A wide spread of algorithms for automatic segmentation of dermoscopic images

Pedro M. Ferreira[1,2], Teresa Mendonça[1,3], and Paula Rocha[2,3]

[1] Faculdade de Ciências, Universidade do Porto, Portugal
meb09015@fe.up.pt
[2] Faculdade de Engenharia, Universidade do Porto, Portugal
[3] Center for Research Development in Mathematics and Applications, Portugal

Abstract. Currently, there is a great interest in the development of computer-aided diagnosis (CAD) systems for dermoscopic images. The segmentation step is one of the most important ones, since its accuracy determines the eventual success or failure of a CAD system. In this paper, different kinds of algorithms for the automatic segmentation of skin lesions in dermoscopic images were implemented and evaluated, namely automatic thresholding, k-means, mean-shift, region growing, gradient vector flow (GVF), and watershed. The segmentation methods were evaluated with three distinct metrics, using as ground truth a database of 50 images manually segmented by an expert dermatologist. Among the implemented segmentation approaches, the GVF snake method achieved the best segmentation performance.

Keywords: Dermoscopy, melanoma, skin lesion, image segmentation

1 Introduction

Skin cancer is among the most commonly diagnosed cancers, of which malignant melanoma is by far its most aggressive form. Although melanoma is less common than other types of skin cancer, its incidence has been quickly increasing over the last years [1]. Fortunately, when melanoma is diagnosed in an early stage, it can easily be treated through a simple excision of the lesion. Therefore, several diagnosis techniques have been explored to improve the early detection of melanoma. Dermoscopy has increasingly become the most important of such diagnosis techniques, since it allows the *in vivo* observation of pigmented skin lesions in a higher magnification, providing a detailed view of their morphological structures [1, 2]. Although dermoscopic images have a great potential in the early diagnosis of melanoma, their interpretation is subjective and significantly depends on the experience of dermatologists [1, 2].

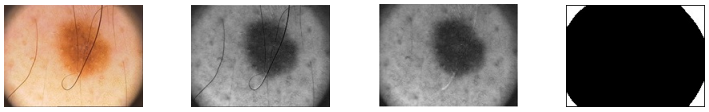
Therefore, several computer-aided diagnosis (CAD) systems for digital dermoscopic images have been introduced in order to support the clinical decision of dermatologists. A CAD system has usually three stages, namely: image segmentation, feature extraction and selection, and lesion classification. The segmentation step is one of the most important ones, since a good segmentation is fundamental for subsequent classification.

In this regard, several segmentation methods have been proposed. These methods can be roughly divided into four main groups, namely thresholding, region-based, edge-based, and clustering-based methods. A thresholding segmentation method is described in [3], where the segmented skin lesion is obtained by a fully automated histogram-based thresholding technique, in which the thresholding is performed in each of the three color planes. Several region-based methods have been used in the segmentation of dermoscopic images. Celebi *et al.* [4] suggest a modified version of the JSEG algorithm for the skin lesion segmentation. In this algorithm the segmentation process is divided into two independent phases: (i) a color quantization, and (ii) a multiscale region growing segmentation. Moreover, in [5] a color image segmentation technique based on region growing and merging, called statistical region merging algorithm, is proposed. An example of an edge-based method can be found in [6], where the skin lesion is segmented either by the geodesic active contours model or the geodesic edge tracing approach. In [7], the gradient vector flow (GVF) snakes method is used. Clustering approaches can be found in [8, 9]. In [8] the skin lesion is segmented using a modified version of the fuzzy c-means clustering technique that takes into account the cluster orientation. In [9], a comparison of the most widely used clustering algorithms, namely median cut, k-means, fuzzy c-means and mean shift, is presented. Although several individual segmentation methods for dermoscopic images have been proposed in the literature, only few works aimed at comparing and evaluating the performance of different segmentation algorithms. A remarkable example can be found in the work of Silveira *et al.* [10], where six different segmentation methods of three classes (thresholding, region-based, and edge-based) are proposed and evaluated. However, the referred work presents two limitative issues: (i) no clustering segmentation approach is used for comparison; and (ii) some of the proposed methods require user intervention for the initialization procedure.

The goal of this paper is to address this comparison issue using a broad range of segmentation methods and to propose new automatic initialization methods for region-based and edge-based approaches. For this purpose, different segmentation methods of the four classes (thresholding, region-based, edge-based, and clustering-based) are presented, implemented, and evaluated. Adaptations were made to some of these methods in order to improve their performance in dermoscopic images and/or to make the segmentation process fully automatic. These adaptations can also be useful in other segmentation applications. Three performance metrics were computed for the quantitative assessment of the segmentation results, using as ground truth a database of 50 images manually segmented by a medical expert.

2 Automatic segmentation methods

In this paper, six different segmentation methods are presented, implemented, and evaluated, including automatic thresholding, k-means, mean-shift, region growing, gradient vector flow (GVF), and watershed.



(a) (b) (c) (d)

Fig. 1: Pre-processing: original image (a), grayscale image (b), filtered image (c), binary mask of the dark regions (d).

Before performing segmentation itself, a pre-processing procedure is applied to dermoscopic images (Figure 1). First, the RGB dermoscopic image is converted into a grayscale image through the selection of the blue color channel, since this is the one that provides the best discrimination between the lesion and the skin. Afterwards, dermoscopic images are filtered with a hair removal filter [11] followed by a median filter for image smoothing. Finally, a binary mask of the dark regions in the four corners of the image is created, since it gives an essential information for the initialization of some of the implemented segmentation methods.

2.1 Automatic thresholding

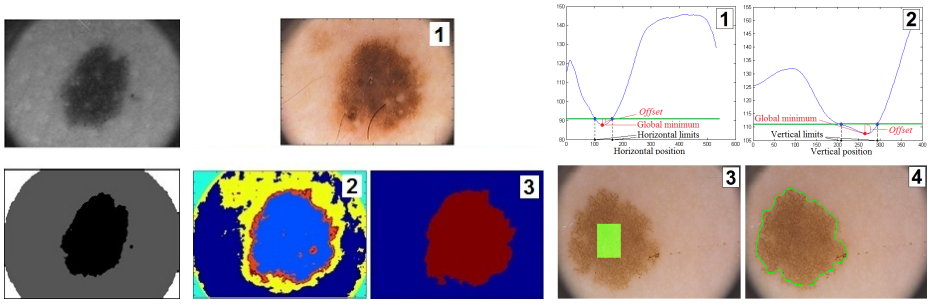
After analyzing the histogram of dermoscopic images, it was observed that most of the images have a bimodal histogram, in which one of these modes corresponds to the lesion and the other to the skin. However, there are a few exceptions. Some images have a unimodal histogram, either when the lesion is very small compared to the skin, or when the lesion is very large and covers almost the entire image.

Therefore, an algorithm for automatic detection of the number of significant histogram peaks (local maxima) was developed. Based on this algorithm an appropriate automatic thresholding method is used to segment the images. If the image histogram has two major peaks then the threshold is obtained through Otsu's method [12], while if the image histogram has a single peak the threshold value is obtained by triangle method [13].

2.2 *k*-means

k-means is a clustering technique that classifies an image into multiple clusters, based on the distances between each pixel and all cluster centroids [14]. In this method, the blue color component from the RGB color space is used as the input data. Another input parameter of k-means is the number of clusters.

Most of the dermoscopic images used in this work have three distinct regions, corresponding to the lesion, to the skin, and to the dark regions in the four corners of the image. However, some images do not present the dark regions in the four corners, and are composed only by two distinct objects. Therefore, the number of clusters used in the k-means algorithm can be either three or two. The definition of the number of clusters is automatically done based on the corner mask that is obtained in the pre-processing step (Figure 1). Figure 2 (a) shows the result of k-means clustering segmentation.



(a) (b) (c)

Fig. 2: (a) k-means segmentation, for $k = 3$; (b) Mean shift segmentation: original RGB image (1), mean shift clustering result (2), final segmentation result (3); (c) Region growing segmentation and initialization: vertical and horizontal projections (1 and 2, respectively), seed region (marked as a green rectangle) superimposed on the original image (3), final segmentation result (4).

2.3 Mean shift

The concept of mean shift is to cluster an image by associating each pixel with a local maximum of the probability density of the image. The local maxima are obtained by an iterative process using a density kernel estimator, and hence the only free parameter of mean shift is its kernel size [14].

For dermoscopic image segmentation, the mean shift algorithm is applied to obtain at least five clusters, since it was observed that considering only three clusters often leads to bad results, specially in the images with a very low contrast between the lesion and the skin. If less than five clusters are obtained, the radius of the used kernel is iteratively decreased and then the mean shift procedure is repeated. This number of clusters (five) has proved to be sufficient to obtain good results after subsequent merging. After clustering, the adjacent clusters are merged based on the mean intensity of each cluster (Figure 2 (b)). Given two adjacent clusters, C_i and C_j, the merging procedure is given by:

$$C_m = C_i \cup C_j, \qquad if \ \left| I_{C_i} - I_{C_j} \right| \leq T_m \tag{1}$$

where C_m is the merged region, I_{C_i} and I_{C_j} are the mean intensities of C_i and C_j. Therefore, adjacent clusters are only merged if the difference between their mean intensities is less than an empirically predefined threshold T_m. This value was easily defined based on the available image dataset.

2.4 Region growing

Region growing is a technique that starts with a pixel or a group of pixels, known as the seeds, and then groups together their neighbouring pixels into larger regions based on homogeneity criteria (e.g., average gray level, texture) [15].

The homogeneity criterion considered and used in this work is the average gray level. An automatic seed finding procedure was developed in order to make the region growing method fully automatic. This procedure is based on the vertical and horizontal projections of the image. The vertical projection of an image, $I(x, y)$, is a function of the horizontal index y:

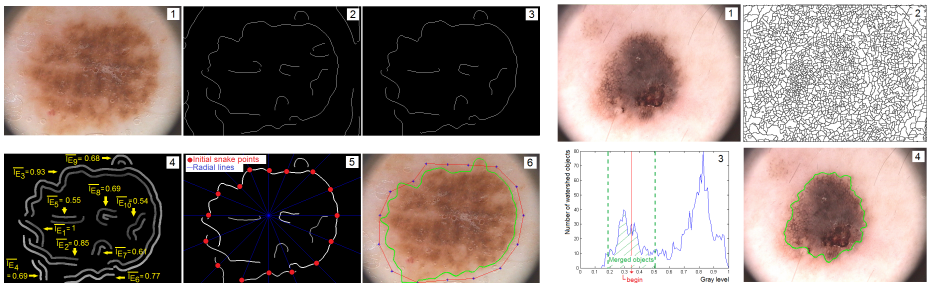
$$P_y = \frac{\sum_{x=1}^{n} I(x, y)}{(n - C_y)} \tag{2}$$

where n is the total number of lines of $I(x, y)$, and x is the vertical index. C_y corresponds to the number of pixels, in column y, belonging to the dark regions in the four corners of the image. The normalization by $(n - C_y)$ is used to reduce the influence of the dark corners in the image projection. A similar procedure is used to compute the horizontal projection P_x. The next step consists in the determination of the global minima of both projections. Then, a predefined offset is added to the global minima of the image projections, and an horizontal line passing through this point is defined. The two points of intersection between the horizontal line and the image projection are computed and used to define the limits of the seed region. The resulting seed region has a rectangular shape. Figure 2 (c) illustrates a segmentation example using region growing as well as the automatic seed finding process.

2.5 GVF snakes

Snakes are deformable curves defined within an image domain that can move towards the desired features, typically edges, under the influence of internal and external forces. The GVF field is used as external force. This field is computed as a diffusion of the image or edge gradient which provides a large capture range and the capability to segment object concavities [15].

An automatic snake initialization method was developed in order to make the segmentation process fully automated. This method can be divided into three main steps: **(i) Edge detection**: first a binary edge map of the image is computed with the canny edge detector. **(ii) Edge validation**: in this step some false positive edge segments are first eliminated (e.g., the edge segments corresponding to the dark corners, and the edges whose length is less than a predefined threshold). Then, two peripheral lines are defined in both sides of each edge. The normalized mean intensity difference between these peripheral lines is computed, $\overline{I_{E_i}}$, to be used as a measure of the relative importance of the edge. The underlying assumption is that this difference is larger in the edges of the lesion. **(iii) Initial curve determination**: first a number of radial lines are computed from a point within the lesion to the exterior. Next, in each radial line, a point is chosen as an initial snake point based on the values of $\overline{I_{E_i}}$ together with the distance to the inner point. Finally, the initial curve is obtained using a linear interpolation of the obtained initial points (Figure 3 (a)).



(a) (b)

Fig. 3: (a) GVF snake segmentation and initialization: original RGB image (1), binary edge map (2), false positive edge segments removal (3), determination of the normalized mean intensity difference between the peripheral lines (4), initial snake points finding (5), initial snake contour (red) and segmentation result (green contour) (6); (b) Watershed segmentation: original RGB image (1), oversegmentation of the lesion (2), histogram of the watershed objects (3), watershed segmentation result after object merging (4).

2.6 Watershed

The watershed transform considers an image as a topographic surface, which is slowly flooded from the holes at each regional minimum. When the rising water coming from two distinct minima is about to merge, a watershed line is created. These watershed lines correspond to the boundaries of image objects [15].

For dermoscopic image segmentation, the watershed transform is applied to the magnitude of the image gradient. However, as gradients of dermoscopic images usually present many regional minima, the result is an oversegmentation of the skin lesion. To overcome this problem, the watershed objects are merged based on the histogram of their mean intensities. The initial snake curve used in the automatic initialization of the GVF snake method (previously described in subsection 2.5) is also employed in the merging procedure. The mean intensity of the initial snake curve mask, L_{begin}, is used to start the merging procedure, and its area is taken as an estimate of the skin lesion area, $A_{estimated}$. Therefore, the merging procedure starts at the gray-level L_{begin} of the watershed objects histogram. Then, the watershed objects are iteratively merged until a stopping condition is achieved. This stopping condition depends both on $A_{estimated}$, and on the mean intensity of the watershed objects, more concretely:

$$0.8 \cdot A_{estimated} < A_{merged} < 1.2 \cdot A_{estimated} \quad \wedge \quad |I_{merged} - I_{objects}| > T_s \quad (3)$$

where A_{merged} and I_{merged} are the area and the mean intensity of the current merged region respectively, $I_{objects}$ is the mean intensity of the watershed objects that are about to be merged, and T_s is a predefined threshold value. Figure 3 (b) shows a segmentation example using watershed.

Table 1: Results of the segmentation methods.

Segmentation method	HM(%)	FPR(%)	FNR(%)	Gross errors(%)
Automatic thresholding	12.58	6.16	2.05	11
K-means	11.20	5.38	2.19	11
Mean shift	10.65	4.21	3.82	7
Region growing	10.35	4.43	3.68	4
GVF snakes	10.14	3.79	2.85	2
Watershed	12.36	2.70	5.77	7

3 Experimental results

The implemented segmentation methods were evaluated on a set of 50 images obtained from the Hospital Pedro Hispano database. The manual segmentation of each image to be used as ground truth in the evaluation of the segmentation methods was performed by an expert dermatologist. Three performance metrics were used to quantify the boundary differences, namely Hammoude distance (HM), false negative rate (FNR), and false positive rate (FPR) [10].

Table 1 shows the median of the performance metrics for the six implemented segmentation methods as well as the percentage of gross errors. This percentage is used as a measure of the segmentation methods robustness, since it corresponds to the rate of segmented images with a Hammoude distance greater than 30%. Analyzing the three performance metrics together, the GVF snake method can be considered as the best segmentation method, since it has the lowest HM (10.14%), the lowest percentage of gross errors (2%), and also the best compromise between the FPR (3.79%) and the FNR (2.85%). We believe that the developed automatic snake initialization procedure is one of reasons for the highest performance of the GVF snake method, since the generated initial snake curves are in general placed very close to the skin lesion boundaries.

Although the six implemented segmentation methods provide acceptable results for the majority of the tested images, there are three main groups of images in which most of the algorithms demonstrate limitations: (i) images in which the lesion is fragmented, (ii) lesions presenting a great variety of colors and textures, and (iii) images with a very low contrast between the lesion and the skin.

4 Conclusion

In this paper, a wide spread of algorithms for the automatic segmentation of dermoscopic images were implemented and evaluated, including the automatic thresholding, region growing, watershed, k-means, mean-shift, and GVF snakes. For some of these methods adaptations were made in order to improve their performance to dermoscopic images and/or to make the segmentation process fully automatic. Among the implemented segmentation methods, the GVF snake method achieved the best results, and has proved to be useful and robust enough for the automatic skin lesion segmentation in a CAD system.

Acknowledgements

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through the project ADDI (PTDC/SAU-BEB/103471/2008), by *FEDER* funds through the Operational Programme Factors of Competitiveness (*COMPETE*), by Portuguese funds through the *Center for Research and Development in Mathematics and Applications* and the FCT, within project PEst-C/MAT/UI4106/2011 with COMPETE number FCOMP-01-0124-FEDER-022690. We thank Dr. Jorge Rozeira from Hospital Pedro Hispano for segmenting the dermoscopic images used as ground truth in this work.

References

1. Argenziano, G., Soyer, H., Giorgio, V.D., Piccolo, D., *et al.*: Dermoscopy, an interactive atlas. EDRA Medical Publishing (2000) http://www.dermoscopy.org.
2. do Carmo, G.C., e Silva, M.R.: Dermoscopy: basic concepts. Int J Dermatol. **47**(7) (July 2008) 712–719
3. Pagadala, P.: Tumor border detection in epiluminescence microscopy images. Master's thesis, University of Missouri-Rolla (1998)
4. Celebi, M.E., Aslandogan, Y.A., Bergstresser, P.R.: Unsupervised border detection of skin lesion images. in Int. Conf. on Information Technology: Coding and Computing **2** (2005) 123–128
5. Celebi, M.E., Kingravi, H.A., Iyatomi, H., Aslandogan, Y.A., *et al.*: Border detection in dermoscopy images using statistical region merging. Skin Research and Technology **14**(3) (2008) 347–353
6. Chung, D.H., Sapiro, G.: Segmenting skin lesions with partial-differential-equations-based image processing algorithms. IEEE Transactions on Medical Imaging **19**(7) (2000) 763–767
7. Erkol, B., Moss, R.H., Stanley, R.J., Stoecker, W.V., Hvatum, E.: Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. Skin Research and Technology **11**(1) (2005) 17–26
8. Schmid, P.: Segmentation of digitized dermatoscopic images by two-dimensional color clustering. IEEE Transactions on Medical Imaging **18**(2) (1999) 164–171
9. Melli, R., Grana, C., Cucchiara, R.: Comparison of color clustering algorithms for segmentation of dermatological images. in *Proc.* of the SPIE Medical Imaging **6144** (2006)
10. Silveira, M., Nascimento, J.C., Marques, J.S., Marçal, A.R.S., *et al.*: Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. IEEE Journal of Selected Topics in Signal Processing **3**(1) (2009) 35–45
11. Barata, C., Marques, J.S., Rozeira, J.: Detecting the pigment network in dermoscopy images: a directional approach. Conf. Proc. IEEE Eng. Med. Biol. Soc. (2011) 5120–5123
12. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst., Man, Cybern. **9**(1) (1979) 62–66
13. Zack, G.W., Rogers, W.E., Latt, S.A.: Automatic measurement of sister chromatid exchange frequency. J Histochem Cytochem. **25**(7) (1977) 741–753
14. Suri, J.S., Wilson, D.L., Laxminarayan, S.: Handbook of Biomedical Image Analysis. Kluwer Academic/Plenum Publishers (2005)
15. Gonzalez, R.C., Woods, R.E.: Digital image processing. 2 edn. Upper Saddle River: Prentice Hal (2002)