A State-of-the-Art Survey on Lesion Border Detection in Dermoscopy Images

M. Emre Celebi
Louisiana State University
Shreveport, Louisiana

Quan Wen
University of Electronic Science and Technology of China
Chengdu, People’s Republic of China

Hitoshi Iyatomi
Hosei University
Tokyo, Japan

Kouhei Shimizu
Hosei University
Tokyo, Japan

Huiyu Zhou
Queen’s University Belfast
Belfast, United Kingdom

Gerald Schaefer
Loughborough University
Loughborough, United Kingdom

CONTENTS

4.1 Introduction .........................................................................................98
4.2 Preprocessing .....................................................................................100
  4.2.1 Color Space Transformation ...................................................100
  4.2.2 Contrast Enhancement ...........................................................101
  4.2.3 Approximate Lesion Localization ...........................................101
  4.2.4 Artifact Removal ....................................................................102
4.3 Segmentation......................................................................................104
4.4 Postprocessing....................................................................................105
4.1 INTRODUCTION

Invasive and in situ malignant melanoma together comprise one of the most rapidly increasing cancers in the world. Invasive melanoma alone has an estimated incidence of 73,870 and an estimated total of 9940 deaths in the United States in 2015 [1]. Early diagnosis is particularly important since melanoma can be cured with a simple excision if detected early.

Dermoscopy, also known as epiluminescence microscopy, is a noninvasive skin imaging technique that uses optical magnification and either liquid immersion and low-angle-of-incidence lighting or cross-polarized lighting, making subsurface structures more easily visible than in conventional clinical images [2]. Dermoscopy allows the identification of dozens of morphological features, such as atypical pigment networks, dots/globules, streaks, blue–white areas, and blotches [3]. This reduces screening errors and provides greater differentiation between difficult lesions, such as pigmented Spitz nevi and small, clinically equivocal lesions [4]. However, it has been demonstrated that dermoscopy may actually lower the diagnostic accuracy in the hands of inexperienced dermatologists [5]. Therefore, in order to minimize the diagnostic errors that result from the difficulty and subjectivity of visual interpretation, the development of computerized image analysis techniques is of paramount importance [6].

Border detection is often the first step in the automated analysis of dermoscopy images [7]. It is crucial for image analysis for two main reasons. First, the border structure provides important information for accurate diagnosis, as many clinical features, such as asymmetry [8], border irregularity [9], and abrupt border cutoff [10], are calculated directly from the border. Second, the extraction of other important clinical features, such as atypical pigment networks [6, 11–25], dots [21, 26–28], globules [6, 28], streaks [16, 19, 29, 30], blue–white areas [19, 31–37], blotches [38–42], regression structures [19, 31, 33, 43, 44], and color variegation [45], critically depends on the accuracy of border detection. Automated border detection is a challenging task for several reasons [46]: (1) low contrast between the lesion and the surrounding skin (Figure 4.1a), (2) irregular (Figure 4.1b) and fuzzy lesion borders (Figure 4.1c), (3) artifacts and intrinsic cutaneous features such as black frames (Figure 4.1d), skin lines, blood vessels (Figure 4.1e), air bubbles (Figure 4.1f), and hairs (Figures 4.1g and 4.1h), (4) perspective distortion (Figure 4.1i) (5) variegated coloring inside the lesion (Figure 4.1j), (6) fragmentation due to various reasons such as regression (Figure 4.1k), (7) presence

of multiple lesions (Figure 4.1l), and (8) lesion larger than the field of view (Figure 4.1j).

Since the late 1990s numerous methods have been developed for automated border detection in dermoscopy images [47–50]. Unfortunately, due to the interdisciplinary nature of the field, the existing literature is scattered among numerous medical and engineering journals. In an earlier study, we reviewed
18 methods published between 1998 and 2008 [46]. The amount of research on border detection has been steadily increasing since 2008. As evidence to this is, 5 out of 11 articles in the 2011 *Computerized Medical Imaging and Graphics* special issue “Advances in Skin Cancer Image Analysis” [47] were devoted to border detection. In this chapter, we update our earlier survey with 32 additional methods published between 2009 and 2014. In Sections 4.2 through 4.4, we review the preprocessing, segmentation, and postprocessing phases, respectively. In Section 4.5, we discuss performance evaluation issues. Finally, in Section 4.6, we compare 50 published border detection methods with respect to various criteria and propose guidelines for future studies in automated border detection.

### 4.2 PREPROCESSING

In this section, we describe the preprocessing steps that facilitate the border detection procedure, namely, color space transformation, contrast enhancement, approximate lesion localization, and artifact removal.

#### 4.2.1 COLOR SPACE TRANSFORMATION

Dermoscopy images are commonly acquired using a digital camera with a dermoscope attachment. Due to the computational simplicity and convenience of scalar (single-channel) processing, the resulting RGB (red–green–blue) color image is often converted to a scalar image using one of the following methods:

- Retaining only the blue channel [51] (lesions are often more prominent in this channel, but typically, this is also the noisiest channel)
- Retaining the channel with the highest entropy [52]
- Applying a fixed luminance transformation, for example, [53, p. 122], such as
  \[
  \text{Luminance (Rec. 601)} = 0.2990 \times \text{Red} + 0.5870 \times \text{Green} + 0.1140 \times \text{Blue}
  \]
  or
  \[
  \text{Luminance (Rec. 709)} = 0.2126 \times \text{Red} + 0.7152 \times \text{Green} + 0.0722 \times \text{Blue}
  \]
- Applying an adaptive luminance transformation [54, 55]
- Applying the Karhunen–Loève (KL) transformation [56] and retaining the channel with the highest variance [57]

In applications where vector (multichannel) processing is desired, the RGB image can be used directly or it can be transformed to a different color space for various reasons, including (1) reducing the number of channels, (2) decoupling luminance and chromaticity information, (3) ensuring (approximate) perceptual uniformity, and (4) achieving invariance to various imaging conditions, such as viewing direction, illumination intensity, and highlights [58]. Common target color spaces [59] in this case include CIELAB, CIELUV, KL, and HSI (hue–saturation–intensity) [60]. Note that there are various formulations of the HSI color space in the literature [61], and therefore it
is good practice to cite an appropriate reference when using this color space or its variants.

4.2.2 CONTRAST ENHANCEMENT

As mentioned in Section 4.1, one of the factors that complicate the detection of borders in dermoscopy images is insufficient contrast. There are two main approaches to address this issue: hardware-based techniques [62–67] and software-based techniques. The former approach is generally preferable to the latter. However, in many cases, for example, web-based systems [68], acquisition cannot be controlled, and thus software-based postacquisition enhancement is the only option.

Gomez et al. [69] proposed a contrast enhancement method based on independent histogram pursuit. This algorithm linearly transforms the input RGB image to a decorrelated color space in which the lesion and the background skin are maximally separated. Given an input RGB image, Celebi et al.’s method [54] determines the optimal weights to convert it to grayscale by maximizing a histogram bimodality measure. The authors demonstrated that their adaptive optimization scheme increases the contrast between the lesion and background skin, leading to a more accurate separation of the two regions using Otsu’s thresholding method [70]. Madooei et al. [55] proposed a physics-based color-to-grayscale conversion method that attenuates shading as well as thin and short hairs. As in Celebi et al.’s study [54], the authors showed that the resulting grayscale image allows for a more accurate segmentation. Iyatomi et al. [71] proposed a color correction method based on the HSV (hue–saturation–value) color space. First, a multiple linear regression model for each of the H, S, and V channels is built using various low-level features extracted from a training image set. Using these regression models, the method then automatically adjusts the hue and saturation of a previously unseen image. Schaefer et al. [72] presented a two-stage scheme that removes color variations and enhances the contrast of the images by combining Grayworld and MaxRGB normalization techniques [73]. Abbas et al. [74, 75] and Norton et al. [76] proposed the use of homomorphic filtering [77] and contrast-limited adaptive histogram equalization [78], respectively. Barata et al. [79] compared the performance of several color constancy algorithms on the task of color normalization. Finally, Lu et al. [80] proposed a no-reference uneven illumination assessment measure based on a variational formulation of Retinex [81].

4.2.3 APPROXIMATE LESION LOCALIZATION

Although dermoscopy images can be quite large, lesions often occupy a relatively small area. An accurate bounding box (the smallest axis-aligned rectangular box that encloses the lesion) might be useful for various reasons [82]: (1) it provides an estimate of the lesion size (certain image
segmentation methods, such as region growing and morphological flooding [83, 84], can use this information as part of their termination criteria), (2) it might improve the accuracy of border detection since the procedure is focused on a region that is guaranteed to contain the lesion (active contour-based segmentation methods can be initialized inside/outside this region [74, 85, 86]), (3) it speeds up border detection since the procedure is performed in a region that is often smaller than the entire image [87], and (4) its surrounding might be used in the estimation of the background skin color, which is useful for various operations, including the elimination of spurious regions that are discovered during border detection [88] and the extraction of dermoscopic features such as streaks and blue–white areas. Several authors employed histogram thresholding methods such as Otsu’s method [70] and Kittler and Illingworth’s method [89] to localize lesions in dermoscopy images [52, 74, 85–87]. Celebi et al. [82], on the other hand, proposed the use of an ensemble of thresholding methods. This method was shown to be significantly more robust than pathological cases (e.g., unimodal histograms) when compared to individual thresholding methods. Wang et al. [83, 84] determined the bounding box by fitting a least-squares quadratic polynomial [90] to each of the horizontal and vertical projections [91, pp. 355–356] of the luminance image.

4.2.4 ARTIFACT REMOVAL

Dermoscopy images often contain artifacts such as black frames, ink markings, rulers, air bubbles, and intrinsic cutaneous features that can affect border detection, such as skin lines, blood vessels, and hairs. These elements complicate the border detection procedure, resulting in loss of accuracy as well as increase in computational time. The most straightforward way to remove these artifacts is to smooth the image using a general-purpose filter such as the Gaussian (GF), mean (MF), median (MF), or anisotropic diffusion (ADF) filter. Several issues should be considered while using these filters:

- **Scalar versus vector processing**: These filters are originally formulated for scalar images. For vector images, one can apply a scalar filter to each channel independently and then combine the results, a strategy referred to as marginal filtering. Although fast, this scheme introduces color artifacts into the output. An alternative solution is to use filters that treat the pixels as vectors [92].
- **Mask size**: The amount of smoothing is proportional to the mask size. Therefore, excessively large masks result in the blurring of edges, which might reduce the accuracy of border detection. Setting the mask size proportional to the image size appears to be a reasonable strategy [88, 93].
- **Computational time**: For the GF, MF, and MF, algorithms that require constant time regardless of the mask size are available [94–96].
As for the ADF, the computational time depends on the mask size and the number of iterations.

An alternative strategy for artifact removal is to use a specialized method for each artifact type. For the removal of rectangular black frames, Celebi et al. [88] proposed an iterative algorithm based on the lightness component of the HSL (hue–saturation–lightness) color space. Similar approaches can be found in [52, 76, 83, 97]. It should be noted that a systematic method to remove circular black frames (see Figure 4.1d) appears to be missing. In most cases, image smoothing effectively removes the skin lines and blood vessels. A method that can remove bubbles with bright edges was introduced in [6], where the authors utilized a morphological top-hat operator [98] followed by a radial search procedure.

Hair detection/removal received the most attention in the literature. An ideal hair detection/removal method should (1) detect/remove light and thin as well as dark and thick hairs [99], (2) deal with intersecting hairs [99], (3) not cause excessive blurring and color bleeding [100], (4) be evaluated both qualitatively and quantitatively using synthetic as well as real images with and without hair [101, 102], and (5) be computationally efficient.

Lee et al. [99] used morphological closing to detect hairs, which were then removed using bilinear interpolation. Schmid-Saugeon et al. [103] transformed the RGB image to CIELUV color space and then detected hairs using morphological closing on the L channel followed by thresholding. The hair pixels were replaced by their values after morphological closing. Wighton et al. [101], Xie et al. [104], and Fiorese et al. [102] also used morphological operators for detection, but used partial differential equation (PDE)–based inpainting [105, 106] for removal. Fleming et al. [6] and Zhou et al. [100] detected hairs using curvilinear structure detection [107] with various constraints. The latter authors removed the detected hairs using exemplar-based inpainting [108]. Nguyen et al. [109] employed matched filtering [110] followed by Gaussian smoothing, entropic thresholding [111], morphological thinning, and curve fitting for detection and bilinear interpolation [99] for removal. Möllersen et al. [64] thresholded the red channel using Otsu’s method [70] and applied a sequence of morphological operators along various directions to detect hairs. Abbas et al. [75] used a two-dimensional derivative of the Gaussian filter [112] for detection and exemplar-based inpainting [108] for removal. Kiani and Sharafat [113] used the radon transform [114] followed by edge detection using the Prewitt operator [115] for detection and interpolation by averaging for removal. Wighton et al. [23] employed linear discriminant analysis [116] and maximum a posteriori estimation [117] for detecting hairs. Barata et al. [24] used a bank of directional filters for detection and PDE-based inpainting [106] for removal. Thon et al. [28] detected hairs using Bayesian multiscale analysis with an intrinsic second-order Gaussian Markov random field (MRF) [118] prior. Afonso and Silveira [119] applied the morphological top-hat operator on the channel with the greatest entropy and then thresholded the result using
Otsu’s method. The hairs were then detected using percolation [120] subject to shape constraints. Abbas et al. [121] used modified matched filtering [122] followed by morphological closing for detection and fast marching-based inpainting [123] for removal. Toossi et al. [124] converted the RGB image to a grayscale one using the KL transform and then smoothed the latter image using the Wiener filter [125]. The edges in the smoothed image were detected using Canny’s algorithm [126] with the high threshold determined using Rosenfeld and de la Torre’s thresholding method [127]. The hairs were then detected using morphological dilation and removed using fast marching-based inpainting [123]. Abbas et al. [128] presented a comparative study of four recent hair detection/removal methods. Mirzaalian et al. [129] presented a method to detect light and dark hairs with varying thickness using dual-channel matched filters [130] and multilabel MRF optimization [131].

4.3 SEGMENTATION

Segmentation refers to the partitioning of an image into disjoint regions that are homogeneous with respect to a chosen property such as luminance, color, and texture [91, p. 178]. Segmentation methods can be roughly classified into the following categories:

- **Histogram thresholding** [52, 64, 74, 76, 85, 86, 97, 132–140]: These methods involve the determination of one or more histogram threshold values that separate the objects from the background [141–145].
- **Clustering** [69, 93, 134, 135, 137, 146–159]: These methods involve the partitioning of a color (feature) space into homogeneous regions using unsupervised clustering algorithms [160–162].
- **Edge based** [75, 97]: These methods involve the detection of edges between the regions using edge operators [163–168].
- **Region based** [52, 57, 87, 88]: These methods involve the grouping of pixels into homogeneous regions using region merging, region splitting, or a combination of the two [169–171].
- **Morphological** [83, 84, 172]: These methods involve the detection of object contours from predetermined seeds using the watershed transform [173–179].
- **Model based** [57]: These methods involve the modeling of images as random fields [180–182] whose parameters are determined using various optimization procedures.
- **Active contours (snakes and their variants)** [52, 74, 85, 86, 147, 152, 183]: These methods involve the detection of object contours using curve evolution techniques [184–186].
- **Fuzzy logic** [52]: These methods involve the classification of pixels using fuzzy rules [187].
- **Supervised learning** [23, 72, 84, 188–191]: These methods involve the application of models obtained by training classifiers such as decision trees, artificial neural networks, and support vector machines.
Several issues should be considered when choosing a segmentation method:

- **Scalar versus vector processing:** Most segmentation methods are designed for scalar images [192-195]. Although numerous vector image segmentation methods have been developed during the past two decades [196-198], their use is hindered by various factors, including excessive computational requirements and difficulty of choosing an appropriate color space [199].

- **Automatic versus semiautomatic:** Some segmentation methods are completely automated, whereas others require human interaction [200]. For example, active contour methods often require the manual delineation of the initial contour, whereas seeded region growing methods [201, 202] require the specification of the initial region seeds. Since there are only a few semiautomatic methods proposed in the literature [52, 203], in this chapter, we focus on automatic methods.

- **Number of parameters:** Most segmentation methods have several parameters whose values need to be determined by the user. In general, the more the number of parameters, the harder the model selection (i.e., the determination of the optimal parameters). Among the methods listed above, histogram thresholding methods are often the simplest ones, as they require very few parameters. In contrast, segmentation methods based on active contours and supervised learning, in particular neural networks, often involve a large number of tunable parameters. Many clustering algorithms, such as k-means and fuzzy c-means, require the number of clusters to be specified by the user [204, 205]. In addition, these algorithms are often highly sensitive to initialization [206-209]. Density-based clustering algorithms [210, 211] such as mean shift [212] and DBSCAN (density based spatial clustering of applications with noise) [213] do not have these limitations. These algorithms, however, involve additional parameters, some of which are difficult to determine.

### 4.4 POSTPROCESSING

The result of the segmentation procedure is typically either a label image or a binary edge map. In order to obtain the lesion border, the segmentation output should be postprocessed. The precise sequence of postprocessing operations depends on the particular choice of the segmentation method. However, certain operations seem to be generally useful. These include the following:

- **Region merging:** The segmentation procedure should ideally produce two regions: the lesion and the background skin. However, since these regions are rarely homogeneous, segmentation methods often
partition them into multiple subregions. In order to obtain a single lesion object, subregions that are part of the lesion should first be identified and then merged. This can be accomplished in several ways:

- **If the black frame of the image has already been removed, the background skin color can be estimated from the corners of the image and the subregions with similar color to the background skin can be eliminated, leaving only those subregions that are part of the lesion [87, 88, 157, 214].**

- **Various color and texture features can be extracted from each region and a classifier can be trained to determine which features effectively discriminate between the regions that are part of the lesion and those that are part of the background skin [6].**

- **The optimal border can be determined by maximizing the normalized texture gradient [215] and minimizing the total similarity [216] between pairs of subregions inside the border and between pairs of subregions outside the border.**

- **Island removal:** Islands (small isolated regions) in the label image can be eliminated using a binary area opening filter [98, p. 113].

- **Border smoothing:** Most segmentation methods produce regions with ragged borders. More natural borders can be obtained by using a variety of operations, including majority filtering [88], morphological filtering [69, 93, 134, 217, 218], and curve fitting [75, 83–85, 97]. Note that it might be better to calculate the border-related features, such as asymmetry, border irregularity, and abrupt border cutoff prior to smoothing.

- **Border expansion:** In several studies, it was observed that the computer-detected borders were mostly contained within the dermatologist-determined borders. This is because the automated segmentation methods tend to find the sharpest pigment change, whereas the dermatologists choose the outmost detectable pigment. The discrepancy between the two borders can be reduced by expanding the computer-detected border using morphological filtering [76, 88], Euclidean distance transform [88], or iterative region growing [68, 136, 139].

### 4.5 EVALUATION

Evaluation of the results seems to be one of the least explored aspects of the border detection task. As in the case of the more general image segmentation problem, there are two major evaluation methods: subjective and objective. The former involves the visual assessment of the border detection results by one or more dermatologists. Since there is no objective measure of quality involved, this technique does not permit parameter tuning or comparisons among automated border detection methods. On the other hand, objective
TABLE 4.1
Definitions of True/False Positive/Negative

<table>
<thead>
<tr>
<th>Detected Pixel</th>
<th>Lesion</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Pixel</td>
<td>True pos. (TP)</td>
<td>False neg. (FN)</td>
</tr>
<tr>
<td>Background</td>
<td>False pos. (FP)</td>
<td>True neg. (TN)</td>
</tr>
</tbody>
</table>

Evaluation involves the quantification of the border detection errors using dermatologist-determined borders. In the rest of this discussion, we refer to the computer-detected borders as automatic borders and those determined by dermatologists as manual borders.

Most of the quantitative evaluation measures are based on the concepts of true/false positive/negative given in Table 4.1 (here actual and detected pixels refer to a pixel in the ground-truth image and the corresponding pixel in the border detection output, respectively). These include the following [219]:

- XOR measure (XOR) \[220\] = \( \frac{FP + FN}{TP + FN} \)
- Sensitivity (SE) = \( \frac{TP}{TP + FN} \) and specificity (SP) = \( \frac{TN}{FP + TN} \)
- False positive rate (FPR) = \( \frac{FP}{FP + TN} \)
- False negative rate (FNR) = \( \frac{FN}{TP + FN} \)
- Precision (PR) = \( \frac{TP}{TP + FP} \)
- F measure (F) = \( 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)
- Accuracy (AC) = \( \frac{TP + TN}{TP + FN + FP + TN} \)
- Error probability (EP) = \( 1 - AC \)
- Jaccard index (J) [221] = \( \frac{TP}{TP + FN + FP} \)
- Hammoude distance (HM)* [222] = \( \frac{FN + FP}{TP + FN + FP} \)

* In his dissertation, Hammoude actually divides by the perimeter of the ground-truth border [222, p. 102]. Dividing by the area of the ground-truth border appears to have been mistakenly attributed to him by Chalana and Kim [223].
In addition, the following relations hold:

- Recall $\equiv$ True positive rate (TPR) $\equiv$ True detection rate (TDR) $\equiv$ Sensitivity
- Sensitivity $= 1 - \text{False negative rate}$
- $\{\text{Specificity} \equiv \text{True negative rate (TNR)}\} = 1 - \text{False positive rate}$
- Precision $\equiv$ Positive predictive value (PPV)
- Hammoude distance $= 1 - \text{Jaccard index}$

In a comprehensive study, Guillod et al. [224] demonstrated that a single dermatologist, even one who is experienced in dermoscopy, cannot be used as an absolute reference for evaluating the accuracy of border detection. In addition, they emphasized that manual borders are not precise, with interdermatologist borders and even borders determined by the same dermatologist at different times showing significant disagreement, so that a probabilistic model of the border is preferred to an absolute gold-standard model. Accordingly, they used 15 sets of borders drawn by five dermatologists over a minimum period of 1 month. A probability image for each lesion was constructed by associating a misclassification probability $p(i, j) = 1 - \frac{n(i, j)}{N}$ with each pixel $(N$, number of observations; $n(i, j)$, number of times pixel $(i, j)$ was selected as part of the lesion). For each automatic border $B$, the detection error was calculated as the mean probability of misclassification over the pixels inside the border, that is, $\sum_{(i,j) \in B} p(i, j)/(TP + FP)$.

Iyatomi et al. [68, 76, 139] modified Guillod et al.’s approach by combining multiple manual borders that correspond to each lesion into one using the majority vote rule. The automatic borders were then compared against these combined ground-truth images. Garnavi et al. [136] used the intersection of the border areas, whereas Garnavi and Aldeen [225] used their union. Celebi et al. [88] compared each automatic border against multiple manual borders independently. Unfortunately, these methods do not accurately capture the variations in manual borders. For example, according to Guillod et al.’s measure, an automated border that is entirely within the manual borders gets a very low error. Iyatomi et al. and Garnavi et al.’s methods discount the variation in the manual borders by reducing them to a single border. On the other hand, Celebi et al.’s approach does not produce a scalar error value, which makes comparisons more difficult.

Celebi et al. [226] proposed the use of an objective measure, the Normalized Probabilistic Rand Index (NPRI) [227], which takes into account the variations in the manual borders. They demonstrated that differences among four of the evaluated border detection methods were in fact smaller than those predicted by the commonly used XOR measure. For a critique of the NPRI, the reader is referred to Peserico and Silletti [228].

Garnavi et al. [229] proposed the weighted performance index (WPI), which is an average of six commonly used weighted measures: sensitivity, specificity, accuracy, precision, similarity, and XOR. Each of these measures
involves a subset of \{TP, FN, FP, TN\} with the following subjective weights: 
\[ w_{TP} = 1.5, \ w_{FN} = w_{TN} = 1, \ \text{and for the calculation of precision, } w_{FP} = 1, \] 
otherwise \[ w_{FP} = 0.5. \] In a follow-up study, Garnavi and Alden [225] utilized 
a constrained nonlinear multivariable optimization scheme to determine these 
weights. Their findings were similar to those of Celebi et al. [226] in that with 
the optimized weights, differences among five automated border detection 
methods were smaller than those predicted using the fixed weights.

Fortina et al. [230] investigated the level of agreement among 12 der-
matologists with differing levels of experience. On a set of 77 images, they 
determined that, on average, the area of disagreement was about 15\% of the 
lesion area. In addition, more experienced dermatologists exhibited greater 
agreement among themselves than with less experienced dermatologists and 
a slight tendency toward tighter borders. They concluded that the agree-
ment among experienced dermatologists could provide an upper bound for 
the accuracy achievable by the automated border detection methods.

None of the above measures quantify the effect of border detection error 
upon the accuracy of the classifier. Loss of classification accuracy due to auto-
matic border error can be measured as the difference between the classification 
accuracy using the manual borders and that using the automatic borders.

4.6 COMPARISONS AND DISCUSSION

Table 4.2 compares recent border detection methods based on their color 
space (and, if applicable, the number of color channels), preprocessing steps, 
and segmentation method. Note that only those methods that are adequately 
described in the literature are included and the postprocessing steps are omit-
ted since they are often not reported in detail. The following observations are 
in order:

- 15/50 methods do not require any preprocessing. It is possible that 
in some studies the authors have omitted the details of this crucial 
phase.
- 25/50 methods operate on multiple color channels. It is, however, 
unclear whether or not the use of color information improves the 
accuracy of border detection [133].
- 25/50 methods use a smoothing filter. Some segmentation methods, 
such as those based on thresholding, are inherently robust against 
noise, while others, for example, active contour-based methods, are 
highly sensitive to noise. Note that smoothing can also be performed 
as part of the hair removal step [99].
- Clustering (19/50) and thresholding (18/50) are the most popular 
segmentation methods, possibly due to the availability of simple and 
robust algorithms.

Table 4.3 compares the border detection methods based on their evalua-
tion methodology: the number of human experts who determined the manual
### TABLE 4.2
Characteristics of Border Detection Methods

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Color Space (# Channels)</th>
<th>Preprocessing</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[132]</td>
<td>2014</td>
<td>KL{RGB} (1)</td>
<td>MF, IC, HR</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[133]</td>
<td>2013</td>
<td>B/RGB (1)</td>
<td>nr</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[147]</td>
<td>2013</td>
<td>Luminance</td>
<td>nr</td>
<td>Active contours + clustering</td>
</tr>
<tr>
<td>[134]</td>
<td>2013</td>
<td>B/RGB (1)</td>
<td>OF</td>
<td>Clustering + thresholding</td>
</tr>
<tr>
<td>[135]</td>
<td>2013</td>
<td>LAB (3)</td>
<td>HR, CR, GF</td>
<td>Clustering + thresholding</td>
</tr>
<tr>
<td>[188]</td>
<td>2013</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>[146]</td>
<td>2013</td>
<td>RGB (3)</td>
<td>HR</td>
<td>Clustering + supervised learning</td>
</tr>
<tr>
<td>[76]</td>
<td>2013</td>
<td>B or G/RGB (1)</td>
<td>MF, IC, CE, BFR</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[86]</td>
<td>2012</td>
<td>J/JCh (1)</td>
<td>CE, IC, HR</td>
<td>Thresholding + active contours</td>
</tr>
<tr>
<td>[97]</td>
<td>2012</td>
<td>LAB (3)</td>
<td>CE, IC, HR, BFR</td>
<td>Thresholding + edge detection</td>
</tr>
<tr>
<td>[148]</td>
<td>2012</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Clustering</td>
</tr>
<tr>
<td>[149]</td>
<td>2012</td>
<td>RGB (3)</td>
<td>GF</td>
<td>Clustering + supervised learning</td>
</tr>
<tr>
<td>[150]</td>
<td>2012</td>
<td>HSI (3)</td>
<td>ADF, CE</td>
<td>Clustering</td>
</tr>
<tr>
<td>[74]</td>
<td>2011</td>
<td>Luminance</td>
<td>MF, CE, IC, HR</td>
<td>Thresholding + active contours</td>
</tr>
<tr>
<td>[75]</td>
<td>2011</td>
<td>Luminance</td>
<td>CE, IC, HR</td>
<td>Edge detection</td>
</tr>
<tr>
<td>[136]</td>
<td>2011</td>
<td>Misc.</td>
<td>GF, CE</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[137]</td>
<td>2011</td>
<td>L/HSL (1)</td>
<td>ALL</td>
<td>Thresholding + clustering</td>
</tr>
<tr>
<td>[72]</td>
<td>2011</td>
<td>Luminance</td>
<td>CE</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>[151]</td>
<td>2011</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Clustering</td>
</tr>
<tr>
<td>[84]</td>
<td>2011</td>
<td>B/RGB (1)</td>
<td>IC, BFR, HR, ALL</td>
<td>Supervised learning + morphological learning</td>
</tr>
<tr>
<td>[66]</td>
<td>2011</td>
<td>LAB (3)</td>
<td>IC, GF</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>[23]</td>
<td>2011</td>
<td>LAB (3)</td>
<td>GF</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>[152]</td>
<td>2011</td>
<td>Luminance</td>
<td>nr</td>
<td>Active contours + clustering</td>
</tr>
<tr>
<td>[189]</td>
<td>2011</td>
<td>LUV (3)</td>
<td>MF, IC</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>[153]</td>
<td>2010</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Clustering</td>
</tr>
<tr>
<td>[183]</td>
<td>2010</td>
<td>L/HSL (1)</td>
<td>OF, ALL</td>
<td>Active contours</td>
</tr>
<tr>
<td>[64]</td>
<td>2010</td>
<td>KL{RGB} (1)</td>
<td>IC, HR</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[83]</td>
<td>2010</td>
<td>B/RGB (1)</td>
<td>IC, BFR, HR, ALL</td>
<td>Morphological learning</td>
</tr>
<tr>
<td>[190]</td>
<td>2010</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>[52]</td>
<td>2009</td>
<td>ENT{RGB} (1)</td>
<td>OF, MF, BFR</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[52]</td>
<td>2009</td>
<td>Luminance</td>
<td>OF, MF, BFR</td>
<td>Thresholding + active contours</td>
</tr>
<tr>
<td>[52]</td>
<td>2009</td>
<td>LAB (3)</td>
<td>OF, MF, BFR</td>
<td>Fuzzy logic + region-based learning</td>
</tr>
<tr>
<td>[154]</td>
<td>2009</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Clustering</td>
</tr>
<tr>
<td>[138]</td>
<td>2009</td>
<td>Luminance</td>
<td>nr</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[155]</td>
<td>2009</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Clustering</td>
</tr>
<tr>
<td>[88]</td>
<td>2008</td>
<td>RGB (3)</td>
<td>MF, BFR</td>
<td>Region-based learning</td>
</tr>
<tr>
<td>[69]</td>
<td>2008</td>
<td>IHP{RGB} (1)</td>
<td>CE, HR</td>
<td>Clustering</td>
</tr>
<tr>
<td>[156]</td>
<td>2008</td>
<td>LAB (3)</td>
<td>nr</td>
<td>Clustering</td>
</tr>
<tr>
<td>[87]</td>
<td>2007</td>
<td>RGB (3)</td>
<td>MF, ALL</td>
<td>Region-based learning</td>
</tr>
<tr>
<td>[139]</td>
<td>2006</td>
<td>B/RGB (1)</td>
<td>GF</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[157]</td>
<td>2006</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Clustering</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 4.2 (Continued)
Characteristics of Border Detection Methods

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Color Space (# Channels)</th>
<th>Preprocessing</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[85]</td>
<td>2005</td>
<td>Luminance</td>
<td>GF, ALL</td>
<td>Thresholding + active contours</td>
</tr>
<tr>
<td>[158]</td>
<td>2003</td>
<td>LUV (3)</td>
<td>MF</td>
<td>Clustering</td>
</tr>
<tr>
<td>[159]</td>
<td>2002</td>
<td>KL{LAB} (2)</td>
<td>GF</td>
<td>Clustering</td>
</tr>
<tr>
<td>[140]</td>
<td>2001</td>
<td>KL{RGB} (1)</td>
<td>MF</td>
<td>Thresholding</td>
</tr>
<tr>
<td>[191]</td>
<td>2000</td>
<td>I/HSI (1)</td>
<td>MF</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>[93]</td>
<td>1999</td>
<td>KL{LUV} (2)</td>
<td>MF</td>
<td>Clustering</td>
</tr>
<tr>
<td>[172]</td>
<td>1999</td>
<td>LAB (3)</td>
<td>ADF, HR</td>
<td>Morphological</td>
</tr>
<tr>
<td>[57]</td>
<td>1998</td>
<td>RGB (3)</td>
<td>nr</td>
<td>Region-based</td>
</tr>
<tr>
<td>[57]</td>
<td>1998</td>
<td>KL{RGB}(1)</td>
<td>nr</td>
<td>Model-based</td>
</tr>
</tbody>
</table>

Note: nr, not reported; KL{C}, KL transform of the color space C; GF, Gaussian filter; MF, mean filter; MF, median filter; ADF, anisotropic diffusion filter; OF, other filter; CE, contrast enhancement; IC, illumination correction; BFR, black frame removal; HR, hair removal; ALL, approximate lesion localization.

borders, the number of images used in the evaluations (and, if available, the diagnostic distribution of these images), the number of automated methods used in the comparisons (only comparisons against published border detection methods are considered), and the measure used to quantify the accuracy of border detection. It can be seen that

- Only 15/50 studies used borders determined by multiple dermatologists.
- Only 26/50 studies reported the diagnostic distribution of their test images. This information is valuable given that not every diagnostic category is equally challenging from a border detection perspective. For example, it is often more difficult to detect the borders of melanomas and dysplastic nevi due to their irregular and fuzzy border structure.
- 16/50 studies did not compare their results to those of other studies. This is partly due to the unavailability of public border detection software, as well as the scarcity of public dermoscopy image databases.
- Recent studies used objective measures to determine the quality of their results, whereas earlier studies relied on visual assessment. XOR is still the most popular quantitative evaluation measure despite the fact that it is biased against small lesions and it is not trivial to extend this measure to capture the variations in manual borders [224, 230, 231].
### TABLE 4.3
Evaluation of Border Detection Methods (b, Benign; m, Malignant; HS, Hausdorff Distance [234])

<table>
<thead>
<tr>
<th>Ref.</th>
<th># Experts</th>
<th># Images (Distribution)</th>
<th># Comp.</th>
<th>Evaluation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>[132]</td>
<td>4</td>
<td>60</td>
<td>2</td>
<td>XOR</td>
</tr>
<tr>
<td>[133]</td>
<td>1</td>
<td>90 (67 b/23 m)</td>
<td>9</td>
<td>XOR</td>
</tr>
<tr>
<td>[147]</td>
<td>3</td>
<td>100 (70 b/30 m)</td>
<td>1</td>
<td>SE, SP, XOR</td>
</tr>
<tr>
<td>[134]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>2</td>
<td>TPR, FPR, EP</td>
</tr>
<tr>
<td>[135]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>3</td>
<td>TPR, FPR, EP</td>
</tr>
<tr>
<td>[146]</td>
<td>1</td>
<td>125 (57 b/68 m), 181 (128 b/53 m)</td>
<td>1</td>
<td>XOR, J, HS</td>
</tr>
<tr>
<td>[76]</td>
<td>4-5</td>
<td>426 (351 b/75 m)</td>
<td>5</td>
<td>PR, RE, F</td>
</tr>
<tr>
<td>[86]</td>
<td>1</td>
<td>270 (80 b/190 m)</td>
<td>0</td>
<td>XOR</td>
</tr>
<tr>
<td>[97]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>3</td>
<td>SE, SP, EP</td>
</tr>
<tr>
<td>[148]</td>
<td>1</td>
<td>50</td>
<td>0</td>
<td>SE, SP</td>
</tr>
<tr>
<td>[149]</td>
<td>1</td>
<td>116</td>
<td>6</td>
<td>SE, SP, AUC</td>
</tr>
<tr>
<td>[150]</td>
<td>3</td>
<td>57</td>
<td>2</td>
<td>Tanimoto [235]</td>
</tr>
<tr>
<td>[74]</td>
<td>1</td>
<td>320 (210 b/110 m)</td>
<td>2</td>
<td>HM, SE, FPR, HS, XOR</td>
</tr>
<tr>
<td>[75]</td>
<td>1</td>
<td>240 (170 b/70 m)</td>
<td>3</td>
<td>SE, SP, EP</td>
</tr>
<tr>
<td>[136]</td>
<td>4</td>
<td>85</td>
<td>3</td>
<td>ACC, PR, SE, SP, AUC</td>
</tr>
<tr>
<td>[137]</td>
<td>3</td>
<td>100 (70 b/30 m)</td>
<td>0</td>
<td>PR, RE, XOR</td>
</tr>
<tr>
<td>[72]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>0</td>
<td>XOR</td>
</tr>
<tr>
<td>[151]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>1</td>
<td>PR, RE, XOR</td>
</tr>
<tr>
<td>[84]</td>
<td>3</td>
<td>100 (70 b/30 m)</td>
<td>4</td>
<td>XOR</td>
</tr>
<tr>
<td>[66]</td>
<td>1</td>
<td>116</td>
<td>6</td>
<td>SE, SP</td>
</tr>
<tr>
<td>[23]</td>
<td>1</td>
<td>118</td>
<td>5</td>
<td>SE, SP</td>
</tr>
<tr>
<td>[152]</td>
<td>3</td>
<td>100 (70 b/30 m)</td>
<td>0</td>
<td>SE, SP</td>
</tr>
<tr>
<td>[189]</td>
<td>3</td>
<td>122 (100 b/22 m)</td>
<td>2</td>
<td>SE, SP, HS</td>
</tr>
<tr>
<td>[153]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>0</td>
<td>PR, RE, XOR</td>
</tr>
<tr>
<td>[183]</td>
<td>1</td>
<td>50</td>
<td>0</td>
<td>PR, RE, XOR</td>
</tr>
<tr>
<td>[64]</td>
<td>3</td>
<td>91 (80 b/11 m)</td>
<td>0</td>
<td>Visual</td>
</tr>
<tr>
<td>[85]</td>
<td>3</td>
<td>100 (70 b/30 m)</td>
<td>3</td>
<td>XOR</td>
</tr>
<tr>
<td>[190]</td>
<td>1</td>
<td>178</td>
<td>3</td>
<td>XOR</td>
</tr>
<tr>
<td>[52]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>4</td>
<td>HM, SE, FPR, HS</td>
</tr>
<tr>
<td>[154]</td>
<td>1</td>
<td>120</td>
<td>1</td>
<td>PR, RE, F, XOR</td>
</tr>
<tr>
<td>[138]</td>
<td>0</td>
<td>nr</td>
<td>0</td>
<td>Visual</td>
</tr>
<tr>
<td>[155]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>3</td>
<td>SE, SP</td>
</tr>
<tr>
<td>[88]</td>
<td>3</td>
<td>90 (67 b/23 m)</td>
<td>4</td>
<td>XOR</td>
</tr>
<tr>
<td>[69]</td>
<td>1</td>
<td>100 (70 b/30 m)</td>
<td>3</td>
<td>XOR</td>
</tr>
<tr>
<td>[156]</td>
<td>1</td>
<td>67</td>
<td>0</td>
<td>XOR</td>
</tr>
<tr>
<td>[87]</td>
<td>2</td>
<td>100 (70 b/30 m)</td>
<td>3</td>
<td>XOR</td>
</tr>
<tr>
<td>[139]</td>
<td>5</td>
<td>319 (244 b/75 m)</td>
<td>1</td>
<td>PR, RE</td>
</tr>
<tr>
<td>[157]</td>
<td>nr</td>
<td>117</td>
<td>3</td>
<td>SE, SP</td>
</tr>
<tr>
<td>[85]</td>
<td>2</td>
<td>100 (70 b/30 m)</td>
<td>1</td>
<td>XOR</td>
</tr>
<tr>
<td>[158]</td>
<td>0</td>
<td>nr</td>
<td>0</td>
<td>nr</td>
</tr>
<tr>
<td>[159]</td>
<td>0</td>
<td>600</td>
<td>0</td>
<td>Visual</td>
</tr>
<tr>
<td>[140]</td>
<td>0</td>
<td>nr</td>
<td>0</td>
<td>nr</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 4.3 (Continued)
Evaluation of Border Detection Methods (b, Benign; m, Malignant; HS, Hausdorff Distance [234])

<table>
<thead>
<tr>
<th>Ref.</th>
<th># Experts</th>
<th># Images (Distribution)</th>
<th># Comp.</th>
<th>Evaluation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>[191]</td>
<td>1</td>
<td>30</td>
<td>0</td>
<td>Visual</td>
</tr>
<tr>
<td>[93]</td>
<td>1</td>
<td>400</td>
<td>0</td>
<td>Visual</td>
</tr>
<tr>
<td>[172]</td>
<td>1</td>
<td>300</td>
<td>0</td>
<td>Visual</td>
</tr>
<tr>
<td>[57]</td>
<td>1</td>
<td>57</td>
<td>5</td>
<td>XOR</td>
</tr>
</tbody>
</table>

Unsolved problems in border detection include the following:

- Incorporation of textural information into the border detection process [87, 156, 232]
- Detection of borders in nonmelanocytic lesions such as seborrheic keratoses and basal/squamous cell carcinomas [76]
- Fusion of multiple border detection methods [133]
- Adaptive expansion of the automatic borders [68, 76, 88, 136, 139]
- Development of clinically oriented evaluation measures that take into account the variations in multiple manual borders

We believe that in a systematic border detection study

1. The image acquisition procedure should be described in sufficient detail.
2. The test image set should be selected randomly from a large and diverse image database.
3. The test image set should be large enough to ensure statistically valid conclusions.
4. The test image set should not be used to train/tune the border detection method.
5. The diagnostic distribution of the test image set should be specified.
6. Algorithms with reasonable computational requirements should be used.
7. The results should be evaluated using borders determined by multiple dermatologists.
8. The results should be compared to those of other published studies.
9. The border detection method should be described in sufficient detail.
10. The implementation of the border detection method should be made publicly available.

Note that all of the aforementioned criteria except (4), (6), (9), and (10) can be satisfied by using a public dermoscopy image set. Therefore, the creation of such a benchmark database should be prioritized in order to improve the
quality of future border detection studies. A promising step in this direction is the publicly available PH² database [233], which contains 200 annotated dermoscopy images.

ACKNOWLEDGMENTS

This publication was made possible by grants from the U.S. National Science Foundation (0959583, 1117457) and National Natural Science Foundation of China (61050110449, 61073120).

REFERENCES


54. M. E. Celebi, H. Iyatomi, and G. Schaefer, Contrast Enhancement in Dermoscopy Images by Maximizing a Histogram Bimodality Measure, in


A State-of-the-Art Survey on Lesion Border Detection in Dermoscopy Images


A State-of-the-Art Survey on Lesion Border Detection in Dermoscopy Images


150. Z. Liu, J. Sun, M. Smith, L. Smith, and R. Warr, Unsupervised Sub- 
151. S. Suer, S. Kockara, and M. Mete, An Improved Border Detection in Der- 
152. H. Zhou, G. Schaefer, M. E. Celebi, F. Lin, and T. Liu, Gradient Vector Flow 
with Mean Shift for Skin Lesion Segmentation, *Computerized Medical Imaging 
153. S. Kockara, M. Mete, V. Yip, B. Lee, and K. Aydin, A Soft Kinetic Data Struc- 
Random Walker Segmentation for Skin Lesions in a Supervised Setting, in *Proceedings of the 12th International Conference on Medical Image Computing 
155. H. Zhou, G. Schaefer, A. Sadka, and M. E. Celebi, Anisotropic Mean Shift 
Based Fuzzy C-Means Segmentation of Dermoscopy Images, *IEEE Journal of 
Spatially Constrained Segmentation of Dermoscopy Images, in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to 
157. R. Melli, C. Grana, and R. Cucchiara, Comparison of Color Clustering 
158. H. Galda, H. Murao, H. Tamaki, and S. Kitamura, Skin Image Segmentation 
159. R. Cucchiara, C. Grana, S. Seidenari, and G. Pellacani, Exploiting Color and 


A State-of-the-Art Survey on Lesion Border Detection in Dermoscopy Images


