

Global versus Hybrid Thresholding for Border Detection in Dermoscopy Images

Rahil Garnavi¹, Mohammad Aldeen¹, Sue Finch², and George Varigos³

¹ NICTA Victoria Research Laboratory, Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, Australia

² Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia

³ Department of Dermatology, Royal Melbourne Hospital, Melbourne, Australia
r.garnavi@ee.unimelb.edu.au, {aldeen,sfinch}@unimelb.edu.au,
george.varigos@mh.org.au

Abstract. In this paper we demonstrate the superiority of the automated hybrid thresholding approach to border detection in dermoscopy images over the global thresholding method through a newly introduced evaluation metric: Performance Index. The approach incorporates optimal color channels into the hybrid thresholding method, which is a combination of global and adaptive local thresholding, to determine the closest border to that drawn by dermatologists. Statistical analysis and optimization procedure are used and shown to be convergent in determining the optimal parameters for the local thresholding procedure in order to obtain the most accurate borders. The effectiveness of the approach is tested on 55 high resolution dermoscopy images of patients, with manual borders drawn by three expert dermatologists, and the union is used as the ground truth. The results demonstrate the significant advantages of the automated hybrid approach over the global thresholding method.

Keywords: Border detection, Histogram thresholding, Dermoscopy, Melanoma, Computer-aided diagnosis.

1 Introduction

Malignant melanoma is one of the most lethal and rapidly increasing cancers. It represents 10% of all cancers in Australia, and its incidence is four times higher than in Canada, the UK and the US, with more than 10,000 cases diagnosed and around 1700 deaths annually [1]. It is well-known that the clinical diagnosis of melanoma is inherently subjective and its accuracy is considered to be limited, especially for equivocal lesions [2]. This has lead to the emergence of the computer-based analysis of dermoscopy images.

Automatic border detection is an important initial step towards development of a computer aided diagnosis of melanoma. The accuracy of the detected border is crucial as it contains important diagnostic features such as asymmetry and border irregularity and also reveals information about homogeneity and dermoscopic patterns. Therefore, failure to include the whole lesion when segmentation

is performed can drastically bias the results of successive steps of the diagnostic system; feature extraction and classification.

Numerous border detection methods have been proposed in the literature [3], including histogram thresholding followed by region growing [4], color clustering [5], statistical region merging [6], k-means++ clustering followed by region merging [7], etc. Melli *et al.* [5] criticized the adaptive thresholding methods for not providing accurate segmentation result due to the lack of sharp bimodal luminance distribution between the surrounding skin and the lesion.

However, in this study we demonstrate that with a proper choice of color channel, adaptive histogram thresholding can produce borders which are highly accordant with those manually drawn by expert dermatologists. In our preliminary study [8] we proposed a hybrid thresholding method, which combines global and adaptive local thresholding and uses color optimization process to detect the lesion borders, and demonstrated its superiority over other state-of-the-art skin lesion segmentation algorithms. Here, we provide a new analytical and experimental framework to justify the necessity of the second stage of the algorithm, the adaptive thresholding, in the border detection process.

The rest of the paper is organized as follows. The border detection method is summarized in Section 2. Experimental results and discussions are presented in Section 3. Section 4 provides the conclusion.

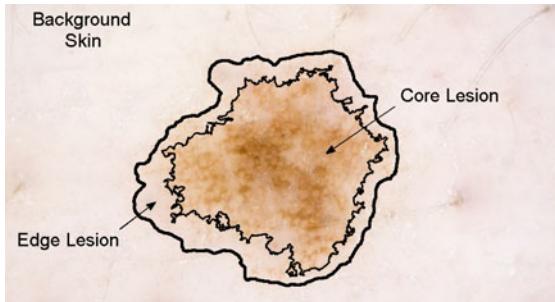


Fig. 1. Three areas generally identified in dermoscopy images: core-lesion, edge-lesion and background skin

2 Hybrid Thresholding Method

Manual borders drawn by dermatologists tend to surround the borders produced by automated method [3,8]. As shown in Figure 1, three areas are generally identified in dermoscopy images; core-lesion, edge-lesion and background skin. The width of the edge-lesion area is variable depending on the skin color, lesion color and the border fuzziness. Existing border detection methods can easily identify the core-lesion area by finding the sharpest pigment change. However, they often fail to precisely detect the edge-lesion area. In contrast, dermatologists prefer to choose the outmost detectable pigment to minimize the risk of incorrect

diagnosis. We tackle the problem by proposing a two-stage histogram thresholding method [8] which successively identifies the core-lesion and edge-lesion areas and here we highlight the success of the method in detecting the edge-lesion area. In the ensuing a summary of the method is provided.

2.1 Forming the Core-Lesion

The core-lesion area is detected by applying global thresholding to the optimal color channel of XoYoR obtained in the color optimization procedure [9]. This step includes the pre-processing operations of hair removal [10], noise filtering using Gaussian low-pass filter (with a 10×10 kernel) and intensity adjustment. This is followed by application of Otsu thresholding method [11] to XoYoR color channel, and performing connected component analysis and morphological operations to obtain the initial border and form the core-lesion area.

2.2 Forming the Edge-Lesion

To expand or shrink the core-lesion boundary to the edge-lesion boundary, an adaptive local thresholding technique based on Otsu method [11] is applied to the X color channel determined as optimal in the color optimization procedure [9]. Starting from an arbitrary point on the core-lesion boundary, the local threshold is calculated over a window of size W . If the local threshold value is less than a defined threshold called T_{expand} the boundary is expanded by one pixel. If it is larger than a defined threshold called T_{shrink} the boundary is shrunk. Otherwise, it is the *No Change* state, in which based on the previous moves, we may make a decision to either laterally move to the adjacent pixel, or expand or shrink the boundary in a radial manner. To define the threshold values for shrinkage and expansion (Equation 2), a bandwidth (Equation 1) is calculated based on background skin and core-lesion pixels values. The Otsu method is applied to the background skin area and core-lesion area to obtain estimates of these values.

$$\text{Bandwidth} = \%B \times (T_{skin} - T_{coreLesion}) \quad (1)$$

where B , the bandwidth factor, is the extent to which the core-lesion area is expanded towards the background skin. The local threshold is calculated for every pixel over its surrounding window and the process is stopped when the initial pixel is revisited. Further details may be found in [8].

$$T_{expand} = T_{skin} - \text{Bandwidth}, \quad T_{shrink} = T_{skin} + \text{Bandwidth} \quad (2)$$

3 Experiments

The proposed method is tested on a set of 55 high resolution dermoscopy images taken by a Canon EOS 450D camera in Royal Melbourne Hospital (24-bit RGB color images, 2000×1334 , TIFF). To validate the automatic borders produced

by the method, manual borders were independently drawn by three dermatologists using Wacom Intuos A4 size Tablet PC. Considering the practical nature of the melanoma diagnosis, which calls for extreme caution when excluding portions of the image, and taking into account the inter-observer and intra-observer variability in borders drawn by dermatologists we calculate the union of the three manually drawn borders for each image and consider that as the ground truth. To quantitatively compare the automatic borders to the manual borders, different metrics have been used [3]. Here, we apply five evaluation metrics of sensitivity specificity, accuracy, similarity and border error given by:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (5)$$

$$Similarity = \frac{2 \times TP}{2 \times TP + FN + FP} \times 100\% \quad (6)$$

$$BorderError = \frac{FP + FN}{TP + FN} \times 100\% \quad (7)$$

where TP, TN, FP, and FN refer to true positive, true negative, false positive and false negative, respectively.

3.1 Standardisation of the Images

The lesion inside a dermoscopy image generally appears in different sizes and locations. Two metrics of accuracy and specificity includes the TN factor which refers to the number of pixels of the surrounding background skin which are truly detected by the automated method. However, the TN factor highly depends on the size of the lesion and its ratio to the whole image, thus the value of accuracy and specificity metrics can be biased in images with small lesions. To the best of our knowledge, this issue has not been addressed in previously published studies. To balance the effect of large TN and normalize the accuracy and specificity metrics, we set a frame of background skin around the lesion, such that the area of the rectangular image frame is twice as big in area as that of an imaginary rectangular encompassing the lesion. This has the effect that the number of background pixel and lesion pixels is roughly the same. Figure 2 shows the segmentation result of a dermoscopy image before and after standardisation.

3.2 Optimal Choice of Window and Bandwidth

Two main parameters are involved in the proposed method: window size (W) and the bandwidth factor (B). In order to determine the optimal value for W and B, a comprehensive set of experiments is performed on the set of 55 dermoscopy images, with W ranging from 30 to 70 and B ranging from 10% to 90% (steps of 10). Different borders can be obtained by using different values of W and B. Consequently, 45 borders are obtained for each dermoscopy image. To evaluate the results, each border is compared with the ground truth in terms of the evaluation metrics expressed in Equations (3–7).

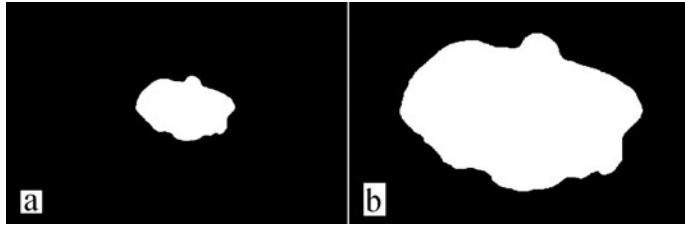


Fig. 2. Segmentation result (a) before and (b) after standardisation

Statistical Analysis. Figures 3 and 4 show the mean value and 95% confidence interval (CI) for metrics of sensitivity, specificity, accuracy, similarity and border error for different values of W and B parameters, respectively. We set levels of acceptability to the lower bound of the confidence interval. Due to the importance of sensitivity, we start the analysis from this metric. As shown in Figure 3(a), given the level of acceptability of 87% for sensitivity, 22 sets of W and B, out of 45, are selected which are marked by filled circles in the graph. With the level of acceptability of 93% for specificity, nine sets from the previous 22 sets are nominated as illustrated in Figure 3(b). Having 93.5% level of acceptability for accuracy, as shown in Figure 3(c), the set is narrowed down to six sets of W and B, i.e. (30,10), (30,20), (40,20), (30,30), (40,30) and (40,40). Having these six sets, two metrics of similarity and border error are studied to make the final decision about the optimal value for B and W. As shown in Figure 4(a and b) according to both of these metrics the pair of (30,30) gains the best results.

Performance Index. It is often the case that different methods provide different results when calculating the five evaluation metrics. The problem that has so far existed is which method provides the best possible result when all of the five metrics are considered simultaneously. To answer this, we introduce the following procedure. First, we define a measure called Performance Index (PI):

$$PI = \frac{SNS + SPC + ACC + SML + (BErr)}{5} \quad (8)$$

where SNS, SPC, ACC, SML and BErr refers to sensitivity, specificity, accuracy, similarity and 100%-borderError, respectively. Second, the index is calculated

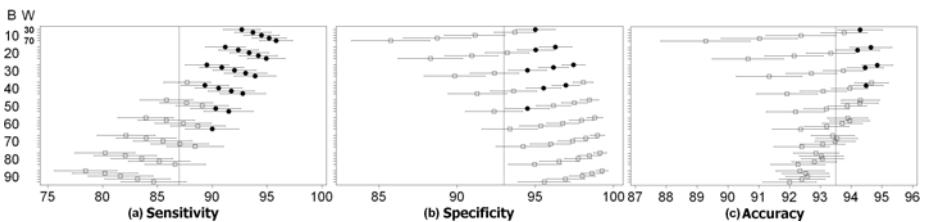


Fig. 3. Mean and 95% confidence interval of (a) sensitivity, (b) specificity and (c) accuracy metrics for different values of B and W over the image set

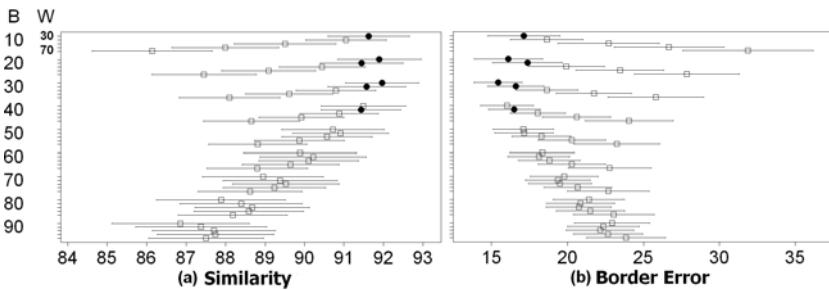


Fig. 4. Mean and 95% confidence interval of (a) similarity and (b) border error metrics for different values of B and W over the image set

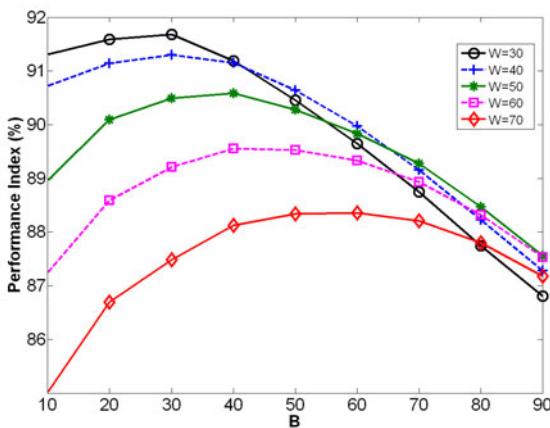


Fig. 5. Performance Index for different values of B and W over the image set

for different sizes of W and B over the image set of 55 dermoscopy images, using the mean value of the metrics. Third, for each W, a family of PIs versus B is plotted, as shown in Figure 5. Fourth, we choose the maximum PI point which reveals the optimal combination of W and B. As shown in Figure 5, the maximum value of PI is obtained for (W=30, B=30). This result is in accordance with the results we obtained from the statistical analysis above. We have chosen (W30, B30) to proceed with. However, there are other combinations for which PI is very similar, e.g. (W30, B20) or (W30, B10). As these combinations provide better sensitivity, they would also be reasonable options to be considered.

Cross Validation. In order to provide a stronger proof for the optimal B and W identified through the statistical analysis and the proposed performance index, we also perform a 11-fold cross validation process, where the data set of 55 images is iteratively partitioned into a 50-image subset for train set and a 5-image subset for test set, resulting in 11 sets with unique combinations of test and train data. For each of the test sets, a family of PI curves for different W and B is plotted, as illustrated in Figure 6, which shows that all train sets converge to

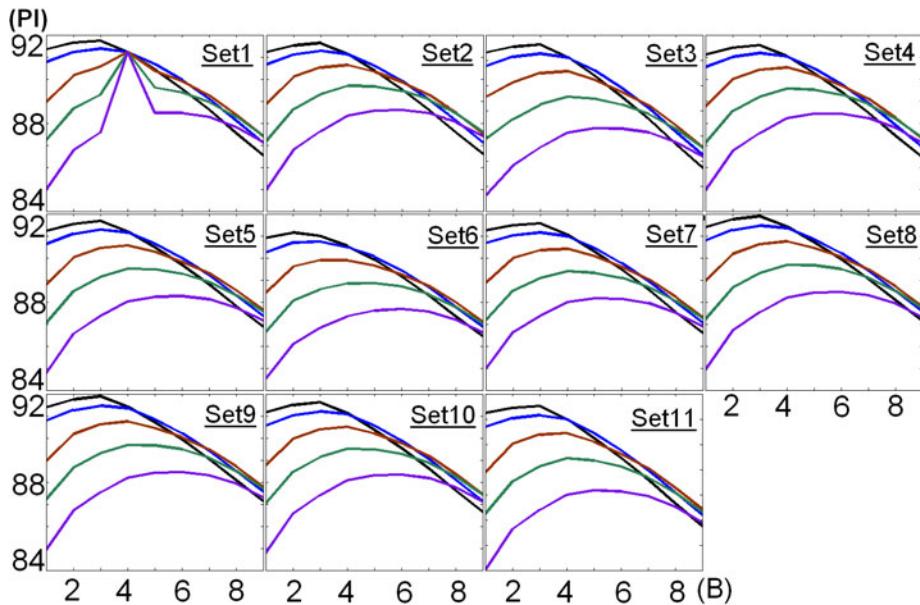


Fig. 6. 11-fold cross validation using performance index evaluation metric

Table 1. Performance Index for different images in each test set

TEST	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10	Set 11
Image 1	90.59	91.47	95.27	94.90	89.53	92.81	92.43	92.96	90.08	93.50	94.99
Image 2	89.62	93.80	94.41	92.58	91.47	90.90	90.53	95.35	96.48	89.52	94.44
Image 3	92.01	91.43	92.26	91.36	91.81	80.83	94.14	92.59	83.89	92.57	86.89
Image 4	91.41	92.58	87.99	95.28	89.37	85.82	85.06	94.19	94.00	90.91	94.35
Image 5	91.63	91.40	91.17	91.12	96.28	92.22	96.67	89.28	83.24	94.95	95.87
Mean	91.05	92.14	92.22	93.05	91.69	88.51	91.77	92.87	89.54	92.29	93.31
Std.	0.95	1.05	2.87	1.94	2.79	5.10	4.37	2.28	5.91	2.13	3.63

the value of (30,30) for B and W, except for set 6, where the optimal PI occurs when W=30 and B=20, which is almost similar to the PI value for (W=30, B=30). Table 1 shows the resultant PI for images and the corresponding mean and standard deviation for each test set using the optimal setting of (W=30, B=30), which demonstrates the acceptability of the identified setting.

3.3 Global versus Hybrid Thresholding

Figure 7 shows the automatic borders obtained by global and the hybrid approach and also the corresponding ground truth border drawn by dermatologists (the union of the three manual borders) for a sample lesion. It shows that the hybrid approach produces a border which is much closer to the ground truth,

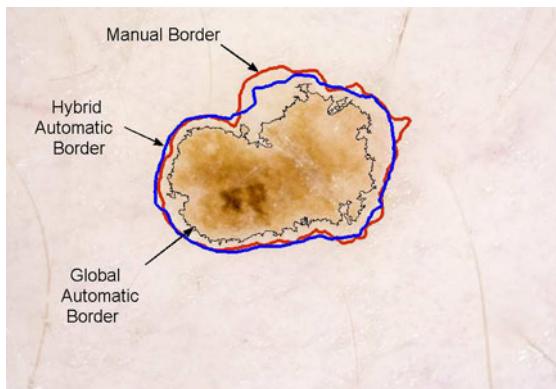


Fig. 7. The global and hybrid borders produced by the automated methods and the manual border drawn by dermatologists

Table 2. Evaluation metrics (mean \pm margin of error) for global and hybrid thresholding methods and the difference between them

	ACC	SNS	SPC	SML	BErr	PI
Global	90.90 ± 0.7	73.35 ± 2.6	99.68 ± 0.1	83.94 ± 1.8	27.24 ± 2.4	84.12 ± 1.5
Hybrid	94.85 ± 0.5	89.52 ± 1.9	97.44 ± 0.7	91.97 ± 0.9	15.4 ± 1.5	91.67 ± 0.9
Hybrid–Global	3.95 ± 0.7	16.17 ± 1.5	-2.23 ± 0.6	8.02 ± 1.4	11.84 ± 2.0	5.6 ± 1.0

compared to the border produced by global thresholding method. In order to objectively show the extent of improvement by evaluation metrics, when the lesion border is expanded from the core-lesion area (the result of global thresholding method) to the edge-lesion area (the result of hybrid thresholding method with the optimal values of $B=30$ and $W=30$), the mean difference between these two methods is calculated.

Table 2 gives the 95% confidence intervals for the mean for each metric, for global and hybrid ($W30$, $B30$) methods, and also the results of mean difference between these two approaches. In the Hybrid column, it is shown that the average values for each metric are clearly in an acceptable range. The mean specificity is 97.4; the 95% confidence interval is 96.9 to 98.1, and we can be confident that sensitivity will be high when the parameters ($W=30$, $B=30$) are used.

Table 2 also demonstrates that with respect to all metrics there is a significant improvement when the hybrid thresholding method is applied, with the exception of specificity. However, in the application of border detection in dermoscopy images, dermatologists are more concerned about not to leave out any part of the lesion, than erroneously including some part of the surrounding normal skin. Thus, in this particular application specificity is less important. Moreover, the mean level of specificity is very high in both approaches. As shown in Table 2 sensitivity, similarity and accuracy are increased by 16.17%, 8% and 3.95%,

respectively and border error is reduced by 11.84% which demonstrate the significant superiority of the hybrid approach. Performance Index is also elevated by 5.6% which proves the better performance of the hybrid thresholding method.

4 Conclusion

In this paper we demonstrate the superiority of the automated hybrid thresholding approach to border detection in dermoscopy images over the global thresholding method, through a newly introduced evaluation metric: Performance Index, which is composed of standard metrics of sensitivity, specificity, accuracy, similarity, and border error. Statistical analysis and optimization procedure are used and shown to be convergent. The effectiveness and validity of the approach is demonstrated by running an experiment on 55 high resolution dermoscopy images. The experimental results show clearly the significant advantages of the proposed hybrid method over the global thresholding approach. In addition to higher performance, other advantageous of the method include simplicity and low computational cost.

Acknowledgments

The assistance of Dr. Gayle Ross and Dr. Alana Tuxen in obtaining the manual borders and Mr. Michael Purves and Ms. Amanda Rebbechi in taking the dermoscopy images, all from Royal Melbourne Hospital, is gratefully acknowledged. This research is supported by NICTA Victoria Research Laboratory, Australia.

References

1. Australia skin cancer facts and figures, <http://www.cancer.org.au/> (accessed September 2009)
2. Perrinaud, A., Gaide, O., French, L.E., Saurat, J.H., Marghoob, A.A., Braun, R.P.: Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? *British Journal of Dermatology* 157, 926–933 (2007)
3. Celebi, M.E., Iyatomi, H., Schaefer, G., Stoecker, W.: Lesion border detection in dermoscopy images. *Computerized Medical Imaging and Graphics* 33, 148–153 (2009)
4. Iyatomi, H., Oka, H., Celebi, M.E., Hashimoto, M., Hagiwara, M., Tanaka, M., Ogawa, K.: An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Computerized Medical Imaging and Graphics* 32, 566–579 (2008)
5. Melli, R., Grana, C., Cucchiara, R.: Comparison of color clustering algorithms for segmentation of dermatological images. In: SPIE Medical Imaging, vol. 6144, pp. 3S1–3S9 (2006)
6. Celebi, M.E., Kingravi, H., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H., et al.: Border detection in dermoscopy images using statistical region merging. *Skin Research and Technology* 14, 347–353 (2008)

7. Zhou, H., Chen, M., Zou, L., Gass, R., Ferris, L., Drogowski, L., Rehg, J.M.: Spatially constrained segmentation of dermoscopy images. In: 5th IEEE International Symposium on Biomedical Imaging, pp. 800–803 (2008)
8. Garnavi, R., Aldeen, M., Celebi, M.E., Finch, S., Varigos, G.: Border detection in dermoscopy images using hybrid thresholding on optimized color channels. To appear in Computerized Medical Imaging and Graphics, Special Issue: Skin Cancer Imaging
9. Garnavi, R., Aldeen, M., Celebi, M.E., Bhuiyan, A., Dolianitis, C., Varigos, G.: Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels. International Journal of Medicine and Medical Sciences 1, 126–134 (2010)
10. Lee, T., Ng, V., Gallagher, R., et al.: Dullrazor: A software approach to hair removal from images. Computers in Biology and Medicine 27, 533–543 (1997)
11. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66 (1979)