*Skin Research and Technology 2011; **0**: 1–11*
Printed in Singapore · All rights reserved
doi: 10.1111/j.1600-0846.2011.00569.x

Skin Research and Technology

Three-phase general border detection method for dermoscopy images using non-uniform illumination correction

Kerri-Ann Norton[1*], Hitoshi Iyatomi[2], M. Emre Celebi[3], Sumiko Ishizaki[4], Mizuki Sawada[4], Reiko Suzaki[4], Ken Kobayashi[4], Masaru Tanaka[4] and Koichi Ogawa[2]

[1]*BioMaPS Institute, Rutgers University, Piscataway, NJ, USA,*
[2]*Department of Applied Informatics, Hosei University, Tokyo, Japan,*
[3]*Department of Computer Science, Louisiana State University, Shreveport, LA, USA and* [4]*Department of Dermatology, Tokyo Women's Medical University, Medical Center East, Tokyo, Japan*

Background: Computer-aided diagnosis of dermoscopy images has shown great promise in developing a quantitative, objective way of classifying skin lesions. An important step in the classification process is lesion segmentation. Many studies have been successful in segmenting melanocytic skin lesions (MSLs), but few have focused on non-melanocytic skin lesions (NoMSLs), as the wide variety of lesions makes accurate segmentation difficult.
Methods: We developed an automatic segmentation program for detecting borders of skin lesions in dermoscopy images. The method consists of a pre-processing phase, general lesion segmentation phase, including illumination correction, and bright region segmentation phase.
Results: We tested our method on a set of 107 NoMSLs and a set of 319 MSLs. Our method achieved precision/recall scores of 84.5% and 88.5% for NoMSLs, and 93.9% and 93.8% for MSLs, in comparison with manual extractions from four or five dermatologists.
Conclusion: The accuracy of our method was competitive or better than five recently published methods. Our new method is the first method for detecting borders of both non-melanocytic and melanocytic skin lesions.

Key words: segmentation – skin cancer – non-melanocytic lesions – melanoma

Accepted for publication 7 July 2011

THE NUMBER of skin cancer cases is on the rise in the United States and other parts of the world, with the incidence being 40% of all new cancer cases (1). Two major types of malignant skin tumors are basal cell carcinoma (BCC) and melanoma (2). BCC is one of non-melanoma skin cancers, whereas melanoma is considered malignant melanocytic tumor as they originate from different cell types (1), see Fig. 1. Melanoma is by far the most proliferative, invasive, and metastatic skin lesions, whereas BCC are seldom metastatic (2). Although advanced-stage melanoma is often fatal, if diagnosed early,melanoma can be cured (3). Therefore, early detection and diagnosis of these skin lesions are important.

Dermoscopy, or epiluminescence light microscopy, was developed to diagnose skin lesions. Dermoscopy was an important advance in skin lesion diagnosis because it makes subsurface areas more visible, allowing for the identification of important dermoscopic features, such as blue-white areas, dots/globules, and blotches (4). Even with these advantages, dermoscopy remains subjective and has low reproducibility (3). Many algorithms have been used in conjunction with dermoscopy for the diagnosis of pigmented skin lesions, such as the ABCD rule of dermoscopy, the ELM pattern analysis, the 7-point checklist, and the CASH algorithm (5). Notwithstanding the variety of algorithms available, diagnosis remains difficult and highly subjective; therefore, there is a need for an objective computer-aided diagnosis method for dermoscopy images.

The overarching goal of this work was to develop a computer-aided diagnosis system that can classify non-cancerous skin lesions

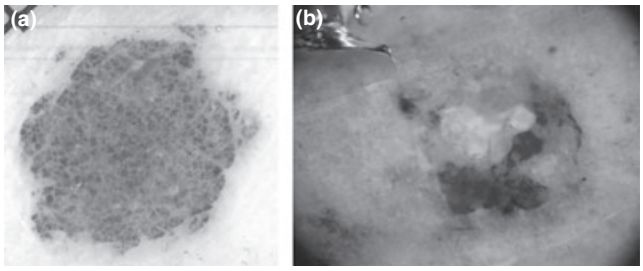

Fig. 1. Dermoscopy images of two types of skin lesions. (a) Melanocytic lesions (MSLs) consist of melanomas, cancers that are known for their invasive nature, and benign melanocytic lesions. Melanomas usually have irregular borders with some color variation, whereas benign melanocytic lesions often have smooth borders and few colors. (b) Non-melanocytic lesions (NoMSLs) come in a wide variety, ranging from basal cell carcinoma to seborrheic keratoses, and tend to be varied in appearance, have irregular borders, and multiple colors and textures within the lesions.

from the cancerous ones. Many systems have been successful in distinguishing benign from cancerous melanocytic skin lesions (MSLs), but few methods have focused on distinguishing these lesions from non-melanocytic skin lesions (NoMSLs). One reason is that melanomas are associated with poor prognosis and resistant to treatments such as chemotherapy, so it is especially important to recognize these lesions early (2). Another reason is that the wide variety of NoMSLs makes detection and classification more difficult.

The border of a skin lesion provides important information for the diagnosis of the disease, and thus an automated imaging method for border detection is a critical step for computer-aided diagnosis. A few methods that focused specifically on the detection and classification of NoMSLs are the use of optical coherence tomography and fluorescence (6–8). Another important method uses the measurements of semi-translucent areas in dermoscopy images to classify BCC from other lesions, but their segmentation method was not automatic (9). Segmentation of NoMSLs is especially important as the diagnostic improvement achieved by dermoscopy is actually higher for NoMSLs than for MSLs (10).

Computed-aided diagnosis of dermoscopy images is becoming a more feasible option for skin lesions, which would allow for an objective and repeatable way of classifying skin lesions. This process usually consists of four components: image acquisition, lesion segmentation or border detection, feature extraction, and classification. Previously, there were many examples where automatic segmentation was successful in detecting the border of skin lesions, for instance, the segmentation of MSLs has been successfully accomplished (11). Inaccuracies in the lesion segmentation component can lead to inaccuracies in classification. Also, improvements in the segmentation of dermoscopy images offer a better classification performance overall (3). Therefore, in this study, we proposed to develop a general lesion segmentation method as a first step in developing a classification method that distinguishes melanomas from both benign melanocytic lesions and NoMSLs.

The segmentation of NoMSLs poses several challenges that make segmentation especially difficult, see Fig. 2. For one thing, there are a large variety of lesion types, such as BCC and seborrheic keratosis (SK), and the borders of these lesions are somewhat ambiguous, making the reliability of the segmentation difficult. There is often low contrast between the lesion and the background skin, as well. In addition, the morphology of the lesions ranges from small spread-out dots to large fractal-like

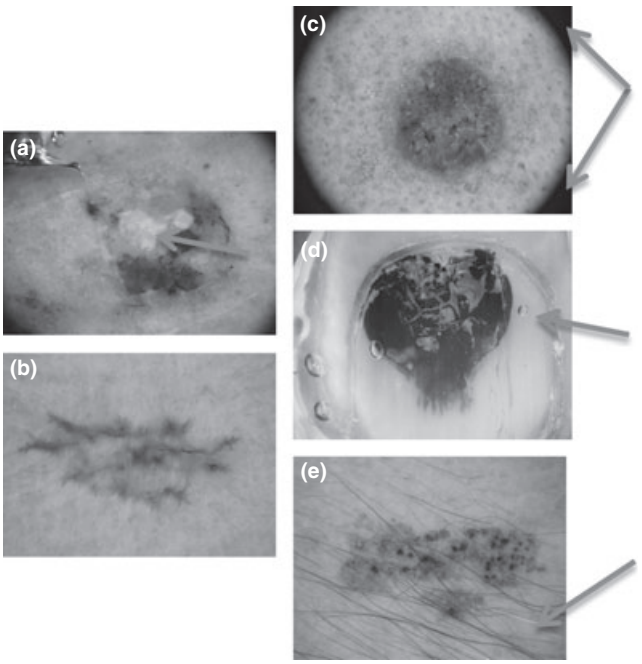

Fig. 2. Segmentation challenges. The variety of non-melanocytic lesions causes them to be difficult to segment because they widely vary in appearance, have irregular borders, and have multiple colors. In general, there are several issues that must be taken into account for accurate segmentation of dermoscopy images. These are the microscope border, bubbles, hairs or other artifacts in the image. Here, we give example images for certain segmentation challenges: (a) color variegation, (b) irregular borders, (c) microscope border, (d) bubbles, and (e) hairs.

lesions, and colors range from dark brown to red to bluish-white lesions. Segmentation is further complicated by the presence of foreign objects such as hairs, air bubbles, and the microscope border.

Notwithstanding these challenges, we have developed a three-step border detection method for the segmentation of NoMSLs. We compare our results with five other segmentation methods from the literature and show that our method is effective for segmenting out NoMSLs. In addition, our method is able to segment out MSLs, including melanoma, thus proving it to be versatile enough to segment skin lesions in general.

Materials

A total of 107 digital NoMSL dermoscopy images, 34 BCC, 59 SK, 26 of which are on the face, three hemangioma, and 11 hematoma, were obtained from the Department of Dermatology, Keio University School of Medicine. The non-melanocytic cases used in this study were clinically diagnosed by several dermatologists. The dimensions of the images were 2272 by 1704.

The previously published (3) melanocytic dataset consisted of a total of 319 digital dermoscopy images: 247 melanocytic nevi, (188 of which are Clark nevi and 59 of which are Reed nevi) and 75 melanomas (including 23 melanoma in situ) from the University of Naples and Graz. All of the melanocytic cases were diagnosed based on histopathological examination of biopsy material by several expert dermatologists. The typical size of the images is 500 by 500, but dimensions of these images varied.

Methods

To segment NoMSLs, we used the following three steps to define the border of these lesions, (See Fig. 3): (i) pre-processing, (ii) general lesion segmentation, and (iii) 'bright' region segmentation. The method was dependent on three parameters, (See Table 1), that were determined experimentally. The structuring element size r, was used mainly for morphology operations such as dilation, opening, closing, etc. and therefore was dependent on the size of the image. The second and third parameters, general object removal ratio s_1 and bright object removal ratio s_2, were used for object removal, (see below for more details).

Pre-processing: (I) and (II)

The goal of the pre-processing phase of this algorithm was to remove pixels that correspond to the microscope border. The border always appears as a gradient with very dark pixels near the edge of the image successively becoming lighter toward the interior of the image. We first needed to identify dark pixels and then expand the segmentation to include lighter areas, as segmenting out these lighter areas

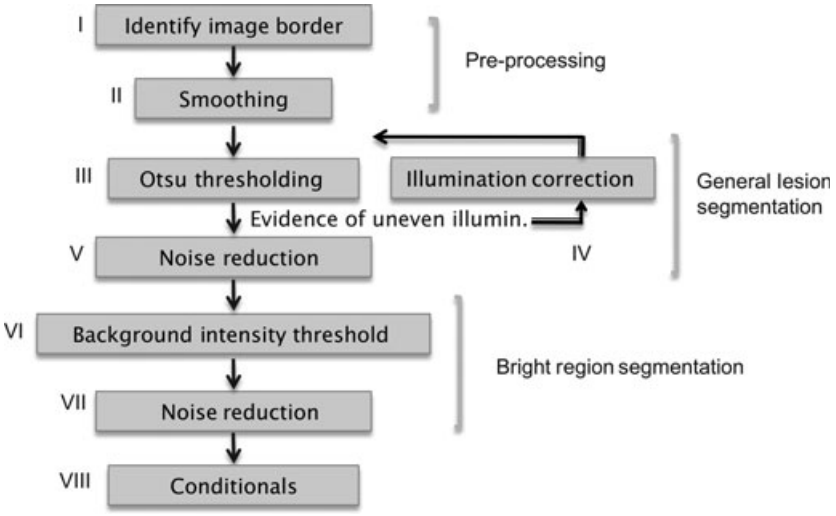

Fig. 3. Flowchart of the segmentation method. The method is broken into three main pieces: pre-processing, general lesion segmentation, and bright region segmentation.

TABLE 1. Parameters for the proposed method

Structuring element size r	General object removal ratio s_1	Bright object removal ratio s_2
l/150	9/100	1/8000

The three parameters, r, s_1, and s_2 were determined experimentally. l is the length (in pixels) of the shortest dimension of the image.

would lead to lesion pixels being identified as the microscope border. First, we converted the RGB (red, green, blue) image to grayscale and identified dark pixels in the image, defined as having intensities <35, (with pixel range between 0 and 255). Next, the segmented dark region was dilated using a disk of size 3*r*, to include lighter areas of the microscope border that would not have been picked up. To prevent any lesion pixels from being falsely included, we removed any objects that are not touching a border of the image, (I). This segmented microscope border was removed from the other lesion segmentations. To segment out the tumor area, we used a thresholding technique using intensity information. The lesions and background skin can have many irregularities such as veins, dirt, and speckles, which break the uniformity of either the background or the lesion. Thus, we needed to smooth out these areas to reduce noise (II). We did this by applying a circular averaging filter of size r to the original RGB image.

General lesion segmentation: (III)–(V)

For the second phase of the border detection method, we segmented out the general lesion using three steps: i) segmentation of tumor area, ii) correction of non-uniform illumination, and iii) noise reduction.

Segmentation of tumor area

Once the smoothing process had taken place, we needed to binarize the image into the 'background skin' and 'lesion' areas. We converted the image into a binary image using Otsu thresholding (12) on the green channel of the filtered image (III), assigning the lower intensity regions to be lesion regions. Otsu's thresholding finds an appropriate threshold by maximizing the between-class variance, and thus it does not specifically account for things such as the microscope border and artifacts. Although the blue channel is commonly used and usually provides good results (14), we found here that the blue channel misses some pixels. By contrast, we found that the red channel tends to include too many pixels. The pixels corresponding to the microscope border region were excluded from the Otsu threshold calculation to avoid bias from these low-intensity pixels.

There were some images in which non-uniform illumination occurs. This usually created large shadowy areas that are darker than the rest of the image background (roughly larger than 1/3 the size of the image). These shadowy areas in the image were often identified as lesion pixels if not previously corrected for. Therefore, to detect these regions, we calculated the area of the largest object in the segmentation. If the area was between 1/3 and 2/3 the size of the total image, we corrected for non-uniform illumination. Thus, we needed to adjust the contrast so that it was relatively uniform across the whole image, which was achieved by using contrast-limited adaptive histogram equalization (CLAHE) (15) on the green channel of the smoothed image, (before thresholding). In this method, the output for each pixel was its rank in its modified histogram of its respective tile. The image was divided into four equal tiles and its 'rank' was calculated as the number of pixels whose intensities were below the pixel of interest in the modified tile histogram. The modified histogram of each tile was found by calculating the original histogram of the tile and the number of pixels were clipped and redistributed uniformly, see Fig. 4. Once the pixel intensities were recalculated based on its rank, the tiles were then merged using bilinear interpolation to prevent unnatural boundaries (IV). After CLAHE, we binarized the image choosing a threshold using Otsu's method, as above (III).

Noise reduction (V)

After the initial binarization, it was necessary to remove areas associated with background artifacts such as hairs, blood vessels, dirt, etc. that we classify as 'noise' in the image, and to avoid any holes in the segmented object. The noise removal step was accomplished using morphological operations (V). Opening was used to merge any small nearby pixels together; for the opening operation, we used a disk of size r. As hairs tend to be long and thin, we used a closing operation with a disk of size 2*r* to remove thin objects. For continuity, we filled in any

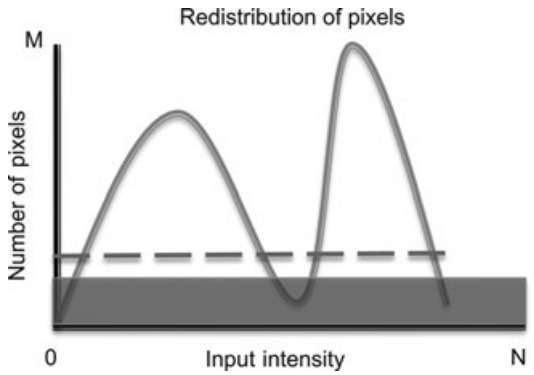

Fig. 4. Contrast-limited adaptive histogram equalization (CLAHE) Histogram. A diagram of the clipping and redistribution of pixels in CLAHE method (13). The intensity of the pixels is redistributed so that intensities with high numbers of pixels are redistributed evenly among the intensity regions. In this way, the number of pixels for a certain intensity region is capped. The dashed bar represents the largest number of pixels allowed for each intensity region, and the red bar represents the redistributed pixels.

holes within the selected region and removed any objects smaller than s_1 times the area of the largest object. Therefore, objects are removed based on the size of the other identified objects, as the magnification may differ between images. It has been seen in previous works that similar automatic segmentation methods tend to result in segmented areas smaller than that of the expert dermatologists (3). Accordingly, we dilated the image using an octagon of size 3*r* to correct for this underrepresentation. Lastly, we filled in any holes that may have occurred due to the dilation process to create a continuous border.

Bright region segmentation: (VI)(VII)

The last phase of the segmentation method identified the 'bright' regions of the lesions. We found that in some lesions, there was a small part that was brighter than the rest of the lesion and not picked up by the general lesion segmentation method. Therefore, we have developed an automatic segmentation method specifically for these regions. We observed that these regions tend to be of higher intensity than the background skin, so first, we found the background intensity in the blue channel, defined by the average intensity of the non-zero pixels excluding the segmented region and the microscope border (VI). Next, we thresholded the image using the intensity value 1.7 standard deviations larger than the average background intensity, taking the regions of higher intensity to be 'bright regions'. We once again used morphology operations to remove noise (VII). First, we opened the image to remove small objects using a disk of size *r*. Next, we filled in any holes, and removed objects smaller than s_2 multiplied by the total area of the image. To include areas that were slightly less bright, we dilated the segmented image using an octagon of size 3*r*. We only included the parts of the 'bright' segmentation that overlaped with the base segmentation, as only these regions should be part of the lesion. Lastly, we added the general lesion segmentation to the 'bright' segmentation and filled in any holes caused by the addition.

Conditionals: (VIII)

There are four conditionals (VIII) that we put on the segmentation to avoid inaccuracies and to single out a continuous border region. (i) First, the 'bright' segmentation was excluded if its area was >20% of the total image area, as these areas are known to be small. The next set of conditionals are aimed at singling out a continuous segmentation region. (ii) For the second conditional, we determined if anything touched the border of the image, and if so, we only kept the object that overlaps the center of the image. If nothing touched the center of the image, the object touching the border whose centroid was closest to the image centroid was included in the segmentation, along with any object not touching a border as well. (iii) The third conditional was aimed at correcting the convexity of the segmentation. It was observed that our computer-aided method tends to result in segmentation being less convex than those of expert dermatologists. To correct for this, the largest region's convexity was measured, defined by the number of pixels in the region divided by the number of pixels in the convex hull of the region. If this region did not have a convexity >85%, a closing operation was performed using an octagon of size 9*r*. This made the segmentation more convex and merged any objects that were close but not touching. (iv) For the last conditional, objects touching the borders of the segmentation were removed if there were more than two objects unless both objects touch the border. Finally, the segmentation was dilated with an octagon of size 3*r* and holes that were formed during these processes

were filled. The entire process takes on average between 1.4 and 4 s to segment an image.

Statistics

To verify the accuracy of our border detection results, we need to have a statistical measure of how precise the segmentation is. There are many statistical metrics to measure accuracy, such as sensitivity, specificity, precision, recall, error probability, and XOR measure (11). We used precision and recall in our studies as they are commonly used in engineering, see Eqs (1) and (2). In general terms, precision is a measure of the accuracy of the segmentation, or whether a chosen pixel is actually part of the lesion. Recall is a measure of how much of the lesion is covered by the segmentation, or how many actual lesion pixels were chosen by the segmentation. In this way, precision and recall balance each other and often if you have a high score in one, you have a low score in the other. For this work, we want to achieve a high score in both, so to compare the precision and recall scores of two different methods, we use the F measure for precision and recall, defined in Eq (3).

The manual segmentation results by several dermatologists are considered to be the 'ground truth.' We determine the number of true positives (TP): this is the number of pixels that were defined as part of the lesion in both the automatic segmentation method and the ground truth. The false positives (FP) are defined as the pixels that are defined as the background in the ground truth, but are identified as part of the lesion in the automatic segmentation. The false negatives (FN) are defined as those pixels that are determined to be lesion pixels in the ground truth, but are identified as background pixels in the automatic segmentation. Lastly, a true negative (TN) pixel is a pixel that is identified as a background pixel in both the manual segmentation and the automatic one.

Precision, recall, and F-measure are defined as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

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

$$F\text{-}measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

We compare our automatic segmentation with manual segmentations by trained professionals. It has been shown that even by trained dermatologists, the deviation of manual extractions is large, with a standard deviation of 8.9% for the tumor area extracted by five dermatologists (3). We showed in previous work that the area in which at least two experts identify a pixel as part of the lesion can be taken as the 'gold standard' (3). One hundred and seven NoMSLs were used in this study, which had been segmented out by four dermatologists to compare the accuracies of the three methodologies. Therefore, for NoMSLs, we used the area in which at least two out of four dermatologists identified a pixel as part of the lesion as the 'gold standard' and for MSLs, we used two of five dermatologists as the 'gold standard,' as was used formerly (3).

Results

Segmentation of NoMSLs

To show the accuracy of our method, we compare the average precision, recall and F-measure scores of our method with five other methods: Garnavi's (16), statistical region merging (SRM) (4), k-means ++ (KPP) (17), unsupervised segmentation of color-texture regions (JSEG) (18), and dermatologist-like (3). The segmentation performance is summarized in Table 2 for NoMSLs. Garnavi's method does a color channel transformation, smoothes the image, and clusters based on thresholding. They keep the two largest areas that represent the background skin and the lesions areas. The SRM method (4) uses statistical region merging and smoothing. They remove light regions and border elements. The KPP method uses spatially constrained clustering, first by converting pixel locations into polar coordinates and clusters using RGB values and radius. They cluster twice and then do region merging based on texture properties. The JSEG method (17) uses color quantization, and then spatial segmentation to reduce the number of colors. Afterward, it applies a region-growing method to merge regions. Lastly, the dermatologist-like method is

TABLE 2. Average precision and recall scores for non-melanocytic lesions

	Our method	Garnavi's method	SRM method	KPP method	JSEG method	Dermatologist-like method
Precision (%)	84.5	95.7	68.6	59.0	86.2	93.7
Recall (%)	88.5	78.6	76.8	65.4	55.1	66.7
F-measure	86.5	86.3	72.5	62.0	67.2	77.9
# Failed	0	10	4	62	14	4
Published	–	2010	2008	2008	2007	2006

Here, we show our results for a set of 107 non-melanocytic lesions and compare it with five recently developed methods. The first three rows are the precision, recall, and F-measure scores. The fourth row states the number of images that were excluded from the calculations because the method failed to extract a lesion. The last row is the year that the algorithm was published. The columns list the names of the methods evaluated.

the oldest and so therefore might benefit from some minor modification (3). In short, the algorithm consists of filtering, thresholding, and merging of small regions. The resulting segments are selected based on boundary conditions and then lastly, a region-growing step is performed.

As can be seen from Table 2, the new method outperforms the other five methods, achieving an average precision score of 84.5%, a recall score of 88.5%, and an F-measure of 86.5. The # failed, or the number of misextractions, was not included in the statistics and so can be considered an independent measure of robustness. Thus, our method attains both robust and accurate results as it was able to extract all of the lesions. The accuracy closest to ours was Garnavi's method, achieving a precision of 84.5%, a recall of 78.6%, and an F-measure of 86.3, but this method failed to extract 10 images, which is around 9% of the images. Therefore, while this method does well for the ones it can extract, it fails to extract a significant number of images.

In Fig. 4, we show a set of sample segmentations for NoMSLs: the dermatologist's 'gold standard', our segmentation method, and the other segmentation method examined. The sample in the first row is a BCC with two types of artifacts that needed to be accounted for: the microscope border and a light region in the lesion. In this example, most of the methods did a good job at segmenting out the lesion except for SRM that only excludes the microscope border and KPP that failed to extract it all together.

The sample in the second row is a benign seborrheic keratosis of the face. The image poses several challenges as there is little contrast between the lesion and the surrounding skin, and there are small hairs and bubbles within the image. Our method achieves good results with this image as well achieving a F-measure of 91, although its recall score is a little lower indicating that our method is missing part of the lesion. This trend is also seen in Garnavi's method, JSEG, and the Dermatologist-like method, where the precision scores are higher than the recall scores, although our new method achieves the best results. This indicates that computer-aided systems have difficulty covering the entire image, which makes sense because there is little contrast between the lesion and the background skin, thus lesion pixels could easily be mistaken for background skin. The SRM and KPP methods once again perform poorly on this lesion.

The sample in the third row is also a benign seborrheic keratosis of the face, with many hairs and multiple colors within the lesion that complicate the segmentation. We achieve good overall results with an F-measure of 93, but we include some background pixels in our lesion indicated by the lower precision score. Garnavi's method actually outperforms our method in this case, achieving a F-measure of 95. Both JSEG and our conventional method achieved F-measure scores of 84. Once again, these methods achieve good precision scores, but low recall. Therefore, they are missing pieces of the lesion. The variation in color within the lesion and background causes the KPP method to achieve low scores both for precision and for recall.

Overall, we were able to segment out the 'bright' regions seen in a few BBC lesions and were also able to achieve good results segmenting out lesions that do not exhibit 'bright' regions. We accurately segment out lesions under different illumination conditions and background color tones and also those including foreign objects such as the microscope border, hairs, bubbles, and stitches. Our method performs better than five other methods as indicated by its higher F-measures. We find that Garnavi's method performs well, but it fails to

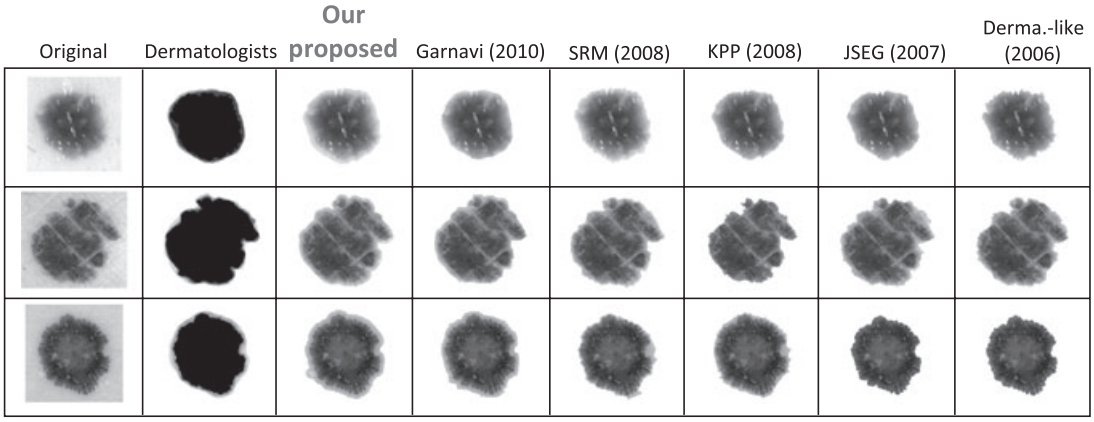

Fig. 5. Non-melanocytic Lesion Segmentations. The left column shows the originals, the second column shows the 'gold standards', the third column shows our segmentation results, and the fourth through eighth column show the other segmentation methods' results, see text for more detail. The precision/recall scores of these regions are as follows: Our method: (a) 91% and 97%, (b) 96% and 86%, and (c) 90% and 96%; Garnavi's method: (a) 98% and 92%, (b) 100% and 71%, and (c) 93% and 97%; SRM method: (a) 35% and 100%, (b) 0% and 0%, and (c) 20% and 100%; KPP method: (a) 0% and 0%, (b) 43% and 46%, and (c) 17% and 45%; JSEG method: (a) 100% and 76%, (b) 100% and 60%, and (c) 100% and 73%; and the dermatologist-like method: (a) 97% and 89%, (b) 99% and 74%, and (c) 99% and 73%.

extract many images. JSEG and the dermatologist-like method perform fairly well, but they tend not to cover the whole lesion areas.

Segmentation of MSLs

We segmented a set of 319 benign melanocytic and melanoma lesions that had been successfully segmented out in previous work (3). Table 3 summarizes the segmentation performances for MSLs; our method achieved an average precision of 93.9%, recall of 93.8%, and an F-measure of 93.9. The best method was the dermatologist-like method with precision of 94.1%, recall of 95.2%, and an F-measure of 94.6. The SRM also outperformed our new method with an F-measure of 94.1. Although these two other methods outperformed our method, the improvement was only minimal. In addition, our method outperformed the three remaining methods. Hence, our method attained competitive or better performance than the other methods tested. We also used our segmentation method on a completely different set of 548 melanocytic lesions (506 melanocytic nevi and 42 melanomas) from Keio University and achieved visually satisfactory results.

In Fig. 6, we show three examples of the six segmentation results for the MSLs. The left column shows the originals, the second column shows the 'gold standards' from the five dermatologists, the third column shows our new segmentation results, and the fourth through eighth column show the other segmentation methods' results: Garnavi's method, SRM, KPP, JSEG, and the dermatologist-like method. All three of our segmentations achieve good results, with the first sample (a) achieving a precision score of 96.2%, a recall score of 98.6%, and an F-measure of 97.4; the second sample (b) achieving a precision score of 96.1%, a recall score of 97.1%, and an F-measure of 96.6; and the last sample (c) achieving a precision score of 90.6%, a recall score of 99.9%, and a F-measure of 95.0. In samples (a) and (b), the precision score and recall score are similar, but in

TABLE 3. Average precision and recall scores for melanocytic lesions

	Our method	Garnavi's method	SRM method	KPP method	JSEG method	Dermatolgist like
Precision (%)	93.9	97.5	95.9	91.8	97.6	94.1
Recall (%)	93.8	88.7	92.4	87.0	88.4	95.2
F-measure	93.9	92.9	94.1	89.3	92.8	94.6
# Failed	0	1	1	28	0	0
Published	2011	2010	2008	2008	2007	2006

Here, we show our results for a set of 319 melanocytic lesions and compare it with five recently developed methods. The first three rows are the precision, recall, and F-measure scores. The fourth row states the number of images that were excluded from the calculations because the method failed to extract a lesion. The last row is the year that the algorithm was published. The columns list the names of the methods evaluated.

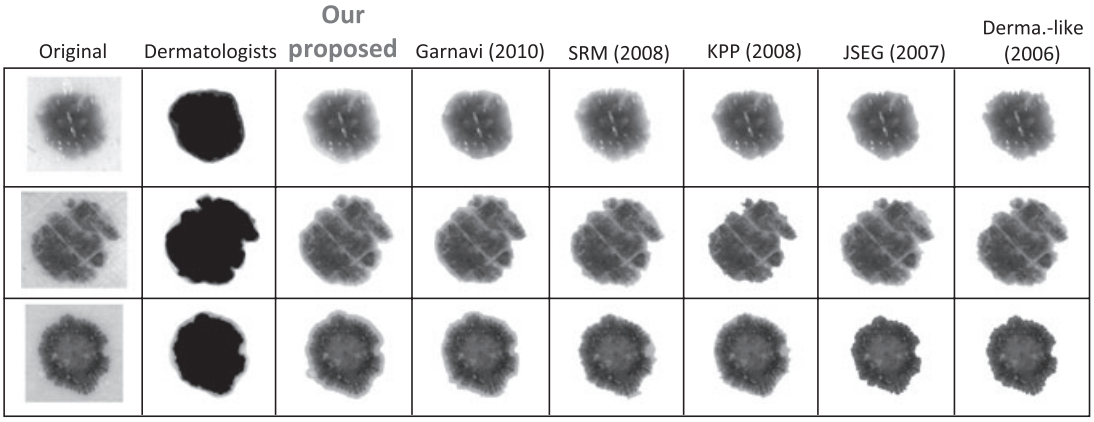

Fig. 6. Melanocytic lesion segmentations. The left column shows the originals, the second column shows the 'gold standards', the third column shows our segmentation results, and the fourth through eighth column show the other segmentation methods' results, see text for more detail. The precision/recall scores of these regions are as follows: Our method: (a) 96% and 97%, (b) 96% and 97%, and 91% and 100%; Garnavi's method: (a) 100% and 89%, (b) 97% and 93%, and (c) 94% and 100%; SRM method: (a) 99% and 97%, (b) 99% and 97%, and (c) 97% and 95%; KPP method: (a) 100% and 86%, (b) 100% and 83%, and (c) 98% and 98%; JSEG method: (a) 100% and 85%, (b) 98% and 96%, and (c) 100% and 82%; and the dermatologist-like method: (a) 93% and 100%, (b) 96% and 98%, and (c) 96% and 98%.

the last example, the recall score is much higher than the precision score. This is because the method overestimated the size of the lesion; it covered it almost completely, but it included too many background pixels as well.

In comparison, for the selected lesions, Garnavi's method usually achieved one score that was higher than ours, but the other that was lower. Garnavi's method was worse than ours for the first two lesions, but better than ours for the last one. SRM outperformed our method for each of the three lesions. KPP also achieved worse results than ours and Garnavi's method for the first two lesions, but achieved better scores than ours for the last lesion. The JSEG method usually had higher precision scores than ours, but also generally had lower recall scores. Lastly, the dermatologist-like extraction method achieves less consistency for the first sample, achieved similar results for the second lesion, and outperformed our method for the last lesion.

Discussion

We have developed a novel, accurate method for segmenting non-melanocytic (NoMSLs) and melanocytic lesions (MSLs). This is the first step toward developing a computer-aided diagnosis system for classifying skin lesions in general. We achieved accurate results on a set of NoMSLs, with precision and recall scores of 85.0% and 88.8%, respectively. For non-melanocytic lesions, our method performs the best of five other recent methods in the literature (3, 4, 16–18). On average, for NoMSLs, our method has a lower precision score than recall score as it tends to cover most of the lesion, but overestimates slightly. In comparison, the dermatologist-like method (3) is very precise, achieving a higher precision score than our method, but having a much lower recall score. This is because this method tends to be very accurate in choosing lesion pixels, but misses parts of the lesion. We feel that this is a result of the design of these methods. Garnavi's method also performs well, but was unable to segment a significant portion of the images. The method also does considerably better at precision than at recall, which means that it tends to miss parts of the lesion. This trend is also true for the dermatologist-like method, which seems to indicate that our new method does a better job of covering the entire lesions in general. We also see a trend in which many of the algorithms fail to extract some of the lesions, which is not the case for our method.

On a set of MSLs, we also achieved good results with scores of 93.9% and 93.8%. Our method is more accurate in segmenting melanocytic lesions than NoMSLs. This is because MSLs are easier to segment out than non-melanocytic lesions due to the lack of color variation and smoother borders. For these cases, our method does equally well for precision and recall. For MSLs, the dermatologist-like method

outperforms our new method, but only slightly. This is logical as the conventional method was designed for MSLs, whereas ours was not. Therefore, the fact that our method is competitive shows that it is a successful method as it performs almost as well as or better than existing methods for melanocytic lesions, which it was not tested on.

Overall, the best methods seem to be our new method, the dermatologist-like method, and Garnavi's method. These methods all had a pre-processing step (filtering), used Otsu's method, and had region selection based on boundary/area conditions. Our method also was the most consistent in terms of precision and recall, their scores always differing the least compared with the other methods, for both non-melanocytic and melanocytic lesions. Our method also had the least amount of failures. The dermatologist-like and Garnavi's method also always achieved good precision scores, while their recall scores varied. This means that they are very accurate in choosing pixels, but may not cover the entire lesion. The dermatologist-like method also had a low failure rate. Garnavi's method achieved lower recall scores for all types of lesions and also had a high failure rate for NoMSLs. KPP and JSEG seemed the least capable of handling NoMSLs of all the other methods. Both of these methods used a clustering-based method, JSEG using color quantization, whereas KPP used clustering based on RGB or Lab color spaces and the radial location. Both methods had the highest failure rates and low recall scores, indicating that they could not detect the whole lesion and sometimes could not detect the lesion at all.

Conclusion

Tumor extraction from dermoscopy images is an important step for computer-aided diagnosis of skin lesions. Our method is the first tumor extraction method to consider both MSLs and NoMSLs. Our method achieved the best scores for NoMSLs (precision: 84.5%, recall: 88.5%), and third best for MSLs (precision: 93.9%, recall: 93.8%) of five recently published methods. Our segmentation method could become the first step in classifying cancerous from benign lesions, and NoMSLs from MSLs, with our overarching goal being to classify melanomas from all other skin lesions using a computer-aided diagnosis system.

References

1. Bowden GT. Prevention of non-melanoma skin cancer by targeting ultraviolet-B-light signalling. Nat Rev Cancer 2004; 4: 23–35.
2. Nakayama K. Growth and progression of melanoma and non-melanoma skin cancers regulated by ubiquitination. Pigment Cell Melanoma Res 2010; 23: 338–351.
3.. Iyatomi H, Oka H, Saito M, et al. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. Melanoma Res 2006; 16: 183–190.
4. Celebi ME, Kingravi HA, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH, Malters JM, Grichnik JM, Marghoob AA, Rabinovitz HS. Border detection in dermoscopy images using statistical region merging. Skin Res Technol 2008; 14: 347–353.
5. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Arch Dermatol 1998; 134: 1563–1570.
6. de Leeuw J, van der Beek N, Neugebauer WD, Bjerring P, Neumann HA. Fluorescence detection and diagnosis of non-melanoma skin cancer at an early stage. Lasers Surg Med 2009; 41: 96–103.
7. Jorgensen TM, Tycho A, Mogensen M, Bjerring P, Jemec GB. Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography. Skin Res Technol 2008; 14: 364–369.
8. Mogensen M, Joergensen TM, Nurnberg BM, Morsy HA, Thomsen JB, Thrane L, Jemec GB. Assessment of optical coherence tomography imaging in the diagnosis of non-melanoma skin cancer and benign lesions versus normal skin: observer-blinded evaluation by dermatologists and pathologists. Dermatol Surg 2009; 35: 965–972.
9. Stoecker WV, Gupta K, Shrestha B, et al. Detection of basal cell carcinoma using color and histogram measures of semitranslucent areas. Skin Res Technol 2009; 15: 283–287.
10. Rosendahl C, Tschandl P, Cameron A, Kittler H. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. J Am Acad Dermatol 2011; 64: 1068–1073.
11. Celebi ME, Iyatomi H, Schaefer G, Stoecker WV. Lesion border detection in dermoscopy images. Comput Med Imaging Graph 2009; 33: 148–153.
12. Otsu N. A threshold selection method for gray-level histograms. IEEE Trans Syst Man Cybern 1979; 9: 62–66.
13. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, ter Haar Romeny B, Zimmerman JB, Zuiderveld K. Adaptive histogram equalization and its variations. Comput Vis Graph Image Process 1987; 39: 355–368.
14. Iyatomi H, Oka H, Celebi ME, Hashimoto M, Hagiwara M, Tanaka M and Ogawa K. An

improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. Comput Med Imaging Graph 2008; 32: 566–579.
15. Zuiderveld K. Contrast limited adaptive histogram equalization. Graph Gems IV San Diego: Academic Press Professional; 1994: 474–485.
16. Garnavi R, Aldeen M, Celebi ME, Varigos G, Finch S. Border detection in dermoscopy images using hybrid thresholding on optimized color channels. Comput Med Imaging Graph 2011; 35: 105–115.
17. Zhou H, Chen M, Gass R, Ferris L, Drogowski L, Rehg JM. Spatially constrained segmentation of dermoscopy images. Biomedical Imaging: From Nano to Macro, 2008 ISBI 2008 5th IEEE International Symposium on 2008; 800–803.
18. Celebi ME, Aslandogan YA, Stoecker WV, Iyatomi H, Oka H, Chen X. Unsupervised border detection in dermoscopy images. Skin Res Technol 2007; 13: 454–462.

Address:
Kerri-Ann Norton
John Hopkins University
611 Traylor Research Bldg.
Baltimore, MD 21205
USA
Tel: +1 443 287 6192
Fax: 410 614 8796
e-mail: knorton4@jhmi.edu