Computerized Medical Imaging and Graphics 35 (2011) 105–115

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Border detection in dermoscopy images using hybrid thresholding on optimized color channels

Rahil Garnavi[a,*], Mohammad Aldeen[a], M. Emre Celebi[b], George Varigos[c], Sue Finch[d]

[a] Department of Electrical and Electronic Engineering, NICTA Victoria Research Laboratory, Universty of Melbourne, Parkville, Melbourne, Victoria 3010, Australia
[b] Department of Computer Science, Louisiana State University, Shreveport, LA, USA
[c] Department of Dermatology, Royal Melbourne Hospital, Australia
[d] Department of Mathematics and Statistics, Universty of Melbourne, Australia

ARTICLE INFO

Article history:
Received 30 September 2009
Received in revised form 16 July 2010
Accepted 7 August 2010

Keywords:
Computer-aided diagnosis
Melanoma
Dermoscopy
Border detection
Thresholding
Color space

ABSTRACT

Automated border detection is one of the most important steps in dermoscopy image analysis. Although numerous border detection methods have been developed, few studies have focused on determining the optimal color channels for border detection in dermoscopy images. This paper proposes an automatic border detection method which determines the optimal color channels and performs hybrid thresholding to detect the lesion borders. The color optimization process is tested on a set of 30 dermoscopy images with four sets of dermatologist-drawn borders used as the ground truth. The hybrid border detection method is tested on a set of 85 dermoscopy images with two sets of ground truth using various metrics including accuracy, precision, sensitivity, specificity, and border error. The proposed method, which is comprised of two stages, is designed to increase specificity in the first stage and sensitivity in the second stage. It is shown to be highly competitive with three state-of-the-art border detection methods and potentially faster, since it mainly involves scalar processing as opposed to vector processing performed in the other methods. Furthermore, it is shown that our method is as good as, and in some cases more effective than a dermatology registrar.

1. Introduction

Malignant melanoma is the deadliest form of skin cancer. The American Cancer Society estimates 68,720 new cases of melanoma in the United States in 2009, with 8650 estimated melanoma deaths (compared to 62,480 and 8420 cases in 2008, respectively) [1]. In Australia, melanoma represents 10% of all cancers and its incidence is four times higher than in Canada, the UK and the US, with more than 10,000 cases diagnosed and around 1250 deaths annually. If melanoma is detected in its early stages, it is highly curable, yet advanced melanoma is lethal. Furthermore, melanoma is more likely to metastasize when compared to other skin tumors [2,3].

Many efforts have been made in the last two decades to improve the clinical diagnosis of melanoma. These include alternative imaging technologies such as dermoscopy (a non-invasive in vivo imaging technique which allows for a magnified and clear visualization of the morphological structures of the skin that are not visible to the naked eye) and several diagnostic algorithms such as pattern analysis, ABCD rule of dermoscopy, Menzies method, 7-point checklist, and the CASH algorithm. A meta-analysis of studies conducted before 2001 shows that using these algorithms along with dermoscopy in expert hands improves the diagnosis of melanoma compared to simple naked-eye examination and traditional clinical examinations by 5–30% [4–7]. However, even with the use of dermoscopy technology and dermoscopic algorithms, clinical diagnosis is still challenging and its accuracy is considered to be limited, especially with equivocal lesions [8]. As opposed to visual assessment, computer-aided diagnosis of melanoma provides quantitative and objective evaluation of skin lesions. It allows for reproducible diagnosis by eliminating the inter-observer and intra-observer variabilities that are inevitably present in dermatologists' examinations.

A system for the computer-aided diagnosis of melanoma is generally comprised of four major components: skin image acquisition, lesion segmentation or border detection, feature extraction, and lesion classification. The accuracy of the segmentation process is extremely important due to the bias it can impose on the subsequent steps of the diagnosis system.

In this paper, a novel automatic border detection method based on color space analysis and clustering-based histogram thresholding is proposed. The method first determines the most effective and discriminative color channels. The lesion image then undergoes a

* Corresponding author. Tel.: +61434862608.
E-mail addresses: r.garnavi@ee.unimelb.edu.au (R. Garnavi), aldeen@unimelb.edu.au (M. Aldeen), ecelebi@lsus.edu (M.E. Celebi), George.varigos@mh.org.au (G. Varigos), sfinch@unimelb.edu.au (S. Finch).

0895-6111/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2010.08.001

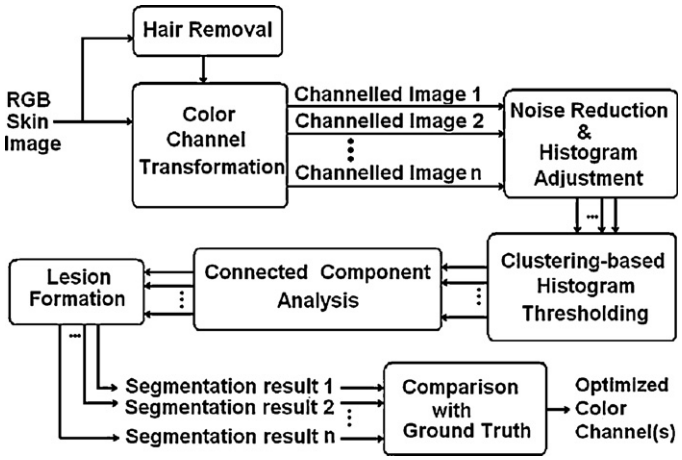

Fig. 1. Color optimization process.

hybrid clustering-based histogram thresholding procedure. An initial border is first determined using global thresholding followed by a set of pixel-based computations and morphological operations, and the boundary is then expanded towards the background skin based on local threshold values.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 describes the process of color channel optimization. Section 4 explains the hybrid histogram thresholding algorithm. Experimental results are discussed in Section 5. Finally, Section 6 gives the conclusion.

2. Related work

Various image features such as shape, color, texture, and luminance have been employed to perform skin lesion segmentation. Numerous border detection methods have been proposed in the literature [9] including histogram thresholding [10,11], color clustering [12–14], JSEG algorithm based on color quantization and spatial segmentation [15], statistical region merging [16], two-stage k-means++ clustering followed by region merging [17], etc. Melli et al. [12] criticized the adaptive thresholding methods for not providing accurate segmentation result due to the problems associated with color calibration and lack of sharp bimodal luminance distribution between the surrounding skin and the lesion. Unsupervised approaches, despite their lower performance compared to supervised trained systems, have been claimed to exhibit higher robustness. In this study, we decided to use a thresholding-based approach for border detection. The rationale behind this choice is the simplicity and low computational cost of scalar processing as described in the following sections.

3. Color channel optimization

Although numerous border detection methods have been developed, few studies have focused on determining the most effective color channels for dermoscopy images. This section presents an automatic border detection method based on color space analysis and clustering-based histogram thresholding. Color channels from various color spaces were investigated to maximize the discrimination between two clusters of pixels within the image, i.e. lesion and background skin pixels. Each color channel undergoes several steps as shown in Fig. 1. These steps will be explained in the following sections. By examining 25 different channel images for each original RGB dermoscopy image, we have determined the color channels that outperform the others in accurately detecting the lesion borders.

Table 1
Color channels used in color space transformation.

	Color channel(s)	Color space
1–12	R, G, B, RGB, RoB, GoB, RoG, RoGoB, RGBoR, RGBoG, RGBoB, RGBoRoGoB	RGB
13	I	HSI
14	V	HSV
15	L	LAB
16	Y	YCbCr
17–23	X, Y, Z, XoY, XoZ, YoZ, XoYoZ	XYZ
24–25	XoYoR, XoYoZoR	XYZ and RGB

3.1. Hair removal

Lesions occluded with dark thick hairs can cause problems in the segmentation process. In such cases, the proposed method starts with a hair removal procedure, which involves the following steps [18]: (1) localizing dark hairs, using a morphological closing operation in the vertical, horizontal and diagonal directions, (2) interpolating the removed hair pixels using nearby non-hair pixels, and (3) smoothing the final result using a median filter to eliminate the remaining thin lines.

3.2. Color space transformation

Color information plays a significant role in dermoscopy image analysis. This step incorporates color information in the dermoscopy image into the segmentation process, where the original RGB image is transformed into various color spaces, and the corresponding color channels are extracted. Despite the existence of numerous color spaces in the literature, the choice of the optimal color space is application-dependent [19]. In this study we investigated the following color spaces [20]: RGB, HSV, HSI, CIE-XYZ, CIE-LAB, and YCbCr. The original RGB dermoscopy image is transformed into a set of 25 color channel images from the above-mentioned color spaces. The selection of these 25 color channels was made after analysing, visually, all possible combinations. As shown in Table 1, these images include single color channel images such as R from the RGB color space and X from the XYZ color space, as well as combinations of them such as XoYoR, where "o" stands for logical OR, which combines the X and Y color channels from the XYZ color space with the R color channel from the RGB color space. For instance, to build the RoGoB color channel the R, G and B color channels are extracted and each undergoes the segmentation process illustrated in Fig. 1. Finally, the three binary segmentation results are combined through the OR operation. The RGB color channel in Table 1 is calculated by forming a weighted sum of the R, G, and B components and converting RGB values to grayscale values.

3.3. Noise filtering

To enhance the accuracy of segmentation and save computational time, it is useful to eliminate the artifacts that might be present in the image. In dermoscopy images, external artifacts include skin lines, air bubbles or other random noise caused by the imaging process. In this study, the dermoscopy images are smoothed using a circular averaging low-pass filter with a radius of 5, using the pillbox point spread function.

3.4. Intensity adjustment

This step is essentially an enhancement process in which the dynamic range of pixel values in the image is mapped into a new range. The purpose is to smooth and stretch the image histogram and increase the contrast of the image in order to determine a more

accurate threshold value in the thresholding step. Intensity adjustment works by rescaling the intensity values in the original image to cover the entire dynamic range.

3.5. Clustering-based histogram thresholding

Thresholding is the process of classifying the pixels of a grayscale image into two classes, so that the image can be converted to binary by assigning each pixel either a 0 or 1, depending on its gray level. The thresholding procedure used in this paper is based on the well-known Otsu's thresholding method [21]. The basic premise of this method is the assumption that an image contains two clusters of pixels, e.g. foreground and background, which, in our case, correspond to the lesion and its surrounding skin, respectively. To identify these two clusters accurately, an algorithm is used to search for an optimal threshold level using discriminant analysis, where zero-th and first-order cumulative moments of the histogram are calculated and used to define a measure of separability between the two clusters. An optimal threshold level separating the two clusters is obtained by minimizing the within-cluster variance (σ_ω^2), which is defined as a weighted sum of variances of the two clusters:

$$\sigma_\omega^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (1)$$

where weights ω_i are the probabilities of the two clusters separated by a threshold t and σ_i^2 denotes the variances of these clusters. It can be shown that minimizing the within-cluster variance (σ_ω^2) is equivalent to maximizing between-cluster variance. The between-cluster variance (σ_b^2) is recalculated in the Otsu method [21] as the following:

$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \quad (2)$$

where μ_i are the mean values of the two clusters. Starting from an initial threshold value of $t = 1$, ω_i and σ_i^2 are updated iteratively and in each iteration $\sigma_b^2(t)$ is calculated. The optimal threshold corresponds to the maximum value of $\sigma_b^2(t)$. The output binary image has values of 1 for all pixels in the input image with luminance greater than the threshold level and 0 for the remaining pixels.

3.6. Connected component analysis

In some of the dermoscopy images, extra objects appear in the surrounding skin area, such as blue marks made by dermatologists when examining the patient's skin. These objects, which have not been eliminated in the noise removal step, appear with intensity values similar to that of the lesion and may be misclassified as lesion. The purpose of this step is to exclude these objects from the segmentation output. To this end, the number of connected objects within the image is determined using the run-length encoding technique [22] and the connected objects are labeled. Finally, the two largest areas (i.e. lesion and surrounding skin) are kept and all other components are discarded.

3.7. Lesion formation

In order to obtain the final lesion object, the holes inside the boundary are filled by performing morphological filling on the binary image. Fig. 2 illustrates the entire segmentation procedure on a sample image. Fig. 3 shows more border detection results.

3.8. Determining the optimal color channel

In order to determine the optimal color channels for border detection, we conducted a pilot study on a set of 30 high resolution dermoscopy images that were provided to us by the Royal Melbourne Hospital. As shown in Fig. 3, the image set contains a variety of dermoscopy images in terms of color, texture, and shape. As the ground truth, for each dermoscopy image, four manual borders were independently drawn by four experts (two experienced dermatologists and two dermatology registrars[1]). Fig. 4 shows the borders drawn by dermatologists for a sample image. To perform a comparison between the automatic borders (generated by the proposed border detection method) and the manual borders (drawn by the dermatologists) four different metrics are used; namely, sensitivity (Eq. (3)), specificity (Eq. (4)), accuracy (Eq. (5)) and similarity (Eq. (6)).

Table 2
Mean ± standard deviation for similarity (%) between dermatologists: E1 and E2, experienced dermatologists; R1 and R2, dermatology registrars.

	E1	E2	R1	R2
E1	–	95.92 ± 1.76	95.09 ± 1.75	94.78 ± 2.25
E2	–	–	94.68 ±1.73	94.78 ± 2.22
R1	–	–	–	95.86 ± 1.25

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

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

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

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

where TP, TN, FP, and FN refer to true positive, true negative, false positive and false negative, respectively. TP and TN represent the number of pixels which are classified correctly as part of the lesion and background skin in both the manual and automatic borders, respectively. FP represents the number of pixels which are classified as part of the lesion in the automatic border, but are labeled as part of the background skin in the manual border. Finally, FN represents the number of pixels which are classified as part of the background skin in the automatic border, but are labeled as part of the lesion in the manual border.

In the first three metrics, the manual borders are taken as the ground truth and the calculated measures are used to quantify the discrepancy between the automatic borders and the manual ones. However, another perspective in the analysis is to evaluate the inter-observer variabilities among the four dermatologists, as demonstrated in Fig. 4, and also to investigate which automatic border is closer to each manual one. To achieve this, a fourth metric, i.e. (Sorensen) similarity index [23], is used to quantify the degree of similarity between any two borders, without taking either of them as the ground truth. This metric (Eq. (6)) has been used in a variety of domains, yet it has not been used in dermoscopy. Table 2 shows the mean and standard deviation values of similarity among the borders drawn by the four dermatologists. The similarity values indicate that there are higher similarities between the two experienced dermatologists and two dermatology registrars, although differences are all less than the standard deviations. The overall similarity index is high which indicates the reliability of the ground truth.

As mentioned earlier, the main purpose of this analysis is to determine the color channels that lead to the most accurate borders. To achieve this aim, 25 different color channels listed in Table 1

[1] A Specialist Registrar or SpR is a doctor in the United Kingdom, Republic of Ireland, and Australia who is receiving advanced training in a specialist field of medicine in order eventually to become a consultant. In surgery, it is also referred to as Higher Surgical Trainee or HST.

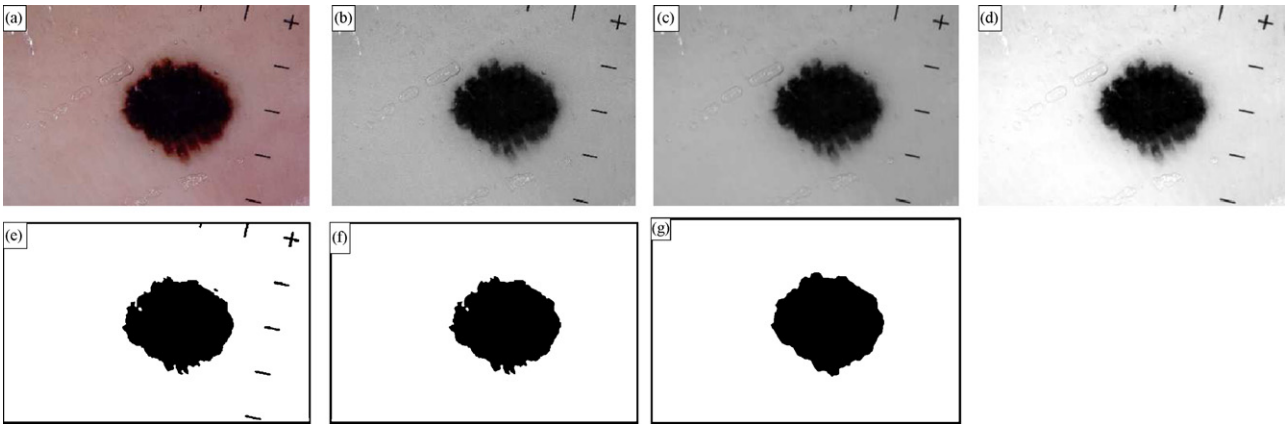

Fig. 2. Segmentation method: (a) original image, (b) color space transformation, (c) noise removal, (d) intensity adjustment, (e) thresholding, (f) connected component analysis, and (g) morphological filling.

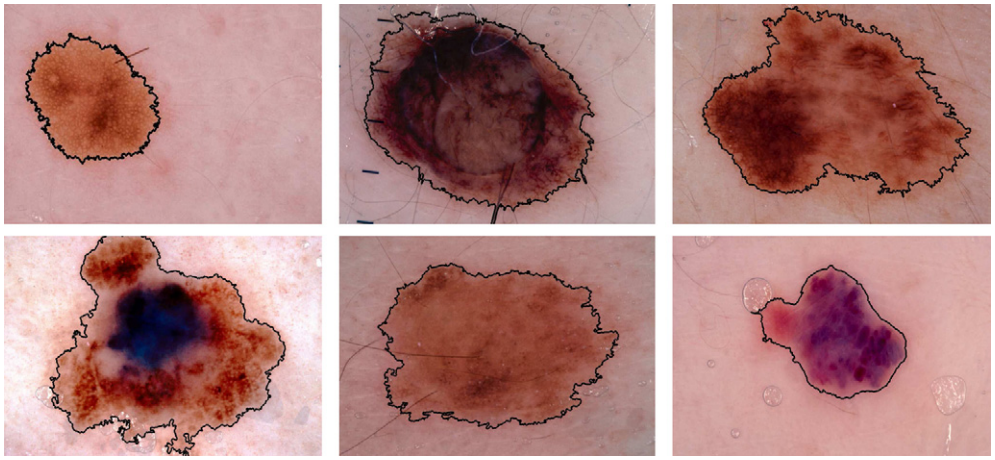

Fig. 3. Sample segmentation results.

Table 3
Color channels with the largest mean for four ground truths: E1 and E2, experienced dermatologists; R1 and R2, dermatology registrars.

	Accuracy	Similarity	Sensitivity	Specificity
E1	XoYoR (96.03)	X (91.45)	XoYoZoR (93.52)	R (99.96)
E2	XoYoR (96.01)	X (91.55)	XoYoZoR (94.33)	R (99.92)
R1	XoYoR (95.14)	XoYoZoR (90.61)	XoYoZoR (90.60)	R (99.99)
R2	XoYoR (94.89)	XoYoZoR (90.56)	XoYoZoR (90.37)	R (99.99)

were extracted from the 30 images, resulting in 25 different borders per image. Each border (out of 25×30), is then separately compared to each of the four manual borders using the abovementioned four metrics. The 30 values of each metric are then averaged, resulting in 25 average measures of sensitivity, specificity, accuracy and similarity. The maximum values of each of these metrics and their corresponding color channels are then identified, resulting in four color channels for each metric, as shown in Table 3. As shown in Table 3, using the average values, the initial 25 color channels are narrowed down to the four color channels of X, XoYoR, XoYoZoR and R. Borders obtained using the R color channel are the smallest with low FP value, thus result in the highest specificity. Borders obtained using XoYoRoZ color channel, which is the result of integrating the X, Y, Z, and R color channels via OR operation, produce a larger border, thus small FN value. Consequently, they result in high sensitivity.

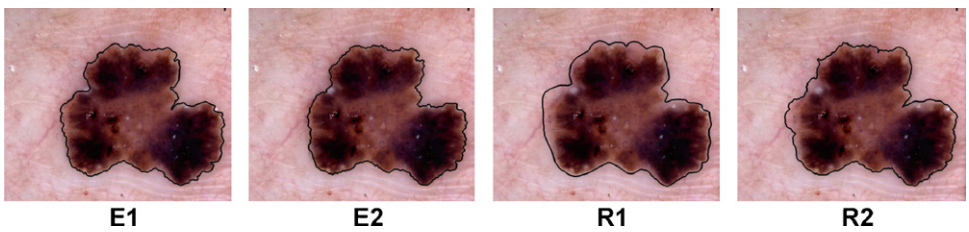

Fig. 4. Manual borders of the same lesion drawn by four dermatologists: E1 and E2, experienced dermatologists; R1 and R2, dermatology registrars.

Table 4
Segmentation results (mean ± margin of error) for optimal color channels, the common manual borders as ground truth.

	Accuracy	Similarity	Sensitivity	Specificity	AUC
X	96.80 ± 0.01	93.18 ± 1.82	90.49 ± 0.03	99.02 ± 0.00	99.8
XoYoR	96.94 ± 0.01	92.89 ± 3.15	92.86 ± 0.02	98.51 ± 0.00	99.8
XoYoZoR	95.15 ± 0.03	91.63 ± 5.15	96.80 ± 0.01	94.47 ± 0.05	99.5
R	92.42 ± 0.03	82.91 ± 5.45	73.40 ± 0.07	99.90 ± 0.00	99.1
JSEG	96.70 ± 0.01	93.23 ± 3.14	92.05 ± 0.04	98.35 ± 0.01	99.7

The results presented in Table 3 were calculated using each manual border as the ground truth, hence there are four sets of results. Another useful analysis is obtained when a single ground truth for each image is used. This ground truth is obtained by finding the intersection of the areas within the four manual borders (in the ensuing, we will refer to these as *common manual borders*). Then the common manual border is used as the ground truth and applied the four color channels shown in Table 3 to the image set. The result of the four metrics are shown in Table 4. Here, we report means and 95% confidence intervals. An observed sample mean provides the best estimate of the true population value; the confidence interval describes a range of values for the true population mean that are consistent with the data that was observed. We report these results as $\bar{x} \pm \text{me}$, where $\bar{x}$ is the sample mean and me is the margin of error; a 95% confidence interval ranges from $\bar{x} - \text{me}$ to $\bar{x} + \text{me}$ [24].

To precisely analyze the sensitivity and specificity metrics, the AUC (Area Under ROC Curve[2]) was obtained by drawing the ROC graph and calculating the corresponding AUC value for R, X, XoYoZ and XoYoRoZ color channels. Table 4 suggests that in terms of the accuracy metric and AUC values, color channels X and XoYoR provide the highest scores. With respect to similarity, X and XoYoR color channels give the highest average. The proposed method is also compared to a recent border detection method; namely JSEG [15]. The last row in Table 4 shows the evaluation results obtained by the JSEG method. The results demonstrate that the proposed method, in spite of its simplicity, with a proper choice of color channels is highly competitive with the well-known JSEG method.

4. Hybrid histogram thresholding

Our experience with dermatologists has consistently shown that manual borders tend to be larger than the corresponding automatic borders. This fact has also been reported in several studies [9]. It is also illustrated in the dermoscopy image shown in Fig. 5, where three areas are identified; namely, core-lesion, edge-lesion and background skin. The width of the edge-lesion area can be quite variable depending on the skin color, lesion color, sharpness or fuzziness of the border, and the lesion pattern. Automated 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 and thus exclude it from the segmentation result. In contrast, experienced dermatologists prefer to choose the outmost detectable pigment to minimize the risk of incorrect diagnosis.

In several studies the discrepancy between the manual and automatic borders have been reduced by expanding the automatic borders using various methods including morphological filtering [16], Euclidean distance transform [16], iterative region growing [25], and gradient information [26]. In this study we introduce a new hybrid histogram thresholding approach and show its remarkable effectiveness when used on dermoscopy images. The proposed hybrid approach is comprised of two stages; in the first stage, we identify the core-lesion using Otsu's thresholding method. In the second stage, the edge-lesion area is identified by applying a local clustering-based histogram thresholding method on the optimal color channels determined earlier. These two stages are detailed next.

4.1. Global thresholding: forming the core-lesion

The core-lesion area is detected by applying a clustering-based histogram thresholding method on the optimal color channels of XoYoR obtained in the color optimization procedure (Section 3, Fig. 1). This step includes the pre-processing operations of hair removal, noise filtering and intensity adjustment. This is followed by application of Otsu's histogram thresholding algorithm to the XoYoR color channel, and performing connected component analysis and morphological operations to obtain the initial border and form the so-called core-lesion area. The histogram analysis procedure is quite fast, since it performs one-dimensional clustering and the results of each X and Y and R color channel are integrated using the logical OR operation.

In this step we have made a new choice for the noise removal filter. In the pilot study, the averaging pillbox filter was used, whereas we have observed that using a Gaussian low-pass filter [27] results in less blurred images and retains more of the important details which can improve the border detection accuracy. The Gaussian filter utilized a 11×11 kernel with a standard deviation (σ) value of 0.5.

4.2. Local thresholding: forming the edge-lesion

To expand or shrink the core-lesion boundary to the edge-lesion boundary, we have applied a local clustering-based thresholding technique based on Otsu's method described in Section 3.5. The histogram thresholding is applied to the X color channel which was determined as optimal (with respect to the common manual borders as the ground truth) in the color optimization procedure (Section 3), after performing noise removal (using a Gaussian filter

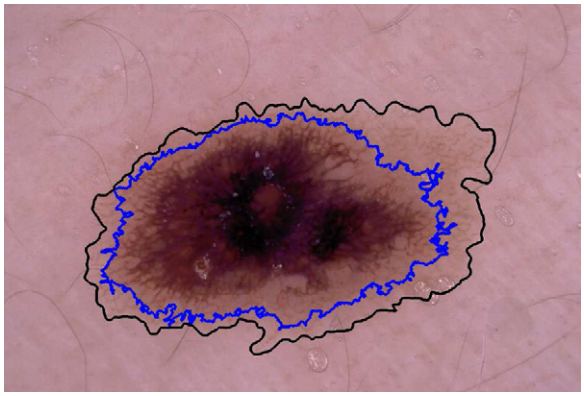

Fig. 5. Three different area appears in dermoscopy images: core-lesion, edge-lesion and background skin.

[2] Receiver operating characteristic (or ROC) curve is a graphical plot of the sensitivity versus (1-specificity).

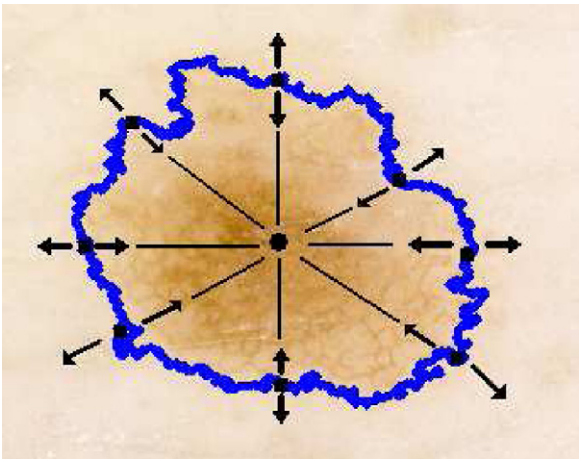

Fig. 6. Shrinkage and expansion is done radially with respect to the centroid of the lesion.

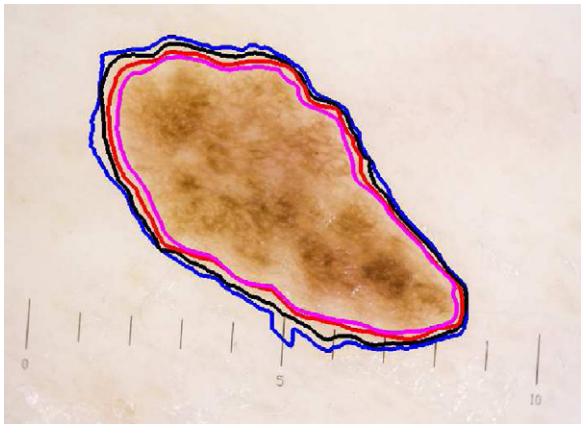

Fig. 8. Different borders produced by different values of W and B parametrs.

with similar parameter settings as above) and intensity adjustment. Starting from an arbitrary point on the initial core-lesion boundary, the local threshold is calculated over a window. Let us define Ws to be a window of size $s \times s$, which is a square window centered on a core-lesion border pixel. If the local threshold value is less than a predefined threshold called T_{expand} the boundary is expanded by one pixel. If it is greater than a predefined threshold called T_{shrink} the boundary is shrunk. Otherwise, we are in the state of *No Change*, which implies that based on the previous moves, we should make a decision to either laterally move to the adjacent pixel, or to expand or shrink the boundary. As shown in Fig. 6 the inward and outward moves (shrinkage and expansion) are done radially (along the line connecting the centroid of the lesion to the pixel on the core-lesion boundary). The lesion centroid is calculated using Eq. (7).

$$(x_c, y_c) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right) \tag{7}$$

where n is the number of pixels inside the lesion, and (x_i, y_i) is the coordinates of the i-th lesion pixel.

To define the threshold values for shrinkage and expansion, a bandwidth is calculated based on background skin value and core-lesion value. We have applied the same thresholding method to the background skin area and core-lesion area to obtain estimates of these values. The bandwidth is calculated by Eq. (8). Both core-lesion and background skin are taken into account when calculating the bandwidth. In this way the Bandwidth is normalized, so that variations in the color of different lesions on the same person can be accounted for. Without this normalization process the lesion border will only expand regardless of the difference between the background skin color and the core-lesion color. The threshold values for expansion and shrinkage are given by Eqs. (9) and (10), respectively.

$$\text{Bandwidth} = \%B \times (T_{\text{skin}} - T_{\text{core-lesion}}) \tag{8}$$

where B, the bandwidth factor, is the percentage we wish to expand from the core-lesion towards the background skin. T_{skin} and $T_{\text{core-lesion}}$ refer to estimates of the background skin and core-lesion

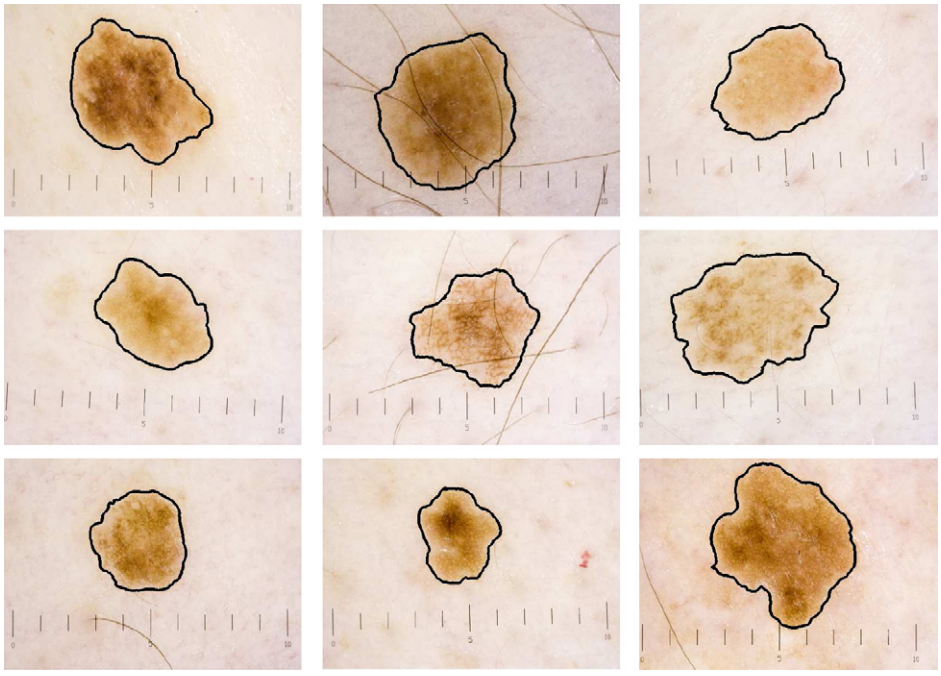

Fig. 7. Sample border detection results using hybrid thresholding on optimized color channels.

Table 5
Comparing automated methods in terms of accuracy, precision, sensitivity, specificity (mean ± margin of error) and AUC, using the experienced dermatologist's manual borders as the ground truth.

	Accuracy	Precision	Sensitivity	Specificity	AUC
W30B60	98.01 ± 0.35	93.91 ± 1.43	89.64 ± 1.90	99.43 ± 0.13	99.86
JSEG	98.11 ± 0.35	95.39 ± 1.36	88.57 ± 2.46	99.40 ± 0.20	99.78
DTEA	97.63 ± 0.49	96.92 ± 0.77	83.84 ± 2.68	99.70 ± 0.07	99.80
KPP	97.52 ± 0.38	97.10 ± 1.32	80.95 ± 2.97	99.63 ± 0.16	99.68

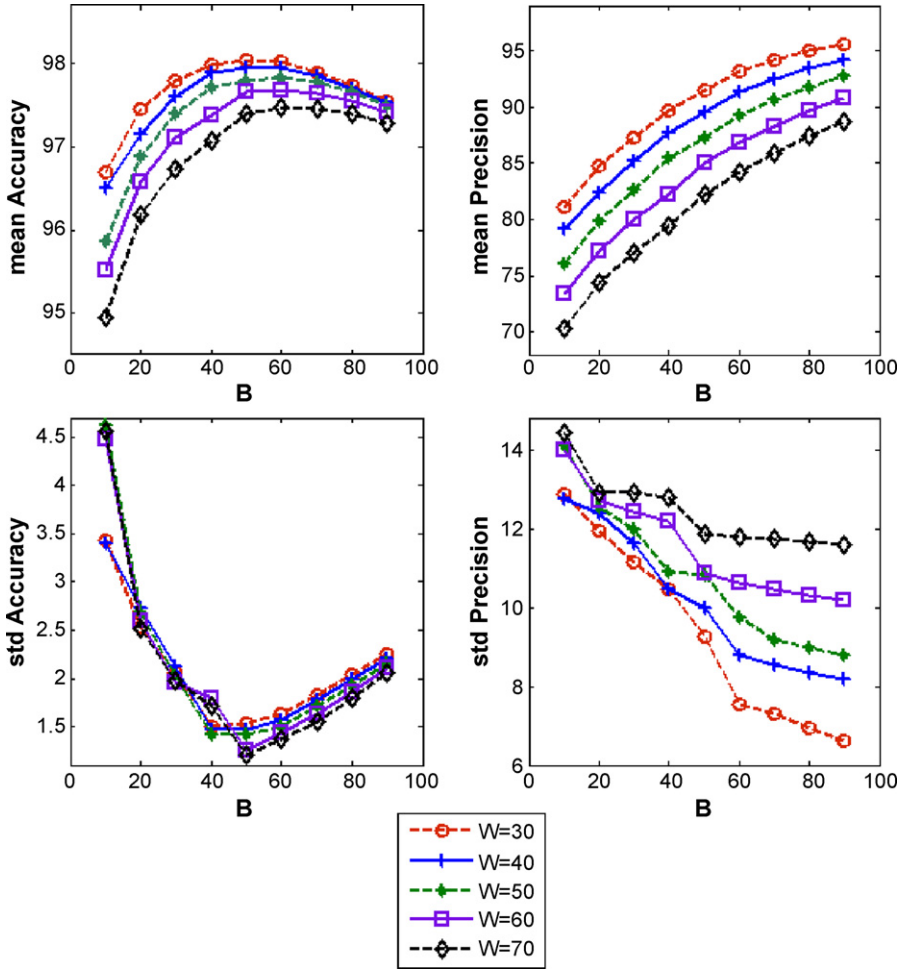

Fig. 9. Mean and standard deviation values for accuracy and precision metrcis for different B and W over the image set compared to experienced dermatologist.

area, respectively.

$$T_{\text{expand}} = T_{\text{skin}} - \text{Bandwidth} \tag{9}$$

$$T_{\text{shrink}} = T_{\text{skin}} + \text{Bandwidth} \tag{10}$$

$$\begin{cases} \text{Expand (move outward)} & \text{if } T_{\text{local}} \le T_{\text{expand}} \\ \text{Shrink (move inward)} & \text{if } T_{\text{local}} \ge T_{\text{shrink}} \\ \text{No Change} & \text{if } T_{\text{expand}} < T_{\text{local}} < T_{\text{shrink}} \end{cases} \tag{11}$$

The local threshold is calculated for each and every window centered on successive pixels along the core-lesion boundary and the process is stopped when a pixel is revisited. Fig. 7 shows samples of the detected borders by the proposed method.

5. Experimental results

The proposed border detection method is tested on a set of 85 high resolution dermoscopy images[3] obtained from Melbourne Royal Hospital. The images were taken by professional photographers using a Canon EOS 450D camera under unified zooming and lighting conditions. They are 24-bit RGB color images with dimensions of 2000 × 1334 pixels. Manual borders were independently drawn by an experienced dermatologist and a first year dermatology registrar, using a Wacom Intuos A4 size Tablet PC. These borders were then used as ground truths for the evaluation of the automatic borders.

[3] The authors would like to have used a larger image set to carry out this experiment but at the time this research was conducted only a set of 85 was available. The use of smaller than desired image set should therefore be taken into account by the readers.

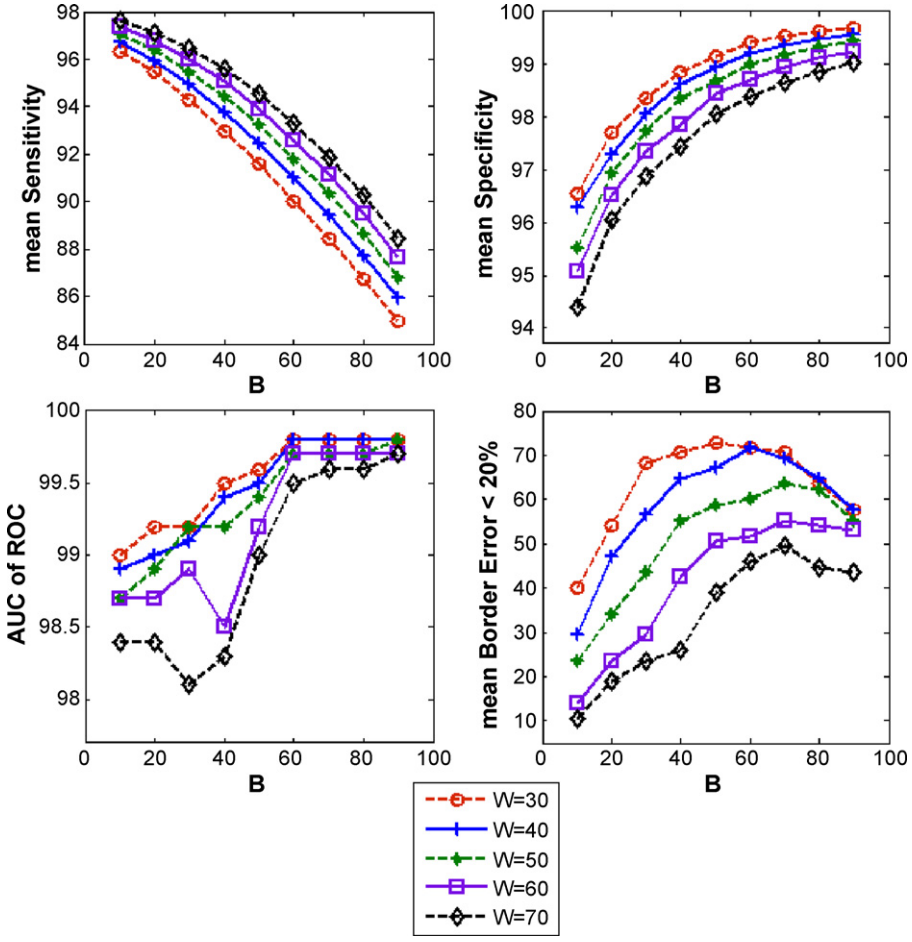

Fig. 10. Mean sensitivity and specificity, the corresponding AUC, the proportion of images with border error ≤ 20%, for different B and W over the image set, compared to experienced dermatologist.

5.1. Optimizing B and W parameters

As mentioned in Section 4.2, two main parameters are involved in the proposed method: window size (W) for calculating the local threshold, and the bandwidth factor (B). The extent to which the core-lesion area is expanded towards the background skin depends on the latter parameter, namely B. A comprehensive experiment is performed on the set of 85 dermoscopy images, with W ranging from 30 to 70 and B ranging from 10% to 90% (steps of 10). Initially, windows of sizes from 10×10 to 90×90 were investigated and the result showed that those below 30×30 and above 70×70 were insignificant, therefore not included in the analysis. As shown in Fig. 8 different borders can be obtained 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 determined by the experienced dermatologist. In addition, each automatic border is compared to the ground truth determined by the dermatology registrar. To quantitatively compare the automatic borders to the manual borders, different metrics have been utilized in the literature [28]. In this work, the comparisons are done in terms of five evaluation metrics of sensitivity, specificity, accuracy, precision and border error. Sensitivity (Eq. (3)) shows the percentage of actual lesion that has been truly detected. Specificity (Eq. (4)) shows the percentage of actual background skin that has been truly detected. Accuracy (Eq. (5)) shows the degree of closeness of the detected border to the actual border, which takes both background skin and lesion pixels into account. Precision or positive predictive value (Eq. (12)) is an index of reproducibility which shows the degree to which further experiments show the same results. Finally, border error or XOR measure (Eq. (13)) measures the discrepancy between automatic and manual borders.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{12}$$

$$\text{Border Error} = \frac{\text{Area}(A \otimes M)}{\text{Area}(M)} \times 100\% \tag{13}$$

where M is the the manually segmented image, $A \otimes M$ represents the differential segmentation obtained by the manual and automatic borders using XOR operation, and $Area(.)$ denotes the number of pixels in the binary images of M and $A \otimes M$.

5.2. Comparison with experienced dermatologist (first ground truth)

Fig. 9 shows the mean and standard deviation (std) values for accuracy and precision metrics for different B and W parameters over the image set, using the experienced dermatologist's borders as the ground truth. As illustrated in the graphs, mean accuracy and precision increase with decreasing window size. The variability around mean (measured by standard deviation) is smaller for smaller W ($W = 30$) in most of the cases. With respect to the accuracy metric, both mean and std graphs suggest that the bandwidth factor (B) is optimal in the medium range of [40..60]%. Precision shows an incremental trend in mean value, along with a decreasing trend of std as B grows, which suggests a value of larger than 50% for B may be optimal.

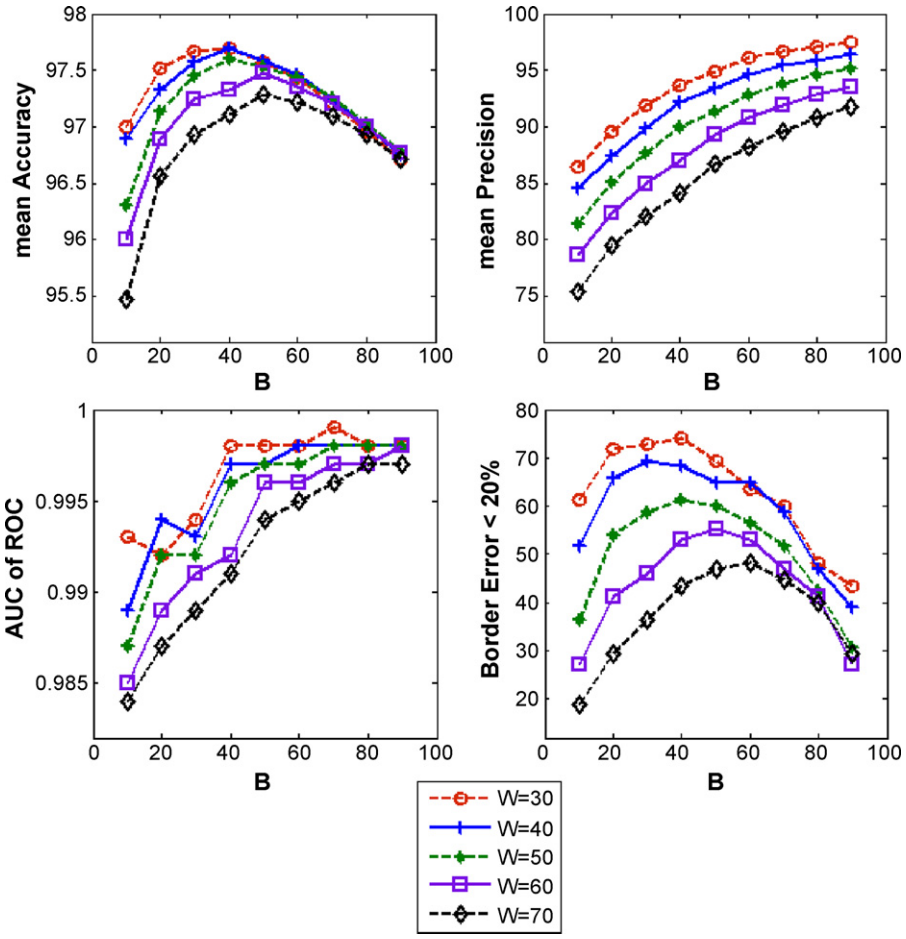

Fig. 11. Mean accuracy and precision, AUC, proportion of images with border error ≤ 20%, for different *B* and *W* over the image set, compared to dermatology registrar.

Fig. 10 shows the mean value for sensitivity and specificity metrics, along with the corresponding AUC value, and also the proportion of images with border error smaller than 20%, for different B and W parameters over the image set, taking the experienced dermatologist's borders as the ground truth. The graphs indicate that the highest AUC value is obtained by a smaller *W* and larger *B* which is in agreement with the results obtained from the accuracy and precision metrics. Overall, the value of 30 for *W* and 60% for *B* seems to be a reasonable trade-off considering all the metrics. Table 5 gives the mean and margin of error for each metric for the proposed method with optimized parameters (W30B60). High mean values (over 93%) were observed for all metrics, except for sensitivity which is close to 90%.

5.3. Comparison with dermatology registrar (second ground truth)

Fig. 11 shows the mean value for accuracy and precision metrics, the AUC calculated from sensitivity and specificity metrics and also the proportion of images with border error smaller than 20%, for different *B* and *W* parameters over the image set, using the dermatology registrar's borders as the ground truth. The trend is similar to that observed in Figs. 9 and 10 with the experienced dermatologist as the ground truth. Large *W* values are ruled out, and similarly, *W* = 30 appears to be optimal across the range examined. However, for accuracy and border error smaller *B* values tend to obtain better results than large *B* values. Considering precision and AUC it can be seen that an intermediate value for *B* in the range of [30..50]% is more likely to produce results that are consistent with the dermatology registrar.

Table 6
Comparing different border detection methods: distribution of border error percentage, using the experienced dermatologist's manual borders as the ground truth.

Border error	≤10	≤15	≤20	≤30	≤40
W30B60	16.25	55.00	72.50	92.50	98.75
JSEG	27.50	51.25	73.75	90.00	95.00
DTEA	7.50	47.50	67.50	88.75	93.75
KPP	13.75	28.75	50.00	78.75	88.75

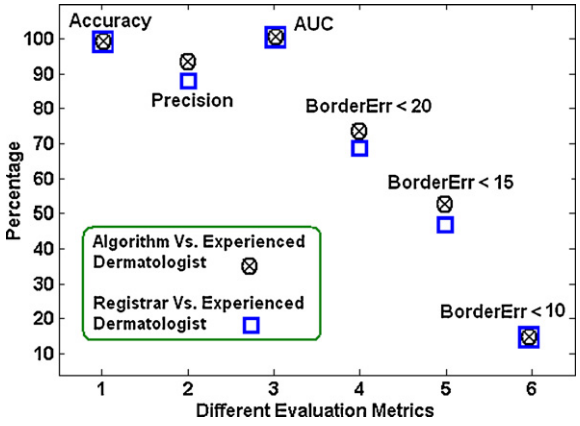

Fig. 12. Comparison between the border detected by our method to manual border drawn by dermatology registrar, using the experienced dermatologist's manual borders as the ground truth.

Table 7
Mean ± margin of error for comparisons of automated methods, using the experienced dermatologist's manual borders as the ground truth.

	Accuracy	Precision	Sensitivity	Specificity	Border error
W30B60–JSEG	−0.09 ± 0.23	−1.47 ± 1.37	1.07 ± 1.95	0.03 ± 0.17	−0.5 ± 1.9
W30B60–DTEA	0.37 ± 0.18	−3.00 ± 1.17	5.80 ± 1.36	−0.26 ± 0.09	2.1 ± 1.7
W30B60–KPP	0.49 ± 0.24	−3.18 ± 1.63	8.69 ± 2.78	−0.19 ± 0.16	5.3 ± 2.3

Furthermore, we have taken the borders drawn by the experienced dermatologist as the ground truth and compared the results of our method (W30B60) to those of the dermatology registrar's. Fig. 12 shows a comparison between our results and dermatology registrar's with respect to different evaluation metrics. From the figure it can be seen that our method yields either equivalent or closer results to those of the experienced dermatologist's than the dermatology registrar's to the experienced dermatologist's. For example, taking the experienced dermatologist as ground truth, the mean precision for our algorithm is 5.3% higher than for the dermatology registrar. The margin of error was 1.9, giving a 95% confidence interval of 3.4–7.2%. We can be reasonably confident that our method is, on average, between 3.4% and 7.2% better than the dermatology registrar. This implies that our method is as good as, and in some cases more effective than the dermatology registrar.

5.4. Comparison with other automated methods

The proposed method with optimized parameters (W30B60) is compared with three state-of-the-art lesion border detection methods; namely, JSEG [15], DTEA [25] and KPP [17]. Tables 5 and 6 show the segmentation evaluation results obtained by JSEG, DTEA, KPP and the proposed thresholding method using XoYoR color channel, in forming core-lesion boundary, and X color channel, in forming edge-lesion boundary, with window size of 30 and bandwidth factor of 60, considering the borders drawn by the experienced dermatologist as the ground truth. Table 5 shows that the proposed method performs better on average in terms of sensitivity and AUC metrics, whereas it is highly competitive with the JSEG method with respect to the accuracy and specificity. With respect to the accuracy and sensitivity, DTEA comes third, followed by KPP and they also yield fairly competitive results in terms of specificity and AUC. However, KPP exhibits the highest precision, followed by DTEA, JSEG and W30B60, respectively. Table 6 shows the number of images with percentage border error less than 10%, 15%, 20%, 30% and 40% for different automated methods, which reflects the same trend as the similarity metric. Table 6 indicates that the proposed method achieves the best scores among the automated methods, while JSEG follows it closely, and DTEA and KPP come next. Table 7 provides estimates and margin of error for the mean differences between W30B60 and the alternative methods. The mean differences between W30B60 and JSEG are generally small and are unlikely to be important. The mean differences between (W30B60, DTEA) and (W30B60, KPP) for sensitivity and border error are relatively large, supporting that W30B60 is performing better. Moreover, the proposed method is potentially faster since it mainly involves scalar processing as opposed to vector processing.

6. Conclusion

This paper presented a novel automatic border detection method based on color space analysis and clustering-based histogram thresholding. The method determines the optimal color channel and applies hybrid thresholding followed by morphological operations to detect the lesion borders. The color optimization process is tested on a set of 30 dermoscopy images, with four sets of dermatologist-drawn borders taken as the ground truth. The hybrid border detection method is tested on a set of 85 high resolution dermoscopy images and the automatic borders are compared to borders drawn by two dermatologists using various evaluation metrics including accuracy, precision, sensitivity, specificity, and border error. The proposed method, which is comprised of two stages, is designed to increase specificity in the first stage and sensitivity in the second stage. The proposed hybrid method is also compared to three state-of-the-art border detection methods and found to perform as well or better in terms of sensitivity, AUC, and border error. In addition, our method is potentially faster since it mainly involves scalar processing as opposed to vector processing performed in the other methods. Taking the borders drawn by the experienced dermatologist as the ground truth, and comparing the automatic borders with those drawn by the dermatology registrar, we have also shown that our method functions closer to the experienced dermatologist than the dermatology registrar, on some metrics. This implies that our method is as good as, and in some cases more effective than the dermatology registrar.

Acknowledgments

The assistance of Dr. Con Dolianitis in obtaining the manual borders and Mr. Michael Purves and Ms. Amanda Rebbechi in taking the dermoscopy images is gratefully acknowledged. This research is supported by NICTA Victoria Research Laboratory, Australia.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.compmedimag.2010.08.001.

References

[1] American Cancer Society, Cancer Facts and Figures. Available from: http://www.cancer.org/downloads/STT/500809web.pdf; 2009 [accessed September 2009].
[2] Australia Skin Cancer Facts and Figures. Available from: http://www.cancer.org.au/cancersmartlifestyle/SunSmart/Skincancerfactsandfigures.htm; 2009 [accessed September 2009].
[3] Australia Skin Cancer Facts and Figures. Available from: http://www.cancer.org.au/Healthprofessionals/cancertypes/melanoma.htm; [accessed September 2009].
[4] Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sera F, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. Journal of the American Academy of Dermatology 2003;48:679–93.
[5] Geller AC, Swetter SM, Brooks K, Demierre M, Yaroch AL. Screening, early detection, and trends for melanoma: current status (2000–2006) and future directions. Journal of the American Academy of Dermatology 2007;57:555–72.
[6] Braun P, Rabinovitz H, Oliviero M, Kopf A, Saurat J. Dermoscopy of pigmented lesions. Journal of the American Academy of Dermatology 2005;52(1):109–21.
[7] Menzies SW. Cutaneous melanoma: making a clinical diagnosis, present and future. Dermatologic Therapy 2006;19:32–9.
[8] Perrinaud A, Gaide O, French L, Saurat J -H, Marghoob A, Braun R. Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems. British Journal of Dermatology 2007;157:926–33.
[9] Celebi ME, Iyatomi H, Schaefer G, Stoecker WV. Lesion border detection in dermoscopy images. Computerized Medical Imaging and Graphics 2009;33(2):148–53.
[10] Iyatomi H, Oka H, Celebi ME, Hashimoto M, Hagiwara M, Tanaka M, et al. An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. Computerized Medical Imaging and Graphics 2008;32(7):566–79.

[11] Hintz-Madsen M, Hansen LK, Larsen J, Drzewiecki K. A probabilistic neural network framework for detection of malignant melanoma. In: Naguib RNG, Sherbet GV, editors. Artificial neural networks in cancer diagnosis, prognosis and patient management. CRC Press; 2001. p. 141–83.
[12] Melli R, Grana C, Cucchiara R. Comparison of color clustering algorithms for segmentation of dermatological images. SPIE Medical Imaging 2006;6144, 3S1–3S9.
[13] Hance G, Umbaugh S, Moss R, Stoecker WV. Unsupervised color image segmentation: with application to skin tumor borders. IEEE Engineering in Medicine and Biology Magazine 1996;15:104–11.
[14] Schmid P. Segmentation of digitized dermatoscopic images by two-dimensional color clustering. IEEE Transactions on Medical Imaging 1999;18(2):164–71.
[15] Celebi ME, Aslandogan YA, Stoecker WV, Iyatomi H, Oka H, Chen X. Unsupervised border detection in dermoscopy images. Skin Research and Technology 2007;13:454–62.
[16] Celebi ME, Kingravi HA, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH, et al. Border detection in dermoscopy images using statistical region merging. Skin Research and Technology 2008;14:347–53.
[17] Zhou H, Chen M, Zou L, Gass R, Ferris L, Drogowski L, et al. Spatially constrained segmentation of dermoscopy images. In: 5th IEEE International Symposium on Biomedical Imaging. 2008. p. 800–3.
[18] Lee T, Ng V, Gallagher R, Coldman A, McLean D. Dullrazor: a software approach to hair removal from images. Computers in Biology and Medicine 1997;27:533–43.
[19] Lucchese L, Mitra S. Color image segmentation: a state-of-the-art survey. In: Proceedings of Indian National Science Academy Part A, PINSA2001. 2001. p. 207–21.
[20] Plataniotis K, Venetsanopoulos A. Color image processing and applications. Springer-Verlag; 2000.
[21] Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 1979;9(1):62–6.
[22] Haralick RM, Shapiro LG. Computer and robot vision. Vol. 1. Addison-Wesley; 1992.
[23] Sorensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. Royal Danish Academy of Sciences and Letters 1948;5:1–34.
[24] Finch S, Cumming G. Putting research in context: understanding confidence intervals from one or more studies. Journal of Pediatric Psychology 2009;34:903–16.
[25] Iyatomi H, Oka H, Saito M, Miyake A, Kimoto M, Yamagami J, et al. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. Melanoma Research 2006;16(2):183–90.
[26] Xu L, Jackowski M, Goshtasby A, Rosemanb D, Binesb S, Yu C, et al. Segmentation of skin cancer images. Image and Vision Computing 1999;17: 65–74.
[27] Shapiro LG, Stockman GC. Computer vision. Prentence Hall; 2001.
[28] Celebi ME, Schaefer G, Iyatomi H, Stoecker WV, Malters JM, Grichnik JM. An improved objective evaluation measure for border detection in dermoscopy images. Skin Research and Technology 2009;15:444–50.

Rahil Garnavi received her BSc degree in Computer Engineering (Software) from Amirkabir University of Technology (Tehran, Iran) in 2003 and her MSc degree in Computer Engineering (Artificial Intelligence) from the University of Isfahan (Isfahan, Iran) in 2005. Since 2007, she has been studying PhD in The University of Melbourne, Department of Electrical and Electronic Engineering (Melbourne, Australia), conducting her research on computer-aided diagnosis of melanoma. Her research interest includes medical image analysis, image processing, computer vision, and machine learning.

Mohammad Aldeen obtained his BAE, MES and PhD from Baghdad University, IRAQ, The University of Michigan, USA, and Brunel University, UK, respectively. He has been with the University of Melbourne since 1985 where he is currently an associate professor. His research areas of interest are image processing, control systems engineering and energy systems.

M. Emre Celebi received his BSc degree in Computer Engineering from Middle East Technical University (Ankara, Turkey) in 2002. He received his MSc and PhD degrees in Computer Science and Engineering from the University of Texas at Arlington (Arlington, TX, USA) in 2003 and 2006, respectively. He is currently an Assistant Professor in the Department of Computer Science at the Louisiana State University in Shreveport (Shreveport, LA, USA). His research interests include medical image analysis, color image processing, and content-based image retrieval.

George Varigos has been Head of The Dermatology Departments of Royal Melbourne Hospital and Royal Children Hospital since 1984 and 1987, obtaining his degree in medicine at The University of Melbourne and his PhD in Immunology/medicine in 1987, and his Australian fellowship in Dermatology in 1978. His research areas include immunology of skin, Porphyria, Acne and more recent Psoriasis Biologics clinical trials. At Royal Melbourne Hospital he also runs an advanced melanoma multidisciplinary clinic weekly and has an interest in Dermoscopy.

Sue Finch studied psychology and statistics as an undergraduate, and has a PhD from the University of Melbourne. She is an Accredited Statistician with the SSAI, and works in the Statistical Consulting Centre at The University of Melbourne. Her research interests include finding better ways of teaching, communicating and understanding statistical concepts, the reform of statistics in the social sciences, and statistical consulting.