com

mm for melanoma skin cancer (ABCD acronym). Rosado *et al.* [7], performed a comparative study between the diagnostic results of computer diagnostic systems and trained clinical expert. Here, higher sensitivity is observed with computer systems, but lower specificity (i.e. more false positive than humans). Unsupervised based pigmented skin border detection using statistical region merging algorithm has been proposed. The method comprises of black frame removal, image smoothing and segmentation skin lesion based on region merging [8]. P. G. Cavalcanti *et al.*, [9] utilized a discriminating 3-channel space obtained by principle component analysis to segment the skin lesion. Sadri *et al.*, proposed a system based on wavelet network for dermoscopic images. Here, fixed-grid network is formed without training. Further, network weights calculation and network optimization is done by using orthogonal least squares algorithm. Here, experimental analysis was done using 30 dermoscopic images [10]. An automated lesion segmentation framework has been proposed which consists of a multi-stage illumination correction and texture-based segmentation [11] [12]. High level intuitive features were also proposed to model ABCD criteria used for melanoma detection. It is mainly designed to overcome the issues related with low-level features [13]. Smartphone based melanoma detection has also been proposed. Here, a set of features were extracted for better assessment of the ABCD rule [14]. For non-dermoscopic images, red color channel image along with otsu thresholding has been used to segment the lesion region [15]. Further, principal component analysis has been applied on RGB color space to discriminate the lesion region [16] [17]. Convolutional neural network based lesion region extraction has also been discussed [18]. Skin lesion segmentation is still one of the challenging tasks due to various issues such as scarcity of the image database [19], non-uniform illumination, system performance varying for different images.

Various approaches such as supervised machine learning based segmentation [18], unsupervised approaches (Fuzzy C-means, PCA, etc) have been introduced to overcome the lesion segmentation issues. Recently, Fuzzy C-means (FCM) clustering has been used for medical image segmentation. Ali *et al.* [20] proposed a fuzzy c-means clustering along with mathematical morphology for melanoma segmentation. The aim of the paper is to develop unsupervised skin lesion segmentation for non-dermoscopic image. In FCM, one of the problems is that numbers of cluster should be predefined. To develop an automatic segmentation system based on fuzzy c-means clustering, an automatic system for number of cluster selection is required. In this paper, we develop an automatic skin lesion segmentation system using fuzzy c-means clustering along with histogram property. Histogram property helps to select the number of cluster automatically. The contributions of the paper are:

1. Number of clusters for FCM has been selected automatically using the histogram property. Histogram has been analysed to detect the distribution model (i.e. binomial or normal). First, the two most prominent local maxima of the number of pixel counts are evaluated and the Euclidean distance between them is evaluated. If the distance between them is greater than some threshold value, then the histogram shows normal distribution. Here, threshold value is set by analysis of the histogram of three different cases of skin, i.e. normal skin, melanoma skin and benign naevus skin.
2. The entropy of the different color channels (Hue, Saturation, Value) has been evaluated to choose the better color channel for Fuzzy c-means clustering.
3. Fuzzy c-means has been applied on value channel image to segment the skin lesion. Further, morphological filtering has been done to remove the unwanted artefacts which have been segmented as the lesion region.
4. Further, a comparative analysis has been performed at different levels such as qualitative as well as quantitative.

III. PROPOSED METHOD

The complete architecture of melanoma segmentation consists of five stages i.e. Color space conversion, channel image extraction, number of cluster selection, contrast enhancement and segmentation as shown in Fig. 1.

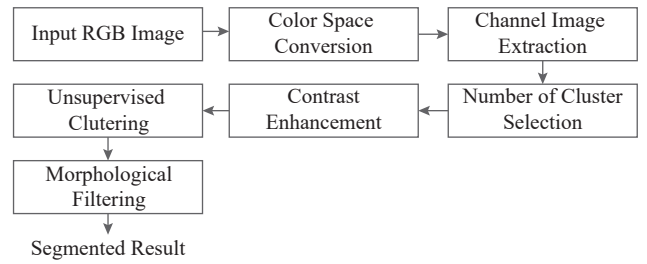


Fig. 1. Proposed method.

A. Image Database

For the experimental analysis, the image dataset of the Department of Dermatology of the University Medical Center Groningen (UMCG) has been used [21] [22]. The images were captured using a Nikon D3 or Nikon D1x body and a Nikkor 2.8/105 mm micro lens maintaining a distance of 33 cm between the lens and the lesion. A total of 170 images (melanoma: 70, naevus: 100) have been used for the experimental purpose. Some of the sample images of the dataset of UMCG are shown in Fig. 2. The sample contains various melanoma cases.

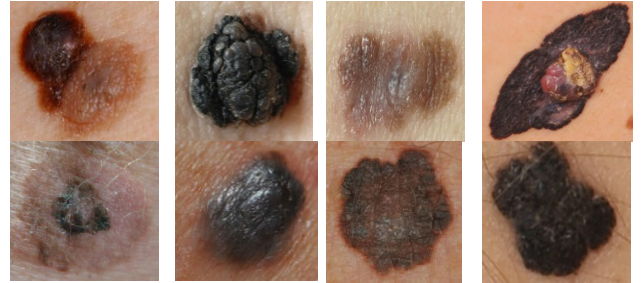


Fig. 2. Sample melanoma images of UMCG dataset.

B. Color Space Conversion

HSV (Hue, Saturation, Value) defines color using familiar comparisons such as color, vibrancy and brightness, with value channel being similar to grayscale image [23]. It is preferred over the RGB color model as it has tendency to perceive color similar to the human eyes. Hue (H), Saturation (S), and Value (V) channel can be defined as

$$H = 60^\circ \times H' \quad (1)$$

$$\text{where } H' = \begin{cases} \text{undefined} & \text{if } c=0 \\ \frac{G-B}{c} \bmod 6, & \text{if } m_2 = R \\ \frac{B-R}{c} + 2, & \text{if } m_2 = G \\ \frac{R-G}{c} + 4, & \text{if } m_2 = B \end{cases}$$

$$V = m_2 \quad (2)$$

$$S = \begin{cases} 0, & \text{if } V = 0 \\ \frac{c}{V}, & \text{otherwise} \end{cases} \quad (3)$$

where $m_1 = \min(R, G, B)$ is minimum component value, $m_2 = \max(R, G, B)$ is maximum component value and $c = m_2 - m_1$ is chroma component value [15] [17].

C. Channel Image Extraction

One of the major issues of lesion segmentation is the poor contrast between the normal skin and lesion region. Therefore, it is required to examine the contrast in different color channels to select the better one. Here, the color channel selection for segmentation is done by entropy [23] [25]. Entropy provides the measure of information content in images. The entropy results of various color channels (i.e. hue, saturation and value) of some sample images are listed in Table I. More entropy indicates the presence of more information which is defined in eq. (4).

$$E = -\sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_{ij} \log_2(P_{ij}) \quad (4)$$

where, P_{ij} represents the probability density function of the image of gray level (i, j) and M represents the total number of gray levels.

TABLE I. ENTROPY RESULTS OF DIFFERENT CHANNEL IMAGES

Sample Images	Entropy		
	Hue	Saturation	Value
1	3.0687	6.2543	6.5193
2	5.0831	6.3505	7.3585
3	5.1328	7.0680	7.6415
4	4.3886	6.6313	6.7071
5	3.7526	6.6496	6.7389
6	4.9115	6.5971	6.6750
7	3.2871	6.1576	6.1019
8	3.2119	6.3167	6.5886
9	3.9622	6.3803	6.6492
10	4.3275	6.4314	6.7713

D. Number of Cluster Selection

In Fuzzy c-means, data clustering is done based on the number of clusters given in advance. Here, the number of clusters is selected automatically using histogram of color channel image. The algorithm used for numbers of cluster selection is shown below:

1. Algorithm for Histogram based Cluster Selection:

Step1: Find the most two prominent local maxima of the histogram

Step2: Calculate the distance (dist) between two prominent local maxima

Step3: if dist <= threshold

Step 4: then,

hist=Binomial distribution, Number of Cluster=2, Perform FCM clustering

Out=Segmented Result

else,

hist= normal distribution

Out=Normal Skin

The threshold value is selected by calculating the Euclidean distance between the two prominent local maxima of the histogram of value channel image. Experimental analysis has been conducted to select the better threshold to separate the normal skin and melanoma skin. From the analysis, the threshold value is set as $Th = 5671$ no. of pixels. If the Euclidean distance between two prominent local maxima is greater than the Th value, then the image will be considered as normal skin image. If less than Th , then the non-dermoscopic image contain melanoma lesion.

E. Contrast Enhancement

Contrast enhancement is used to map the intensity of an image to a specific range. It helps to improve the quality of the image. The

contrast of an image can be decrease or increase depending upon the range of the data [22] [24].

F. Fuzzy C-Means Clustering

An unsupervised Fuzzy C-Means (FCM) clustering has been extensively used for medical imaging systems mainly for segmentation [26] [27] [28]. Here, FCM has been used to segment the melanoma skin from non-dermoscopic images. FCM algorithm consists of two different activities. In the first activity, the number of cluster should be provided. In the second activity, data clustering is performed by iteratively searching for a set of fuzzy clusters. The number of cluster selection is done automatically based on the histogram analysis. A membership value v_{ij} is used to indicate the membership degree of the i th data point to the j th cluster. The main aim of FCM is to minimize the cost function J as

$$J = \sum_{i=1}^M \sum_{j=1}^C v_{ij}^n \|x_i - c_j\|^2 \text{ where, } n \in [1, \infty] \text{ subject to the constraints,} \quad (5)$$

$$\begin{cases} \sum_{i=1}^C v_{ij} = 1, & \text{where } 0 \leq v_{ij} \leq 1 \\ \sum_{j=1}^m v_{ij} > 0, & \text{for } i = 1, 2, \dots, C \end{cases} \quad (6)$$

where, m is the number of data points, $\|x_i - c_j\|$ is the Euclidean distance between the data point and cluster center. C is the number of clusters ($C \leq m$). The membership function v_{ij} and centroids v_j are updated until minimum J is acquired. The algorithm of FCM algorithm is shown below:

Step 1: Select the cluster centers randomly

Step 2: Calculate the fuzzy membership v_{ij}

Step 3: Compute the fuzzy centers

Step 4: Repeat step 2 and 3 until the minimum objective function is obtained

G. Morphological Filtering

Morphological filtering is a non-linear operation which is related to morphological features in an image. It helps to remove as well as recover the unwanted portion of the image segmented as foreground or background in binary image. In binary image, morphological operation creates a new binary image in which the pixel has a non-zero value. Some of the basic morphological operations are erosion, dilation, opening and closing. Here, morphological opening is used to recover lesion region segmented as background [23].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Experimental analysis for lesion segmentation has been done for both quantitative and qualitative analysis. The analysis has been performed using the UMCG dataset which is publicly available. The image dataset consists of melanoma (70 images) and benign naevus skin (100 images). The proposed system has been implemented in MATLAB R2017b, on a computer with Intel Core i3 processor, and 4GB RAM. To reduce the runtime for segmentation, the image has been resized into 250x250 pixels. The proposed system took around 1.7s to segment the lesion region from image having 250x250 pixels. The proposed system consists of color space conversion, channel image extraction, number of cluster selection, contrast enhancement, unsupervised clustering and morphological filtering. Here, RGB image has been converted into HSV image as the HSV color space has capability to perceive the color similar to human eye. From HSV color space, the most preferable color space has been chosen based on the entropy of the various color spaces. Value channel has higher entropy in comparison to other color channels, so it has been used for lesion

segmentation. Further, number of clusters has been selected based on the histogram property as shown in Fig. 3.

From the analysis of the histogram of the various skin conditions, it has been observed that the distance between the most two prominent local maximum of the histogram is less in the skin having lesion. In case of normal skin, the distance between the local maxima is higher in comparison to lesion skin. The threshold value has been used to decide whether the histogram is binomial or not. This threshold value has been set based on the analysis of the histogram of various skin conditions. Moreover, the non-uniform illumination and low contrast of the non-dermoscopic image has been corrected by using contrast enhancement

technique. The preprocessed image is segmented by using FCM with automatically selected clusters number to segment the lesion from normal skin. The performance of the proposed method is evaluated by metrics such as sensitivity, specificity and accuracy. The segmentation has been done as a classification problem. Here, the pixels in the images are classified as lesion or normal skin pixels.

The ground truth of the segmented lesion region has been created as shown in Fig. 4 left column. Fig. 4 shows the comparative analysis of the qualitative measure of various segmentation results. The segmentation result of the proposed method has been compared with Otsu_R method [15] in both qualitative and quantitative case. From

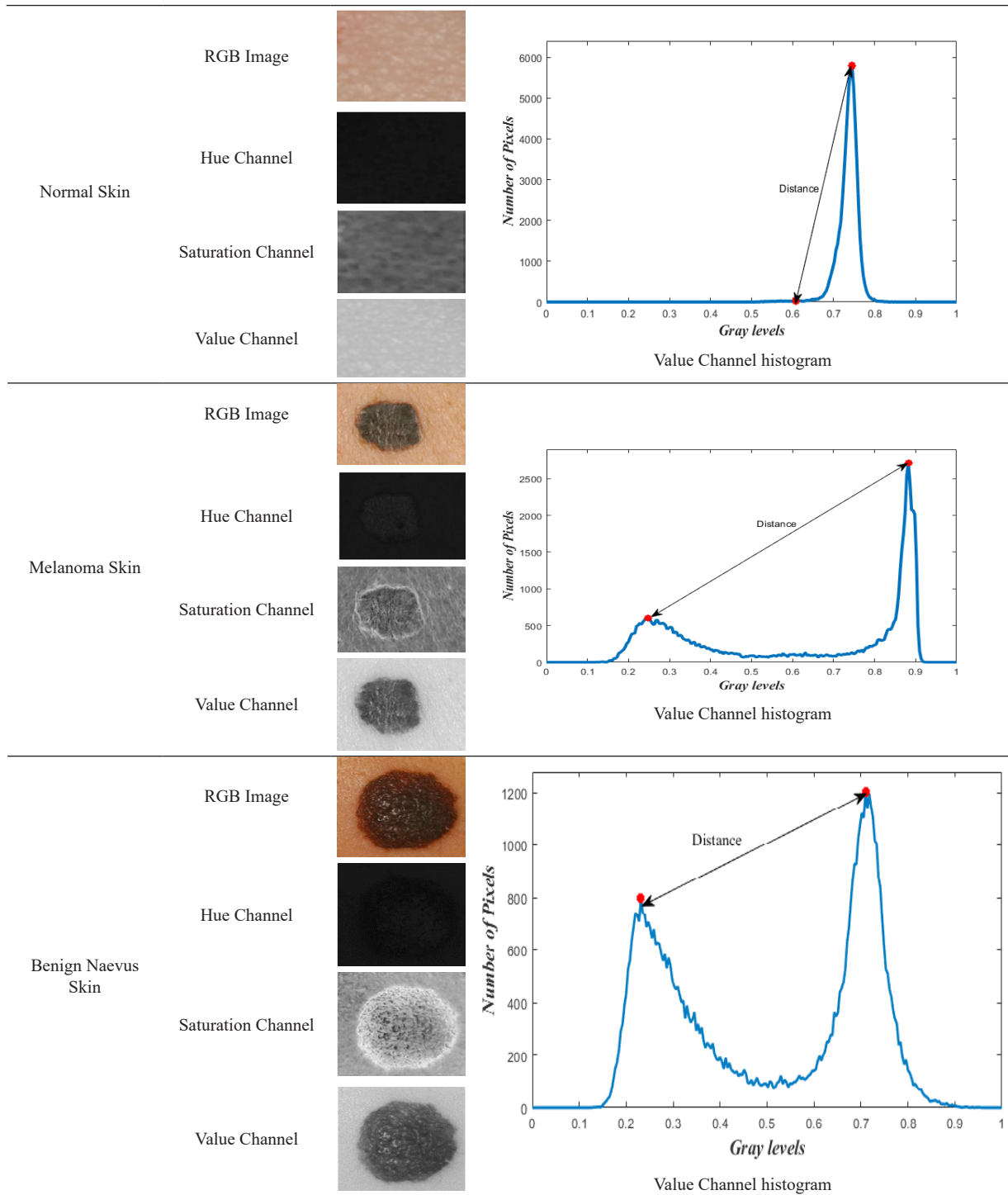


Fig. 3. Details of the histogram for various skin condition.

Table II, it has been observed that the proposed method provides a sensitivity of 90.02%, specificity of 99.15% and accuracy of 95.69%. Our proposed method outperforms the existing method with sensitivity by 11.3%, specificity by 12.78% and accuracy by 12.18%.

TABLE II. COMPARATIVE ANALYSIS OF VARIOUS TECHNIQUES USING UMGCG DATASET

Segmentation Techniques	Image type	Performance statistics (%)		
		Sensitivity	Specificity	Accuracy
Otsu_R [15]	Non-dermoscopic	78.90	86.37	83.51
<i>Proposed method</i>	Non-dermoscopic	90.02	99.15	95.69

V. CONCLUSION AND FUTURE WORK

Computer assisted skin lesions segmentation using non-dermoscopic images is of great importance. Due to various factors such as non-uniform illumination and low contrast in non-dermoscopic images, skin lesion segmentation becomes a challenging task. In our proposed method, the information contained in different color channel images is analyzed using entropy. From the analysis, value channel image provides the better results. So, value channel image has been further preprocessed using contrast enhancement to segment the lesion region. Further, pixels clustering has been done using FCM. For FCM, the number of clusters has been selected using local maxima and Euclidean distance property to detect the histogram distribution. Experimental results show that the proposed segmentation method outperforms existing methods, providing an accuracy of 95.69% and sensitivity of 90.02%. In future, lesion classification based on various supervised classifier can be performed. The importance of feature extraction can also be analyzed. The proposed system is mainly designed for non-dermoscopic images. As the images (i.e.

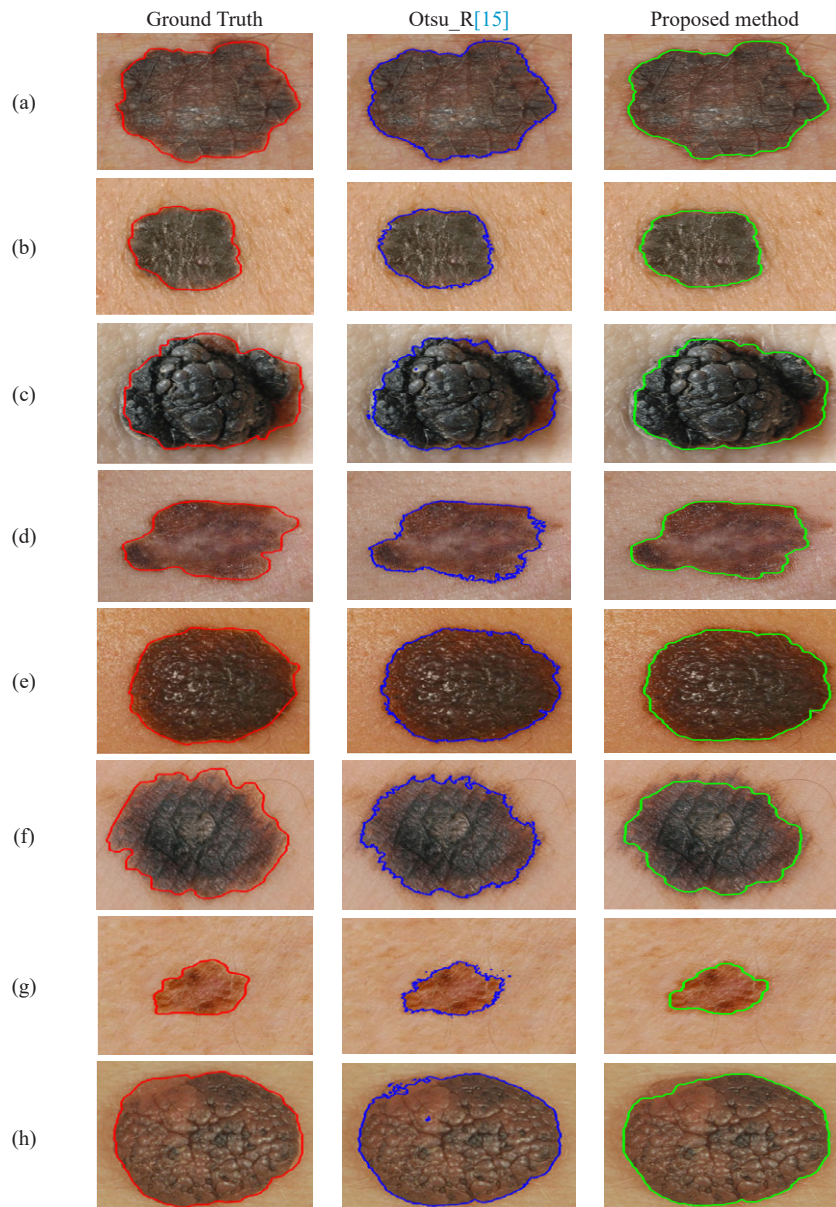
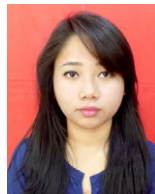


Fig. 4. Segmentation results of skin lesion. *left column*: input images with ground truth (red line); *middle column*: segmentation results of otsu_R [15]; *right column*: segmentation results of proposed method.

non-dermoscopic and dermoscopic) acquired different properties, the proposed system may not be applicable for dermoscopic images, in the present form. With certain customization, the system may also be used for lesion segmentation in dermoscopic images.

REFERENCES

- [1] American Cancer Society: Cancer facts & figures 2016, Atlanta [online], Available: <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html>
- [2] H. C. Engasser and E. M. Warshaw, "Dermatoscopy use by US dermatologists: a cross-sectional survey," *J. Amer. Academy Dermatol.*, 63(3), pp. 412–419, 2010.
- [3] L. Belkhdja, D. Hamdadou, "IMCAD: Computer Aided System for Breast Masses Detection based on Immune Recognition," *Int. J. Interactive Multimedia and Artificial Intelligence*, vol. 5(5), pp. 97-108, 2019.
- [4] M. Celebi et. al., "Border detection in dermoscopy images using statistical region merging," *Skin Res. Technol.*, vol. 14, pp.347–353, 2008.
- [5] M. Celebi, H. Iyatomi, G. Schaefer, W. Stoecker, "Lesion border detection in dermoscopy images," *Comput Med Imaging Graph.*, vol. 33, pp.148–153, 2009.
- [6] R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," *CA: A Cancer J. Clinicians*, vol. 35, pp. 130–151, 1985.
- [7] B. Rosado, S. Menzies, A. Harbauer, H. Pehamberger, K. Wolff, M. Binder, "Accuracy of computer diagnosis of melanoma: A quantitative meta analysis," *Archives of Dermatology*, vol. 139, pp. 361–367, 2003.
- [8] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. Alp Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malter, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, and S. W. Menzies, "Border Detection in Dermoscopy Images using Statistical Region Merging," *Skin Res. and Tech.*, vol. 14, pp. 347-353, 2008.
- [9] P. G. Cavalcanti and J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," *Comput Med Imag and Graph*, vol. 6, pp. 481-491, 2011.
- [10] A. R. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari and F. Kolahdouzan, "Segmentation of Dermoscopy Images using Wavelet Networks," *IEEE Trans. Biomed Engg.*, vol. 60(4), pp. 1134-1141, 2013.
- [11] J. Glaister, "Automatic segmentation of skin lesions from dermatological photographs," M.S. thesis, Dept. of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada, 2013.
- [12] J. Glaister, A. Wong and D. A. Clausi, "Segmentation of Skin Lesions From Digital Images Using Joint Statistical Texture Distinctiveness," *IEEE Trans. Biomed. Engg.*, vol. 61(4), pp. 1220-1230, 2014.
- [13] R. Amelard, J. Glaister, A. Wong and D. A. Clausi, "High-Level Intuitive Features (HLIFs) for Intuitive Skin Lesion Description," *IEEE Trans. Biomed. Engg.*, 62(3), pp. 820-831, 2015.
- [14] M. H. Jafari, S. Samavi, N. Karimi, S. M. R. Soroushmehr, K. Ward, K. Najarian, "Automatic Detection of Melanoma Using Broad Extraction of Features from Digital Images," in *Proc. 38th Annu. Int. Conf. IEEE Engg. in Medicine and Biology Society (EMBC)*, Florida, 2016, pp. 1357-1360.
- [15] P. Cavalcanti, Y. Yari, J Scharcanski, "Pigmented skin lesion segmentation on macroscopic images," in *Proc. 25th Int. Conf. on Image Vision Computing*, 2010, pp 1–7.
- [16] P. Cavalcanti, J. Scharcanski, C Lopes C, "Shading attenuation in human skin color images," in *Proc. 6th Int. Symp. on Advances in Visual Computing*, USA, 2010, pp. 190–198.
- [17] Otsu N, "A threshold selection method from gray-level histograms," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 9(1), pp. 62–66, 1979.
- [18] M. H. Jafari, E. Nasr-Esfahani, N. Karimi, S. M. Reza Soroushmehr, S. Samavi, K. Najarian, "Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma," *Int. J. CARS*, 2017, DOI 10.1007/s11548-017-1567-8
- [19] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial Intel. in Medicine*, 2012, pp. 69-90.
- [20] A. Alil, M. S. Couceiro, and A. E. Hassenian, "Melanoma Detection using Fuzzy C-Means Clustering Coupled with Mathematical Morphology," in *Proc. 14th Int. Conf. on Hybrid Intelligent Systems*, Kuwait, December, 2014, pp.73-78.
- [21] Dermatology database used in MED-NODE [online]. Available: http://www.cs.rug.nl/~imaging/databases/melanoma_naevi/
- [22] I. Giotis, N. Molders, S. Land, M. Biehl, M.F. Jonkman and N. Petkov, "MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Systems with Applications*, vol. 42, pp.6578-6585, 2015.
- [24] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "Digital Image Processing using Matlab," 2nd edition, McGraw Hill Education, 2009.
- [25] Q. Wu, F. Merchant, and K. R. Castleman, "Microscopic Image Processing," Academic Press, 2008.
- [26] R. M. Rangayyan, "Biomedical Image Analysis," CRC Press, London, 2004.
- [27] J. Somasekar, and B.E. Reddy, "Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging," *Comput Electr Eng*, vol. 45, pp. 336-351, 2015.
- [28] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *J. Cybernetics*, vol. 3, pp. 32-57, 1973.
- [29] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithm," Plenum Press, New York, 1981.



Salam Shuleenda Devi

Salam Shuleenda Devi has completed her PhD from National Institute of Technology Silchar, India and M.Tech from KIIT University, Odisha India. She is currently working as Assistant Professor in the Department of Electronics and Communication Engineering at National Institute of Technology Meghalaya, India. Her research interests include Biomedical Image Processing, Machine

Learning, and Pattern Recognition. e-mail: shuleenda26@gmail.com



Ngangbam Herojit Singh

Ngangbam Herojit Singh has completed his PhD from National Institute of Technology Manipur, India. He is currently working as Assistant Professor (temporary) in the Department of Computer Science Engineering at National Institute of Technology Mizoram, India. His major research interests are Artificial Intelligence, Robotics, and Machine Learning. e-mail: herojitng@gmail.com



Rabul Hussain Laskar

Rabul Hussain Laskar has completed his PhD from National Institute of Technology, Silchar, India and his M.Tech from Indian Institute of Technology, Guwahati. He is currently working as Associate Professor in the Department of Electronics and Communication Engineering at National Institute of Technology Silchar, India. His major research interests are in Speech processing, Image processing,

Digital Signal Processing. e-mail: rhlaskar@ece.nits.ac.in