

A Fully Automatic Random Walker Segmentation for Skin Lesions in a Supervised Setting

Paul Wighton^{1,2,3}, Maryam Sadeghi^{1,3}, Tim K. Lee^{1,2,3}, and M. Stella Atkins¹

¹ School of Computing Science, Simon Fraser University, Canada
pwighton@sfu.ca

² Department of Dermatology and Skin Science, University of British Columbia and Vancouver Coastal Health Research Institute, Canada

³ Cancer Control Research Program and Cancer Imaging Department, BC Cancer Research Centre, Canada

Abstract. We present a method for automatically segmenting skin lesions by initializing the random walker algorithm with seed points whose properties, such as colour and texture, have been learnt via a training set. We leverage the speed and robustness of the random walker algorithm and augment it into a fully automatic method by using supervised statistical pattern recognition techniques. We validate our results by comparing the resulting segmentations to the manual segmentations of an expert over 120 cases, including 100 cases which are categorized as difficult (i.e.: low contrast, heavily occluded, etc.). We achieve an F-measure of 0.95 when segmenting easy cases, and an F-measure of 0.85 when segmenting difficult cases.

1 Introduction

The segmentation of skin lesions is a crucial step in the process of automatically diagnosing melanoma. Inaccurate segmentations will affect all downstream processes such as feature extraction, feature selection and even the final diagnosis. Accurate segmentations are especially crucial for features that measure properties of the lesion border. Additionally, a recent study found that all commercially available automated systems evaluated had difficulty segmenting when the contrast between the lesion and skin was low[1]. We seek to improve skin lesion segmentations by employing supervised techniques to automatically initialize the random walker (RW) algorithm[2], which has been shown useful for interactive segmentations where boundaries are not clearly defined.

2 Previous Work

2.1 Skin Lesion Segmentation

The closest related work in skin lesion segmentation is by Celebi et. al.[3]. They reduce a dermoscopic image to 20 distinct colour groups, and assign labels to

pixels based on the group to which they belong. They then define a metric, the *J-value*, which measures the spatial separation or isolation of each group. The *J-value* is derived from the separability criterion used by Fisher in Linear Discriminant Analysis[4]. Next, they define a localized *J-value* for a specific pixel by computing the metric over a neighbourhood around the pixel. By varying the neighbourhood size, they create several of these *J-images*. Multiscale methods are used to combine the images into a final segmentation. By creating a class map via color reduction, and employing various neighbourhood sizes they are incorporating, on some levels, textural information into their segmentation. In a follow-up study [5] Celebi et. al. apply their segmentation method to a set of dermoscopic images. Images are excluded if 1) the entire lesion is not visible, 2) the image contains too much occluding hair or 3) there is insufficient contrast between the lesion and surrounding skin. In total, 596 images are segmented, 32 of which are deemed to be unsatisfactory.

As will be seen in section 3, our method employs two of these concepts from [3]: the use of textural information in segmentation and the use of Fisher’s separability criterion. Our application of these concepts, however, is substantially different.

2.2 Random Walker

The RW algorithm[2] is a general purpose, interactive, multi-label segmentation technique where a user labels the image with ‘seed points’ which denote the ground truth label for that pixel. Then, for an arbitrary pixel, the probability of a random walker reaching a seed of a specific label (before reaching seeds of any other label) is computed. However, the RW algorithm is sensitive to the exact placement of seeds and to the number of seeds placed[6]. While the RW algorithm is fast, intuitive and robust, it has been determined that a large number of seed points (up to 50% of the image) is required to reproduce a segmentation with only minor differences[6].

We have adopted the RW method described above into a novel framework, to automatically segment skin lesions from dermoscopic images.

3 Method

In this paper we present an approach to leverage the advantages of RW for automatic skin lesion segmentation. We initialize the RW algorithm automatically with seed points generated by ‘learning’ (by means of a training set) the difference between the properties of ‘skin lesion pixels’ and ‘healthy skin pixels’.

3.1 Supervised Probabilistic Segmentation

We begin with a set of 120 expertly segmented dermoscopic images taken from atlases[7][8]. Each pixel is assigned either the label ‘inside’ (l_1) or ‘outside’ (l_2)

based on the ground truth segmentation. In this stage we aim to learn the difference between these two groups. Images are converted to $L^*a^*b^*$ space, and each channel is filtered with a set of Gaussian and Laplacian of Gaussian filters. Let m denote the number of filters employed. Pixels are then represented as a $1 \times 3m$ vector since each filter is applied to each of the 3 image channels. Linear Discriminant Analysis (LDA)[4] is then used to determine the linear combination of filters that best discriminate ‘inside’ and ‘outside’ pixels. LDA is similar to Principal Component Analysis (PCA), but where PCA is an *unsupervised* technique that reduces dimensionality while maintaining variance, LDA is a *supervised* technique that reduces dimensionality while maintaining class separability. This is achieved through an eigenvalue decomposition of an $3m \times 3m$ scatter matrix, which represents the separability of the classes with respect to each filter. Since this is a 2-class problem, we consider only the principle eigenvector. This eigenvector results in a linear combination of the filtersets for each image channel. Since the filterset employed is a series of low-pass (Gaussian) and high-pass (Laplacian of Gaussian) filters, the resulting ‘eigenfilters’ can be interpreted as either a high, low, or multiple-band-pass filters. We are therefore not only learning the colour difference between these two groups of pixels, but also the difference in the spatial variation of colours. This process is illustrated in Figure 1.

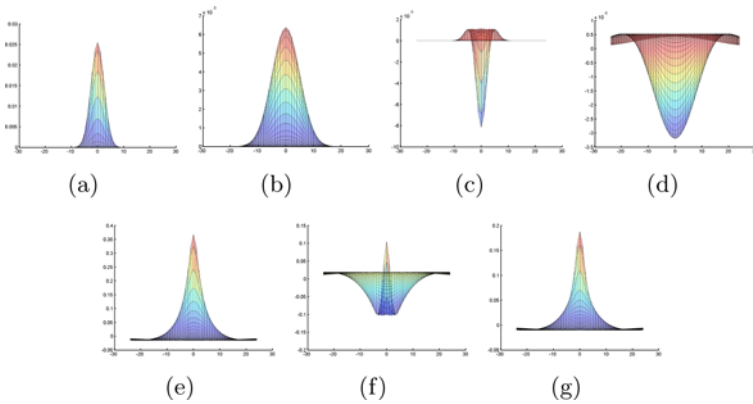


Fig. 1. Learning the difference between pixels inside and outside the segmentation. a)-d): Some filters from the filterset applied to each channel of each image. The filterset consists of Gaussian filters (a,b) and Laplacian of Gaussian filters (c,d) and the ‘eigenfilters’ as a result of LDA for the L^* , a^* and b^* channels respectively (e,f,g).

Next, the response of the pixel groups (‘inside’ and ‘outside’) along this eigenvector are modeled as Gaussian distributions

$$P(p|l_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \tag{1}$$

We create probability maps for unseen images by filtering the image with the resulting eigenfilters from LDA, and for each pixel p , assigning it a normalized probability that the pixel is inside the lesion

$$P = \frac{P(p|l_1)}{P(p|l_1) + P(p|l_2)} \quad (2)$$

The creation of a probability map is illustrated in Figure 2

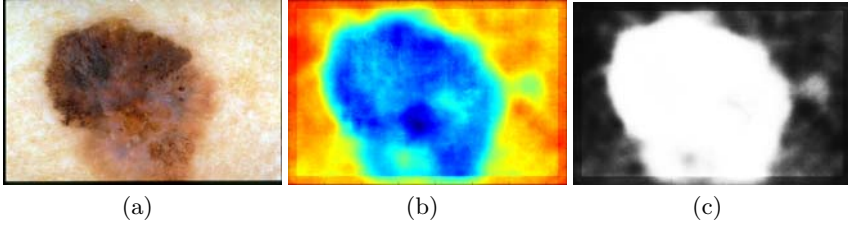


Fig. 2. The creation of a supervised probabilistic segmentation. a) The original dermoscopic image b) The image’s response to the ‘eigenfilter’ from Figure 1 c) The resulting probability map by applying equation 2. Note the high response to the photodamaged skin to the right of the lesion. This is due to the fact that this pattern (known as a *pigment network*) usually occurs within lesions.

3.2 Initializing the Random Walker Algorithm

The original RW algorithm is an interactive segmentation which requires the user to place seed points. In our proposed automatic RW approach, there is no user interaction and the object and the background seeds are automatically determined from the probability map generated in section 3.1. To generate seed points, two thresholds must be determined. Let T_S represent the skin threshold and T_L represent the lesion threshold. Once these thresholds are determined, an arbitrary number of seeds can be automatically generated as long as the thresholding constraints are satisfied. Let $P(p)$ represent the probability a pixel p is a part of the lesion, as determined by equation 2. A pixel is a candidate for a background seed if $P(p) < T_S$. Similarly, a pixel is a candidate for an object seed if $P(p) > T_L$.

To determine T_S and T_L , we analyze the histogram of the probability map (shown in Figure 3(b),(f)). We fit a Gaussian Mixture Model to the histogram and extract the dominant Gaussians that represent the skin and lesion[9]. Let μ_S and μ_L represent the means of the ‘skin’ and ‘lesion’ Gaussians respectively. Similarly, let σ_S and σ_L represent the variances. Thresholds are then determined by:

$$T_S = \mu_S + 3\sigma_S \quad (3)$$

$$T_L = \mu_L - 3\sigma_L \quad (4)$$

Now, let $F(x)$ represent the cumulative histogram of the probability map. We then define two metrics α_H and β_H , using the subscript H ('histogram') to differentiate from the β parameter of the RW algorithm:

$$\alpha_H = \frac{F(T_L) - F(T_S)}{F(T_S)} \quad (5)$$

$$\beta_H = \frac{F(T_L) - F(T_S)}{F(1) - F(T_L)} \quad (6)$$

Low values for both α_H and β_H imply an easy to segment, high contrast image, as shown in Figure 3(a)-(d). The area shaded red in Figure 3(b) denotes the amount of pixels for which a label cannot be determined with certainty.

If however, either α_H , β_H or both are above a certain threshold, then the contrast between the lesion and skin is poor, and the segmentation is more difficult. Empirically, this threshold has been defined as 2.0. If α is above 2.0 then we define a new skin threshold T'_S as the median of the uncertainty range (the range between T_S and T_L). Similarly, if β is above 2.0 we define T'_L as the median of the uncertainty range. If both α and β are above 2.0, we take the larger value to determine which threshold to shift. This threshold adaptation is illustrated in Figure 3(e)-(h). Initially α_H and β_H are computed in Figure 3(e). The amount of uncertain pixels is large (grey and red shaded are) which is reflected in the high value $\beta_H = 7.88$. Since $\beta_H > 2.00$, we define $T'_L = 0.42$, which reduces the uncertain region (red) considerably.

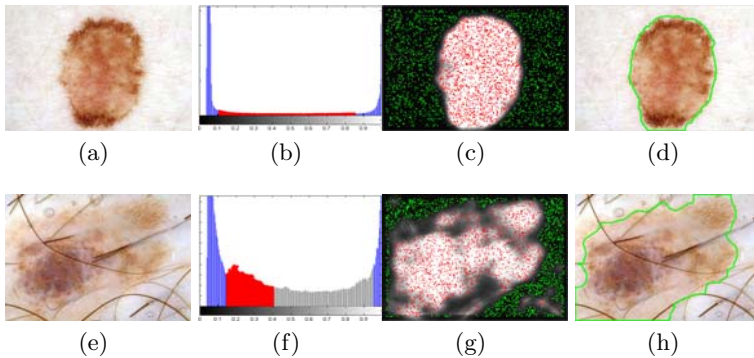


Fig. 3. Automatically initializing the RW algorithm. First row: A high contrast, easy to segment image. a) The initial image. b) The histogram of the image's probability map as generated by section 3.1. The blue area denotes candidate seed pixels ($\alpha_H = 1.27$, $\beta_H = 1.74$, $T_S = 0.10$, $T_L = 0.85$). c) Seed pixels randomly selected. d) The resulting segmentation. Second row: A difficult low-contrast lesion with occluding hair. The original parameters ($\alpha_H = 0.77$, $\beta_H = 7.88$, $T_S = 0.18$, $T_L = 0.97$) indicate its difficulty since, $\beta_H > 2.0$. T'_L is therefore set to 0.42 (reducing the uncertainty area to the red shaded region).

After determining the thresholds for the skin (T_S or T'_S) and the lesion (T_L or T'_L) pixels, seed points can now be chosen according to these thresholding constraints. We randomly choose 3% of pixels as seeds. Since spatial filtering methods are inaccurate near image borders (as can be seen in Figure 2(b) we impose an additional constraint and do not consider pixels in proximity to the image border as seed point candidates.

After placing seeds in the areas of high certainty, RW segments the image. RW gracefully handles the uncertain area in the probability map along the lesion border. We initialize the RW graph edge weights using a Gaussian function of the image intensity as Grady does[2]. The Gaussian width in this function, which we denote as β_{RW} , is a free parameter that determines the degree to which two intensities are considered similar. Throughout this paper, this parameter has been fixed at 30. Finally, after applying the RW algorithm, the segmentations undergo morphological post-processing to fill holes and break isthmuses.

4 Results

We tested our method on a dataset of images taken from [7] and [8]. We begin by selecting 100 images that pose a challenge to segmentation methods, and call this imageset ‘challenging’. These represent images that are often excluded from other studies[3]. An image is considered challenging if one or more of the following conditions is met: 1) the contrast between the skin and lesion is low, 2) there is significant occlusion by either oil or hair, 3) the entire lesion is not visible, 4) the lesion contains variegated colours or 5) the lesion border is not clearly defined. Next, we select 20 images that do not meet any of the above conditions, and call this imageset ‘simple’. We merge these two imagesets, calling the resulting imageset ‘whole’. Finally, we create an imageset to measure the intraobserver agreement of our expert. We randomly select 10 images from the ‘challenging’ imageset. These images undergo a random rotation of 90, 180 or 270 degrees, and some are randomly inverted along the X and/or Y axes. This is done to reduce the likelihood that the dermatologist would recognize the duplicate image while performing the segmentation task. We call this imageset ‘intra’.

Probability maps for all images are generated as described in section 3.1 using ten-fold cross validation. Seeds are placed automatically as described in section 3.2. The results are summarized in Table 1. We also compare our results to the Otsu thresholding method[10] and measure the intra-observer variability of the expert. Segmentations obtained from our modified random walker algorithm, the Otsu method and the dermatologist are denoted as ‘MRW’, ‘Otsu’ and ‘Derm’ respectively. For all comparisons we compute precision, recall, F-measure[11], and border error[12].

As can be seen in Table 1, while the Otsu method consistently achieves a higher precision, its recall is much worse. This implies that the Otsu method consistently underestimates the lesion border, labeling many pixels as ‘skin’ that ought to be labeled as ‘lesion’. When examining the more comprehensive metrics such as F-measure or border error, it is apparent that our modified random

Table 1. Comparing the results of our modified random walker segmentation algorithm (MRW) to that of Otsu’s thresholding method[10] (Otsu), and a dermatologist’s manual segmentation which acts as ground truth (Derm). Comparisons are performed over simple and challenging imagesets taken from [7] and [8]. See Section 4 for a description of these imagesets.

Comparison	Imageset	n	Precision	Recall	F-measure	Mean BE	Std BE
MRW vs. Derm	simple	20	0.96	0.95	0.95	0.079	0.024
MRW vs. Derm	challenging	100	0.83	0.90	0.85	0.31	0.19
MRW vs. Derm	whole	120	0.87	0.92	0.88	0.24	0.18
Otsu vs. Derm	simple	20	0.99	0.86	0.91	0.15	0.083
Otsu vs. Derm	challenging	100	0.88	0.68	0.71	0.44	0.40
Otsu vs. Derm	whole	120	0.91	0.74	0.78	0.34	0.36
Derm vs. Derm	intra	10	0.95	0.91	0.93	0.085	0.036

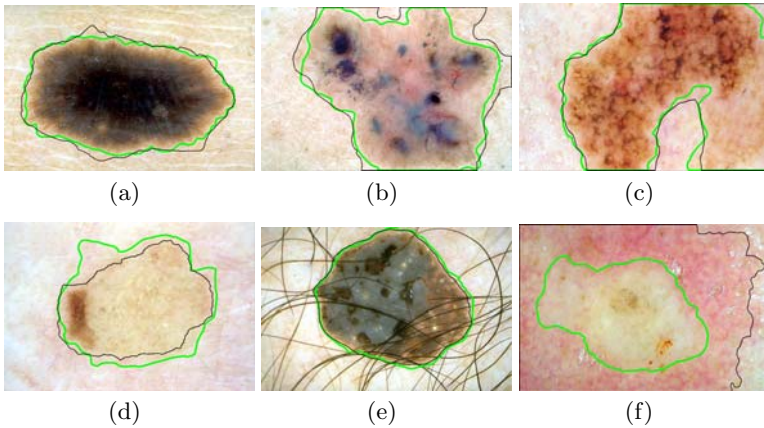


Fig. 4. Sample segmentation results for our method (denoted in black) compared to ground truth (denoted in green). a) A typical, easy to segment lesion. b) A lesion with variegated colours. c) An example of the entire lesion not being visible. Also, the lesion border is unclear in the bottom right hand side. d) A low contrast lesion. e) A lesion occluded significantly by hair. f) A difficult case where our method fails.

walker outperforms Otsu’s method. The poorer F-measure and border error results for the Otsu method on the challenging image set indicates its degree of difficulty. This is also born out by the results of the intra-observer agreement of the expert dermatologist on the ‘intra’ imageset.

Figure 4 shows sample results of the segmentations obtained from our method (denoted in black) as well as the ground truth segmentation (denoted in green) for a variety of lesions, including several difficult lesions.

5 Conclusion

We have developed a fully automatic method for segmenting unseen skin lesions by leveraging knowledge extracted from expert ground truth, and the random

walker algorithm. Our method uses colour as well as texture to perform the segmentation and adapts itself to handle difficult, low-contrast images. Clinically, this is the first step towards an automated skin lesion diagnosis system. Future work will refine the method, and validate it on a larger dataset.

Acknowledgments. This research was funded by NSERC, CIHR and MITACS. The authors would also like to thank their collaborators Dr. Harvey Lui, Dr. David McLean and Dr. Bernardita Ortiz-Policarpio for their guidance and segmentations.

References

1. Braun, R.P., Rabinovitz, H.S., Oliviero, M., Kopf, A.W., Saurat, J.H.: Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? *British Journal of Dermatology* 157(5), 926–933 (2007)
2. Grady, L.: Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(11), 1768–1783 (2006)
3. Celebi, M., Aslandogan, et al.: Unsupervised border detection in dermoscopy images. *Skin Research and Technology* 13(4), 454–462 (2007)
4. Fitzpatrick, J.M., Sonka, M.: *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis* (SPIE Press Monograph Vol. PM80). 1s edn. SPIE–The International Society for Optical Engineering (June 2000)
5. Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H.: A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 362–373 (2007)
6. Grady, L., Schiwietz, T., Aharon, S., Munchen, T.U.: Random walks for interactive organ segmentation in two and three dimensions: Implementation and validation. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 773–780. Springer, Heidelberg (2005)
7. Argenziano, G., Soyer, H., et al.: *Interactive Atlas of Dermoscopy* (Book and CD-ROM). Edra medical publishing and new media (2000)
8. Soyer, H., Argenziano, G., et al.: *Dermoscopy of Pigmented Skin Lesions. An Atlas based on the Consensus Net Meeting on Dermoscopy*. Edra medical publishing and new media (2000)
9. Huang, Z., Chau, K.: A new image thresholding method based on Gaussian mixture model. *Applied Mathematics and Computation* 205(2), 899–907 (2008)
10. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics* 9(1), 62–66 (1979)
11. Makhoul, J., Kubala, F., Schwartz, R., Weischedel, R.: Performance measures for information extraction. In: *Broadcast News Workshop 1999*, p. 249 (1999)
12. Hance, G., Umbaugh, S., Moss, R., Stoecker, W.: Unsupervised color image segmentation: with application to skin tumor borders. *IEEE Engineering in Medicine and Biology Magazine* 15(1), 104–111 (1996)