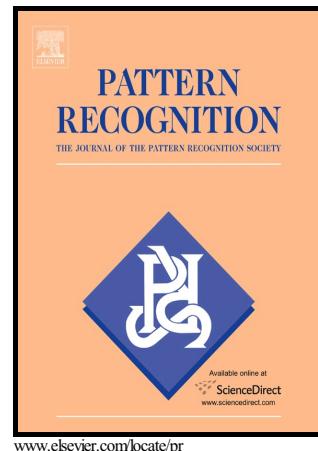


Author's Accepted Manuscript

A Simple Weighted Thresholding Method for the Segmentation of Pigmented Skin Lesions in Macroscopic Images

Maciel Zortea, Eliezer Flores, Jacob Scharcanski



PII: S0031-3203(16)30344-2
DOI: <http://dx.doi.org/10.1016/j.patcog.2016.10.031>
Reference: PR5939

To appear in: *Pattern Recognition*

Received date: 21 January 2016

Revised date: 11 August 2016

Accepted date: 27 October 2016

Cite this article as: Maciel Zortea, Eliezer Flores and Jacob Scharcanski, A Simple Weighted Thresholding Method for the Segmentation of Pigmented Skin Lesions in Macroscopic Images, *Pattern Recognition* <http://dx.doi.org/10.1016/j.patcog.2016.10.031>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain

A Simple Weighted Thresholding Method for the Segmentation of Pigmented Skin Lesions in Macroscopic Images

Maciel Zortea^a, Eliezer Flores^a, Jacob Scharcanski^{a,*}

^a*Instituto de Informática, Universidade Federal do Rio Grande do Sul.
Avenida Bento Gonçalves 9500, Porto Alegre 91501-970, RS, Brazil*

Abstract

This work proposes a simple and yet effective thresholding method to segment pigmented skin lesions in macroscopic photographs automatically. Segmentation is one of the first steps in computer-aided diagnosis of skin cancers. Therefore, an accurate segmentation may play an important clinical role. We develop an algorithm that searches for a thin rectangular-shaped region near the image borders that is likely to contain mostly skin pixels. Segmentation is obtained by adapting Otsu's thresholding method by combining independent threshold estimates computed from histograms of different parts of a new intensity image designed to discriminate lesions from background skin. The proposed approach exploits the fact that the object of interest is approximately centered in the input photograph. A cross-diagonal sampling scheme helps to balance the size of the classes when the area of the lesion and the area of the surrounding skin are very different. A post-processing stage that includes morphological filtering and a weighted scheme to select the most salient object follows. The experimental results suggest that the method potentially can be used successfully to segment atypical nevi and melanomas in lesions with a highly heterogeneous background skin. The proposed algorithm is of interest for use in clinical settings as part of a CAD system.

Keywords: Segmentation, Thresholding, Pigmented skin lesions, Melanoma, Computer-aided diagnosis.

1. Introduction

The World Health Organization estimates between 2 and 3 million non-melanoma skin cancers and 132,000 melanoma skin cancers occur globally each year. One in every three cancers diagnosed is a skin cancer, and the global incidence of melanoma continues to rise¹. Early detection is of paramount importance for melanoma treatment. At this stage, the prognosis for the patient is excellent as it can be cured by simple excision. However, melanoma detection in early stages is difficult because they resemble common

*Corresponding author. Tel.: +55 51 3308 7128.

Email addresses: maciel.zortea@ufrgs.br (Maciel Zortea), esflores@inf.ufrgs.br (Eliezer Flores), jacobs@inf.ufrgs.br (Jacob Scharcanski)

¹<http://www.who.int/uv/faq/skincancer/en/index1.html>, accessed May 2015.

nevi. Algorithms and computer-aided diagnosis systems (CADs) have been developed to support health care professionals in this challenging task.

Dermoscopy, a non-invasive examination technique that uses a hand-held magnifier illuminated by a light source, allows doctors to see and evaluate colors and microstructures deeper in the skin lesion, not visible to naked eye. This technique has proven to increase diagnostic accuracy in hands of trained practitioners. Currently, most CADs systems use digital dermoscopy [1]. These CADs systems, however, have yet to demonstrate their benefit for experienced dermatologists in their clinical settings [2].

Diagnosis accuracy based on dermoscopy decreases in hands of non trained personal [3]. Therefore, having CADs that would be able to analyze macroscopic (clinical) digital photographs may complement dermoscopy-based systems and promote early melanoma detection. The use of standard digital cameras for acquiring images is appealing. Cheap and portable, they are increasingly accessible to the general population. A recent example is the use of cameras in smartphones to promote health applications. While experts remain cautious about the utility and safety of such “apps”, Kassianos et al. [4] noted that several of these have “the potential to take and store images of skin lesions, either for review by a dermatologist, or for self-monitoring to identify change, an important predictor of melanoma”. Some also provide risk assessment to patients, estimating the probability that a lesion was malignant or benign.

Image segmentation is a key component of CADs for the analysis of medical images [1, 5]. In case of skin lesions, segmentation amounts to finding the border that separates the lesion area from the surrounding skin. Once the lesion is detected, features describing the asymmetry, border, color, and structures present in the lesion can be computed and used to train statistical models to predict the diagnosis. Therefore, obtaining an accurate segmentation of the lesion is important, especially to quantify shape and border features [6].

1.1. Related work

Many algorithms have been proposed to segment pigmented skin lesions automatically. Morphological differences in the appearance of pigmented skin lesions in macroscopic and dermoscopic images directly influence the choice of method for lesion segmentation [1]. Otsu’s thresholding [7] is a simple and fast approach and frequently has been used as a component in several algorithms designed for analysis of macroscopic photographs [8, 9, 10, 11, 12]. Thresholding methods determine a cutting value to turn a grayscale representation of an image into a binary image [13]. As an illustration, Ruiz et al. [8] applied Otsu’s thresholding to grayscale images as part of a decision support system for the diagnosis of melanoma.

Alcón et al. [14] argue that thresholding does not suffer the drawbacks of edge detection-based algorithms that perform poorly when applied to skin images, mainly because of the presence of fine details, hair, or both on the skin. Snakes or active contours [15] are an alternative to thresholding to detect the complex shape of the suspicious melanoma lesions. However, the performance of snakes or active contours methods may be more sensitive to the presence of hair.

Considering the aforementioned observations, Cavalcanti and Scharcanski [11] proposed a method based on Otsu's thresholding applied to a color space projection that attempts to maximizes the separability between non-lesion and lesion pixels, followed by the Chan-Vese method. The algorithm was tested on the same data set used in this study. This approach was shown to provide segmentation results that were more accurate than previous methods, including a multi-direction gradient vector flow (GVF) snake-based scheme [16], a thresholding method inspired by Otsu's algorithm [14, 17], thresholding a multichannel image representation to distinguish the skin lesion from the background areas [10], independent component analysis (ICA) followed by segmentation using the Chan-Vese method [9], and k-means on the image patches projected in a learned dictionary obtained via non-negative matrix factorization (NMF) [18].

Recently, Flores and Scharcanski [19] used a dictionary-based technique to represent a compact description of image patches. The authors proposed an unsupervised version of the information-theoretic dictionary learning (ITDL) method [20], and segmented the skin lesions using normalized graph cuts [21].

In addition to color information, another interesting alternative is to incorporate textural information in the segmentation process [1]. For instance, Glaister et al. [12] used the Otsu's thresholding method combined with statistical region merging [22] to segment skin lesions in macroscopic images.

1.2. Objective and structure of this work

We propose a new automated algorithm for the segmentation of pigmented skin lesions in macroscopic photographs, namely *Simple Weighted Otsu Thresholding* (SWOT), which adapts the original Otsu's thresholding functional considering (i) all samples in the image and (ii) samples located in the cross-diagonals of the image. The resulting threshold is weighted with an independent estimate obtained from a peripheral region containing mainly skin pixels. While strikingly simple, SWOT is shown to yield segmentation accuracy in data set used by Alcón et al. [14] that are very similar to or better than previous methods. Competitive results also have been obtained on the Dermquest data set. Furthermore, it is fast and could operationally run on equipment with limited processing and memory resources. In addition, SWOT is easy to use and straightforward to extend to multispectral images.

The remaining of the paper is organized as follows. Section 2 reviews Otsu's thresholding method. Section 3 briefly reviews the preprocessing algorithm used to attenuate shadows in the input photographs. Section 4 presents the novel contributions of this work, which includes (i) a description of an automated procedure to select a thin region near the border of the image, that is likely to contain mostly skin pixels. Samples from this region are used to estimate the color of the background skin, used to compute the proposed intensity image, that is the input to the (ii) new thresholding algorithm. Section 5 describes the post-processing scheme used to eliminate possible artifacts, selecting the continuous object that is larger and closer to the center of the image. Section 6 investigates how the parameters setting affect the accuracy of the proposed segmentation, and how the accuracy compares with state-of-the-art algorithms. After discussing

75 results in Section 7, we conclude in Section 8.

2. Threshold selection

Given the pixels of a grayscale image, represented by L gray levels $\{1, 2, 3, \dots, L\}$, segmentation aims at splitting the pixels into two classes C_1 and C_2 (e.g. background and foreground). In this context, C_1 denotes the pixels with levels $\{1, 2, 3, \dots, t'\}$ and C_2 denotes the pixels with levels $\{t' + 1, \dots, L\}$.

Otsu's thresholding method is a popular method to solve such a problem. Working on the histogram of an image, Otsu's method exhaustively searches for the partition threshold t' that minimizes the intra-class variances of the classes. This is equivalent to maximizing the inter-class variance:

$$t' = \operatorname{argmax}_{t \in \{1, \dots, L\}} \sigma^2(t) \quad (1)$$

where

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

and $\omega_i(t)$ and $\mu_i(t)$ are the prior class probability, and mean value when the threshold is placed at t . These quantities are computed from the normalized histogram bin counts $p(i)$ as:

$$\omega_1(t) = \sum_{i=1}^{t'} p(i), \quad (3)$$

$$\mu_1(t) = \frac{1}{\omega_1(t)} \sum_{i=1}^{t'} p(i)x(i). \quad (4)$$

80 Similarly, $\omega_2(t)$ and $\mu_2(t)$ are computed using the histogram bins $x(i) > t$, with $\sum_{i=t+1}^L p(i) = 1$.

The ability of Otsu's and other alternatives such as Kittler and Illingworth's (K&I) method [23] to produce a good segmentation depends on the ratio of the size of the classes, and their relative variances. According to Kurita et al. [24], when the number of samples of each class is about the same and the classes can be assumed to have equal variance, then Otsu's method is better because the number of free parameters 85 of the model to be estimated is smaller than the others, and the estimator is more stable. Conversely, Kittler and Illingworth's method should be used if the populations of each one of the classes are different, and the distributions have different variances. These observations, derived from ideal Gaussian scenarios, have practical implications for the segmentation of skin lesions. Challenges arise when the skin lesion cover a small fraction of the image (i.e., the prior for the lesion is much smaller than the prior for the background 90 skin), as in the example shown later in Fig. 2. Bimodal distribution may not describe well the histogram of the lesions, and deviations from Gaussianity are expected for the proposed intensity image described in Section 4.2. Therefore, we base our method on Otsu's thresholding that has less free parameters to be optimized, for its simplicity, and reasonable robustness to deviation from Gaussianity.

3. Pre-processing: shadow attenuation

Illumination variations hamper the segmentation of skin lesions in macroscopic photographs. Therefore, the input images to be segmented using all methods tested in Section 6 were first corrected for shadow attenuation using the approach proposed in [25], which compensates for non-uniform illumination. The algorithm assumes that the Value component $V(x, y)$ of the images in the HSV color space can be modeled by a quadratic formulation $z(x, y)$, as a function of the spatial coordinates (x, y) of the image:

$$z(x, y) = P_1x^2 + P_2y^2 + P_3xy + P_4x + P_5y + P_6. \quad (5)$$

The coefficients P_i for $i = 1, \dots, 6$ are obtained by minimizing the error term $\varepsilon(x, y) = [V(x, y) - z(x, y)]^2$, computed using only the pixels at the four small patches located at the corners of the image. A shadow-attenuated Value competent $V'(x, y)$ is computed as:

$$V'(x, y) = \frac{V(x, y)}{z(x, y)}, \quad (6)$$

and the new component $V'(x, y)$ replaces the original $V(x, y)$ component in the HSV image. The resulting image is converted back to RGB.

4. SWOT segmentation

The proposed SWOT segmentation is characterized by five processing steps:

1. Automatic selection of candidate skin pixels;
2. Computation of an intensity image for thresholding;
3. Noise reduction using median filtering;
4. Threshold estimation;
5. Post-processing.

Fig. 1(a) shows a flowchart of the proposed SWOT segmentation method showing the main processing steps. The boxes in gray show the novel contributions of our study. Sections 4.1– 4.3 detail the processing steps, which are applied to images corrected by shadow attenuation, followed by the motivation in Section 4.4. Two worked examples, explaining details of the calculation of the thresholds, are included in Figs. 2–3. The final segmentation mask is obtained using the post-processing presented in Section 5.

4.1. Automatic selection of candidate skin pixels

We search for a thin rectangular-shaped region \mathcal{R}_s with its sides parallel to the outer borders of the image (like the regions delimited by black lines shown in Fig. 2(a) and Fig. 3(a)) that is likely to contain mostly

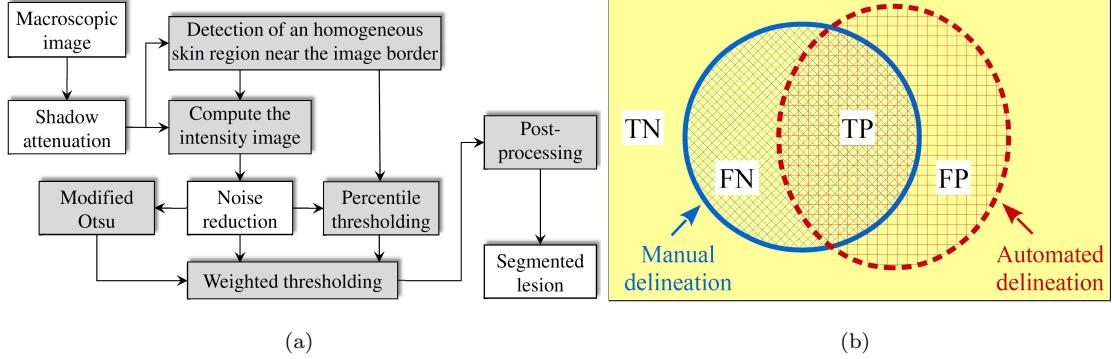


Figure 1: Schematics of the proposed skin lesion segmentation method: (a) Flowchart of the proposed SWOT segmentation method showing the main processing modules. The boxes in gray show the novel contributions. (b) Areas corresponding to true positive (TP) and false positive (FP), true negative (TN), and false negative (FN) detections given a manual and automated segmentation.

skin pixels. This borderline region is used to estimate a vector representing the color of the background skin and later to estimate the threshold (Section 4.3).

Aiming at selecting a homogenous region containing mainly skin pixels, we measure iteratively the ratio

115 between the standard deviation σ and the mean μ of all pixels in the tested borderline region, and select the location that minimizes the total coefficient of variation $\sum_{i=1}^3 \sigma_i/\mu_i$ in the CIELAB color channels $i = \{1, 2, 3\}$.

In some cases, the lesion area can be as homogeneous as the skin. Therefore, the search is constrained to avoid wrongly placing the desired skin region inside the lesion borders. We use an iterative search method 120 that starts from the border and shrinks the search area up to η (where $\eta = 1/4$ in our experiments) of the distance of the smaller side of the image, moving in non-overlapping steps of size $s = 0.02\mathcal{L}$, where \mathcal{L} is the size of the larger side of the image. The center of the region \mathcal{R}_s remains fixed. The search area is changed by shrinking the borderline region towards the center of the image. This scheme proved to be a simple, fast, and reliable. Additional precautions would be needed if the lesions covered most of the imaged area².

125 4.2. Computation of an intensity image for thresholding

Once the region \mathcal{R}_s containing mainly skin is found, a rough estimate of the skin color $[l_s \ a_s \ b_s]^T$ is obtained by computing the component-wise median of the pixels located in \mathcal{R}_s in the CIELAB color space. With this estimate, the intensity image \mathcal{I} , that will be the input to our thresholding method, is computed

²SWOT is not intended to be used when the lesion extends beyond the image borders. In such case, the user should be advised to take another shot by holding the camera at a greater distance from the skin surface.

as the Euclidian distance of each CIELAB pixel $[l \ a \ b]^T$ located at (x, y) to the skin color:

$$\mathcal{I}(x, y) = [(l(x, y) - l_s)^2 + (a(x, y) - a_s)^2 + (b(x, y) - b_s)^2]^{1/2}. \quad (7)$$

- ₁₃₀ We choose the CIELAB because of its relative perceptual uniformity. Large (small) differences between any two colors correspond approximately to long (short) Euclidian distances between the colors in the three-dimensional CIELAB space. The samples in \mathcal{R}_s will also be used to compute the threshold in Eq. 11.

Median filtering is applied to the intensity image $\mathcal{I}(x, y)$ to reduce spatial noise. Median filtering attenuates the presence of hair and strong highlights in the skin lesion image during image acquisition. A squared filter of side $ws = 0.01\mathcal{L}$ was used in our experiments.

4.3. Threshold estimation

Otsu's thresholding algorithm often provided overestimated thresholds when applied directly to our intensity images, resulting in undersegmented lesions. Besides, the threshold estimates tend to be unreliable when the area of the lesion covers only a small fraction of the input photographs. Therefore, here the proposed threshold t' is a weighted combination of independent scores t_h , that are computed by modifying Otsu's functional to be more robust to class imbalances (differences in prior probabilities of the lesion and background skin areas), and t_s , which is an independent and conservative ("one-class") threshold based on the estimated background skin pixels:

$$t' = \alpha \cdot t_h + (1 - \alpha) \cdot t_s, \quad (8)$$

where the constant α is expected to balance between the overestimated and conservative thresholds.

- To attenuate the effect of class imbalances, we reformulate Otsu's functional proposing that the choice of the threshold should simultaneously optimize the inter-class variance σ^2 not only of the histogram of all ₁₄₀ pixels of the image (in such case the entire lesion is considered but the prior of the lesion might be too small in comparison with the prior of the surrounding skin, since their areas can be very different), but also the histograms of a set of regions $k = 1, \dots, n$ that should be adequately chosen to provide a better balance between the number of lesion and surrounding skin samples. Specifically,

$$t_h = \operatorname{argmax}_{t \in \{1, \dots, L\}} \frac{\sigma_1^2(t)}{\phi_1} + \frac{\sigma_2^2(t)}{\phi_2} + \dots + \frac{\sigma_n^2(t)}{\phi_n}, \quad (9)$$

where the inter-class variance σ_k^2 is computed as in Eq. 2, and the normalizing constant ϕ_k is given by the scaled Euclidian norm:

$$\phi_k = n \left\| [\sigma_k^2(1), \dots, \sigma_k^2(L)] \right\|, \quad (10)$$

which adjusts for different priors in the sample sets (sample sizes).³

³Note that ϕ_k in Eq. 9 is constant for t . In general, $\phi_1 \neq \phi_2 \neq \dots \neq \phi_n$.

145 Several strategies may be formulated for choosing the number and location of the n regions in Eq. 9. Without loss of generality, we will limit our discussion to the case of $n = 2$ regions. For convenience, the index $\{k = 1\}$ in Eq. 9 will correspond to using all the intensity pixel in the image. Similarly, $\{k = 2\}$ will refer to using only the subset of pixels located in the cross-diagonals of the image (see Figs. 2 and 3 for two examples). The use of $n = 2$ components in Eq. 9 assures that all pixels in the image are considered in
150 the optimization schema while the contribution of the pixels located in the cross-diagonals are emphasized. Our motivation for the use of cross-diagonals will be provided later in Section 4.4. In short, Eq. 9 aims at increasing the balance between classes by re-sampling, decreasing the number of pixels in the over-sampled class (i.e., surrounding skin), while increasing relatively the number of pixels in the under-sampled class (i.e., skin lesion) to improve the balance between the lesion and surrounding skin classes.

The other threshold t_s , needed in Eq. 8, is modeled using the intensity pixels located in the region \mathcal{R}_s automatically selected near to the border of the image (see Section 4.1):

$$t_s = \Gamma_{0.50} + \beta \cdot (\Gamma_{0.50} - \Gamma_{0.05}), \quad (11)$$

155 where Γ_ν for $\nu = \{0.05, 0.50\}$ are the 5th and 50th percentiles of the intensity image \mathcal{I} in \mathcal{R}_s . These conservative percentiles are used to model the dispersion of the \mathcal{I} intensity values, since lesion pixel values below the median $\Gamma_{0.50}$ are unlikely. This is because the intensity values of the skin pixels typically are lower in \mathcal{I} than those in the lesion area, and \mathcal{R}_s is designed to contain mainly skin pixels.

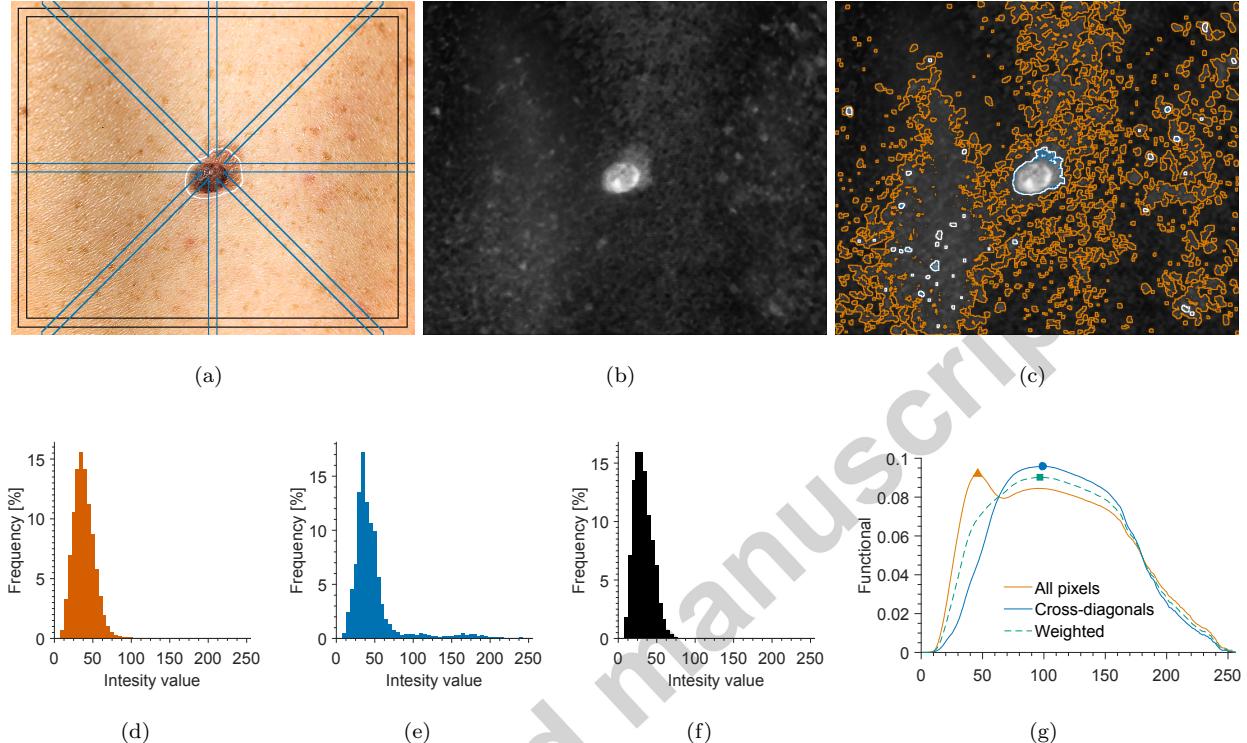
160 The constants α and β add flexibility to the thresholding process, correcting the usually higher values obtained by applying the default Otsu's thresholding to the proposed intensity image \mathcal{I} , which has an intensity distribution that deviates from the normal distribution. The α and β can be determined using a separate training set. These free constants may also help to adapt the threshold for different equipments.

Depending on the choice of β and the eventual presence of artifacts contaminating the image, t_s might result greater than t_h . This condition seems physically implausible using the proposed intensity image \mathcal{I} .
165 Therefore, if $t_s > t_h$, we set $\alpha = 1$, and an eventually unreliable t_s estimate does not affect Eq. 8.

4.4. Motivation and comments

The problem of segmentation of skin lesions has an interesting particularity that is exploited by the proposed SWOT algorithm: the lesions can be assumed to be approximately located in the central portion of the image (if minimum care is taken during image acquisition). Although the shape and size of the lesions
170 are a priori unknown, if we sample the full image, or around a few orthogonal lines passing through the center of the image (cross-diagonals), we get histograms with different class balances in terms of the number of skin and lesion pixels due to the irregular shape of the lesion regions.

We use the re-sampling strategy to increase the balance between classes, decreasing proportionally the number of pixels from the background skin class, which is typically over-represented, and favoring the



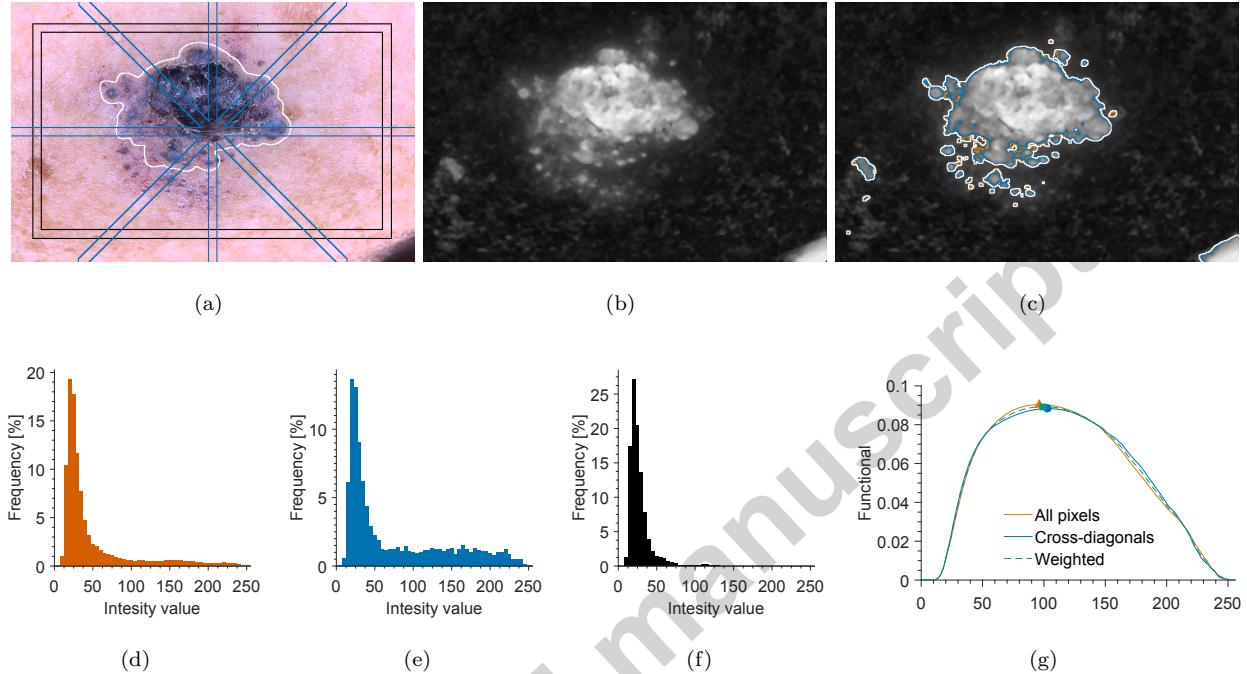


Figure 3: A worked example of the proposed analysis and calculation of thresholds for a macroscopic photograph of a melanoma. (a) The input image corrected for shadow attenuation. The cross-diagonal region is outlined in blue, and the automatically selected region containing mainly skin pixels is in black color. (b) The proposed intensity image \mathcal{I} after median filtering. (c) Contours of intensity image \mathcal{I} segmented using the thresholds $t' = 95$ (in orange), $t' = 102$ (in blue), and $t' = 84.8$ (in white). (d) Histogram of the intensity values of all image pixels, Otsu's estimated threshold is $t' = 95$. (e) Histogram of the intensity values only from the cross-diagonals pixels, Otsu's estimated threshold is $t' = 102$. (f) Histogram of the intensity values only from the peripheral skin pixels, and the proposed $t_s = 23.47 + 1.0 \cdot (8.76) = 32.2$ using Eq. 11 with $\beta = 1.0$. (g) Otsu's functionals normalized to a length of 1 with the respective peaks showing the position of the estimated thresholds ($t' = 95, t' = 102$). The weighted functional indicated by the dashed line is computed using Eq. 9, with a peak at $t_h = 98$. Finally, the proposed SWOT threshold is computed combining the estimates t_s from (f) and t_h from (g): $t' = 0.8 \cdot 98 + (1 - 0.8) \cdot 32.2 = 84.8$ (Eq. 8 with $\alpha = 0.8$). Using this estimate, the resulting border of the segmented lesion, after the post-processing step, is the white contour shown in (a).

¹⁷⁵ under-sampled skin lesion class. This idea is illustrated in Fig. 4(a), that shows how the sampling rate of
 a circular toy object represented as one class and, centered in an image of fixed size, varies in relation to
 the background represented as another class. The figure shows that the number of pixels collected from a
 tiny object covering only 1.5% of the image using the proposed cross-diagonal sampling scheme increases
 substantially to 10%, which is beneficial for histogram and thresholding analysis. Similarly, a balance of 50%
¹⁸⁰ between the background and object classes is achieved for a small object that covers only 10% the image.
 As expected, for larger objects, the cross-diagonal sampling becomes less relevant because sampling the full
 image area would be preferable. Nevertheless, the sampling of an object covering 50% of the image area
 still receives about 28% of samples from the under-sampled background class using the proposed approach
 based on cross-diagonals as opposite to 50% using the full image data (Fig. 4(a)).

¹⁸⁵ Ideally, the threshold calculated using all the image pixels and the threshold obtained with the subset
 of pixels located in the cross-diagonals of the skin lesion image should be about the same. In practice,
 especially for difficult cases, Otsu's threshold can differ substantially (see the example in Fig. 2, where
 the estimated threshold for the histogram (d) is 45, and completely misses the lesion mode resulting in a
¹⁹⁰ significant oversegmentation of the lesion, but the threshold is 98 for (e), where the skin mode becomes
 more apparent). This may happen because when applying Otsu's thresholding to the intensity pixels from
 all the \mathcal{I} image, the lesion mode is less evident. When applying Otsu's thresholding only to the pixels in
 the cross-diagonals, the lesion mode becomes more evident, but pixels from part of the lesion are not used,
 and therefore part of the information is missed (like the pixels outside the cross-diagonals in Fig. 2(a) and
 Fig. 3(a)). Therefore, we propose the simple functional in Eq. 9 that balances both contributions. Fig. 4(b)
¹⁹⁵ shows that on average Otsu's threshold using the cross-diagonal pixels are usually higher than the respective
 threshold estimates using all image pixels in our test set. These higher values tend to undersegment the
 lesions using the proposed intensity image \mathcal{I} . In our experiments, a few cases with larger differences in
 threshold values occurred for small lesions.

5. Post-processing stage

²⁰⁰ The binary segmentation mask, produced applying the threshold t' (Eq. 8) to the intensity image \mathcal{I}
 (Eq. 7), might contain several disjoint objects that could be confused with the lesion region.⁴ The main
 objective of the post-processing stage is to eliminate small/thin objects, that might be remaining of hair or
 other artifacts, and select the single connected region that is more likely to be the actual lesion.

We start by applying a morphological opening [26] to the binary input mask. The resulting image is
²⁰⁵ then dilated, and element-wise multiplied by the input mask to preserve the shape of the lesion border.
 For morphological operations, we use as structuring element a disk of radius $r = \gamma\mathcal{L}$ ($\gamma = 0.01$ in our

⁴We assume that lesion pixels are those having values above the threshold.

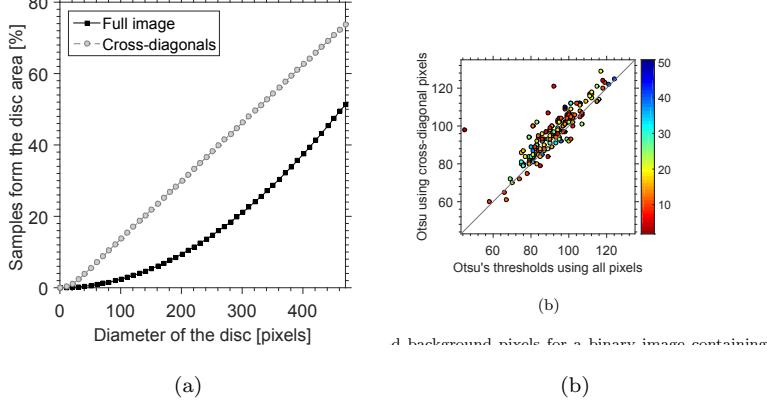


Figure 4: (a) Ratio between the number of foreground and background pixels for a binary image containing a toy disk of variable diameter (1–470 pixels) centered on an image of 720 x 470 pixels, computed using the full image, or only the pixels in the cross-diagonals with a width of 13 pixels (like in Fig. 2(a)). (b) Otsu’s thresholds using all image pixels and using only the cross-diagonals pixels, for each of the 152 lesions in the Alcón et al. data set. The color bar indicates the relative size of the lesion to the image area (%). Otsu’s values are in the range 0–255, and tend to be higher by re-sampling using the cross-diagonals. The solid 1:1 line is shown for visual reference.

experiments). After that, eventual holes in the binary image are filled. This is important if the image has bright and noisy pixels (like the reflections in Fig. 7(f)).

In the sequence, we compute a region score n_i for each candidate segment i . The region score n_i increases with size, intensity and proximity of the segment i center to the center of the image, penalizing segments that are small and/or off center. If the segment i is labeled as \mathcal{R}_i , the region score is given by:

$$n_i = \sum_{x,y \in \mathcal{R}_i} f(x,y)s(x,y), \quad (12)$$

where $f(x,y)$ is the two-dimensional Gaussian function:

$$f(x,y) = \exp \left\{ - \left[\frac{(x-x_o)^2}{2\sigma_x^2} + \frac{(y-y_o)^2}{2\sigma_y^2} \right] \right\}, \quad (13)$$

where (x_o, y_o) are the coordinates of the center of the image, and (σ_x, σ_y) are the bandwidth along x and y .
 210 The use of the Gaussian weight $f(x,y)$ acts as a spatial prior on the location of the segments, and proved helpful in [27]. Here we combined $f(x,y)$ with the “strength” $s(x,y)$, which is a likelihood function that returns higher values for confident lesion pixels than for background skin pixels. For convenience, we set $s(x,y) = \mathcal{I}(x,y)$. In our experiments, we set $\sigma_x = \sigma_y = 0.10\mathcal{L}$. We retain the region with the highest score n_i , discarding all other objects. Finally, we smooth the borders of the retained object using a Gaussian filter
 215 and close eventual remaining holes in the object.

6. Experimental results

6.1. Macroscopic images

Two data sets were used for experiments. The first was the data set used by Alcón et al., that consists of 152 macroscopic photographs of pigmented skin lesions, with benign and malignant diagnosis. Out of 220 these, 45 images correspond to atypical nevi, and the remaining 107 were melanoma. Each 24 bits RGB color image has a size of 720 pixels on the larger side, and between 439 to 706 pixels on the smaller side. Figs. 2–3 show two examples. Some images were duplicates from the same lesion, suggesting acquisitions under different conditions. This data set was proposed in [14] and comes from the *Dermnet* database⁵. Dermquest, consisting of 76 macroscopic photographs of melanomas and 61 benign nevi, was the second data 225 set analyzed⁶. Before starting the analysis, all images and reference segmentation masks in the Dermquest data set were first downsampled at 50% of its original sizes, resulting in image sides between 433 and 820 pixels. This reduction made the lesions easier to process and compare to those of the data set used by Alcón et al..

The presence of hair or acquisition artifacts were not used as an image exclusion criteria. No special 230 pre-processing was used for hair removal. One reference “ground truth”, manually delineated by an expert dermatologist, was available for each photograph. The area of the lesions in the data set used by Alcón et al. cover between 1.8% to 51% of the image (mean value is 17.3%). These figures decrease to 0.4% to 34.2% in the Dermquest data set (mean value is 6.3%).

6.2. Accuracy of the proposed method

Fig. 5 shows the segmentation accuracy of the proposed method when testing several combinations of the free parameters α and β . A few selected cases for the data set used by Alcón et al. are detailed in Table 1, summarized by the mean, median, and standard deviation (SD). The segmentation accuracy was measured using the XOR error, the sensitivity, and the specificity, which are computed using the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) detections (see Fig. 1(b)):

$$\begin{aligned} \text{XOR error} &= (\text{FP} + \text{FN}) / (\text{TP} + \text{FN}), \\ \text{Sensitivity} &= \text{TP} / (\text{TP} + \text{FN}), \\ \text{Specificity} &= \text{TN} / (\text{FP} + \text{TN}), \end{aligned} \quad (14)$$

235 The XOR error gives the ratio of the number of pixels for which the automatic and ground truth masks disagree to the area of the lesion in the ground truth. A drawback of this measure is that it tends to

⁵Dermnet Skin Disease Image Atlas, www.dermnet.com.

⁶The Dermquest data set was proposed by the authors of [28]. It is available from the authors upon request, and for download at: '<http://vip.uwaterloo.ca/demos/skin-cancer-detection>'.

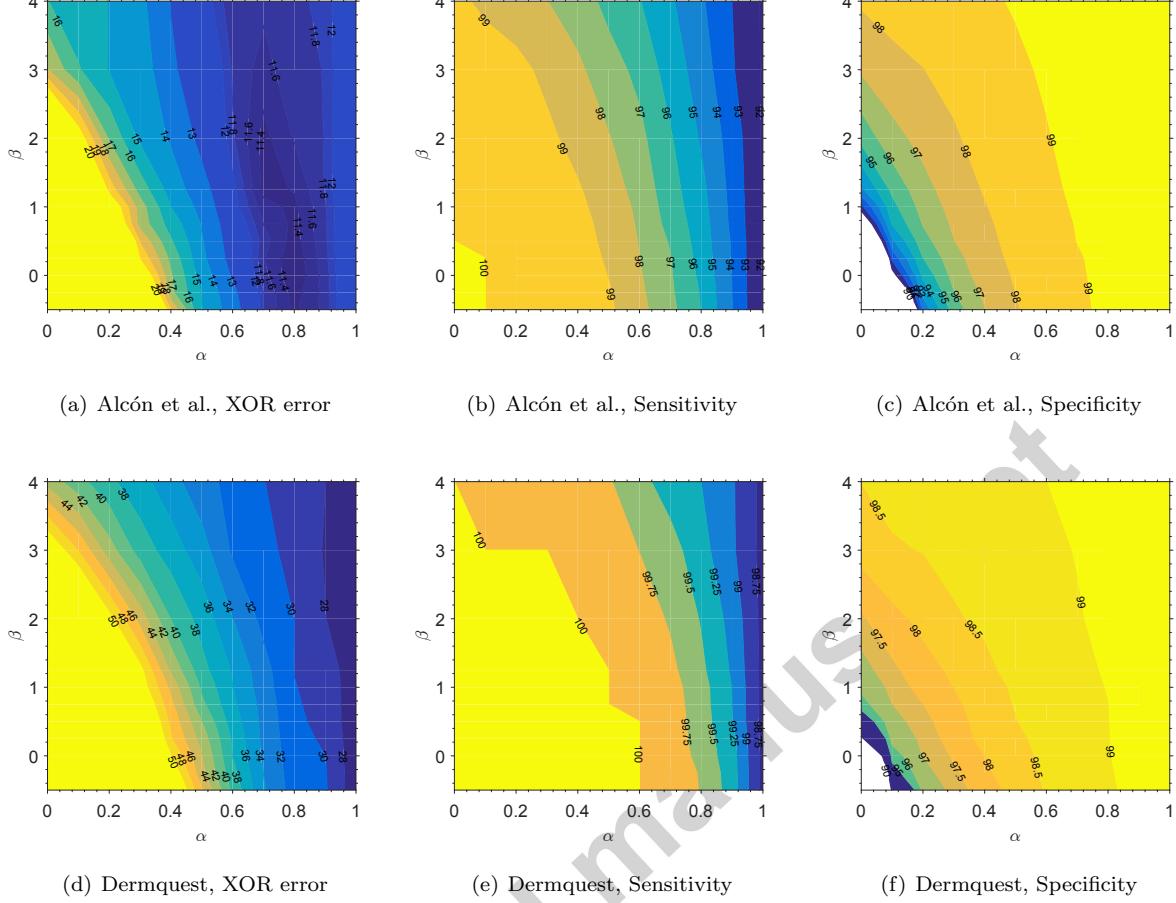


Figure 5: Segmentation scores of the proposed algorithm varying the two weighting parameters α and β for the data sets used by Alcón et al. and Dermquest. Values correspond to the median computed using all lesion images in the test set. Specificity below 90% in (c) are shown in white color.

penalize small lesions. Sensitivity and specificity correspond to the percentage of correctly detected lesion and background skin pixels, respectively. While the sensitivity and specificity are bounded to [0,100%], the XOR error is susceptible to large tail values (see Fig. 6).

Although the α and β parameters control the estimated threshold (see Eq. 8 and Eq. 11), the results in Fig. 5 show that the segmentation accuracies are more sensitive to α . Best accuracies were obtained with $\alpha = 0.80$ around. Wide variations of $-0.50 \leq \beta \leq 4$ did not change much the segmentation results, although the XOR and sensitivity accuracies slightly decrease with $\alpha = 0.80$ and $\beta = 4$ (see Table 1). We recommend using $\alpha = 0.80$ and $\beta = 1.00$, since these parameters were found with an independent validation set of melanocytic lesions. Ideally the lesions in the validation data set should be acquired under the same conditions, and have similar colors and structures as those in the test sets. Table 1 includes the analysis of two extreme cases: $\alpha = 0$, meaning only percentile-based thresholding is used; and $\alpha = 1$, meaning that

Table 1: Segmentation results of the proposed algorithm varying the two weighting parameters α and β . Best scores measured in the 152 test images of the Alcón et al. data set are in bold.

α	β	XOR error (%)			Sensitivity (%)			Specificity (%)		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.00	1.00	71.77	49.20	80.76	99.11	100.00	3.24	88.96	90.70	8.37
0.00	2.00	40.84	25.53	57.61	97.23	99.95	7.05	94.14	95.30	5.68
0.00	3.00	30.75	18.04	47.01	94.74	99.73	11.54	96.09	97.12	4.50
0.00	4.00	26.42	16.03	38.53	92.82	99.24	13.64	97.12	98.12	3.70
0.80	-0.50	14.93	11.36	10.22	91.87	95.97	10.87	98.98	99.17	0.80
0.80	-0.25	14.92	11.30	10.23	91.74	95.91	10.98	99.01	99.18	0.77
0.80	0.00	14.95	11.22	10.28	91.61	95.78	11.08	99.02	99.20	0.76
0.80	0.25	15.00	11.24	10.37	91.46	95.62	11.22	99.04	99.22	0.75
0.80	0.50	15.05	11.30	10.43	91.32	95.52	11.32	99.05	99.24	0.75
0.80	0.75	15.09	11.34	10.49	91.18	95.47	11.43	99.07	99.25	0.74
0.80†	1.00†	15.14	11.35	10.56	91.04	95.35	11.54	99.09	99.26	0.73
0.80	1.25	15.21	11.42	10.67	90.87	95.27	11.70	99.10	99.27	0.72
0.80	2.00	15.70	11.62	11.42	90.13	94.89	12.61	99.14	99.31	0.70
0.80	3.00	15.95	11.76	11.72	89.58	94.52	13.03	99.19	99.39	0.68
0.80	4.00	16.17	11.77	11.90	89.08	94.14	13.30	99.24	99.42	0.64
1.00	–	16.81	12.57	12.33	87.28	91.89	13.89	99.42	99.55	0.48

† Suggested choice of α and β based on an independent validation set.

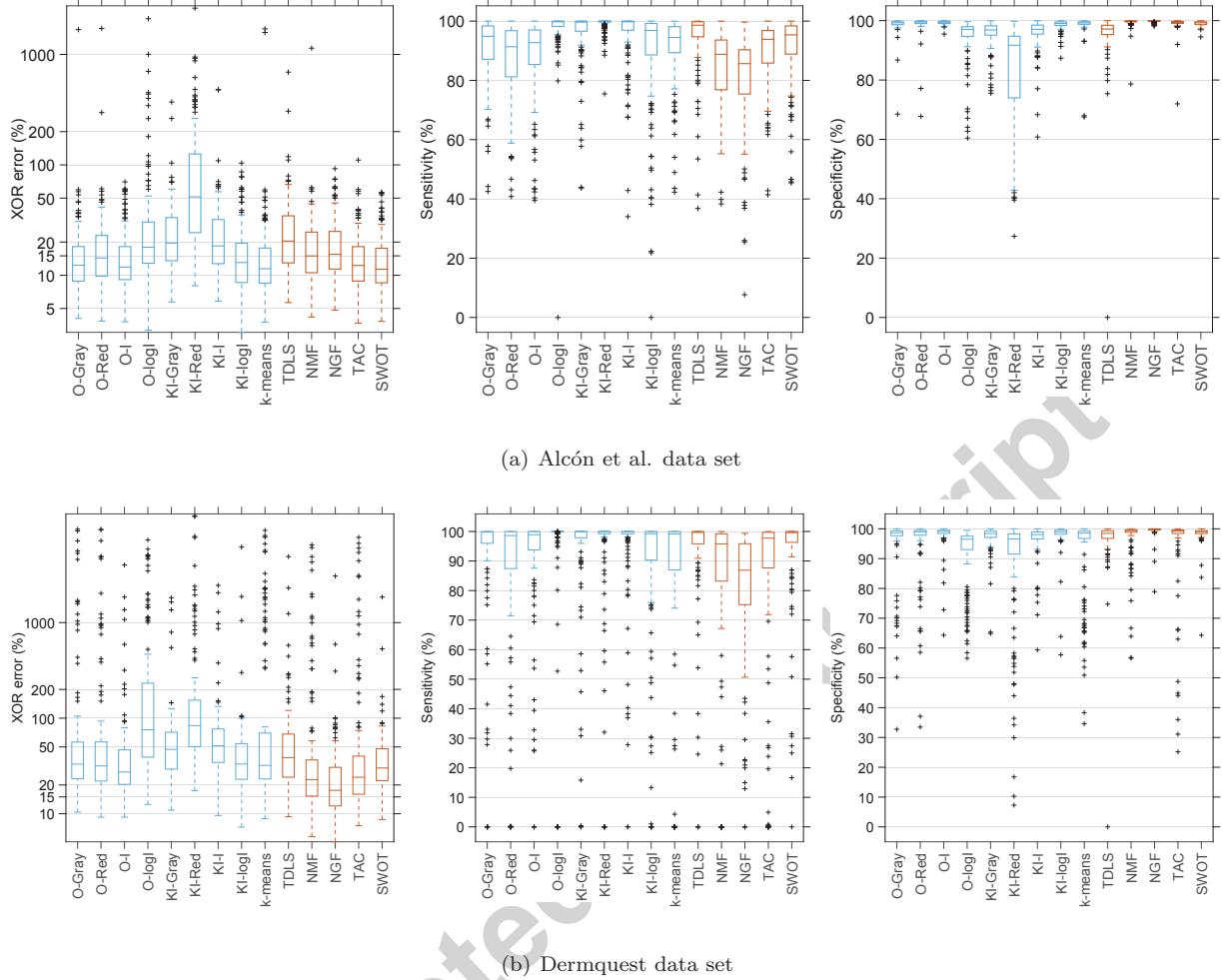
only the weighted average of Otsu's functional (Eq. 9) is considered. In both cases, the accuracies are lower, and the weighted scheme proposed in SWOT improved results.

250 *6.3. Comparison with unsupervised methods*

Fig. 6 shows the dispersion of the segmentation accuracies for the proposed method when compared to alternative approaches, including the most accurate results for the same data set published earlier [11]. The XOR error, sensitivity and specificity values of the 152 lesions in the Alcón et al. data set are summarized in Table 2. The respective scores for the 137 lesions in the Dermquest data set are summarized 255 in Table 3. Methods that are simpler in nature, such as thresholding and clustering, but run using the same post-processing proposed in Section 5 were included in the analysis. Specifically, we focus on the original Otsu's and K&I algorithms, tested on different input channels, and k-means. The specific recent approaches designed for segmentation of macroscopic photographs of skin lesions, namely thresholding on a image representation followed by active-contour (TAC) [11], segmentation using non-negative matrix factorization (NMF) [18], segmentation based on normalized graph cuts (NGC) [19], and texture distinctiveness 260 (TDLS) [12] have their own implementations.

For the benchmark with k-means, we set $k = 2$ and use standard Euclidian distances in the CIELAB color space. The “darker” cluster, i.e., the one with the lower average in the lightness component, is assumed to correspond to the lesion. To allow a fair comparison with our proposal, median filtering was applied to all 265 color components. In the post-processing, the strength term $s(x, y)$ in Eq. 12 is set to $(d_s)/(d_l + d_s)$, where d_l is the Euclidian distance from the pixel (i, j) to the centroid of the lesion cluster, and d_s is the distance to the centroid of the skin cluster, both measured in the feature space. Similarly, the red, grayscale, and intensity channels used as input to the Otsu's and K&I algorithms were filtered for noise reduction using the same kernel size as in the proposed method. For convenience, in case of the grayscale and the red channels, 270 their complementary channels were taken as the strength term in Eq. 12, because the lesion pixels usually have lower value in such channels.

The analysis of the Alcón et al. data set reveals that except for a failure in 2 out of the 152 lesions the simple k-means (Fig. 6) have a median XOR error (11.47%) that is only beaten by the proposed SWOT method (11.35%, Table 2). Otsu applied to the grayscale also comes very close (12.34%) but fails in one lesion 275 (the atypical nevi case shown in Fig. 2(a)). Both the SWOT and the active-contour method (TAC) were more robust in the sense that such failures were avoided. Similar behavior was observed in the sensitivity and specificity scores. A simple Otsu's thresholding on the proposed intensity image already provided good results avoiding large XOR errors, but had lower sensitivity. K&I did not improve such results in this data set, but when we took the logarithm of the proposed intensity image, making its distribution resemble more 280 Gaussian, it provided better results than the other channels tested (using the red and grayscale channels).



O-Gray	Otsu's thresh. method on grayscale [8]	k-means	k-means on the Lab color space
O-Red	Otsu's thresh. method on the red channel	TDLS	Texture distinctiveness [12]
O-I	Otsu's thresh. method on intensity	NMF	NMF Segmentation [18]
O-logI	Otsu's thresh. method on log. of intensity	NGC	Normalized Graph Cuts [19]
KI-Gray	K&I's thresh. method on grayscale	TAC	Thresholding followed by active-contour [11]
KI-Red	K&I's thresh. method on the Red channel	SWOT	Proposed segmentation method
KI-I	K&I's thresh. method on intensity		
KI-logI	K&I's thresh. method on log. of intensity		

Figure 6: Dispersion of the accuracies of the lesion segmentations, quantified by XOR error, sensitivity and specificity, showing how the proposed SWOT method compares to several unsupervised segmentation methods. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points that are considered not to be outliers, and the outliers are plotted individually.

Table 2: Comparison of segmentation results for the Alcón et al. data set using unsupervised methods. Significantly better (+) and worse (*) marks indicating that the Wilcoxon signed-rank test at 1% significance level rejects the null hypothesis of zero median in the difference between the proposed SWOT ($\alpha = 0.8, \beta = 1.00$) and the alternative methods.

Method	XOR error (%)			Sensitivity (%)			Specificity (%)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
O-Gray	26.37	12.34*	135.26	91.20	94.88*	10.47	98.82	99.28	2.77
O-Red	30.85	14.33*	140.21	86.93	91.35*	13.04	98.99	99.53+	3.22
O-I	16.60	11.87*	12.92	88.33	92.74*	13.64	99.34	99.48+	0.57
O-logI	61.17	17.95*	203.99	97.67	99.75+	8.55	94.80	97.05*	6.81
KI-Gray	29.36	19.58*	37.83	95.88	99.64+	9.52	95.79	96.87*	4.43
KI-Red	126.73	51.31*	258.57	99.16	100.00+	2.62	83.20	91.73*	17.19
KI-I	29.91	18.46*	54.32	95.82	99.76+	9.54	95.93	97.07*	4.98
KI-logI	18.04	13.05*	16.40	89.63	96.86	17.27	98.71	99.04*	1.55
K-means	36.22	11.47	186.61	91.26	94.47*	10.87	98.69	99.28+	3.69
TDLS	31.80	20.41*	61.28	94.87	98.61+	9.96	95.70	97.17*	8.58
NMF	25.99	14.97*	91.69	84.20	88.80*	13.17	99.51	99.80+	1.79
NGC	21.12	15.48*	15.05	80.27	85.65*	15.92	99.83	99.92+	0.28
TAC	15.60	12.28	12.48	90.00	93.86*	10.52	99.12	99.41+	2.35
SWOT	15.14	11.35	10.56	91.04	95.35	11.54	99.09	99.26	0.73

On the Dermquest data set, the XOR error of all methods tested was higher, suggesting that segmentations were less accurate. The NGC algorithm had the lowest median XOR error of 17.78% with SWOT reaching 30.08% (Table 3). Alternatively, considering the mean XOR error, NGC and SWOT perform similarly (53.55% vs 54.30%). Overall, the sensitivity and specificity scores in Table 3, and the respective dispersion of values shown in Fig. 6, suggest that SWOT provides competitive tradeoff of accuracies also for the Dermquest data set.

We used the Wilcoxon signed-rank test to check the null hypothesis of zero median in the difference between the segmentation scores of the proposed SWOT and the alternative methods. Table 2 shows that for the Alcón et al. data set, in the case of the XOR error, at 1% significance level the null hypothesis was most often rejected (11 cases) than accepted (2 cases), suggesting that SWOT is equal or better than most of the alternative methods. Furthermore, the average and standard deviation scores of the proposed method consistently outperformed the benchmark methods. On the Dermquest data set, the advantage of SWOT is less evident when analyzing only the median scores. Although the null hypothesis was also rejected on 11 cases, seven cases favored SWOT and four cases favored alternative methods (Table 3). For both data

Table 3: Comparison of segmentation results for the Dermquest data set using unsupervised methods. Significantly better (+) and worse (*) marks indicating that the Wilcoxon signed-rank test at 1% significance level rejects the null hypothesis of zero median in the difference between the proposed SWOT ($\alpha = 0.8, \beta = 1.00$) and the alternative methods.

Method	XOR error (%)			Sensitivity (%)			Specificity (%)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
O-Gray	406.21	33.01*	1473.24	89.50	99.78	25.17	95.30	98.67*	10.59
O-Red	399.36	31.67	1455.15	85.86	98.61*	26.74	95.41	99.00	10.87
O-I	101.78	27.32+	400.53	89.49	98.89*	23.13	98.17	99.16+	4.23
O-logI	544.66	75.67*	1288.11	93.77	100.00+	22.47	91.38	96.52*	10.65
KI-Gray	94.05	47.12*	242.94	92.53	100.00+	20.91	97.15	98.42*	4.68
KI-Red	542.91	83.61*	1850.49	92.54	100.00+	22.31	88.55	96.58*	19.11
KI-I	112.71	51.47*	303.93	94.46	100.00+	17.17	96.52	98.07*	5.61
KI-logI	107.03	33.12*	553.07	86.33	99.27*	27.36	98.09	98.98	4.91
K-means	582.34	32.06	1572.74	81.36	99.16*	35.22	92.17	98.56	14.09
TDLS	120.37	38.68*	470.47	92.57	99.67	19.97	96.70	98.47*	8.92
NMF	288.02	22.78+	1001.05	79.80	95.82*	33.29	97.03	99.37+	7.42
NGC	53.55	17.59+	266.24	79.55	86.89*	23.95	99.46	99.84+	2.09
TAC	346.80	24.09+	1220.14	81.92	97.73*	32.56	95.20	99.23+	13.56
SWOT	54.30	30.08	162.02	93.18	99.60	17.29	98.37	99.01	3.44

²⁹⁵ sets, in most cases, the sensitivity and specificity results were either statistically better or worse, but no alternative method simultaneously had both better or worse.

6.4. Comparison with supervised methods

Here we further investigate whether the proposed intensity image \mathcal{I} can reasonably replace the full color information available in macroscopic images for the purpose of skin lesion segmentation by thresholding, and ³⁰⁰ weather the use of thresholding provide competitive results if compared to more advanced classifiers in the context of skin lesion segmentation. Because of the availability of ground truth information (the manually drawn lesion borders), these questions are easier to answer in a supervised setting.

In this study, we focus on well-known binary classification (segmentation) methods as a comparison to our simple thresholding scheme, including parametric models such as the Linear (LDA) and Quadratic ³⁰⁵ Discriminant Analysis (QDA), and non-parametric alternatives such as the Classification and Regression Trees (CART), the k-Nearest Neighbors (k-NN) algorithm, and the Support Vector Machines using the Radial Basis Function kernel (SVM-RBF). An excellent treatment of such methods is provided in [29]. Initially, equal priors and miss-classification costs for the lesion and the background classes were assumed ³¹⁰ for estimating the posterior probability of a pixel being lesion. Then, we make a fine adjustment dividing the posterior probability of a pixel being in the class lesion, in the range [0,1], into 255 intervals (thresholds), selecting the value that, together with the post-processing, minimizes the XOR error. This simulates a hypothetical scenario in which, in addition to selecting training samples, the user provides feedback by manually adjusting the probability level that visually resulted in the accurate segmentation of each lesion. Similarly, for benchmarking the results with supervised thresholding, we choose the threshold value that ³¹⁵ minimizes the segmentation error using the intensity image \mathcal{I} .

The supervised models were trained for each macroscopic image in the data set. For each image, we randomly selected 2,000 pixels from the reference ground truth mask, for the lesion and background skin classes. The pixel values at the 3-dimensional CIELAB color space were used as input features for classification. At the pixel level, the estimated posterior probability for the lesion class was filtered using the same ³²⁰ median filter as in the proposed method. This was to avoid that strong reflections, seen as white patches in the image, ended up “breaking” the lesion in small objects that were later discarded by the post-processing. In the post-processing stage, the “strength” in Eq. 12 is the estimated posterior for the class lesion, which for convenience was also scaled in the range [0,255].

Tables 4–5 show that in the best case, despite simple, the thresholding approach, applied to the proposed ³²⁵ intensity image, provides accuracies that are similar to the ones obtained using parametric and non-parametric methods trained using the full color information. This suggests that using the intensity image \mathcal{I} tends to lead to good segmentations of the skin lesions present in the test set. Furthermore, it can be observed based on the Alcón et al. test set that unsupervised methods have an untapped potential

Table 4: Comparison of segmentation results for the Alcón et al. data set using supervised methods.

Method	XOR error (%)			Sensitivity (%)			Specificity (%)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
LDA	9.42	8.79	3.82	96.09	96.57	2.53	98.84	99.17	0.94
QDA	11.54	10.13	5.38	96.72	97.21	2.83	98.49	98.68	0.97
CART	11.35	10.24	4.30	97.55	97.96	2.17	98.37	98.55	0.83
3NN	12.43	11.31	5.07	97.62	98.52	2.80	98.21	98.37	0.92
SVM-RBF	11.10	9.97	4.59	97.06	97.51	2.57	98.50	98.66	0.84
Thresholding	9.64	9.14	3.96	95.85	96.66	3.01	98.79	99.12	1.19

Table 5: Comparison of segmentation results for the Dermquest data set using supervised methods.

Method	XOR error (%)			Sensitivity (%)			Specificity (%)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
LDA	23.40	19.66	12.94	94.38	96.29	9.73	99.13	99.37	0.77
QDA	36.57	31.06	21.63	95.09	99.68	15.05	98.60	98.74	0.89
CART	36.93	30.17	22.34	95.90	99.90	15.14	98.65	98.70	0.70
3NN	39.96	33.93	22.53	94.49	100.00	20.54	98.46	98.52	0.79
SVM-RBF	32.55	26.84	19.85	94.10	98.04	15.28	98.86	99.01	0.72
Thresholding	21.07	18.23	14.35	90.96	94.60	14.74	99.25	99.51	0.83

to reduce about 5% the average XOR error and about 4.5% the average sensitivity (compare Table 2 vs
330 Table 4). The improvement can be even higher for the Dermquest data set, where the average XOR error
can potentially be reduced from about 54% to 21% (compare Table 3 vs Table 5).

Fig. 7 presents the segmentation XOR error, sensitivity and specificity for benign and malignant cases
separately, for selected supervised and unsupervised methods, and shows examples of segmentation using the
proposed SWOT method. The average segmentation accuracy of melanomas were lower than for atypical
335 nevi in the Alcón et al. data set, for all methods tested, except for the slightly higher sensitivity to melanomas
using LDA. A possible cause for the accuracy difference between benign lesions and melanomas may be due
to the more complex patterns of malignant lesions in this data set.

7. Discussion

Ease of use is an important property of the proposed SWOT algorithm. It has two free parameters
340 (α, β) that can be easily estimated using a quantitative or qualitative assessment on a separate data set. For

convenience, all the remaining secondary parameters, like the filter sizes and the width of the cross-diagonal region area, are set in terms of the relative size of the image (we did not attempt to optimize such secondary parameters, and eventually better results could be obtained using fine-tuning). The SWOT algorithm proved to be well suited for the segmentation of skin lesions in the considered test sets.

345 For simplicity, the cross-diagonals used in the proposed method were assumed to pass through the center of the input photographs. We also explored an alternative version (not shown) where the cross-diagonals were centered on a pre-segmented mask obtained applying Otsu's method to the proposed intensity image \mathcal{I} . The final segmentation scores were similar in both cases.

SWOT is challenged by cases like the lesion in Fig. 7(h) in the Alcón et al. data set. In this case, 350 the variability of the skin near the border of the region is substantially lower than near the center. When looking at the other lesions in the Alcón et al. data set, this seems to be rather an exception than a rule. Lesions like Fig. 7(k) challenge the current post-processing. In this example, SWOT chooses a threshold larger than it would be desirable, segmenting the lesion area in two large pieces, separated by the lighter regression area in the middle, and selecting only one (the larger) object as the lesion. Cases like Fig. 7(l) are 355 very challenging for automated segmentation, and also for dermatologists. SWOT performs reasonably well under severe hair occlusion such as in Fig. 7(j), even if the utility of segmentation remains questionable in such cases. SWOT also provided good results for lesions with a highly heterogeneous background skin and strong reflections (white patches) inside the lesion area, like in the example shown in Fig. 7(i). Fig. 7(g) illustrates the challenge of dealing with lesions that are spatially close to each other, and it is debatable if 360 they should be grouped into a single segment or not.

As expected, supervision improves segmentation accuracies. However, from a practical perspective, users 365 of CAD systems may prefer fully automated segmentation algorithms. The main reason is that selecting training samples for each image, at least as densely as done in this study, would be time-consuming. For the sake of completeness, we report that during the validation of SWOT, a semi-supervised alternative formulation was also tested. The idea was to model the threshold for segmentation of the proposed intensity image \mathcal{I} as a weighted average of two components: the unsupervised SWOT estimate, and a threshold estimated by minimizing the overlap between the probability densities of the lesion and the surrounding skin classes in a pooled set of training images. This alternative approach did not seem to improve the results already obtained using the unsupervised SWOT algorithm.

Our current quantitative assessment of segmentation accuracy comes with a caveat. The “ground truth” 370 is subjective and based on the opinion of a single dermatologist. However, the ground truth visually seems to fit the contours perceived as the lesion borders (some examples are shown in Fig. 7). It is important to keep in mind that previous studies using dermoscopic images have demonstrated that dermatologists have difficulty reproducing measurements precisely [30].

The final diagnosis of CAD systems may not differ significantly by choosing good segmentation algo- 375

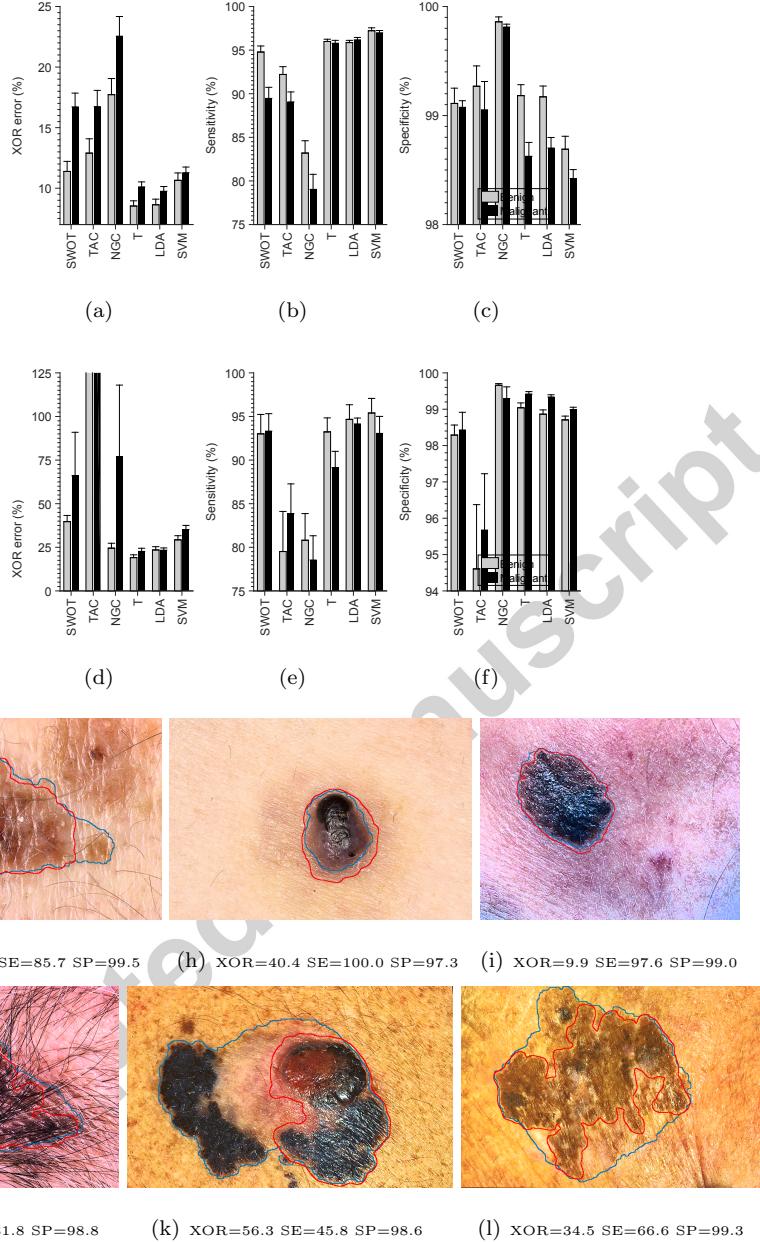


Figure 7: Mean segmentation XOR error, sensitivity and specificity for 45 benign and 107 malignant cases in the Alcón et al. data set (a–c), and 61 benign and 76 melanomas in the Dermquest data set (d–f). The bar indicates the standard error of the mean (s.e.m); “SWOT” is the proposed unsupervised method; “TAC” is the thresholding followed by active-contour segmentation; “T” corresponds to manual (supervised) thresholding, picking-up of the best value in the proposed intensity image. The XOR error for TAC in (d) is above the limit of the axis; (g–l) show examples of borders detected using the proposed method (red contour) and the reference by the dermatologist (in blue). All lesions are melanomas, and the images have been pre-processed for shadow attenuation. The XOR error, the sensitivity (SE), and the specificity (SP) scores are in percentages.

rithms [17]. Therefore, developing simple and accurate segmentation methods like SWOT is of practical interest.

The proposed segmentation algorithm is implemented in MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States) and run on a computer with an Intel Dual Core 2.53 GHz processor with 2GB RAM. The average runtime to segment a 720×470 image using our code, which is not fully optimized, is about 2 seconds. Most of this processing time is spent on the 9×9 median filtering and on the morphological operations used in the post-processing phase, filtering small segments that need to be eliminated. For a 1440×940 it takes about 12 seconds, with most of the time spent computing the 15×15 median filtering. These rough estimates using non-optimized code of processing times suggest that the SWOT algorithm is suitable for practical use in clinical settings as part of CAD systems running on devices with limited processing capabilities.

8. Conclusions

Segmentation is an important component in the computer assisted diagnosis of skin cancers. We proposed a new method for segmentation of macroscopic photographs of pigmented skin lesions, SWOT, which is based on a modification of the Otsu's thresholding method, analyzing separately samples from different locations in the image. SWOT provided good segmentation results for most of the 152 and 137 melanocytic lesions in the Alcón et al. and Dermquest test sets, respectively, including benign and malignant cases.

Besides, we provided an empirical comparison showing that the accuracy of the proposed method compares favorably with state-of-the-art algorithms designed specifically for segmentation of macroscopic photographs of skin lesions. We measured how the automatically generated segmentation compares to the reference border manually drawn by a dermatologist, quantifying the XOR error, the sensitivity, and the specificity. Alternative methods, based on thresholding and clustering, but running through the proposed post-processing step, were also included in the analysis.

SWOT runs on a novel intensity image where each pixel measures its Euclidian distance to the automatically estimated color of the background skin. Furthermore, we validated the suitability of the replacing color images by the intensity image in a hypothetic supervised scenario, showing that simple thresholding, using the proposed intensity image, provides segmentation accuracies similar to those of LDA, or more advanced classifiers such as SVM-RBF, trained using the full color information in the CIELAB space. In the data set proposed by Alcón et al.[14], which we use in the experiments, unsupervised segmentation methods have an untapped potential to reduce about 5% the average XOR error and about 4.5% the average sensitivity if compared to supervised segmentation. The reduction potentially can be even greater in the Dermquest data set.

The rebalancing of the classes proposed in this work provides a strategy that may be exploited in other

segmentation problems, under that assumption that the object of interest lies approximately in the center
 410 of the image. In such case, re-sampling across the diagonals of the image improves the representability of the undersampled class.

Finally, we emphasize the simplicity and accuracy of SWOT, indicating that the algorithm is a useful part of a practical CAD system, specially because the CAD system diagnosis capability is not impacted significantly by small inaccuracies in lesion boundary detection [17], increasing the attractiveness of simpler
 415 segmentation methods. Further research is needed to evaluate its performance in clinical settings.

Acknowledgments

This work was supported in part by the Brazilian National Council for Scientific and Technological Development (CNPq).

References

- 420 [1] K. Korotkov, R. Garcia, Computerized analysis of pigmented skin lesions: a review, *Artificial intelligence in medicine* 56 (2012) 69–90.
- [2] A. Perrinaud, O. Gaide, L. French, J.-H. Saurat, A. Marghoob, R. Braun, Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? a study comparing the results of three systems, *British Journal of Dermatology* 157 (2007) 926–933.
- 425 [3] H. Kittler, H. Pehamberger, K. Wolff, M. Binder, Diagnostic accuracy of dermoscopy, *The lancet oncology* 3 (2002) 159–165.
- [4] A. Kassianos, J. Emery, P. Murchie, F. Walter, Smartphone applications for melanoma detection by community, patient and generalist clinician users: a review, *British Journal of Dermatology* (2015).
- 430 [5] W. G. Flores, W. C. de Albuquerque Pereira, A. F. C. Infantosi, Improving classification performance of breast lesions on ultrasonography, *Pattern Recognition* 48 (2015) 1125 – 1136.
- [6] L. Ma, R. C. Staunton, Analysis of the contour structural irregularity of skin lesions using wavelet decomposition, *Pattern Recognition* 46 (2013) 98–106.
- 435 [7] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man and Cybernetics* 9 (1979) 62–66.
- [8] D. Ruiz, V. Berenguer, A. Soriano, J. Martin, A cooperative approach for the diagnosis of the melanoma, in: *Engineering in Medicine and Biology Society, 2008. EMBC 2008. 30th Annual International Conference of the IEEE*, IEEE, 2008, pp. 5144–5147.
- 440 [9] P. G. Cavalcanti, J. Scharcanski, L. Di Persia, D. Milone, An ICA-based method for the segmentation of pigmented skin lesions in macroscopic images, in: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 5993–5996.
- [10] P. G. Cavalcanti, J. Scharcanski, Automated prescreening of pigmented skin lesions using standard cameras, *Computerized Medical Imaging and Graphics* 35 (2011) 481–491.
- 445 [11] P. G. Cavalcanti, J. Scharcanski, A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images, *Computer methods and programs in biomedicine* 112 (2013) 684–693.
- [12] J. Glaister, A. Wong, D. A. Clausi, Segmentation of skin lesions from digital images using joint statistical texture distinctiveness, *Biomedical Engineering, IEEE Transactions on* 61 (2014) 1220–1230.

- [13] M. Sezgin, et al., Survey over image thresholding techniques and quantitative performance evaluation, *Journal of Electronic imaging* 13 (2004) 146–168.
- [14] J. F. Alcón, C. Ciuhu, W. Ten Kate, A. Heinrich, N. Uzunbajakava, G. Krekels, D. Siem, G. De Haan, Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis, *Selected Topics in Signal Processing, IEEE Journal of* 3 (2009) 14–25.
- [15] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, *International journal of computer vision* 1 (1988) 321–331.
- [16] J. Tang, A multi-direction GVF snake for the segmentation of skin cancer images, *Pattern Recognition* 42 (2009) 1172–1179.
- [17] P. G. Cavalcanti, J. Scharcanski, Macroscopic pigmented skin lesion segmentation and its influence on lesion classification and diagnosis, in: M. E. Celebi, G. Schaefer (Eds.), *Color Medical Image Analysis*, volume 6 of *Lecture Notes in Computational Vision and Biomechanics*, Springer Netherlands, 2013, pp. 15–39.
- [18] P. G. Cavalcanti, J. Scharcanski, C. E. Martinez, L. E. D. Persia, Segmentation of pigmented skin lesions using non-negative matrix factorization, in: 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, IEEE, 2014.
- [19] E. S. Flores, J. Scharcanski, Segmentation of pigmented melanocytic skin lesions based on learned dictionaries and normalized graph cuts, in: Proceedings of the 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, IEEE Computer Society, 2014, pp. 33–40.
- [20] Q. Qiu, V. M. Patel, R. Chellappa, Information-theoretic dictionary learning for image classification, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36 (2014) 2173–2184.
- [21] J. Shi, J. Malik, Normalized cuts and image segmentation, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22 (2000) 888–905.
- [22] R. Nock, F. Nielsen, Statistical region merging, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26 (2004) 1452–1458.
- [23] J. Kittler, J. Illingworth, Minimum error thresholding, *Pattern recognition* 19 (1986) 41–47.
- [24] T. Kurita, N. Otsu, N. Abdelmalek, Maximum likelihood thresholding based on population mixture models, *Pattern Recognition* 25 (1992) 1231–1240.
- [25] P. G. Cavalcanti, J. Scharcanski, C. B. Lopes, Shading attenuation in human skin color images, in: *Proceedings of the 6th international conference on Advances in Visual Computing, Volume Part I*, Springer, 2010, pp. 190–198.
- [26] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, Inc., Orlando, FL, USA, 1983.
- [27] M. Zortea, S. O. Skrøvseth, T. R. Schopf, H. M. Kirchesch, F. Godtliebsen, Automatic segmentation of dermoscopic images by iterative classification, *Journal of Biomedical Imaging* 2011 (2011) 1–19.
- [28] R. Amelard, A. Wong, D. A. Clausi, Extracting high-level intuitive features (hlif) for classifying skin lesions using standard camera images, in: *Proceedings of the 2012 Ninth Conference on Computer and Robot Vision, Canadian Image Processing and Pattern Recognition Society (CIPPR)*, 2012, pp. 396–403. doi:<http://dx.doi.org/10.1109/CRV.2012.59>.
- [29] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2001.
- [30] G. Joel, S.-S. Philippe, G. David, C. Jean Philippe, B. Ralph, K. Joakim, S. Jean-Hilaire, K. Murat, Validation of segmentation techniques for digital dermoscopy, *Skin Research and Technology* 8 (2002) 240–249.