PH^2 - A Public Database for the Analysis of Dermoscopic Images

TERESA MENDONÇA

Faculdade de Ciências
Universidade do Porto
Porto, Portugal
Email: tmendo@fc.up.pt

PEDRO M. FERREIRA

Faculdade de Engenharia
Universidade do Porto
Porto, Portugal
Email: meb09015@fe.up.pt

ANDRÉ R. S. MARÇAL

Faculdade de Ciências
Universidade do Porto
Porto, Portugal
Email: andre.marcal@fc.up.pt

CATARINA BARATA, JORGE S. MARQUES

Institute for Systems and Robotics
Instituto Superior Técnico
Lisboa, Portugal
Email: ana.c.fidalgo.barata@ist.utl.pt, jsm@isr.ist.utl.pt

JOANA ROCHA, JORGE ROZEIRA

Hospital Pedro Hispano
Matosinhos, Portugal
Email: joanacgomesrocha@hotmail.com, jrozeira@me.com

1.1 Introduction

Skin cancer represents a serious public health problem because of its increasing incidence and subsequent mortality. Among skin cancers, malignant melanoma is by far the most deadly form. As the early detection of melanoma significantly increases the survival rate of the patient, several non-invasive imaging techniques, such as dermoscopy, have been developed to aid the screening process [1]. Dermoscopy involves the use of an optical instrument paired with a powerful lighting system, allowing the examination of skin lesions in a higher magnification. Therefore, dermoscopic images provide a more detailed view of the morphological structures and patterns as compared to the normally magnified images of the skin lesions [1, 2]. However, the visual interpretation and examination of dermoscopic images can be a time consuming task and, as shown by Kittler *et al*. [3], the diagnosis accuracy of dermoscopy significantly depends on the experience of the dermatologists. Several medical diagnosis procedures have been introduced in order to guide dermatologists and other health care professionals, such as the pattern analysis, the ABCD rule, the 7-point checklist, and the Menzies method. A number of dermoscopic criteria (i.e. asymmetry, border, colors, differential structures) have to be assessed in these methods to produce the final clinical diagnosis. However, the diagnosis of skin lesions is still a challenging task even using these medical procedures mainly due to the subjectivity of clinical interpretation and lacking of reproducibility [1, 2].

Several computer-aided diagnosis (CAD) systems for digital dermoscopic images have been proposed to assist the clinical evaluation of skin lesions. Such CAD system are usually based on three stages: image segmentation, feature extraction/selection, and lesion classification. Each of these stages has challenges and therefore needs to have a proper evaluation and validation, which requires reliable reference (or ground ground truth) data. The reference data has to be prepared and validated by expert dermatologists. The

process is however time consuming, particularly in what refers to the manual segmentation of the lesion in each dermoscopic image. For this reason there is currently a lack of public ground truth dermoscopic image databases available for public use (possibility limited to research and educational purposes). This was recognised in the ADDI project (Automatic computer based Diagnosis system for Dermoscopy Images) [4] and thus one of its main objectives was to prepare and made available a dermoscopic reference database.

A dermoscopic image database is presented here - (PH^2). This database can be used as ground truth in the evaluation and validation of both segmentation and classification algorithms. The PH^2 database contains a total of 200 melanocytic lesions, including 80 common nevi, 80 atypical nevi, and 40 melanomas. The rather small number of melanomas, compared with the other two types of melanocytic lesions, can be explained by two main reasons. First of all, the number of real cases of melanomas is actually much smaller than the other ones. In addition, as melanomas are usually not completely inserted in the image frame and present many image artifacts, they are not always suitable to be used as ground truth in the evaluation of CAD systems.

For each image in the database, the manual segmentation and the clinical diagnosis of the skin lesion as well as the identification of other important dermoscopic criteria are available. These dermoscopic criteria include the assessment of the lesion asymmetry, and also the identification of colors and several differential structures, such as pigment network, dots, globules, streaks, regression areas and blue-whitish veil. The PH^2 database is freely available for research and educational purposes, following a brief registration process at the website [4].

The size of the PH^2 database (200 images) might seem small, particularly when compared with a traditional machine learning ground truth databases, which may have hundreds or thousands of annotated images. However, it is important to highlight that the annotation of dermoscopic images is not just a binary issue (benign or malign). The annotation of each image requires a large amount of time and effort, since several dermoscopic features have to be assessed to perform the lesion diagnosis. Moreover, the skin lesion and the color classes present in each image have to be manually segmented by expert clinicians. The PH^2 database can also be used for medical training. For instance, dermatologist trainees can test their skills by comparing their own diagnosis and evaluation with the ground truth available in the PH^2 database.

1.2 PH^2 Database

The creation of the PH^2 database was made possible by a joint collaboration between the Universidade do Porto and Universidade de Lisboa in conjunction with the Dermatology service of Hospital Pedro Hispano in Matosinhos, Portugal. The PH^2 database was mainly created to make available a common dataset that may be used for the performance evaluation of different computer aided diagnosis systems of dermoscopic

images.

The database currently consists of an image set of 200 dermoscopic images along with the corresponding medical annotations, comprising 80 common nevi, 80 atypical nevi, and 40 malignant melanomas. The dermoscopic images were carefully acquired using a magnification of 20$\times$ under unchanged conditions. They are 8-bit RBG color images with a resolution of 768×560 pixels. The set of images available in the PH^2 database was selected with some constraints, regarding their quality, resolution, and dermoscopic features, so that they are suitable enough to be used in a dermoscopic reference database. An illustrative collection of the images that can be found in the PH^2 database is presented in Fig.1.1.

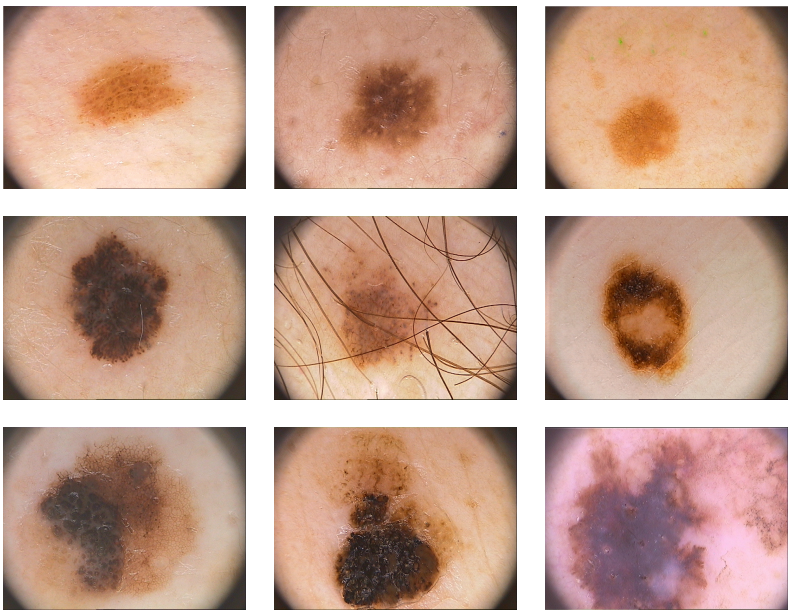


Figure 1.1: A sample of images from PH^2 database, including common nevi (1st row), atypical nevi (2nd row) and melanomas (3rd row).

Each image of the database was visually inspected and manually annotated by an expert dermotologist with regard to three main parameters, including:

1. Manual segmentation of both skin lesion and colors;

2. Clinical and histological diagnosis;

3. Dermoscopic criteria, which include the assessment of the following features: asymmetry, colors, pigment network, dots/globules, streaks, regression areas, blue-whitish veil).

The PH^2 dataset is available in [4] together with an interactive image viewing tool, called PH^2 Browser. This tool allows the visualization of the medical segmentations and classifications of each image of the

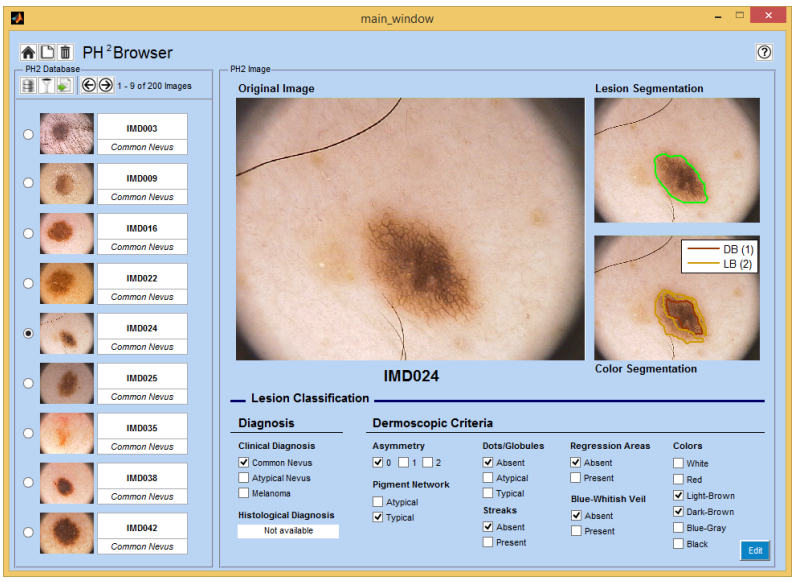


Figure 1.2: Graphical user interface of the PH^2 Browser.

database. Furthermore, it allows searching dermoscopic images according to the desired criteria as well as the exportation of the images, along with the corresponding segmentation and classification, into several formats. Fig.1.2 shows the graphical user interface of the PH^2 Browser.

1.2.1 Manual segmentation of the skin lesion

The availability of manually segmented skin lesions, performed by expert dermatologists, is of crucial importance since they give essential information for the evaluation of the segmentation step of a CAD system.

The manual segmentation of each image of the database is available in a binary format, more concretely as a binary mask with the same size of the original image. The pixels of the skin lesion are labeled with the intensity value of 1, whereas all pixels of the background region have values of 0. This output format was chosen because it can be easily used to extract the boundary position of the skin lesion. An example of a dermoscopic image and the corresponding manual segmentation is presented in Fig. 1.3.

1.2.2 Clinical diagnosis

Regarding the nature of melanocytic lesions, they can be broadly classified into benign nevi and malignant melanomas. Benign nevi can be further classified into common and atypical nevus, concerning their atypia degree. This classification procedure was adopted and followed during the assessment of the PH^2 images, and hence each image of the database is classified into common nevus, atypical nevus, or melanoma. Although the clinical diagnosis is available for all images of the database, the histological diagnosis is just available for a subset of them, since the histological test is just performed for those lesions considered highly

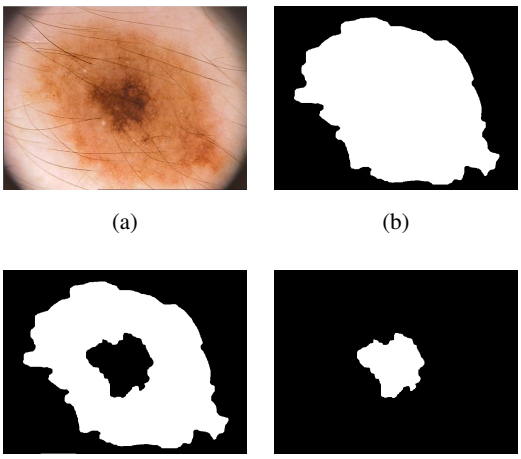


Figure 1.3: Manual segmentation of both skin lesion and color classes of a dermoscopic image: original image (a), manual segmentation of the skin lesion (b), manual segmentation of the light-brown color class (c), and manual segmentation of the dark-brown color class (d).

suspicious by dermatologists. Examples of each type of skin lesion are illustrated in Fig.1.1.

1.2.3 Dermoscopic criteria

The dermoscopic features available in the PH^2 database are: asymmetry, colors, pigment network, dots, globules, streaks, regression areas and blue-whitish veil. These dermoscopic features were selected for two main reasons. First, this set of dermoscopic criteria corresponds to those features that are commonly used by dematologists to perform a clinical diagnosis. In addition, this set of features comprises the majority of the dermoscopic features that have to be assessed in the most widely used medical diagnosis procedures, such as the ABCD rule, the 7-point checklist and the Menzies method. The definition of each of these features along with the description of its assessment process are described in the following.

- Asymmetry

The asymmetry plays an important role in the diagnosis of a melanocytic lesion. For instance, it is the largest weight factor in the ABCD rule of dermoscopy.

In the PH^2 database, the lesion asymmetry was evaluated by the dermatologist according to the ABCD rule. Therefore, the asymmetry of a lesion is assessed regarding its contour, colors, and structures distribution simultaneously. Moreover, there are three possible labels for this parameter: 0 for fully symmetric lesions; 1 for asymmetric lesions with respect to one axis; and 2 for asymmetric lesions with respect to two axes.

- Colors

Overall, six different colors are taken into account during the diagnosis of a melanocytic lesion. The set of color classes comprises the white, red, light-brown, dark-brown, blue-gray, and black [1].

Each image of the database was evaluated by a dermatologist in order to identify the presence, as well as the location, of the six color classes. The location of each color in an image was recorded as a binary mask, manually segmented by the dermatologist. An example is presented in Fig.1.3, where two color classes were identified.

- Pigment network

The pigment network appears in melanocytic lesions as a grid-like network consisting of thin pigmented lines and hypopigmented holes, creating a shape similar to a honeycomb [1]. Pigment network is one of the most important structures in dermoscopy, since its presence allows the distinction between different skin lesions classes as well as the identification of the lesion diagnosis, since the presence of an atypical network is usually a sign of a malignant lesion. This structure was visually evaluated by dermatologists, and classified as typical or atypical. The presence of atypical pigmented network in a skin lesion is illustrated in Fig.1.4.

- Dots/Globules

As illustrated in Fig.1.4, dots/globules are spherical or oval, variously sized, black, brown or gray structures (dots are usually smaller than globules). The presence of these dermoscopic structures has also an important role in the distinction between skin lesions classes [1].

These structures were visually evaluated by dermatologists, and categorically classified as present or absent in each image of the PH^2 database. When dots/globules are present in a given lesion, these structures are further classified as regular or irregular concerning their distribution in the lesion.

- Streaks

Streaks are finger-like projections of the pigment network from the periphery of the lesion. Instead of both pigment network and dots/globules, the presence of streaks in a skin lesion is by itself a sign of malignancy [1]. Therefore, these structures are just classified as present or absent in each image of the database.

- Regression areas

Regression areas are defined as white, scar-like depigmentation often combined with pepperlike regions (speckled blue-gray granules) [1]. In the PH^2 database, this parameter is classified in two main groups

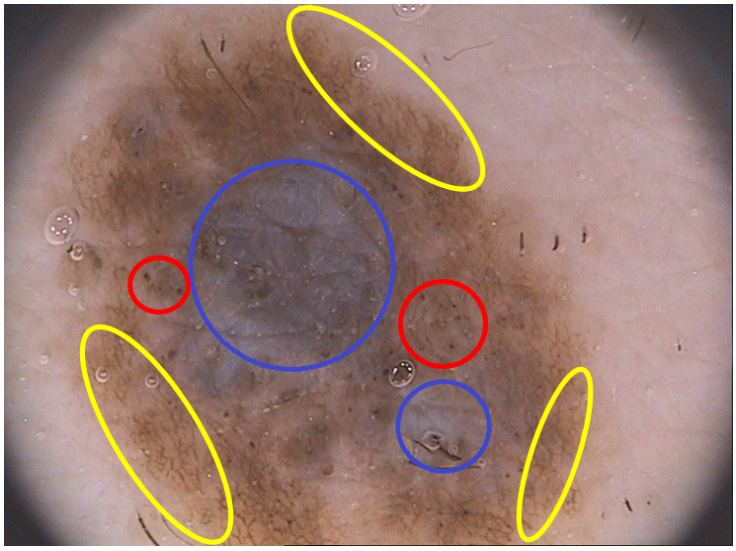


Figure 1.4: Dermoscopic features identification: atypical pigment network (yellow ellipses), atypical dots/globules (red circles), and blue-whitish veil (blue circles).

(present or absent) concerning its presence in the skin lesion.

- Blue-whitish veil

The blue-whitish veil can be defined as a confluent, opaque, irregular blue pigmentation with an overlying, white, ground-glass haze. Its presence is a strong malignancy indicator [1]. This dermoscopic structure is categorically labeled as present or absent, in each image of the database. The presence of the blue-whitish veil in a melanocytic lesion is illustrated in Fig.1.4.

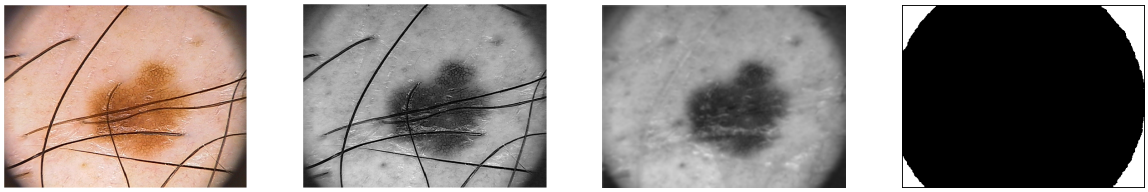
1.3 Dermoscopy Image Analysis

1.3.1 Image segmentation

Image segmentation is one of the most important tasks in image processing, since its accuracy determines the eventual success or failure of computerized analysis procedures. In the dermoscopy image analysis field, segmentation is used in order to automatically extract the pigmented skin lesion from the surrounding skin. However, dermoscopic images are a great challenge for segmentation algorithms, because there are a great diversity of lesion shapes, boundaries, and colors along with several skin types and textures. Moreover, dermoscopic images usually contain some intrinsic skin features such as hairs, blood vessels, and air bubbles. Therefore, several segmentation algorithms have been suggested to overcome these difficulties. A comprehensive survey of the methods applied to the segmentation of skin lesions in dermoscopic images is provided in [5].

Herein, the segmentation problem is addressed using gradient vector flow (GVF) snakes [6]. Snakes (or active contours) are deformable curves that are pushed towards nearby edges under the influence of external and internal forces computed from the curve itself and the image data, respectively. Active contours algorithms are often very sensitive to initialization. The initial snake points can be manually defined by an operator or automatically determined. In this work, an automatic snake initialization method is proposed, in order to make the segmentation process fully automated.

Before performing segmentation itself, a pre-processing procedure is applied to dermoscopic images (Figure 1.5). 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 [7] 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 to remove the influence of these regions on the segmentation results.



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

Figure 1.5: Pre-processing: original image (a), grayscale image (b), filtered image (c), binary mask of the dark regions (d).

Gradient Vector Flow (GVF) background

The gradient vector flow (GVF) snakes method is an extension of the classic snakes algorithm [8]. The snake is defined by a parametric curve $v(s) = (x(s), y(s))$, where x and y are the coordinates along the contour, and $s \in [0, 1]$ is the parametric domain. The evolution of the snake from an initial position to the object boundaries is expressed as an energy minimization process that can be written as:

$$E_{snake} = \int_0^1 P(v(s)) + \mathbf{g}(v(s))ds \tag{1.1}$$

where $P(x, y)$ denotes the internal energy that controls the contour shape and size, and hence the way the contour can stretch and curve; and $\mathbf{g}(x, y) = [u(x, y), v(x, y)]$ is the GVF field. The GVF field, $\mathbf{g}$, is computed as a diffusion of the gradient vectors of a gray-level or binary edge map derived from the image, which is obtained by minimizing the following energy function:

$$\epsilon = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 \, |\mathbf{g} - \nabla f|^2 \, dxdy \tag{1.2}$$

where f is the edge map of a given image, and μ is a regularization parameter that controls the degree of smoothness of the GVF field, and hence should be defined according to the amount of noise present in the image [6].

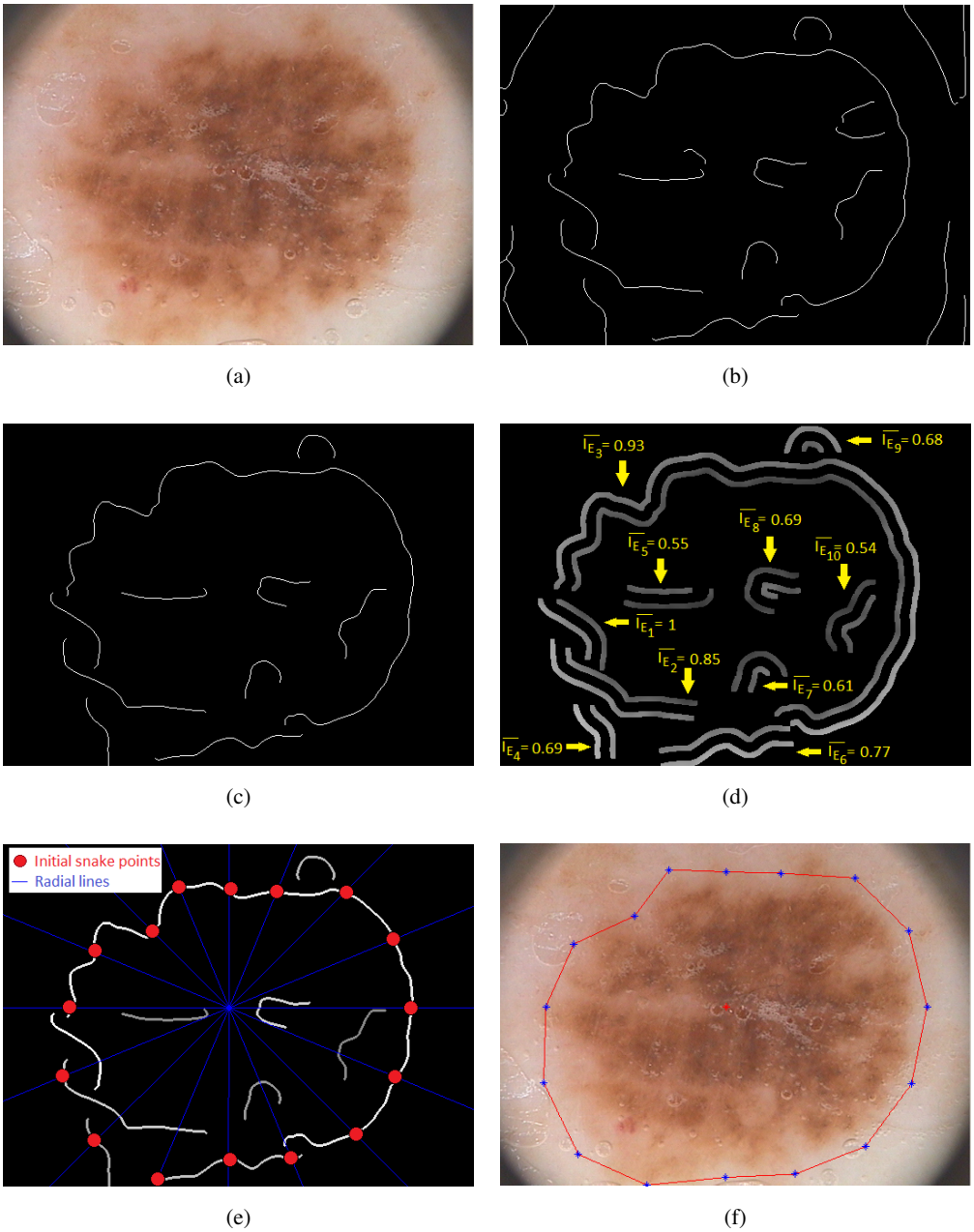


Figure 1.6: Automatic snake initialization method: (a) Original RGB image; (b) Edge map obtained through the Canny edge detector; (c) Edge map after removing some false positives edge segments; (d) Determination of the normalized mean intensity difference between the peripheral regions; (e) Initial snake points finding process; and (f) Initial snake curve.

Automatic snake initialization

The proposed automatic snake initialization method is mainly based on the information obtained from the Canny edge detector [9], and can be described into three main steps, namely (i) edge detection; (ii) edge validation; and (iii) initial curve determination.

(i) Edge detection

The aim of this step is to create a binary edge map from the gray-level image. To accomplish this purpose, the Canny edge detector algorithm is used. The binary edge map obtained through the Canny edge detector is visible in Figure 1.6(b), in which each pixel is labeled as either an edge point (value 1) or a nonedge point (value 0).

(ii) Edge validation

At this stage the edge map includes a large number of false positives edge segments. These false positives are usually resulting from the presence of the dark regions in the four corners of the image, and also from pigment network segments, skin lines, and even hairs when these artifacts have not been completely removed in the pre-processing step. Therefore, the edge segments corresponding to the dark corners are first eliminated. Then, since edges of the skin lesions are larger than most of noisy edges, the length is used as a criterion in order to eliminate the edges whose length is less than a predefined threshold. The effect of this step is illustrated in Figure 1.6(c).

The next step aims to quantify the relative importance of each edge. To accomplish this purpose, the peripheral regions of each edge are identified (Figure 1.6(d)). These two regions in both sides of the edges are obtained with the application of a morphological dilation to each edge individually. It is important to note that the pixels immediately adjacent to the edges are not considered in the peripheral regions in order to reduce the relative importance of the edges created by small transitions (i.e. skin lines, hairs, etc).

The difference between the mean intensity of the peripheral regions is computed as a measure of the relative importance of each edge. The underlying assumption is that this difference is larger in the edges of the skin lesion. Given n edge segments E_i, $i = 1, ..., n$, the measure of the importance of each edge is given by:

$$I_{E_i} = \left| P_{E_{i_1}} - P_{E_{i_2}} \right| \tag{1.3}$$

where $P_{E_{i_1}}$ and $P_{E_{i_2}}$ are the mean intensities of the peripheral regions. The maximal mean intensity difference is used to normalize I_{E_i}, thus yielding:

$$\overline{I_{E_i}} = \frac{I_{E_i}}{\max_i I_{E_i}} \tag{1.4}$$

Then, every pixel of a given edge E_i is assigned with the value of the respective normalized mean intensity difference between the peripheral regions $\overline{I_{E_i}}$. $\overline{I_{E_i}}$ values range from 0 to 1 and, as expected, the edges of the skin lesion have the highest $\overline{I_{E_i}}$ values (Figure 1.6(d)).

(iii) Initial curve determination

In this step the initial curve to be used in the initialization of the GVF method is automatically defined, by first determining a set of initial points which are then connected to form a closed curve. To accomplish this purpose, a number of radial lines R_{θ_j} are drawn from a point within the lesion to the exterior, each of them with a particular orientation $\theta_j \in [0, ..., 2\pi[$, $j = 1, ..., 16$. The center point, $C(x_c, y_c)$, is automatically computed based on the vertical and horizontal projections of the image. As skin lesions are darker than the surrounding skin, the global minimizers of the image projections are used to provide the coordinates of C.

Then, an initial snake point is defined in each radial line R_{θ_j} as follows. First, take the intersection of this line with the edges E_i, $i = 1, ..., n$. Let P_{θ_j} be the set of all edge points detected along the radial line R_{θ_j}, i.e., $P_{\theta_j} = \left\{p_j^1, ..., p_j^{N_j}\right\}$ and let $Q_{\theta_j} = \left\{q_j^1, ..., q_j^{N_j}\right\}$ be the set of values of the mean intensity difference between peripheral regions, $\overline{I_E}$, associated to each edge point p_j^k, $k = 1, ..., N_j$. Then a subset S_{θ_j} of P_{θ_j} containing the edge points with highest Q_{θ_j} values, is defined as:

$$S_{\theta_j} = \left\{p_j^k \mid q_j^* - q_j^k \leq T_E\right\} \tag{1.5}$$

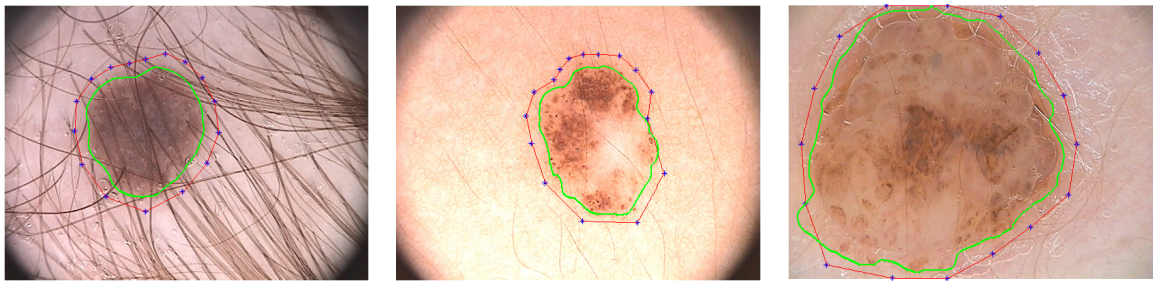
where

$$q_j^* = \max_{k=1,...,Nj} q_j^k \tag{1.6}$$

and T_E is a predefined threshold value. If S_{θ_j} only has one element, then this point is defined as the initial snake point, s_j, along the radial line R_{θ_j}. In case there are more than one point in the subset S_{θ_j}, the initial snake point is the point s_j^* whose distance to the inner point C is larger, provided that the distance between s_j^* and the point p_j^*, corresponding to the maximum value q_j^*, is not larger than a certain threshold T_d. Figure 1.6(e) illustrates the detection process of the initial snake points positions. It is important to note that when a given radial line R_{θ_j} does not intersect any edge ($P_{\theta_j} = \emptyset$), no initial snake point is defined in that line.

After detecting the initial snake points s_j, a curve is obtained using a linear interpolation of these points. Finally, in order to obtain the initial snake curve, this curve is uniformly expanded in all outward directions by 20 pixels to ensure that it contains the skin lesion (Figure 1.6(f)).

Figure 1.7 illustrates the robustness of the automatic snake initialization method, since it works well even in dermoscopic images with a large amount of hairs, in images with fragmented skin lesions, and also in images with skin lesions with different colors and textures. Furthermore, the final segmentation result is achieved after few iterations, since the initial snake curves are in general placed very close to the skin lesion boundaries.



(a) (b) (c)

Figure 1.7: GVF snake segmentation in difficult dermoscopic images: (a) Presence of hairs; (b) Fragmented skin lesion; and (c) Skin lesion with multiple colors. In these images the dotted red contour represents the initial snake curve, whereas the green contour corresponds to the final segmentation.

Segmentation Results

In this section, the experimental results of the implemented segmentation method are presented. As the proposed automatic initialization method is mainly based on the gradient information, it is not suitable enough to produce good results when the skin lesions are not completely contained in the image. Therefore, the segmentation algorithm was not tested on the entire PH^2 database, but only in the images of the database in which the skin lesion is contained in the image domain. This results in a subset of 174 images for test.

Three performance metrics were used to quantify the boundary differences, namely the border error rate (Err), the false negative rate (FNR), and the false positive rate (FPR). Err is given by $\frac{\#FP+\#FN}{\#GT}$, where GT represents the manually segmented area.

Table 1.1 shows the mean, the median, and the standard deviation (Std) of the performance metrics obtained by the implemented segmentation method. The segmentation method achieved an average Err, FPR, and FNR of 11.9%, 5.15%, and 2.05%, respectively. These results demonstrate that the implemented segmentation algorithm provide good results for the majority of the tested images. However, there are three main groups of images in which the algorithm may 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.

Table 1.1: Performance of the implemented segmentation algorithm.

	Err (%)	FPR (%)	FNR (%)
Mean	11.9%	5.15%	2.05%
Median	9.2%	3.48%	1.00%
Std	7.1%	5.46%	3.30%

1.3.2 Color labeling

Color plays a major role in the analysis of dermoscopy images [1]. For example, the ABCD rule of dermoscopy considers six admissible colors (blue-gray, black, white, dark and light brown, red) that medical doctors try to detect in melanocytic lesions. The number of colors observed in an image is a malignancy measure that is combined with other criteria (border, symmetry, differential structures) [10].

The detection of colors is a subjective task which requires considerable training. Nonetheless, the development of computational methods for this task is a desired goal. This problem is addressed in [11], assuming that the feature vector **y**, associated to a patch and a color c is a random variable described by a mixture of Gaussians

$$p(\mathbf{y}|c,\theta^c) = \sum_{m=1}^{k_c} \alpha_m^c \; p(\mathbf{y}|c,\theta_m^c) \;\; , \tag{1.7}$$

where k_c is the number of Gaussians in the mixture and $\theta_m^c = (\alpha_m^c, \boldsymbol{\mu}_m^c, \boldsymbol{R}_m^c)$ denotes the set of parameters (weight, mean, covariance matrix) of the $m-th$ Gaussian mode.

The number of Gaussians, k_c, and their parameters, θ_m, should be estimated from a set of dermoscopy images anotated by an expert. This problem can be solved since the PH^2 database contains 29 images with color segmentations performed by an experienced dermatologist. The estimation of the mixture order and parameters was carried out using the algorithm proposed in [12]. This procedure is repeated for each of the five colors considered in this chapter (blue-gray, black, white, dark and light brown). The red color was excluded due to the lack of examples associated to this color in the anotated images.

After estimating the five mixtures, each pixel is then classified into the most probable color. This is achieved by the Bayes law

$$p(c|\mathbf{y}) = \frac{p(\mathbf{y}|c,\widehat{\theta}^c)p(c)}{p(\mathbf{y}|\widehat{\theta})} \qquad c = 1,\ldots,5, \tag{1.8}$$

where $\widehat{\theta}^c = (\widehat{\theta}_1^c,\ldots,\widehat{\theta}_{k_c}^c)$, $\widehat{\theta} = (\widehat{\theta}^1,...,\widehat{\theta}^5)$ are the estimates of the mixture parameters. A label c is assigned to the pixel **y** using the following decision rule. First, the degrees of membership to each of the five colors are sorted. The highest and second highest values are denoted as $c_1$and c_2, respectively. Then, this information is used to either label the patch or reject it, as follows

1. If $p(c_1|\mathbf{y}) \geq \delta$ and $p(c_1|\mathbf{y}) - p(c_2|\mathbf{y}) > \epsilon$, where δ and ϵ are empirically determined thresholds, the patch is labeled according to color c_1.

2. If $p(c_1|\mathbf{y}) \geq \delta$ and $p(c_1|\mathbf{y}) - p(c_2|\mathbf{y}) \leq \epsilon$, the patch receives a label which expresses doubt between c_1 and c_2.

3. If $p(c_1|\mathbf{y}) < \delta$, the patch is rejected.

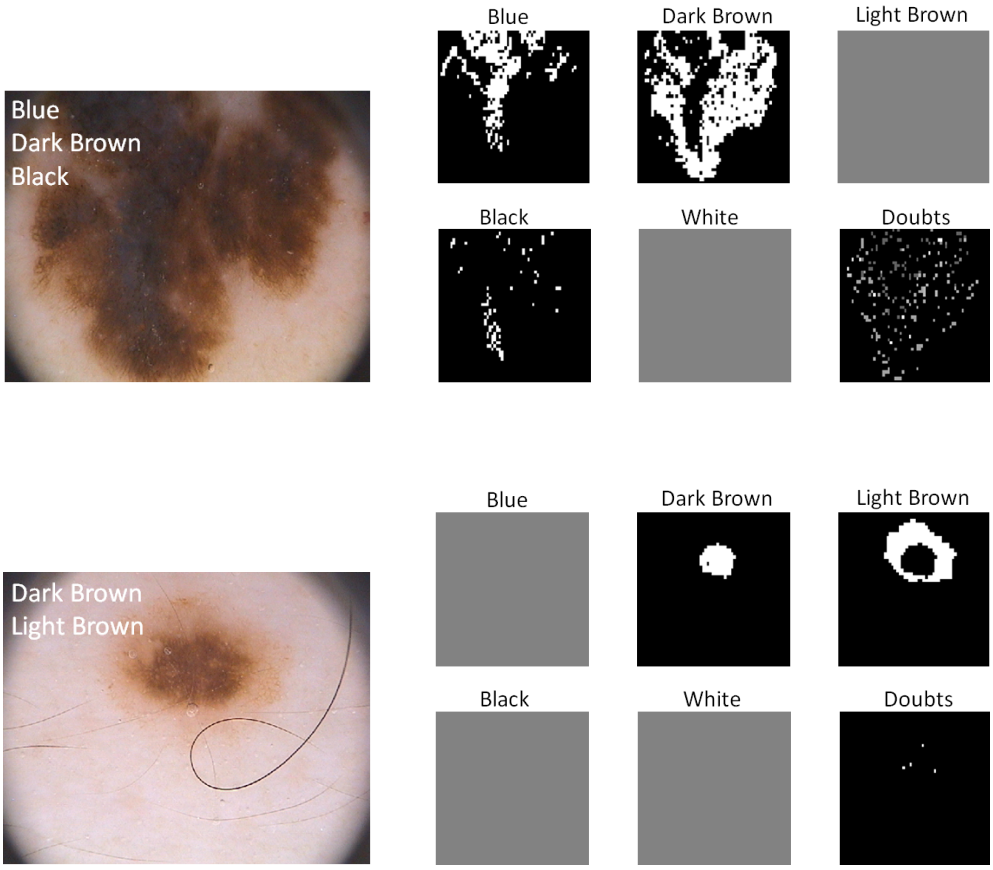


Figure 1.8: Examples of color detection in melanoma (top) and benign lesions (bot).

Finally, we assign color labels to the test images. This task is performed using the patches previously labeled with one of the five colors. We do not consider the patches that have doubt labels in this process. For each color, we compute the area of the patches with its label (i.e the number of pixels) and compare it with an empirically determined area ratio threshold. Each color is validated only if its area ratio is above a specific threshold, where the area ratio (λ_c) for color c is defined as

$$\lambda_c = \frac{A_{patches^c}}{A_{lesion}} \quad . \tag{1.9}$$

For each color we have a different area ratio threshold, which were empirically determined.

We will now present experimental results obtained with the PH^2 database. The five mixture models were estimated using twenty-nine anotated images from the PH^2 database (see section 1.2). Each of these images is associated to five binary masks defining examples of each color. Then, we selected another 123 images (test set) to evaluate the performance of the algorithm. The test images were classified according to the learned models and a set of labels was automatically computed for each image and compared with the colors annotated by a medical specialist. Table 1.2 compares the performance of the proposed system with the labeling performed by a specialist. We used the following metrics to evaluate the algorithm. Each

Table 1.2: Evaluation of color labeling algorithm in the PH^2 database

	CD	DF	CND	FA	SE	SP	ACC
Blue-Gray	24	2	62	15	92.3%	80.5%	86.4%
Dark-Brown	74	4	16	9	94.9%	64.0%	79.4%
Light-Brown	71	7	19	6	91.0%	76.0%	83.5%
Black	24	4	54	21	85.7%	72.0%	78.9%
White	6	1	71	25	85.7%	74.0%	79.8%
Total	199	18	222	76	91.7%	74.5%	81.7%

color correctly identified in an image is considered a *correct detection* (CD), a missed color is a *detection failure* (DF), a detected color that has no corresponding label is a *false alarm* (FA), and each color that is correctly non detected is a *correct non detection* (CND). Additionally, we also compute three different statistics: sensitivity (SE), specificity (SP), and accuracy (ACC). It is possible to see that a good match was obtained in this data set. Figure 1.8 shows two examples of color detection.

1.3.3 Melanoma detection

Melanoma is the most aggressive type of skin cancer. The early detection of melanomas is therefore a major goal in dermoscopy analysis. Several methods have been proposed in the literature to perform this task [13, 14, 15, 16, 17]. Most of them fit into a three step structure: *i) lesion segmentation*; *ii) feature extraction* and *iii) feature classification*. First, the lesion boundary is estimated in the input images. Then, a set of features (color, texture, shape, symmetry) is extracted from each lesion. Finally, a classifier (*e.g.,* support vector machine (SVM)) is trained to distinguish benign lesions from malign ones.

We developped computer diagnosis systems for the detection of melanomas, based on two classification strategies: i) global classification strategy using global features extracted from the lesion [17] and ii) local strategy, representing the lesion by a set of local patches , each of them described by local features; the decision is obtained using a bag-of-features (BoF) classifier [17, 18].

Several types of features were considered in the global approach: color features, texture features, shape and symmetry features. The BoF approach also considered color and texture features associated to lesion patches. See the different descriptors in Table 1.3. In each approach, the color features were computed using multiple color spaces (RGB, HSV,CIE La*b* and Opponent).

The global system used a k-nearest neighbor (KNN) classifier with $k \in \{3, 5, .., 25\}$ neighbors. The BoF system used a dictionary of $\{100, 200, 300\}$ words; the decision is made by a kNN classifier, as well, with $k \in \{3, 5, .., 25\}$. Details can be found in [17, 18].

Table 1.3: Considered Features

Features	Global	Local
Color	Color Histograms	Color Histograms
	Mean Color Vectors	Mean Color Vectors
Texture	Gradient Histograms	Gradient Histograms
	GCLM	GLCM
	Laws Masks	Laws Masks
	Gabor Filters	Gabor Filters
Shape	Simple Shape Descriptors	-
	Fourier Descriptors	-
	Wavelet Descriptors	-
	Moment Invariants	-
Symmetry	Shape symmetry	-
	Color Symmetry (histograms and mean color vector)	-

The algorithms were evaluated by 10 fold cross validation. The set of 200 images is split into 10 subsets (folds) of 20 images each; 9 folds are used for training and 1 fold for testing. This procedure is repeated 10 times with different test folds. The results are shown in Table 1.4 for both classification strategies and different types of features. Very good performance is achieved by several configurations. The best type of features are color and color symmetry features. Both strategies (global and local) lead to very good results. We prefer local classifiers since they take into account local properties of the dermoscopy image as it is done in the examination performed by medical experts. The global methods are however much faster during the training phase.

Table 1.4: Performance of melanoma detection algorithms for different types of features and classification strategies

	Global			Local		
Features	Sensitivity	Specificity	Descriptor	Sensitivity	Specificity	Descriptor
Color	90%	89%	HSV histogram	93%	84%	La*b* histogram
Texture	93%	78%	Gradient histogram	88%	76%	Laws Masks
Shape	81%	88%	Wavelets	-	-	-
Symmetry	92%	85%	HSV Mean Color	-	-	-

Figures 1.9 and 1.10 show examples of melanocytic lesions extracted from the PH^2 database together with the decision provided by the two systems described in this section. Color features were used in both cases to characterize the lesion or its patches. We observe that the systems provide correct decisions in most cases and solve some difficult examples in which the classification of the skin lesions is not simple.

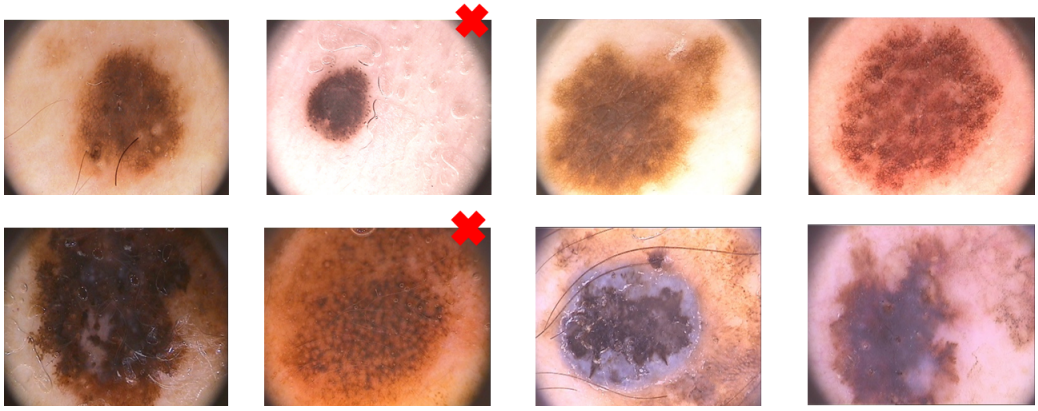


Figure 1.9: Examples of benign (top) and malign (bottom) lesions classified by global system; $\times$ an error

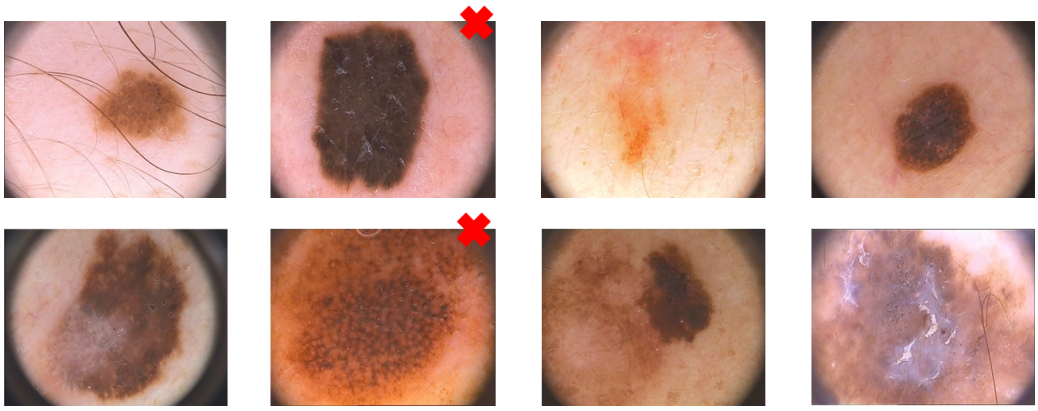


Figure 1.10: Examples of benign (top) and malign (bottom) lesions classified by local system; $\times$ an error

1.4 Conclusions

This chapter presents a dataset of 200 dermoscopic images with medical anotations, publicly available at *www.fc.up.pt/addi*. The images and medical anotations were provided by the Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal, and comprise both benign nevi (160) and malign melanomas (40).

The PH^2 dataset aims to provide a benchmarking tool for the comparison of Computer Aided Diagnosis Systems for the analysis of dermoscopic images. This chapter includes examples of such systems trained

with the PH^2 database and tested using 10 fold cross validation. The results presented in this chapter for illustrative purposes allow a direct comparison with other systems.

The PH^2 database was made available online in September 2013. Since the release date, the PH^2 database counts with 51 downloads so far. The database downloads were performed by several research groups from 21 countries worldwide. These download statistics are a strong indicator of the impact of the PH^2 database in the research community due to the lack of public ground truth databases of dermoscopic images.

Bibliography

[1] G. Argenziano, H. P. Soyer, V. De Giorgi, D. Piccolo, P. Carli, M. Delfino, A. Ferrari, V. Hofmann-Wellenhog, D. Massi, G. Mazzocchetti, M. Scalvenzi, and I. H. Wolf, "Interactive atlas of dermoscopy," 2000.

[2] G. C. do Carmo and M. R. e Silva, "Dermoscopy: basic concepts," *Int J Dermatol.*, vol. 47, pp. 712–719, July 2008.

[3] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The Lancet Oncology*, vol. 3, pp. 159–65, March 2002.

[4] "Addi project: Automatic computer-based diagnosis system for dermoscopy images." http://www.fc.up.pt/addi.

[5] M. E. Celebi, G. Schaefer, H. Iyatomi, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 33, pp. 148–153, March 2009.

[6] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transaction on image processing*, vol. 17, no. 3, 1998.

[7] C. Barata, J. S. Marques, and J. Rozeira, "Detecting the pigment network in dermoscopy images: a directional approach," *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5120–5123, 2011.

[8] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal Computer Vision*, pp. 321–331, 1988.

[9] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–714, 1986.

[10] W. Stolz, A. Riemann, and A. B. Cognetta, "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma," *European Journal of Dermatology*, vol. 4, pp. 521–527, 1994.

[11] C. Barata, M. Figueiredo, M. Celebi, and J. S. Marques, "Color identification in dermoscopy images using gaussian mixture models," in *Accepted for presentation at ICASSP2014 - Bio Imaging and Signal Processing*, 2014.

[12] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.

[13] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 233–239, 2001.

[14] M. E. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker, and R. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, pp. 362–373, 2007.

[15] H. Iyatomi, H. Oka, M. E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, "An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 566–579, 2008.

[16] Q. Abbas, E. C. M, I. F. Garcia, and W. Ahmad, "Melanoma recognition framework based on expert definition of abcd for dermoscopic images," *Skin Research and Technology*, vol. 19, pp. e93–e102, 2013.

[17] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *accepted for publication in IEEE Systems Journal*, 2013.

[18] C. Barata, M. Ruela, T. Mendonça, and J. S. Marques, "A bag-of-features approach for the classification of melanomas in dermoscopy images: The role of color and texture descriptors," in *Computer Vision Techniques for the Diagnosis of Skin Cancer*, pp. 49–69, 2014.