Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features

Catarina Barata, Margarida Ruela, Mariana Francisco, Teresa Mendonça, and Jorge S. Marques

***Abstract*—Melanoma is one of the deadliest forms of cancer; hence, great effort has been put into the development of diagnosis methods for this disease. This paper addresses two different systems for the detection of melanomas in dermoscopy images. The first system uses global methods to classify skin lesions, whereas the second system uses local features and the bag-of-features classifier. This paper aims at determining the best system for skin lesion classification. The other objective is to compare the role of color and texture features in lesion classification and determine which set of features is more discriminative. It is concluded that color features outperform texture features when used alone and that both methods achieve very good results, i.e., Sensitivity $= 96\%$ and Specificity $= 80\%$ for global methods against Sensitivity $= 100\%$ and Specificity $= 75\%$ for local methods. The classification results were obtained on a data set of 176 dermoscopy images from Hospital Pedro Hispano, Matosinhos.**

***Index Terms*—Bag of features (BoF), color, computer-aided diagnosis, dermoscopy, melanoma, texture.**

I. Introduction

THE American Cancer Society estimates that more than 76 000 new cases of melanoma will be diagnosed in 2012. This type of skin cancer is the less common, accounting for less than 5% of all skin cancer cases. However, it is by far the most aggressive since it is more likely to metastasize than other skin tumors. This characteristic makes melanoma the deadliest form of skin cancer (it is estimated that more than 75% of deaths related with skin cancer in 2012 will be from melanoma) [1]. Melanoma incidence rates have been significantly increasing in the last decades, which makes this one of the cancers that has been receiving attention both from the public health field, with medical prevention campaigns, and from the cancer research field.

The ultimate goal for physicians is to diagnose melanoma in its earliest stage, since it is less probable that it has already metastasized, thus greatly increasing the probability of survival. Nowadays, a technique used by dermatologists to diagnose skin lesions and, consequently, to detect melanomas is dermoscopy. This is a noninvasive procedure used for *in vivo* observation of skin lesions. The physician places gel on the skin lesion and inspects it with a magnification instrument (dermatoscope, stereomicroscope, or a digital imaging system), which amplifies the lesion 6–100×, depending on the instrument used [2]. This magnification allows the recognition of several surface and subsurface structures, which are not visible to the naked eye, and that can be used to diagnose a skin lesion using one of the several medical diagnostic algorithms, e.g., the ABCD rule [3], Menzies method [4], and the seven-point checklist [5]. These three algorithms have a common first step, i.e., the identification of the inspected lesion regarding its origin as melanocytic or nonmelanocytic; melanoma is a melanocytic lesion, since it is a malignant proliferation of melanocytes. In their second step, they distinguish between benign melanocytic lesions and melanoma using different approaches. The ABCD rule assigns a score to a lesion. This score is a combination of the scores for four different features, namely, (A)symmetry, (B)order, number of different (C)olors, and atypical (D)ifferential structures (atypical pigment network, dots, streaks, etc.). A score higher than 5.45 diagnoses a lesion as melanoma [3]. Menzies method identifies two types of dermoscopic features, i.e., negative (symmetrical pattern, single color) and positive (blue-white veil, atypical dots and network, etc.). The presence of positive features signs a melanoma [4]. Finally, the seven-point checklist also scores a lesion. However, this algorithm only inspects the lesion for the presence of atypical differential structures. The scores for different dermoscopic structures can be 2 or 1 and are added up to give the final score. A score higher than 3 indicates melanoma [5]. Fig. 1 illustrates a melanoma and some of the common dermoscopic features considered by dermatologists.

It has been reported that dermoscopy can actually increase the sensitivity of melanoma detection by 10%–27% [6]. However, it has been demonstrated that dermoscopy only increases the diagnostic performance if the dermatologists have received formal training [7]. Moreover, even with the use of the described diagnostic algorithms, which aim to make the diagnosis more reliable and reproducible, the diagnosis of a skin lesion by a dermatologist is still subjective since it depends on human vision and on clinical experience. Computerized dermoscopy image analysis systems can be used to tackle this problem. In these systems, a computer is used as a diagnostic tool and to follow up on suspicious skin lesions.

Manuscript received July 30, 2012; accepted May 10, 2013. This work was supported by FCT under Grant SFRH/BD/84658/2012 and projects PTDC/SAUBEB/103471/2008 and PEst-OE/EEI/LA0009/2011.

C. Barata, M. Ruela, M. Francisco, and J. S. Marques are with the Institute for Systems and Robotics, Instituto Superior Técnico, 1049-001 Lisbon, Portugal (e-mail: ana.c.fidalgo.barata@ist.utl.pt).

T. Mendonça is with the Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSYST.2013.2271540

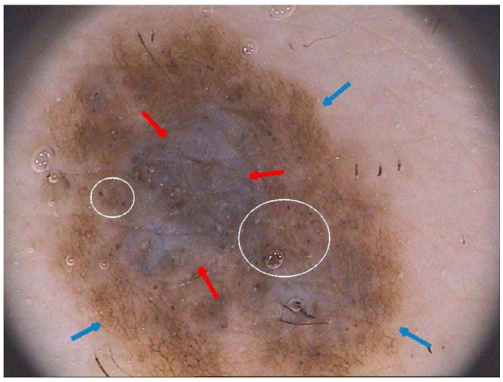

Fig. 1. Melanoma: (blue arrows) atypical pigment network; (red arrows) blue-white veil; (white circles) atypical dots; more than one color; and asymmetric.

This paper proposes two systems for the automatic classification of melanocytic skin lesions. The first system uses global methods to classify skin lesions. This approach evolves in three sequential steps. First, the lesion is segmented using an automatic segmentation method. Then, a set of features from the ABCD rule (color and texture features) is extracted and used to train a classifier to perform binary classification as *melanoma* or *benign*. The second system uses local features, following a recent trend in image analysis and recognition. This paper is organized as follows: Section II gives an overview of several systems proposed in the literature. Section III provides an overall description of global and local strategies. Sections IV and V describe the proposed systems and the type of features extracted. The results obtained and their discussion are performed in Section VI. Section VII concludes this paper with some remarks on future work.

II. Related Works

In the last decade, several systems for melanoma detection have been proposed. Some systems try to mimic the performance of dermatologists by detecting and extracting several dermoscopic structures, such as pigment network [8]–[10], irregular streaks [11], granularities [12], regression structures [11], blue-white veil [13], and blotches [14]. These structures can then be used to score a lesion in a similar way to the one adopted by dermatologists. A CAD system that mimics the performance of the seven-point checklist method can be found in the literature [15].

However, the vast majority of melanoma detection systems found in the literature follows a pattern recognition approach [16]–[25]. Most of these works use global methods to classify skin lesions [16]–[20], [22], [24]. The described systems usually consist of three/four consecutive steps, i.e., lesion segmentation, feature extraction, feature selection, and lesion classification using a trained classifier. These systems are usually inspired in the ABCD rule [3], and the extracted features aim to reproduce each one of the accounted scores. The most common features used in these studies include shape features (e.g., compactness, aspect ratio, and maximum diameter), which represent both asymmetry and border; color features in several color spaces (e.g., mean and standard deviation); and texture features (e.g., gray-level cooccurrence matrix) [22]. Interesting results have been obtained by different authors with different sets of features and classifiers. Ganster *et al.* [16] extracted shape and color features and used a k-Nearest Neighbor (kNN) classifier to distinguish between melanoma and benign nevi. They used a large data set with more than 5300 dermoscopy images and achieved a Sensitivity (SE) of 87% and a Specificity (SP) of 92%. Rubegni *et al.* [18] also used shape and color features and added texture features. They achieved a $SE = 96\%$ and a $SP = 93\%$ in a data set with 217 melanomas and 588 images, using an artificial neural network classifier. Celebi *et al.* [22] proposed a system based on a support vector machine (SVM) classifier to identify melanomas. They used a large feature vector, which contained shape, color, and texture features, and achieved $SE = 93\%$ and $SP = 92\%$. The system proposed by Iyatomi *et al.* [24] is an Internet-based melanoma screening system that uses a feature vector of over 400 features, divided in shape, texture, and color features. Their data set contained more than 1200 dermoscopy images, and they achieved SE and SP values both equal to 86%. Local methods have been also recently proposed to classify skin lesions. Situ *et al.* [25] described an algorithm for lesion classification that uses the bag-of-features (BoF) approach. They represent each image as a set of several patches sampled from a 16×16 regular grid placed on the lesion. To describe each patch, they use wavelets and Gabor-like filters, leading to a total of 23 features. Two different classifiers were compared, i.e., naive Bayes and SVM classifiers; and their best performance is 82% on a data set of 100 dermoscopy images, 30 of which were melanomas.

The aforementioned works consider that the three commonly used types of features (shape, color, and texture) are equally relevant for the automatic detection of melanomas. Other works use a single class of features as a descriptor for melanoma detection [21], [23], [26]–[28]. Color distribution in the RGB color space (mean RGB distance, variance, and maximum distance) was used to distinguish between melanomas, atypical nevi, and benign nevi by Seidenari *et al.* [21]. They achieved an accuracy of 86% and observed significant difference in pigment distribution between the three studied populations. Relative color histograms have been also used to distinguish between melanomas and benign or atypical nevi [23], [26], both with SE greater than 80%. The gray-level cooccurrence matrix and other texture descriptors, such as Laws energy masks and Gabor-like filters, were applied to classify skin lesions using texture information only [27], [28]. The most promising results were achieved by Sheha *et al.* [28], who obtained an accuracy of 92% on a data set of 102 images (51 melanomas). Despite being very promising and demonstrative that both types of features are good discriminators, these cannot be compared since different data sets and experimental conditions are employed. Therefore, it is not possible to determine the relative importance of each feature for skin lesion classification. To the best of our knowledge, no work can be found in the literature that compares the relevance of each type of features. Furthermore, most works describe the lesion using global features, although it has been shown that local features perform better in several image analysis problems (e.g., object recognition and image retrieval) [29].

This paper addresses two distinct objectives. First, it aims at determining which is the best strategy for skin lesion

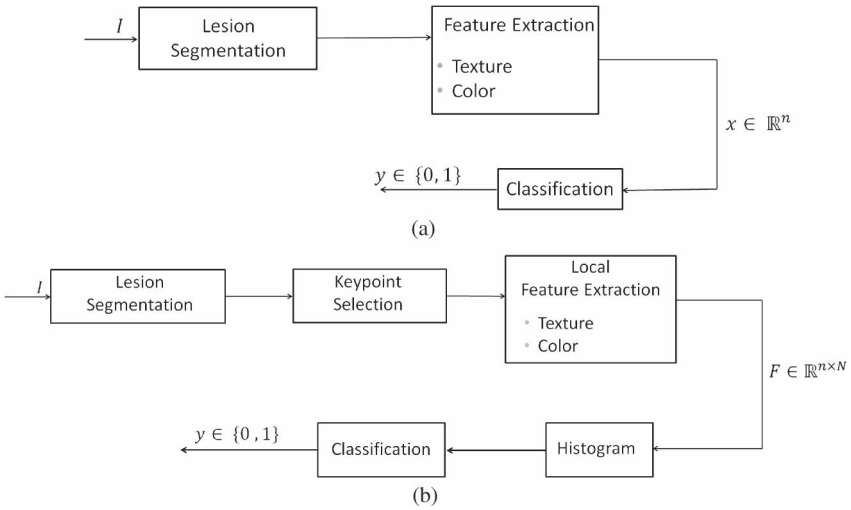

Fig. 2. Block diagram of two melanoma detection systems, using (a) global and (b) local features.

classification: global or local? Second, it aims at comparing the performance of two classes of features (color and texture) in melanoma identification and determining which feature is the best discriminator in this specific problem.

III. System Overview

The analysis of dermoscopy images is an object recognition problem under very specific conditions. Several approaches have been proposed to deal with this kind of problems. Some of them characterize the object by a set of global features (e.g., color moments and gradient histograms) and use these features to discriminate the object from the background. However, these techniques fail to recognize objects in more complex settings, e.g., when the object boundary is unknown and cannot be easily found in the image or when the object shape, color, and texture exhibit severe changes and cannot be easily described by global measurements extracted from the image. A recent trend consists in selecting local patches (small regions) in the image and describing each patch by a set of local features. This approach does not require a global model for the object and can easily cope with changes in shape, color, and texture, achieving good results in very difficult problems [30].

In this paper, we will compare two different strategies for melanoma detection and build a recognition system based on each of them. The first system describes the dermoscopy image by a set of global features and uses a classifier to discriminate melanomas from nonmelanoma lesions [see the block diagram in Fig. 2(a)]. This is a supervised system since the classifier learns to detect the melanoma lesions using a training set of images, which is labeled by an expert. Each training image, i.e., $I^{(k)}$, $k \in 1, \ldots, L$, is characterized by feature vector $\mathbf{x}_k \in \mathbb{R}^n$ and by binary label $y_k \in \{0, 1\}$. The classifier is trained to discriminate both types of images using the training data $(\mathbf{x}_k, y_k)$, $k = 1, \ldots, L$.

The second system characterizes the dermoscopy image by using a BoF approach [see Fig. 2(b)] [29], [31]–[34]. First, a set of keypoints is selected inside the lesion region. Then, each keypoint is characterized by a vector of local features. This feature vector represents color and texture properties in a local patch centered at the keypoint. Since the number of keypoints and local features varies from image to image, we cannot directly feed a classifier with these data. Instead, all local features associated with all the training images are gathered and used to compute a smaller set of prototypes (centroids) denoted as *visual words*. Then, the local features of each dermoscopy image are assigned to the nearest visual word, and a histogram is computed. The histogram counts the number of times each visual word was selected. A statistical classifier is then trained to discriminate melanoma lesions from nonmelanoma ones, using the histogram of visual words as input.

This paper tries to determine which of these strategies performs best in melanoma detection. The answer is not simple. If a skin lesion is homogeneous, the first approach will probably be the best since it is able to describe the lesion by a set of global features that use all the available information. However, since skin lesions often have differential structures (pigment network, vascular network, dots, streaks, etc.) that are localized and appear in specific regions, these structures may not be well described by global features, and local methods may represent them better.

Both systems assume that the dermoscopy image is segmented and that the lesion region is separated from healthy skin. Several segmentation methods were proposed for this problem [35]–[39]. A comparison among several methods can be found in [40]. However, there is no consensus about which one performs best. We adopted a simple adaptive thresholding technique (see the description of the method in the Appendix), and the output of the segmentation step was verified by an expert and edited when necessary. This procedure was adopted in order to make the system performance independent of the specific algorithm used for segmentation. Otherwise, it would be difficult to interpret the system decisions since errors might be caused by insufficient information provided by the features or by segmentation errors.

IV. Global Lesion Classification System

The global classification system consists of three processing steps [see Fig. 2(a)]. First, the skin lesion is segmented. Then, a

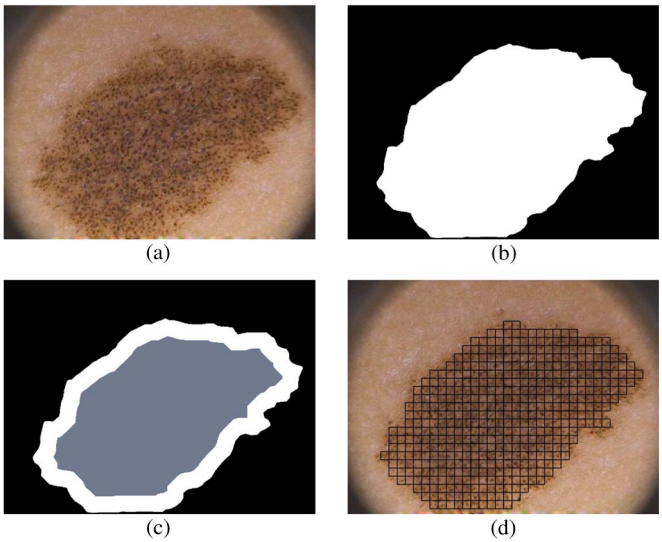

Fig. 3. (a) Original image. Computation of global features for (b) R and (c) $R_1 + R_2$. (d) Computation of local features.

vector of global color and texture features is extracted from the skin region. Finally, a statistical classifier is trained using a set of images, which is labeled by an expert (training set). Three classifiers are considered in this work, i.e., AdaBoost classifier [41], SVM classifier [42], and kNN classifier.

Concerning the image features, medical doctors use several visual cues, such as differential structures, number of colors, symmetry, and border transitions. Two strategies will be considered in this paper. In the first case, image features are computed inside the skin lesion, i.e., R [see Fig. 3(b)], whereas in the second case, we split R into an inner part, i.e., R_1, and the border, i.e., R_2 [see Fig. 3(c)]. R_1 is obtained by eroding R with a disk of radius r, which was empirically chosen to be 1/10 of the lesion smallest axis after comparing several values. This strategy tries to capitalize the available knowledge about medical analysis of dermoscopy images, i.e., special attention is assigned to the border analysis.

In this paper, we will focus on two types of image features (color and texture), and we will try to assess the role played by each of them in the final decision: what performance is achieved by color features only, by texture features only, and by combining both of them. The features used in this work are presented in the next sections.

A. Texture Features

Image texture represents the spatial organization of intensity and color in an image, and it can be characterized in many different ways. Some methods use pixel statistics. A classic approach consists in computing the statistics of pairs of neighboring pixels, using the cooccurrence matrix [43]. This idea has been modified and improved in many different ways. For example, local binary patterns perform a binary classification of the pixels in the vicinity of each pixel and compute the statistics of the neighboring pixel configurations [44]. Texture has been also described by applying a Fourier transform to the image and by characterizing the spectral energy in different frequency bands [45]. Other image transforms have been also used, e.g., wavelets [46], Laplacian pyramids [47], or linear filters (e.g., Law texture features) [48].

Many works use the gradient histograms to characterize texture, and these features have achieved excellent results in several challenging problems. We will use two gradient histograms in this work. First, we convert the RGB image into a gray-level image by selecting the color channel with the highest entropy [40]. In order to compute the image gradient, we filter the gray-level image using a Gaussian filter with $\sigma = 2$ and then compute the gradient vector at each point $g(x) = [g_1(x)\ g_2(x)]^T$ using Sobel masks. The gradient magnitude and orientation are then computed as

$$\|g(x)\| = \sqrt{g_1^2(x) + g_2^2(x)}, \quad \phi(x) = \tan^{-1}\left(\frac{g_2(x)}{g_1(x)}\right). \quad (1)$$

The histograms of the gradient amplitude and orientation are defined as follows:

$$h_a(i) = \frac{1}{N}\sum_{x\in R} b_i\left(\|g(x)\|\right), \quad i = 1, \ldots, M_a \quad (2)$$

$$h_\phi(i) = \frac{1}{N}\sum_{x\in R} \tilde{b}_i\left(\phi(x)\right), \quad i = 1, \ldots, M_\phi \quad (3)$$

where R denotes the set of pixels that were classified as lesion, $N = \#R$ is the number of lesion pixels, and M_a and M_ϕ are the number of bins used in the amplitude and orientation histograms; $b_i(.)$ and $\tilde{b}_i(.)$ are the characteristic functions of the ith histogram bin, defined by

$$b_i(a) = \begin{cases} 1, & \text{if } a \text{ belongs to the } i\text{th amplitude bin} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$\tilde{b}_i(\phi) = \begin{cases} 1, & \text{if } \phi \text{ belongs to the } i\text{th orientation bin} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

When lesion region R is split into inner region R_1 and border R_2, we compute separate histograms for each of these regions using the same expressions and replacing R by R_i and N by N_i.

Fig. 4 shows the histograms of amplitude and orientation for a set of dermoscopy melanoma images (left) and nonmelanoma images (right). There is a significant overlap among the features of both classes, in all the histogram bins. This means that the binary decision is not simple.

B. Color Features

Color information is used by medical doctors in the classification of skin lesions (see Section I). Therefore, color features have been explicitly used in most CAD systems. The most popular features used in dermoscopy analysis are color statistics, such as the mean color and color variance [16], [24]. Most works compute these statistics for the RGB color components. However, other color representations have also been adopted [22].

The RGB color space represents a target color as a mixture of three primary colors, i.e., red (R), green (G) and blue (B).

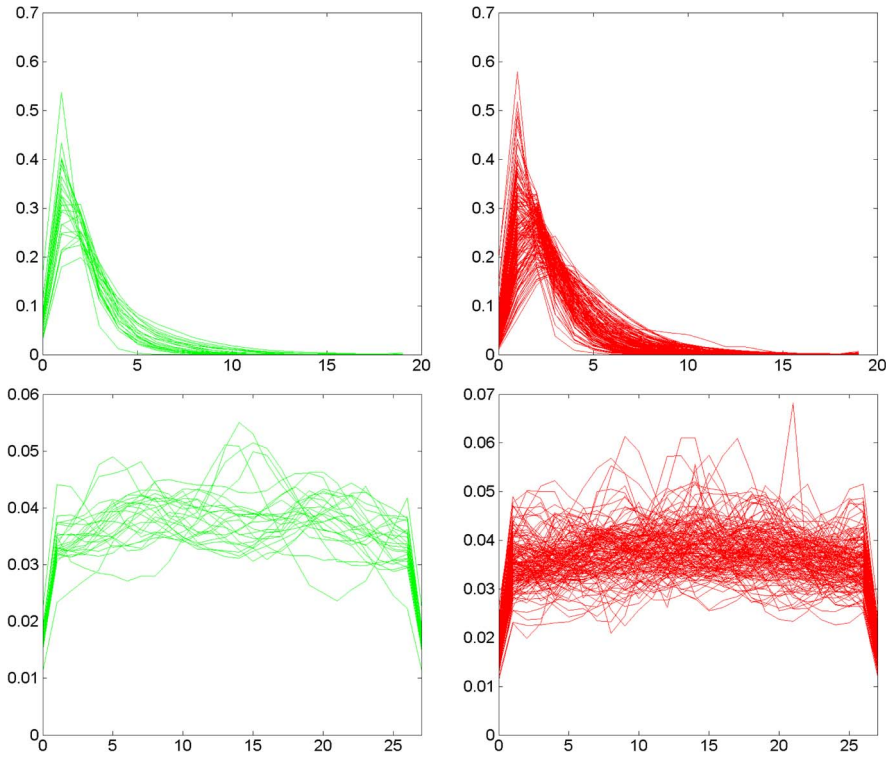

Fig. 4. Histograms of (top) amplitude and (bottom) orientation for (left) melanoma and (right) nonmelanoma lesions.

The mixture coefficients are denoted as color components. However, this color space has a number of drawbacks: it is not perceptually uniform, it depends on the acquisition setup, and it exhibits a high correlation among the three color channels [49]. To overcome these difficulties, other color representations have been proposed, e.g., biologically inspired color spaces such as the opponent color space (Opp) [50], [51], color spaces related to the human description of color such as hue saturation and brightness (HSV and HSI) [49], [52], or perceptually uniform color spaces such as the CIE La*b* and L*uv [49]. Although there is a 1-1 map between all pairs of color spaces, the choice of the color space modifies the recognition results since their components present different statistics (e.g., some are more correlated than others) and different histograms.

In this paper, we have used all six color spaces mentioned above. We characterize the color distribution in lesion region R (or R_1, R_2) using a set of three color histograms, each of them with M_c bins. The color histogram associated to the color channel, i.e., $I_c(x)$, $c \in \{1, 2, 3\}$, is given by

$$h_c(i) = \frac{1}{N} \sum_{x \in R} b_c\left(I_c(x)\right) \quad i = 1, \ldots, M_c \tag{6}$$

where $N = \#R$ denotes the number of pixels in region R, and i is the histogram bin. The bins are defined by splitting the color component range (which depends on the color space) into M_c subintervals with the same length.

Fig. 5 shows the concatenation of the histograms h_c for four different color spaces and splitting region R into R_1 and R_2. The differences between melanoma and nonmelanoma features are more noticeable in these cases.

V. Detection With Local Features

BoF methods have recently been used in complex image analysis problems, e.g., image retrieval and object-class recognition in large databases [53], [54]. Since complex objects cannot be represented by a global model, BoF methods represent them by a collection of local models, e.g., image patches. In a first stage, a set of keypoints, i.e., $p_i \in \mathbb{R}^2$, is defined in the image. In the case of dermoscopy images, we assume that keypoints are nodes of a regular grid of size $\Delta \times \Delta$ defined in the image domain. Second, each keypoint is described by a feature vector $x_i \in \mathbb{R}^n$, which represents local information conveyed in the image patch of size $\Delta \times \Delta$ centered at the ith keypoint [see Fig. 3(d)]. The size of the patches is equal to the size of the grid to avoid overlapping. Patches for which the area is more than 50% outside the lesion are discarded. Given an input image I, we will denote the family of local features associated to I by

$$F = \{x_1, \ldots, x_N\}, \quad x_i \in \mathbb{R}^n \tag{7}$$

where N is the number of keypoints inside the lesion (see Fig. 3, bottom right).

A. Training

The training phase aims to learn the parameters of a BoF classifier using a set of labeled dermoscopy images. Let $\mathcal{T} = \{I^{(1)}, \ldots, I^{(L)}\}$ be a set of L images used to train the classifier and let

$$\mathcal{F} = \bigcup_{k=1}^{L} F^{(k)} \tag{8}$$

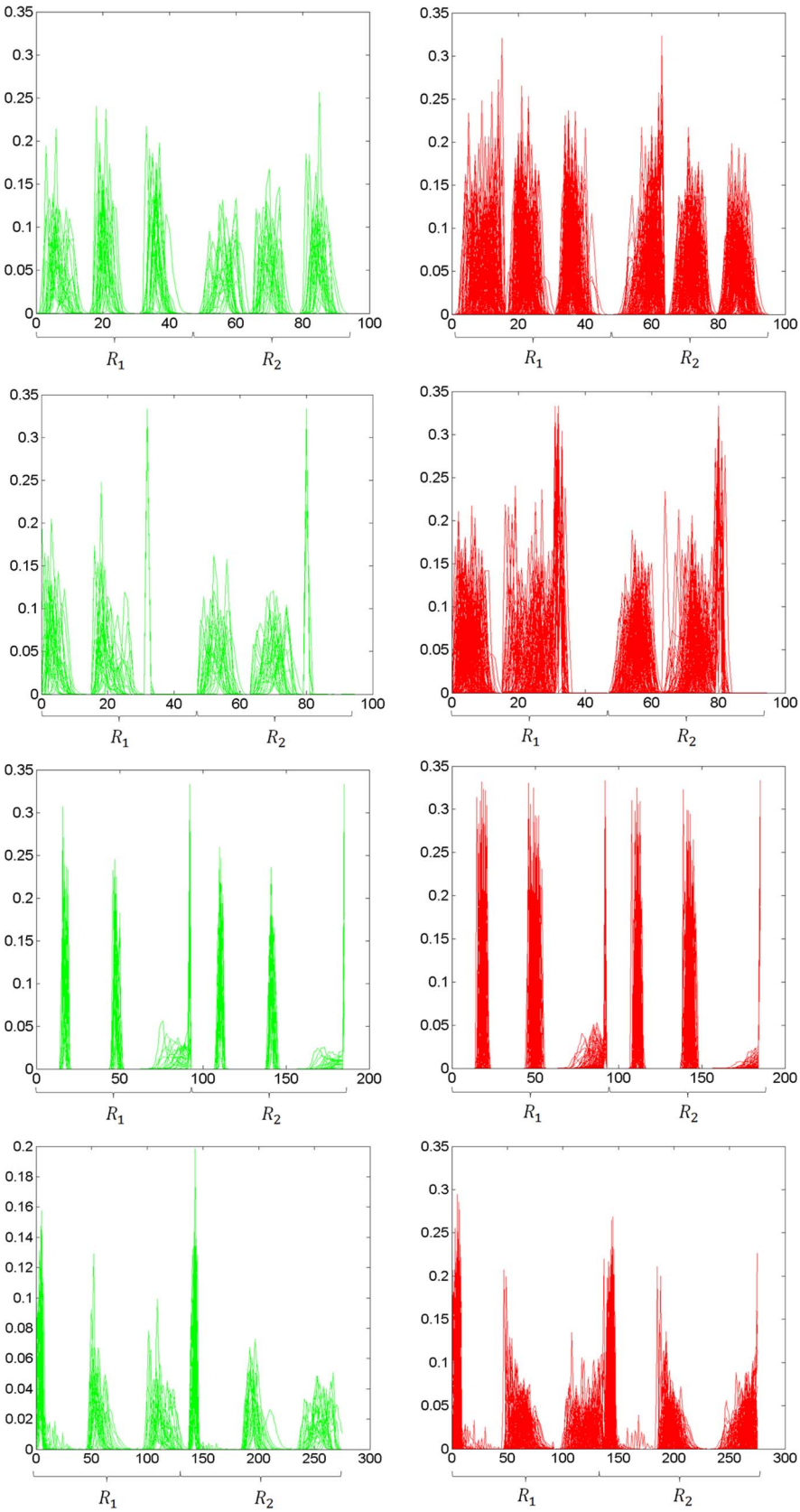

Fig. 5. Concatenated histograms of (first row) RGB, (second row) L*uv, (third row) Opp, and (fourth row) HSV for (left) melanoma and (right) nonmelanoma lesions, computed in R_1 and R_2.

be the set of all local features detected in the training set. First, we will approximate the feature vectors in $\mathcal{F}$ by a set of prototypes (centroids) $c_1, \ldots, c_K$. This is done using a clustering algorithm (e.g., k-means algorithm). The prototypes are known as *visual words* to emphasize the similarity with the bag-of-words method used in text analysis. After obtaining a

set of prototypes $c_1, \ldots, c_k$, all feature vectors in the training set are classified in the nearest prototype and a label

$$l_i^{(k)} = \arg\min_j \left\| x_i^{(k)} - c_j \right\| \tag{9}$$

is assigned to each feature vector $x_i^{(k)}$.

Then, a histogram that counts the occurrence of each of the prototypes is computed for each of the training images I^k. Therefore

$$h^{(k)}(l) = \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \delta\left(l_i^{(k)} - l\right) \tag{10}$$

where $\delta(.)$ denotes the Kronecker delta ($\delta(x) = 1$, if $x = 0$; $\delta(x) = 0$, otherwise).

Each training image $I^{(k)}$ is characterized by a histogram of visual words $h^{(k)}$ with K bins. The histogram is, therefore, a feature vector that characterizes the image. Since the images in the training set have been classified by an expert, we can train a classifier to predict the labels (melanoma versus nonmelanoma) given the histogram of visual words.

B. Test

In the test phase, we wish to classify a new image using the previously learned model. Given an input image I, we compute the keypoints inside the lesion and the corresponding local features. Then, we classify the local features using the dictionary of visual words and build the histogram of visual words. Finally, the histogram is classified using the classifier learned during the training phase.

C. Local Features

Each image patch is characterized by color and texture features similar to the ones described in Section IV, i.e., we use gradient histograms and color histograms to characterize the image in each patch.

VI. Experimental Results and Discussion

The systems described in the previous sections were evaluated using a data set of 176 dermoscopy images (25 melanomas and 151 nevi). Each image was classified by an experienced dermatologist. These images were taken from the database of Hospital Pedro Hispano, Matosinhos, and were obtained by dermatologists during clinical exams using a dermatoscope with a magnification of 20×. They are stored in *BMP* and *JPEG* formats, and their average resolution is 573 × 765. The performance of each system was evaluated for each type of feature: T1) gradient magnitude and T2) gradient orientation; C1) RGB, C2) HSV, C3) La*b*, C4) HSI, C5) L*uv, and C6) Opp. We have also performed tests using all the color features (C), both texture and color (T + C) and the best texture (T_B) and color (C_B) feature. Since there is always a tradeoff between sensibility (SE) and specificity (SP), we evaluate each system configuration using a cost function defined by

$$\mathcal{C} = \frac{c_{10}(1 - SE) + c_{01}(1 - SP)}{c_{10} + c_{01}} \tag{11}$$

TABLE I
Best Experimental Results Using Global Features and a kNN Classifier

	R			$R_1 + R_2$		
	SE	*SP*	$\mathscr{C}$	*SE*	*SP*	$\mathscr{C}$
T_1	88%	52%	0.263	88%	62%	0.226
T_2	84%	44%	0.319	76%	48%	0.353
C_1	88%	81%	0.149	84%	82%	0.168
C_2	92%	75%	0.146	100%	66%	0.135
C_3	88%	77%	0.165	92%	76%	0.143
C_4	92%	72%	0.162	96%	69%	0.149
C_5	96%	70%	0.146	**96%**	**79%**	**0.109**
C_6	**100%**	**71%**	**0.117**	**100%**	**72%**	**0.111**
C	92%	72%	0.159	96%	63%	0.172
$T + C$	96%	65%	0.164	96%	63%	0.172
$T_1 + C_5$	96%	76%	0.119	88%	79%	0.157
$T_1 + C_6$	96%	75%	0.125	96%	74%	0.130

where c_{10} is the cost of an incorrectly classified melanoma [false negative (FN)], and c_{01} is the cost of an incorrectly classified nonmelanoma [false positive (FP)]. Since we consider that an incorrect classification of a melanoma is the gravest error, we assume that $c_{10} = 1.5c_{01}$ and $c_{01} = 1$ in this paper.

In the following subsections, we will address the specific procedures used to evaluate each one of the systems as well as the results obtained.

A. Assessment of the Global Lesion Classification System

The data set used has a limited number of images. Therefore, we have tested the global system using a leave-one-out approach. This approach requires that for each set of features, 176 classifiers are trained, each one using all the images except one, which is used for testing. The final performance measures are computed by averaging the individual measures obtained for each classifier. Class unbalance is another issue that must be addressed. To tackle this problem, we have repeated the melanoma features belonging to each training set until we obtained the same number of examples for both classes. To prevent having equal examples on the training set, we have added Gaussian noise ($w \sim N(0, \sigma_n)$), with $\sigma = 0.0001$, to each repeated feature vector.

The performance of the classifier for each feature was evaluated using different combinations of parameters, both in the feature extraction process, by varying the number of bins of the histograms ($n_{bins} \in \{15, 16, \ldots, 50\}$), and in the classification process, by performing tests with different classification algorithms (kNN, SVM, and AdaBoost). For each classifier, we varied specific parameters, namely, the number of neighbors (k) and distance ($dist$) used by the kNN algorithm to compare feature vectors, the type of kernel used by the SVM, and the number of combined weak classifiers (W) by the AdaBoost. We have considered all the combinations of these parameters and all the possible features and tested more than 30 000 different configurations. For each configuration, we trained 176 classifiers using the leave-one-out procedure.

The best results obtained using the kNN classifier, selected using cost criterion (11), are summarized in Table I. We tested all the n_{bins} for different combinations of $k \in \{5, 7, \ldots, 50\}$. For each k used, we computed the results for three different distances, i.e., Euclidean, Kolmogorov, and Kullback–Leibler.

TABLE II
BEST PARAMETERS FOR THE kNN CLASSIFIER

	R			R_1+R_2		
	n_{bins}	*dist*	k	n_{bins}	*dist*	k
T_1	19	*Kullback – Leibler*	35	17	*Euclidean*	7
T_2	37	*Kolmogorov*	9	29	*Kullback – Leibler*	21
C_1	29	*Kullback – Leibler*	25	15	*Kolmogorov*	11
C_2	24	*Euclidean*	5	23	*Euclidean*	15
C_3	19	*Kolmogorov*	17	20	*Kullback – Leibler*	31
C_4	34	*Euclidean*	7	21	*Kolmogorov*	35
C_5	16	*Kullback – Leibler*	23	15	*Kullback – Leibler*	19
C_6	20	*Kullback – Leibler*	11	20	*Euclidean*	33
C	25	*Euclidean*	7	25	*Euclidean*	29
$T+C$	23	*Euclidean*	33	25	*Euclidean*	31
T_1+C_5	50	*Kolmogorov*	15	28	*Euclidean*	7
T_1+C_6	33	*Kullback – Leibler*	15	34	*Kolmogorov*	5

TABLE III
BEST EXPERIMENTAL RESULTS USING GLOBAL FEATURES AND AN RBF-SVM CLASSIFIER

	R			R_1+R_2		
	SE	*SP*	$\mathscr{C}$	*SE*	*SP*	$\mathscr{C}$
T_1	64%	68%	0.346	56%	73%	0.373
T_2	84%	28%	0.382	100%	10%	0.360
C_1	88%	60%	0.231	84%	68%	0.223
C_2	88%	60%	0.234	88%	60%	0.234
C_3	84%	74%	0.199	76%	80%	0.223
C_4	88%	60%	0.231	88%	58%	0.242
C_5	100%	40%	0.241	100%	40%	0.241
C_6	**84%**	**78%**	**0.183**	**84%**	**79%**	**0.181**
C	88%	74%	0.178	84%	74%	0.202
$T+C$	84%	76%	0.191	72%	79%	0.250
T_1+C_6	**88%**	**77%**	**0.165**	**92%**	**72%**	**0.149**

The parameters that led to the results in Table I can be seen in Table II.

Table I shows that very good results can be achieved using the kNN classifier and that each color feature outperforms the texture features, which is an interesting result. Another interesting result is that in most of the cases, the performance of the algorithm tends to improve when lesion region R is split into two disjoint regions (R_1 and R_2), which suggests that the periphery and the inner part of a lesion contain complementary information that allows better discrimination between melanomas and nonmelanomas. This has been also observed in [22]. It is not clear which of the three distances used is best for this specific problem, since all of them are selected and considered to be the best for some of the features. We tried to improve the results by combining the two types of features. However, the system performance decreases when compared with the system performance for the best single color features, i.e., C_5 ($SE = 96\%,\ SP = 79\%$) and C_6 ($SE = 100\%$ and $SP = 72\%$). This change is explained by the fact that we are increasing the number of features by more than ten times. The number of training patterns is, therefore, insufficient to obtain good generalization.

To overcome these difficulties, we have used two additional classifiers that exhibit better generalization performance when the number of features is increased. First, we have tested SVM using two different kernels, i.e., linear and a Gaussian radial basis function (RBF). RBF-SVM greatly outperformed the linear one; therefore, we will only show results for the second (see Table III).

The SVM classifier does not perform as well as the kNN one. Nonetheless, the obtained results are still interesting and provide some information. As before, the color features outperform the texture ones, which reinforce the idea that this kind of features is more discriminative. In this case, it is not possible to determine if the use of a single region (R) is better than two (R_1, R_2) since the results are not as clear as before. Despite the low performance of the texture features, we are able to improve the performance of the algorithm by combining T_1 and C_6, i.e., the best texture and color features. This feature concatenation leads to the best classification results with SVM: $SE = 92\%$ and $SP = 72\%$.

TABLE IV
BEST EXPERIMENTAL RESULTS USING GLOBAL FEATURES AND AN ADABOOST CLASSIFIER

	R			R_1+R_2		
	SE	*SP*	$\mathscr{C}$	*SE*	*SP*	$\mathscr{C}$
T_1	**96%**	**59%**	**0.188**	84%	67%	0.228
T_2	80%	30%	0.350	76%	55%	0.324
C_1	84%	84%	0.160	76%	85%	0.202
C_2	72%	90%	0.208	**96%**	**80%**	**0.103**
C_3	88%	79%	0.157	96%	78%	0.111
C_4	84%	74%	0.202	**96%**	**80%**	**0.103**
C_5	80%	85%	0.178	88%	64%	0.215
C_6	**96%**	**77%**	**0.117**	**92%**	**85%**	**0.106**
C	96%	77%	0.117	92%	80%	0.127
$T+C$	96%	77%	0.117	92%	79%	0.130
T_1+C_2	72%	90%	0.208	96%	80%	0.103
T_1+C_4	84%	78%	0.183	96%	80%	0.103
T_1+C_6	96%	77%	0.117	92%	85%	0.106

TABLE V
BEST PARAMETERS FOR THE ADABOOST CLASSIFIER

	R		R_1+R_2	
	n_{bins}	W	n_{bins}	W
T_1	44	5	25	5
T_2	44	2	15	5
C_1	20	2	28	2
C_2	35	5	45	2
C_3	33	2	37	2
C_4	46	2	45	2
C_5	48	5	25	2
C_6	24	2	30	2
C	24	2	45	2
$T+C$	24	2	45	2
T_1+C_2	33	5	45	2
T_1+C_4	46	2	45	2
T_1+C_6	24	2	30	2

AdaBoost has the ability to select appropriate features for a given problem and to train a strong classifier from a combination of weak classifiers [55]. Therefore, this kind of classifier was deemed appropriate for our classification problem since it is able to select a set of informative features for this problem, and it is able to discard features that do not improve the classifier performance. AdaBoost was trained using different numbers of weak classifiers $W \in \{2, 100\}$, achieving the best overall results ($SE = 96\%$ and $SP = 80\%$ for C_2/C_4 and $SE = 92\%$ and $SP = 85\%$ for C_6) with combinations of only two to five weak classifiers (see Tables IV and V). Moreover, we also obtained the best global classification results for a single texture feature T_1 ($SE = 96\%, SP = 59\%$). Since each weak classifier uses only one bin to classify the data, the experimental results suggest that only a very small number of features is enough to achieve the best scores. This is a remarkable result, which suggests that some of the features are much more important than others. It also suggests that best generalization is achieved with a small number of features than with a large feature set.

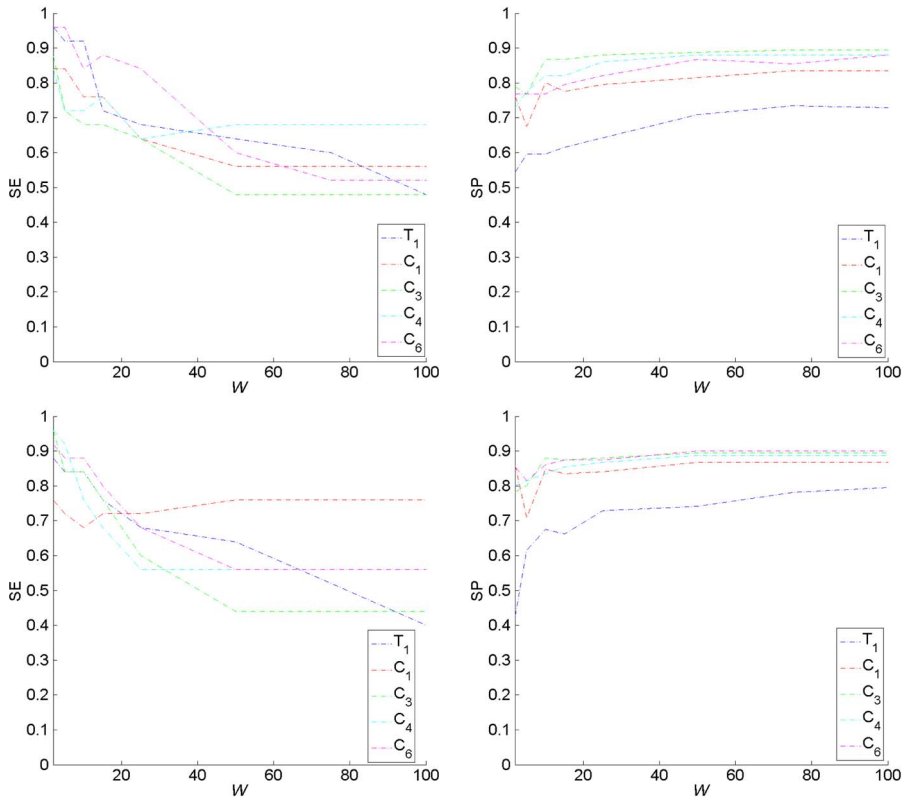

Fig. 6. Variation of SE and SP for (first row) R and (second row) $R_1 + R_2$, using $W \in [2, 100]$.

We sought to determine the impact that increasing the number of weak classifiers (number of selected bins) would have on the classification results. To do so, we varied the value of W, while maintaining n_{bins} constant and equal to the one of the best configuration for each single feature (see Table V). The curves obtained are shown in Fig. 6. It is clear that the performance of the classifier depends on the number of features used. This reinforces the idea that most of the bins contain noisy information and have a marginal influence on the classifier output.

The characteristics of AdaBoost allows to further explore the data set. We performed tests with the best single color features (C_2, C_4, and C_6) and their combinations with the best texture feature (T_1) to determine if some histogram bins are more selected than others. This would mean that these bins are very discriminative. The results are shown in Fig. 7. All the tests were performed using the best configuration (see Table V) and considering the two regions R_1 and R_2, since it is clear that extracting features separately from inner and border improves the classification results. By inspecting the results, it is possible to notice that some bins are indeed selected most of the time. These bins correspond to specific channels of each color space, e.g., for C_6, the two most selected bins correspond to bins of the histogram of the second channel (O_2), and for C_2 and C_4, the most selected bin is from the histogram of the second channel [Saturation (S)], whereas the two other correspond to the third channel [Value (V) and Intensity (I), respectively]. An interesting result is that when a fusion of two best features is performed, the classifier only selects bins from color histograms, which demonstrates that color features have more discriminative information. The classifier selects information from R_1 and R_2, which suggests that both regions are informative. We have computed the mean and median values for the two most selected bins for each of the three best color features (see Table VI). The most selected bin is labeled as "1" and the second most selected bin as "2" (recall Fig. 7, to identify the corresponding bins for each color feature). These bins have values considerably different for the two classification classes, which demonstrates that these bins can be used to distinguish between melanomas and nonmelanomas.

Fig. 8 shows four examples classified by the proposed system using Adaboost, C_6, and the best combination of parameters for this feature. These examples correspond to the four possible cases, i.e., true negative (TN)—correctly classified nonmelanoma, false positive (FP)—incorrectly classified nonmelanoma, false negative (FN)—incorrectly classified melanoma, and true positive (TP)—correctly classified melanoma.

Although we do not use the same features and data sets as the other works found in the literature, a tentative comparison can be performed. Faziloglu *et al.* [26] reported the following values: $SE = 84.3\%$ and $SP = 83.0\%$, and Stanley *et al.* [23] reported the following values: $SE = 87.7\%$ and $SP = 74.9\%$. Both groups use only color features, and their data sets are larger than ours (258 and 226 images, respectively), possessing an equal proportion of melanomas and nonmelanomas. With our best configurations and classifier, we achieve results similar to the ones reported by these two groups, i.e., $SE = 96.0\%$ and $SP = 80.0\%$ using C_2/C_4 and $SE = 92.0\%$ and $SP = 85\%$ using C_6. Our classification results are also similar to the ones

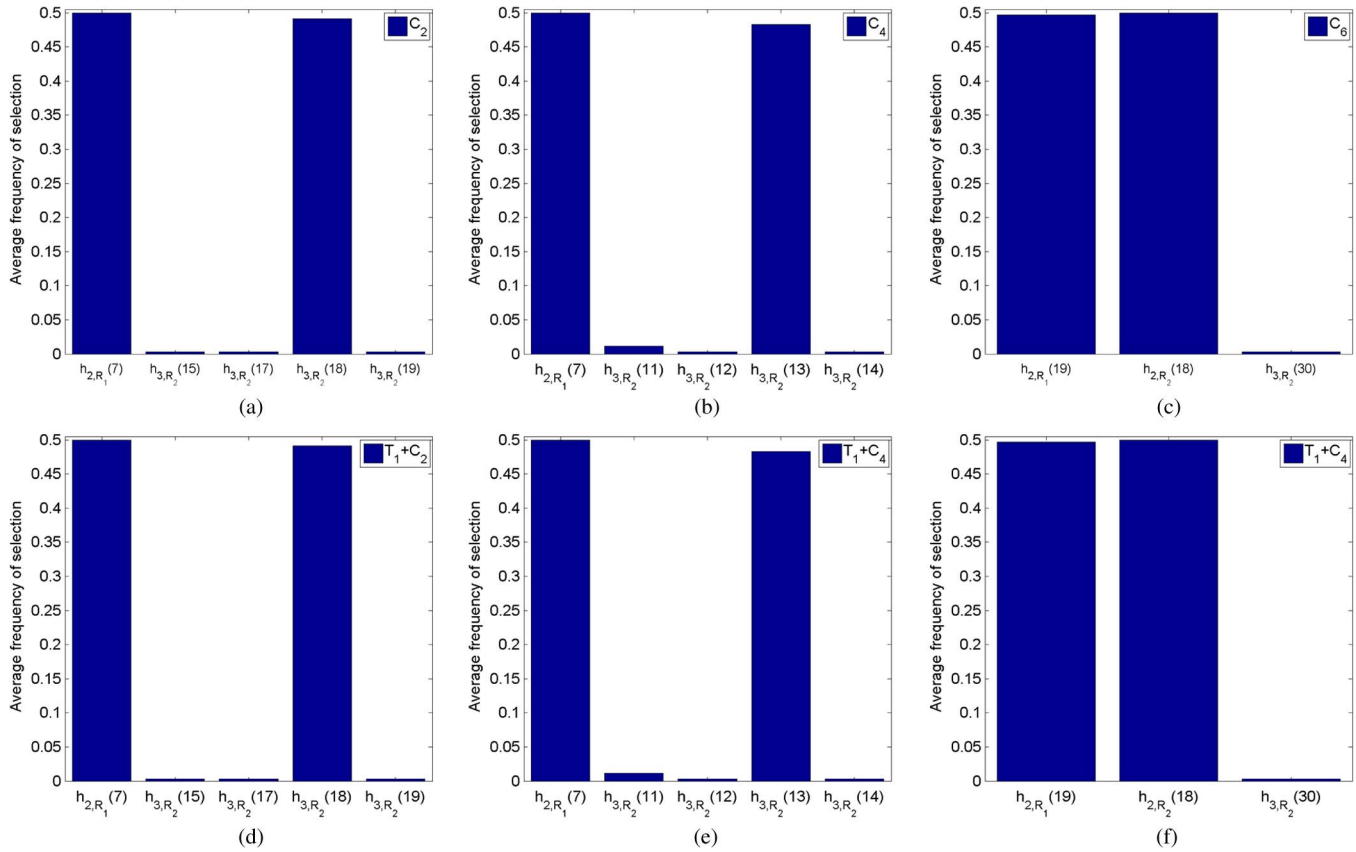

Fig. 7. Average frequency of selection of the most selected histogram bins for (a) C_2, (b) C_4, (c) C_6, (d) $T_1 + C_2$, (e) $T_1 + C_4$, and (f) $T_1 + C_6$.

TABLE VI
MEAN AND MEDIAN VALUES FOR THE TWO MOST SELECTED BINS (THE MOST SELECTED BIN IS "1" AND THE SECOND MOST SELECTED IS "2") FOR C_2, C_4, AND C_6

		C_2		C_4		C_6	
		Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*
1	*Melanoma*	0.027	0.023	0.027	0.023	0.094	0.093
	Non – Melanoma	0.005	0	0.005	0	0.039	7.23×10^{-4}
2	*Melanoma*	0.009	0.008	0.011	3.42×10^{-4}	0.063	0.054
	Non – Melanoma	0.002	0	0.001	0	0.010	0

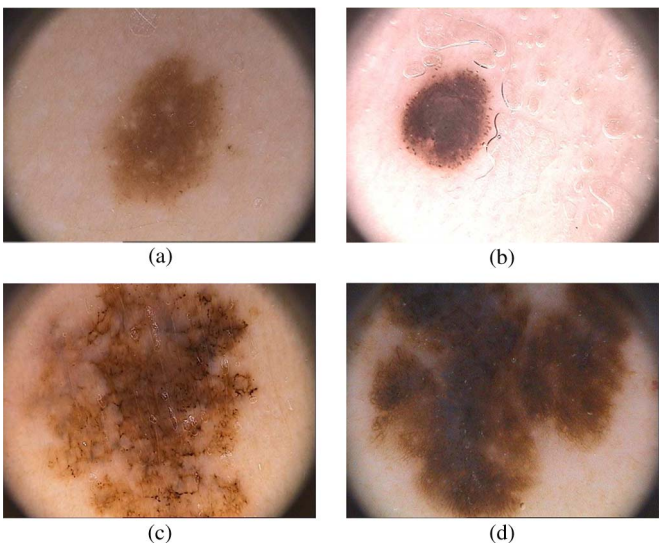

Fig. 8. Examples of classification cases. (a) TN. (b) FP. (c) FN. (d) TP.

reported by Seidenari *et al.* [21] ($SE = 87.5\%, SP = 85.7\%$). The performance achieved using the best texture feature T_1 (64.2% for $SE = 96.0\%$ and $SP = 59\%$) is similar to the ones reported by Yuan *et al.* [27] (accuracy of 70.0%).

B. Assessment of the Local Lesion Classification System

A similar procedure was used to optimize and test the CAD system based on local features. The same database of images was used in this case. However, the algorithm was evaluated using stratified tenfold cross validation, since the increase in computational time associated with the local methods turned leave-one-out evaluation unfeasible. The melanoma images were evenly distributed by all the folders, and their features were replicated as before, in order to balance both classes.

Since the feature extraction operation and the BoF classifier [see Fig. 2(b)] depend on a number of parameters, we have explicitly varied each of them in order to achieve the best

TABLE VII
BEST EXPERIMENTAL RESULTS USING LOCAL FEATURES AND BOF

	SE	*SP*	*Ć*
T_1	33%	76%	0.495
T_2	**90%**	**45%**	**0.281**
C_1	100%	72%	0.110
C_2	87%	88%	0.128
C_3	**93%**	**85%**	**0.104**
C_4	95%	79%	0.115
C_5	**100%**	**75%**	**0.099**
C_6	93%	68%	0.172
C	100%	71%	0.116
$T+C$	100%	70%	0.118
T_1+C_3	91%	86%	0.114
T_1+C_5	**98%**	**79%**	**0.100**

TABLE VIII
BEST PARAMETERS FOR THE BOF CLASSIFIER

	Δ	n_{bins}	K	*dist*	k
T_1	60	15	200	*Kullback – Leibler*	9
T_2	80	25	300	*Kullback – Leibler*	25
C_1	80	25	300	*Kullback – Leibler*	13
C_2	80	25	300	*Kullback – Leibler*	25
C_3	80	25	300	*Kullback – Leibler*	19
C_4	80	25	300	*Kullback – Leibler*	13
C_5	80	15	300	*Kullback – Leibler*	13
C_6	80	25	200	*Kullback – Leibler*	17
C	80	25	200	*Kullback – Leibler*	15
$T+C$	80	25	200	*Kullback – Leibler*	15
T_2+C_3	60	25	300	*Kullback – Leibler*	15
T_2+C_5	100	15	300	*Kullback – Leibler*	13

classification scores. In the case of features, we changed the grid size $\Delta \in \{20, 40, \ldots, 100\}$ and the number of histogram bins in the range $n_{bins} \in \{15, 25, 35, 45\}$. The visual words are obtained by applying the k-means clustering algorithm to the training vector features. The number of visual words was varied in the range $K \in \{100, 200, 300\}$. The test image is classified by the kNN algorithm as being melanoma or nonmelanoma. Three distances were considered (Euclidean, Kolmogorov, and Kullback–Leibler), and the number of neighbors was varied in the range $k \in \{5, 7, \ldots, 25\}$. We have tested all the possible combinations of these five parameters for each type of feature. This amounted to 44 928 configurations, each of them involving the training of ten classifiers, using the stratified tenfold cross validation.

Tables VII and VIII summarize the results achieved in this study. Table VII shows the best results achieved by the BoF classifier for each kind of features, and Table VIII shows the parameters of the best configurations. In the case of texture features, the best results are obtained with the histogram of orientation T_2. The role of the orientation seems to be much more important when computed in local patches than when it was obtained from the whole lesion. Color features achieve better results as before, leading to remarkable scores, e.g., $SE = 100\%$ and $SP = 75\%$ or $SE = 93\%$ and $SP = 85\%$. The concatenation of texture and color features lead to excellent results ($SE = 98\%,\ SP = 79\%$), but it does not improve the best color feature configuration.

Table VIII also allows to draw interesting conclusions. The best distance for these features is the Kullback–Leibler divergence. This is not surprising since all the features considered in this study are normalized histograms. The best block size is 80 for most features considered in these experiments. This is a bit surprising since we would expect that smaller block sizes would allow a better representation of local differential structures (e.g., pigment network, dots, streaks, and blue whitish veil). These differential structures are difficult to characterize even when the analysis is made by medical experts trained in dermoscopy and may be insufficiently described by the automatically extracted features.

TABLE IX
COMPUTATIONAL TIME FOR THE BEST GLOBAL AND LOCAL SYSTEMS PER IMAGE

	Global System	**Local System**
Feature Extraction	2 Regions - 0.13s	0.39s
Classification	AdaBoost - 0.48s	Histogram+kNN - 0.022s

Fig. 9 shows the variations of SE and SP for the best single color and texture features (C_3, C_5, and T_2), using different values of K (first row), Δ (second row), n_{bins} (third row), and the three possible distances (fourth row). The values that were kept unchangeable are the ones of the best configuration for each single feature (see Table VIII). It is clear that the number of visual words K influences the performance of the classifier. Furthermore, the size of the grid Δ seems to also have a great impact on the performance of the classifier.

Although most of previously published systems use global features, we can compare these results with the ones presented in [25], which achieved $SE = 80\%$ and $SP = 78\%$ in a database of 100 dermoscopy images with 30 melanomas. We stress that these results were obtained using texture features only (Gabor filters and wavelets) with a small block size ($\Delta = 16$).

C. Comparison Between Global and Local Classification Systems

Comparing local and global strategies for the detection of melanomas (Tables IV and VII), we conclude that both strategies achieve very good results in this database. The local features with BoF achieve slightly better results. In both cases, the color features perform significantly better than texture features. This is a bit surprising since differential structures play an important role in medical diagnosis of dermoscopy images, and therefore, we would expect that this behavior could be captured by texture descriptors.

The computational time for feature extraction and classification of an image using the two best systems can be seen in Table IX. Computing local features takes more time than computing global ones; this was expected since in the case of the local system, features are extracted for each patch, whereas in the case of a global system, the features are only extracted twice (the shown values are for the case of splitting the lesion in the inner part and the border).

VII. CONCLUSION

This paper has compared two different strategies for the detection of melanomas in dermoscopy images based on local

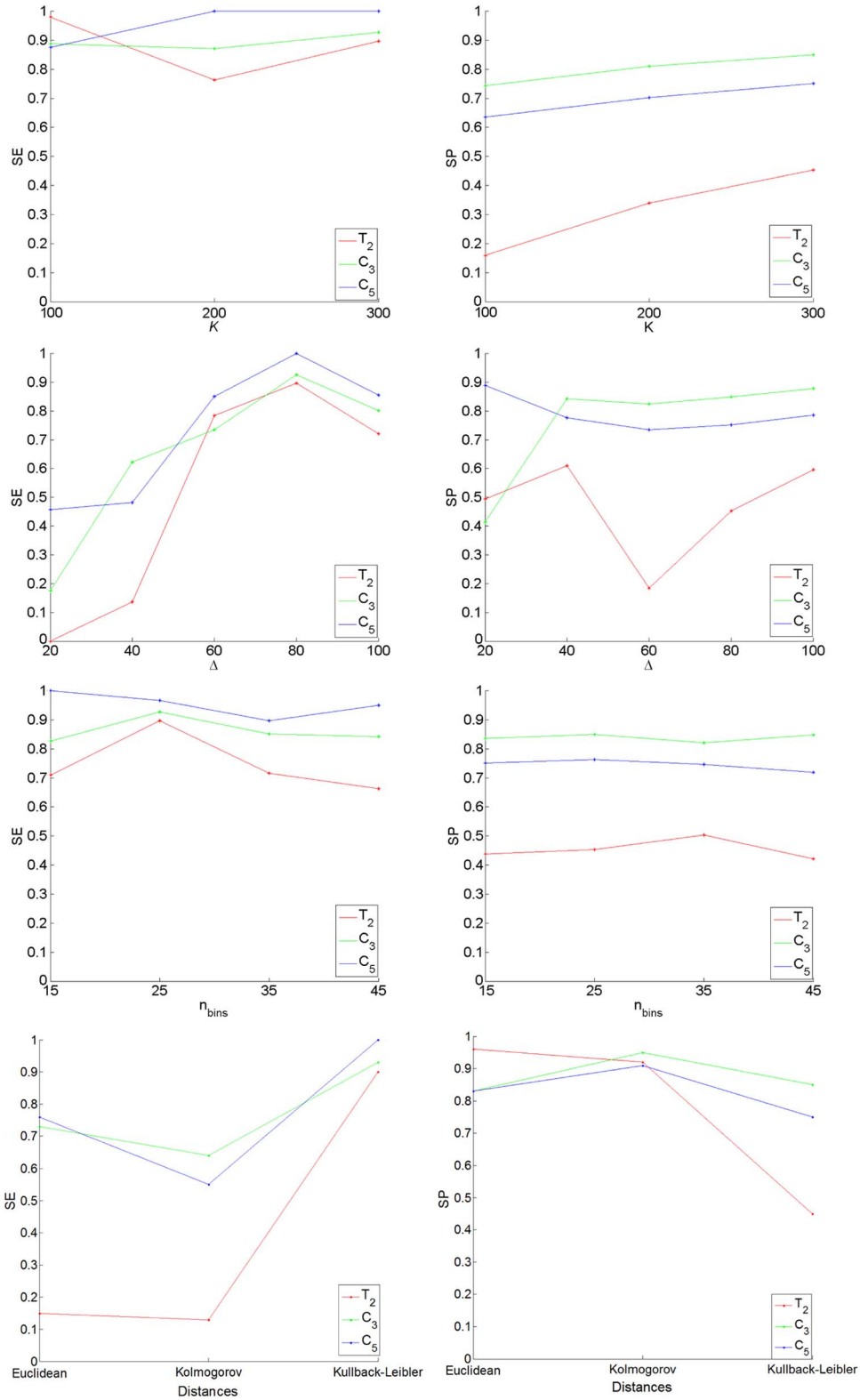

Fig. 9. Variation of SE and SP using (first row) $K \in \{100, 200, 300\}$, (second row) $\Delta \in \{20, 40, \ldots, 100\}$, (third row) $n_{bins} \in \{15, 25, 35, 45\}$, and the (fourth row) three tested distances.

and global features. Most previous works use global features (texture, shape, and color) associated to the whole lesion followed by a binary classifier trained from the data. However, local features are gaining increasing importance in many image analysis problems (e.g., image retrieval and object recognition) and may be seen as a promising alternative.

A second driving idea associated with this paper is the evaluation of the role played by color and texture features in the decision. Most works consider both types of features together, but they do not attempt to clarify if one of these features plays a more relevant role. This question has been explicitly addressed in this paper.

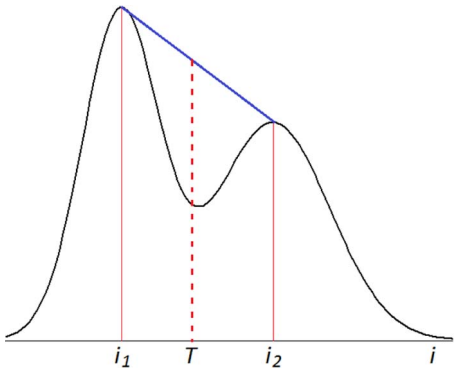

Fig. 10. Threshold computation from the histogram in the case of (right) two peaks.

Both issues were tested using a data set of dermoscopy images from Hospital Pedro Hispano, and the system parameters were tuned by exhaustive testing of many thousands of classifiers. It was concluded that color features perform much better than texture features alone achieving classification scores of $SE = 93\%$ and $SP = 85\%$. Concerning the use of global versus local features and classification strategies, both systems achieve very good results. The best global system results were achieved when the lesion region was split into two subregions (inner and border), leading to $SE = 96\%$ and $SP = 80\%$. However, the local system performs slightly better in terms of the classification cost used in this study $(SE = 100, SP = 75\%)$.

APPENDIX
LESION SEGMENTATION

Lesion segmentation is obtained using an adaptive thresholding algorithm, which comprises three steps: 1) histogram computation; 2) peak detection; and 3) threshold estimation. These three steps are described in the sequel.

First, the input image is converted into a gray-level image by selecting the channel with the highest entropy [40]. Then, the histogram of intensity, i.e., $h(i)$, $i = 0, \ldots, 255$, is computed using all the pixels inside a rectangular window, which is located at the center of the image. The histogram is then low-pass filtered with a Gaussian filter. The filter impulse response is $G(i) = C \exp\{-i^2/(2\sigma^2)\}$, where $\sigma = 5$, and C is a normalization constant.

We assume that the analysis intersects the lesion, and one of the histogram modes is associated to the lesion. We therefore extract the most significant peaks of the histogram (with a maximum 2). If the histogram has a single peak located at i_1, the threshold is defined by $T = i_1 + \Delta T$. The ΔT offset was empirically tuned and set to 15. If two significant peaks are extracted, we choose the threshold that corresponds to the maximum valley. The valley is defined as follows: Given two peaks located at $i_1, i_2, (i_2 > i_1)$, we define the valley at i as the difference between the straight line defined by the peaks and the histogram amplitude at i (see Fig. 10).

The only information missing is how to select the most significant peaks in a histogram if more than two peaks are detected (see Fig. 11). This involves two steps. First, close peaks are merged and replaced by a single peak, i.e., the largest one. We consider that two peaks located at i_1 and i_2 are close if $|i_1 - i_2| < \Delta i$ $(\Delta i = 8)$. If the number of peaks is larger than 2 after this step, the highest peak is selected. Then, all the remaining ones are tested. The second selected peak is the one that corresponds to the largest depth of the valley between two histogram peaks.

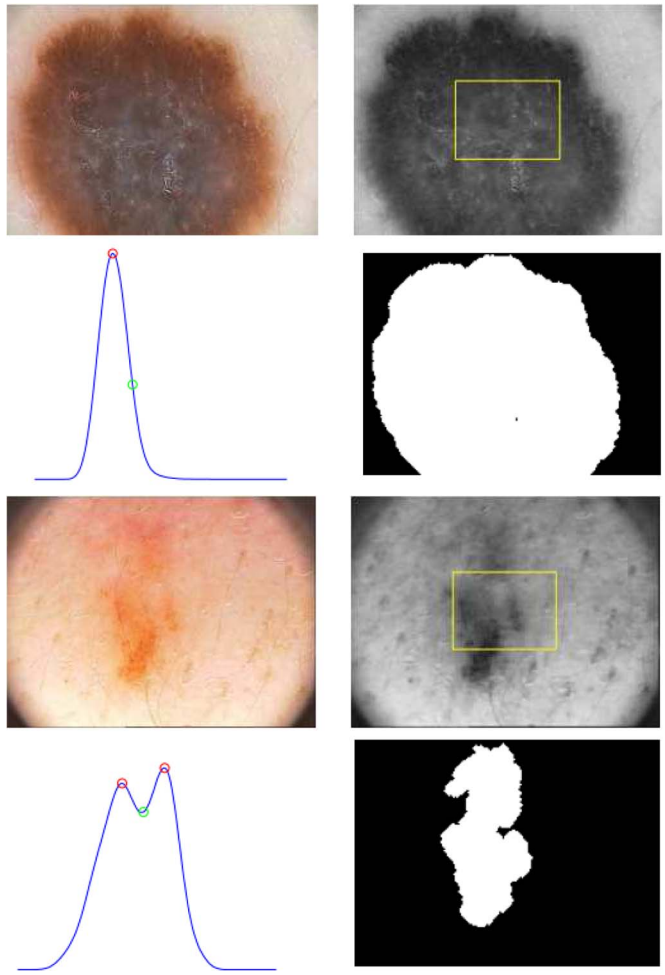

Fig. 11. Threshold computation from the histogram in the case of (right) two peaks.

REFERENCES

[1] "Cancer facts and figures 2012," American Cancer Society, Atlanta, GA, USA, 2012, Tech. Rep.
[2] Dermoscopy Tutorial 2003. [Online]. Available: http://www.dermoscopy.org/atlas/base.htm
[3] W. Stolz, A. Riemann, and A. B. Cognetta, "ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma," *Eur. J. Dermatol.*, vol. 4, no. 7, pp. 521–527, 1994.
[4] S. Menzies, C. Ingvar, K. Crotty, and W. McCarthy, "Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features," *Arch. Dermatol.*, vol. 132, no. 10, pp. 1178–1182, Oct. 1996.
[5] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Arch. Dermatol.*, vol. 134, no. 12, pp. 1563–1570, Dec. 1998.

[6] J. Mayer, "Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma," *Med. J. Aust.*, vol. 167, no. 4, pp. 206–210, Aug. 1997.

[7] M. Binder, M. Puespoeck-Schwartz, A. Steiner, H. Kittler, K. Wolff, M. Muellner, and H. Pehamberger, "Epiluminescence microscopy of small pigmented skin lesions: Short-term formal training improves the diagnosis performance of dermatologists," *J. Amer. Acad. Dermatol.*, vol. 36, no. 2, pp. 197–202, Feb. 1997.

[8] G. Di Leo, A. Paolillo, P. Sommella, and C. Liguori, "An improved procedure for the automatic detection of dermoscopic structures in digital elm images of skin lesions," in *Proc. 2008 IEEE Comput. Soc. VECIMS*, 2008, pp. 190–194.

[9] M. Sadeghi, M. Razmara, P. Wighton, T. K. Lee, and M. S. Atkins, "A novel method for detection of pigment network in dermoscopic images using graphs," *Comput. Med. Imaging Graph.*, vol. 35, no. 2, pp. 137–143, Mar. 2011.

[10] C. Barata, J. Marques, and J. Rozeira, "Detecting the pigment network in dermoscopy images: A directional approach," in *Proc. 33rd IEEE EMBS Annu. Int. Conf.*, Sep. 2011, pp. 5120–5123.

[11] G. Fabbrocini, G. Betta, G. Di Leo, C. Liguori, A. Paolillo, A. Pietrosanto, P. Sommella, O. Rescigno, S. Cacciapuoti, F. Pastore, V. Vita, I. Mordente, and F. Ayala, "Epilluminescence image processing for melanocytic skin lesion diagnosis based on 7-point check list: A preliminary discussion on three parameters," *Open Dermatol. J.*, vol. 4, pp. 110–115, 2010.

[12] W. Stoecker, M. Wronkiewicz, R. Chowdhury, R. J. Stanley, J. Xu, A. Bangert, B. Shrestha, D. A. Calcara, H. S. Rabinovitz, M. Oliviero, F. Ahmed, L. A. Perry, and R. Drugge, "Detection of granularity in dermoscopy images of malignant melanoma using color and texture features," *Comput. Med. Imaging Graph.*, vol. 35, no. 2, pp. 144–147, Mar. 2011.

[13] M. E. Celebi, H. Iyatomi, W. Stoecker, R. Moss, H. Rabinovitz, G. Argenziano, and H. Soyer, "Automatic detection of blue-white veil and related structures in dermoscopy images," *Comput. Med. Imaging Graph.*, vol. 32, no. 8, pp. 670–677, Dec. 2008.

[14] W. Stoecker, K. Gupta, R. Stanley, R. Moss, and B. Shrestha, "Detection of asymmetric blotches in dermoscopy images of malignant melanomas using relative color," *Skin Res. Technol.*, vol. 11, no. 3, pp. 179–184, Aug. 2005.

[15] G. Di Leo, A. Paolillo, P. Sommella, and G. Fabbrocini, "Automatic diagnosis of melanoma: A software system based on the 7-point checklist," in *Proc. 43rd Hawaii Int. Conf. Syst. Sci.*, 2010, pp. 1818–1823.

[16] H. Ganster, A. Pinz, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imaging*, vol. 20, no. 3, pp. 233–239, Mar. 2001.

[17] M. Elbaum, "Computer-aided melanoma diagnosis," *Dermatol. Clin.*, vol. 20, no. 4, pp. 735–747, Oct. 2002.

[18] P. Rubegni, G. Cevenini, M. Burroni, R. Perotti, G. Dell'Eva, P. Sbano, and C. Miracco, "Automated diagnosis of pigment skin lesions," *Int. J. Cancer*, vol. 101, no. 6, pp. 576–580, Oct. 2002.

[19] A. Blum, H. Luedtke, U. Ellwanger, R. Schwabe, G. Rassner, and C. Garbe, "Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions," *Brit. J. Dermatol.*, vol. 151, no. 5, pp. 1029–1038, Nov. 2004.

[20] M. Burroni, P. Sbano, G. Cevenini, M. Risulo, G. Dell'eva, P. Barbini, C. Miracco, L. Fimiani, M. Andreassi, and P. Rubegni, "Dysplastic naevus versus in situ melanoma: Digital dermoscopy analysis," *Brit. J. Dermatol.*, vol. 152, no. 4, pp. 679–684, Apr. 2005.

[21] S. Seidenari, G. Pellacani, and C. Grana, "Pigment distribution in melanocytic lesion images: A digital parameter to be employed for computer-aided diagnosis," *Skin Res. Technol.*, vol. 11, no. 4, pp. 236–241, Nov. 2005.

[22] M. E. Celebi, H. A. 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, no. 6, pp. 362–373, Sep. 2007.

[23] R. Stanley, W. Stoecker, and R. Moss, "A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images," *Skin Res. Technol.*, vol. 13, no. 1, pp. 67–72, Feb. 2007.

[24] 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," *Comput. Med. Imaging Graph.*, vol. 32, no. 7, pp. 566–579, Oct. 2008.

[25] N. Situ, X. Yuan, G. Chen, and J. Zouridakis, "Malignant melanoma detection by bag-of-features classification," in *Proc. 30th IEEE EMBS Annu. Int. Conf.*, 2008, pp. 3110–3113.

[26] Y. Faziloglu, R. Stanley, R. Moss, and W. S. R. McLean, "Colour histogram analysis for melanoma discrimination in clinical images," *Skin Res. Technol.*, vol. 9, no. 2, pp. 147–156, May 2003.

[27] X. Yuan, Z. Yang, and G. Zouridakis, "SVM-based texture classification and application to early melanoma detection," in *Proc. 28th IEEE EMBS Annu. Int. Conf.*, 2006, pp. 4775–4778.

[28] M. Sheha, M. S. Mabrouk, and A. Sharawy, "Automatic detection of melanoma skin cancer using texture analysis," *Int. J. Comput. Appl.*, vol. 42, no. 20, pp. 22–26, Mar. 2012.

[29] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp. 1470–1477.

[30] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Conf. Comput. Vis. Patter Recog.*, 2007, pp. 1–8.

[31] G. Qiu, "Indexing chromatic and achromatic patterns for content-based colour image retrieval," *Pattern Recog.*, vol. 35, no. 8, pp. 1675–1686, Aug. 2002.

[32] C. Walravem, B. Caputo, and A. Graf, "Recognition with local features: The kernel recipe," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp. 257–264.

[33] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2006, pp. 2169–2178.

[34] F. S. Khan, J. van de Weijer, and M. Vanrell, "Top-down color attention for object recognition," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 979–986.

[35] M. E. Celebi, G. Schaefer, and H. Iyatomi, "Objective evaluation of methods for border detection in dermoscopy images," in *Proc. 30th IEEE EMBS Annu. Int. Conf.*, 2008, pp. 3056–3059.

[36] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "Border detection in dermoscopy images using statistical region merging," *Skin Res. Technol.*, vol. 14, no. 3, pp. 347–353, Aug. 2008.

[37] M. Burroni, L. Alparone, and F. Argenti, "Comments on 'a new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions'," *IEEE Trans. Med. Imag.*, vol. 25, no. 12, pp. 1655–1656, Dec. 2006.

[38] H. Iyatomi, H. Oka, M. Saito, A. Miyake, M. Kimoto, J. Yamagami, S. Kobayashi, A. Tanikawa, M. Hagiwara, K. Ogawa, G. Argenziano, H. P. Soyer, and M. Tanaka, "Quantitative assessment of tumour area extraction from dermoscopy images and evaluation of the computer-based methods for automatic melanoma diagnostic system," *Melanoma Res.*, vol. 16, no. 2, pp. 183–190, Apr. 2006.

[39] M. E. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Res. Technol.*, vol. 19, no. 1, pp. e252–e258, Feb. 2013.

[40] M. Silveira, J. Nascimento, J. Marques, A. Marçal, T. Mendonça, S. Yamauchi, and J. Maeda, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE J. Sel. Topics Signal Processing.*, vol. 3, no. 1, pp. 35–45, Feb. 2009.

[41] Y. Freund and R. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[42] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[43] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man Cybern.*, vol. SNC-3, no. 6, pp. 610–621, Nov. 1973.

[44] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recog.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[45] R. Azencott, J. Wang, and L. Younes, "Texture classification using windowed Fourier filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 148–153, Feb. 1997.

[46] S. Arivazhagana, L. Ganesanb, and S. Priyala, "Texture classification using Gabor wavelets based rotation invariant features," *Pattern Recog. Lett.*, vol. 27, no. 16, pp. 1976–1982, Dec. 2006.

[47] D. Heeger and J. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. SIGGRAPH*, 1995, pp. 229–238.

[48] T. Randen and J. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.

[49] M. Tkalcic and J. Tasicl, "Colour spaces: Perceptual, historical and applicational background," in *Proc. IEEE Region 8 EUROCON_Computer as a Tool*, 2003, vol. 1, pp. 304–308.
[50] M. Bratkova, S. Boulos, and P. Shirley, "orgb: A practical opponent color space for computer graphics," *IEEE Comput. Graph. Appl.*, vol. 29, no. 1, pp. 42–55, Jan./Feb. 2009.
[51] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1593, Sep. 2010.
[52] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recog.*, vol. 40, no. 3, pp. 1106–1122, Mar. 2007.
[53] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, Jun. 2001.
[54] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2003, pp. II-264–II-271.
[55] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

Catarina Barata received the E. Biomedical and M.Sc. degrees from Technical University of Lisbon, Lisbon, Portugal, in 2009 and 2011, respectively. She is currently working toward the Ph.D. degree in the Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon.

Her research interests include image processing, pattern recognition, and dermoscopy.

Margarida Ruela received the E. Biomedical and M.Sc. degrees from Technical University of Lisbon, Lisbon, Portugal, in 2010 and 2012, respectively.

She is currently a Researcher with the Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon. Her research interests include biomedical image and signal processing and pattern recognition.

Mariana Francisco received the E. Biomedical and M.Sc. degrees from Technical University of Lisbon, Lisbon, Portugal, in 2010 and 2012, respectively.

She is currently with the Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon.

Teresa Mendonça received the Applied Math and Ph.D. degrees from the University of Porto, Porto, Portugal, in 1980 and 1993, respectively.

She is currently an Assistant Professor with the Department of Mathematics, Faculty of Sciences, University of Porto, where she is a Researcher with the Center for Research and Development in Mathematics and Applications. She has authored or coauthored more than 120 papers in international journals and conferences. Her current research interests include systems modeling and linear control, biomedical applications, and statistical image processing.

Jorge S. Marques received the E.E. and Ph.D. degrees and the Aggregation Title from Technical University of Lisbon, Lisbon, Portugal, in 1981, 1990, and 2002, respectively.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Lisbon, where he is also a Researcher with the Institute for Systems and Robotics. He has authored or coauthored over 150 papers in international journals and conferences and has authored a book entitled *Pattern Recognition: Statistical and Neural Methods* (IST Press, 2005, 2nd ed., in Portuguese). His current research interests include statistical image processing, shape analysis, and pattern recognition.

Dr. Marques was the Cochairman of the IAPR Conference IbPRIA 2005, the President of the Portuguese Association for Pattern Recognition from 2001 to 2003, and an Associate Editor of the *Statistics and Computing Journal* (Springer).