Bagged Textural and Color Features for Melanoma Skin Cancer Detection in Dermoscopic and Standard Images

Naser Alfed[1], Fouad Khelifi[1]

Abstract

Early detection of malignant melanoma skin cancer is crucial for treating the disease and saving lives. Many computerized techniques have been reported in the literature to diagnose and classify the disease with satisfactory skin cancer detection performance. However, reducing the false detection rate is still challenging and preoccupying because false positives trigger the alarm and require intervention by an expert pathologist for further examination and screening. In this paper, an automatic skin cancer diagnosis system that combines different textural and color features is proposed. New textural and color features are used in a bag-of-features approach for efficient and accurate detection. We particularly claim that the Histogram of Gradients (HG) and the Histogram of Lines (HL) are more suitable for the analysis and classification of dermoscopic and standard skin images than the conventional Histogram of Oriented Gradient (HOG) and the Histogram of Oriented Lines (HOL), respectively. The HG and HL are bagged separately using a codebook for each and then combined with other bagged color vector angles and Zernike moments to exploit the color information. The overall system has been assessed through intensive experiments using different classifiers on a dermoscopic image dataset and another standard dataset. Experimental results have shown the superiority of the proposed system over state-of-the-art techniques.

Keywords: Malignant melanoma, skin cancer diagnosis, dermoscopic images, standard skin images, textural and color features.

1. Introduction

Malignant melanoma is one of the most aggressive form of skin cancer and its incidence has been rapidly increasing over the last few decades, causing the majority of deaths related to skin cancer (Korotkov and Garcia, 2012; Silveira et al., 2009). Fortunately, if melanoma skin cancer is detected at its early stages, it can be curative for the patient. However, distinguishing melanoma in its early stages from other pigmented skin lesions is still challenging. Many techniques have been used to predict and classify melanoma skin cancer. Image processing tools for skin cancer detection usually require preprocessing operations for enhancing images and segmenting the regions of interest to extract efficient features. Dermoscopic and standard images captured from skin usually have some noisy artifacts such as applied oil and hair and this should be removed prior to segmentation. In this context, Dull Razor medical software was first developed by Lee et al. (1997) to remove hairs from pigmented areas. However, the system has been criticized for disrupting the normal skin pattern over the area covered by the hairs (She et al., 2006). The fast median filtering was later adopted to remove noise from the acquired skin images (Tanaka et al., 2004). In (Kiani and Sharafat, 2011), an improved version of the Dull Razor medical software, called E-shaver, has been proposed. The technique mainly enhances hair detection and removals by identifying light-colored hairs in addition to dark hairs. Barata et al. (2012) proposed two important steps for reflective artifacts and hair detec-

[1]Department of Computer and Information Sciences, Northumbria University at Newcastle, NE2 1XE, UK. E-mail: {naser.alfed, fouad.khelifi}@northumbria.ac.uk

tion and removals using a bank of directional filters. Once a skin lesion image is enhanced via artifacts and hair removal techniques, the lesion area is segmented (Lee and Chen, 2015). Many segmentation techniques have been developed in the literature for melanoma skin cancer diagnosis in dermoscopic images (Silveira et al., 2009). Among these techniques, the region-based approach selects a set of seed points and from each point a region grows up if the neighbouring pixels have similar properties to that of the seed point. Region growing often generates irregular boundaries and small holes (Zhu and Yuille, 1996). It also has some drawbacks such as sensitivity to noise and often results in over segmentation (Tobias and Seara, 2002). In (Nock and Nielsen, 2004), a statistical region merging (SRM) algorithm has been proposed.

One of the common features used by dermatologists for diagnosing melanoma skin cancer is the rule-based approach, called ABCD, and based on morphological analysis of lesions in dermoscopic images. This is used to distinguish between melanoma skin cancer and non melanomas (Nachbar et al., 1994). It is a medical diagnostic method that is based on four criteria, i.e., asymmetry, border irregularity, color variegation, and different structure. The 7-point checklist is another medical diagnostic method that is widely used by researchers and is based on a set of different characteristics depending on color, shape, and texture (Argenziano et al., 1998). These features can be categorized into color and texture features. In (Barata et al., 2014a), the authors investigated the role of color and texture features for skin melanoma cancer detection and concluded that color features are more efficient than texture in dermoscopic images. They also showed that the features which are locally extracted from the images bring more information about the lesion than global features. In (Barata et al., 2014b), the authors demonstrated, via experiments, that color descriptors deliver better performance in detecting melanoma skin lesions than texture descriptors. The authors in (Barata et al., 2012) adopted a set of directional filters and a connected component analysis to extract five different features for pigment network detection in dermoscopic images. In (Barata et al., 2013), the role of key-point sampling in a bag of features approach was investigated. The authors suggested that performance of the system can be influenced by the number of detected key-points. In (Barata et al., 2015), color constancy has been explored to overcome the problem of changes that may occur during the skin image acquisition process. The authors in (Riaz et al., 2014) proposed a combination of texture and color features for the classification of melanoma and non-melanoma skin images. A variation of the local binary patterns (LBP) was used for the texture features to extract scale adaptive patterns. As for the color information, the histograms of the HSV (Hue, Saturation, Value) color space was adopted. More recently, Ruela et al. (2015) have explored the importance of shape and symmetry features in Melanoma diagnosis in order to determine the type of features that play a crucial role in classification. In (Abuzaghleh et al., 2015), the authors proposed a combination of Lesion Variation Pattern Features (LVPF) with some extracted shape, color and texture features including the pigment network feature set, the lesion shape feature, the lesion orientation feature, the lesion margin feature, the lesion intensity pattern feature, and the lesion variation pattern feature. In (Vasconcelos et al., 2015), color features have been derived from the ABCD rule where the authors proposed a clustering approach to adjust the system to different datasets and image types. In (Kruk et al., 2015), different texture and statistical features were adopted, including the numerical descriptors based on the Kolmogorov-Smirnov (KS) statistical distance, the classical Haralick descriptors and fractal texture analysis-based descriptors. In (Giotis et al., 2015), physician annotations for skin lesions, referred to as visual diagnostic attributes, were combined with lesion color and lesion texture for melanoma skin detection in non-dermoscopic images. Very recently, three types of features have been used in (Chakravorty et al., 2016), namely, geometry features, color features, and finally structural features. More recently, the authors in (Oliveira et al., 2016) adopted asymmetry, border, color and texture features followed by an SVM classifier for the classification of pigmented skin lesions in macroscopic (standard) images.

In this paper, an automatic skin cancer diagnosis system that combines different textural and color fea-

tures is proposed. New textural and color features are introduced in a bag of features approach for efficient and accurate skin cancer detection. We particularly claim that the Histogram of Gradients (HG) and the Histogram of Lines (HL) are more suitable for the analysis and classification of dermoscopic and standard skin images than the conventional Histogram of Oriented Gradient (HOG) and the Histogram of Oriented Lines (HOL), respectively. This is because the orientation of melanoma edges and texture is not a discriminating feature when compared to non melanoma lesions. Therefore, the use of edge and line orientation in skin images reduces the inter class dissimilarity, and this causes an adversary effect on classification. The HG and HL are bagged separately using a codebook for each and then combined with other bagged Color Vector Angles (CVA) and Zernike moments to exploit the color information. Experimental results demonstrate the efficiency of the proposed texture and color features as well as the superiority of the overall system over state-of-the-art melanoma skin cancer detection techniques. The rest of the paper is organized as follows. Section 2 describes the proposed system and the features extracted whereas section 3 provides a discussion of experimental results obtained. Conclusions are drawn in section 4.

2. Proposed system

The proposed system consists of six main stages at the training phase, namely: preprocessing, key-point detection, segmentation, patch extraction and selection (region of interest), feature extraction, codebook generation, histogram building, features concatenation, and classification. Fig. 1 illustrates the proposed approach.

2.1. Pre-processing

For efficient feature extraction, a process of image enhancement is first conducted. The reason for this is that the original acquired images may have some air bubbles and artifacts caused by gel applied before the capture of the images in addition to hairs and other noise. The same procedures that have been implemented in (Alfed et al., 2015) which is based on method proposed by Barata et al. (2012) which

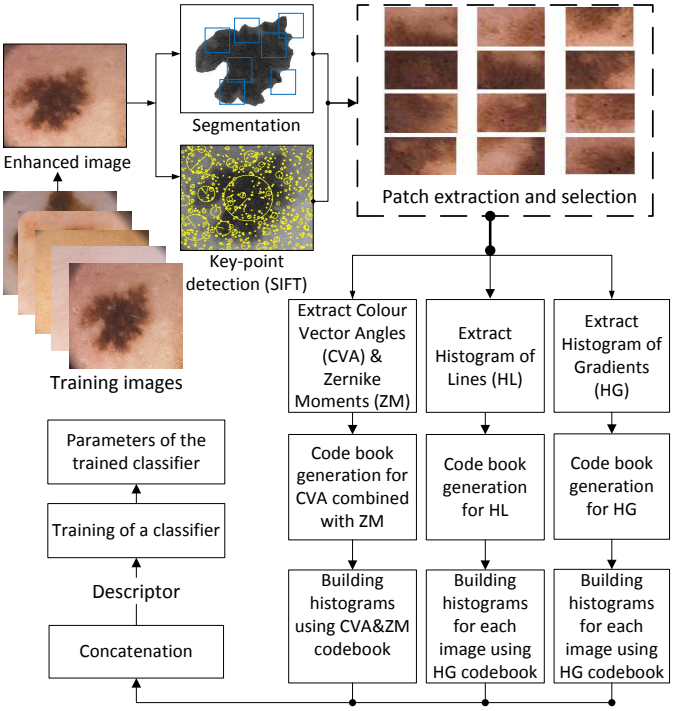

Figure 1: The training phase of the proposed system

consists of detecting and removing two undesirable patterns: reflective artifacts and hair. To detect artifacts reflection, a simple thresholding algorithm is adopted where every pixel with a certain brightness in the image is compared against its neighbourhood. If the pixel of interest is distinguished from its neighbourhood with a clearly larger luminance, the pixel is said to be an artifact reflection. Once artifacts are detected, an inpainting operation is applied accordingly. Fig. 2 illustrates the process of converting the original color image (a) into a grey-scale image (b) and then removal of artifacts is obtained in (c).

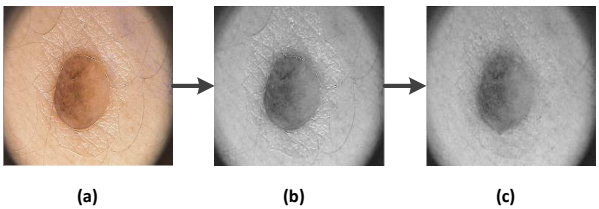

Figure 2: Pre-processing for artifacts removal; (a):original image, (b):grey-scale image, (c):after artifacts removal

Hair detection and removal is another big challenge especially when hair lines are not accurately detected. The directional Gabor filters are first ap-

plied and then followed by finding the local maximum at each pixel location. The parameters of the Gaussian filters used here have been adopted from (Barata et al., 2012). The hair detection algorithm uses a bank of 64 directional filters to perform the detection. Again, once hair is detected, an inpainting operation is applied on the corresponding area. Fig. 3 illustrates hair detection and removal using the directional Gabor filers.

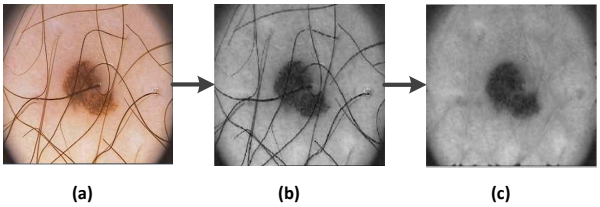

Figure 3: Pre-processing for artefacts and hair removal; (a): original image, (b): blue component image, (c): after hair detection and removal

2.2. Key-points detection

Lowe (2004) presented an efficient method for detecting key points from images that are invariant to image scale and rotation to perform reliable matching between different views of an object. The proposed Lowe's approach was named the Scale Invariant Feature Transform (SIFT). The SIFT method can be used for transforming image data into scale-invariant coordinates relative to local features. It generates a good number of features over the whole range of scales and locations that fully describe the image (Lowe, 1999). Cascade filtering is widely used for Key-points detection. It uses efficient algorithms to identify the locations of the feature-point candidates, subsequently; these candidates are subjected to further analysis. Thus, searching for stable key-points through all continuous possible scales is an important step to determine the locations, which are known as scale invariant features. First, denote by $L(x, y, \sigma)$ the convolution of an image $I(x, y)$ with a Gaussian kernel $G(x, y, \sigma)$ at scale SIGMA as:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

where $*$ denotes the convolution process and the Gaussian operator is defined by:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

With the standard deviation σ of the Gaussian distribution. Lowe proposed searching for the scale-space extrema in a difference-of-Gaussian function (DoG) to detect stable key-point locations efficiently. DoG can be computed by finding the differences of the adjacent scales of image I and separated by a constant multiplicative factor k which is defined as:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (3)$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4)$$

The detection of local extrema in the DoG image can be obtained by comparing each sample pixel to its 26 neighbours: 8 at the same image scale and 9 pixels at the above and below immediately scales (see Fig. 4). If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate key-point.

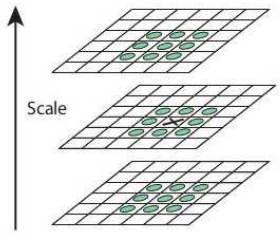

Figure 4: Key-point detection, Copyright OpenCV-Python

Once candidate key points are detected, the key points with low contrast are discarded to maintain only stable points. This is performed by comparing the second-order Taylor expansion of the DoG at the corresponding key points location against a certain threshold.

2.3. Segmentation

In this work, it is meant by segmentation the detection of the lesion area in a skin image. Here, the Statistical Region Merging (SRM) algorithm has been adopted. This is an unsupervised learning approach for border detection proposed by Nock and Nielsen (2004). This method has become well known and is widely used in the segmentation process due to

its efficiency, simplicity, and accurate performance without the use of quantization or colour space transformations. The SRM algorithm is used to evaluate the pixel values within a regional area and grouped them based on homogeneous properties attaining a smaller list. Two important components that to define this algorithm are merging (test) and order in merging. It basically accumulates couples of adjacent pixels in the image I to be in one set S_I. These couples are then sorted in an ascending order of a real function $f(p, p')$, with p and p' being pixels of I and traversing this order only once. The current region to which p belongs to, is taken as the test. Then the $R(p)$ and $R(p')$ are merged if the corresponding function f returns true. Fig. 5 illustrates skin lesion detection from a dermoscopic image.

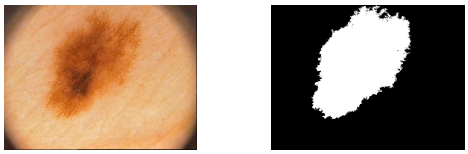

Figure 5: Image lesion detection (ROI). First column: Original images. Second column: Skin lesion.

2.4. Patches extraction and selection

In this section, the patches are extracted based on detected key-points. Each of these key-points is to be considered as the center of a candidate square patch. The selection of only the patches with areas of more than 50% inside the segmented lesion is performed. The selected patches that are extracted from the enhanced image have been used for the next step of extracting local features.

2.5. Feature extraction

In this work, a number of efficient features have been extracted from each selected patch in order to describe skin lesions efficiently. This consists of the HG, HL, CVA, and the 3^{rd} order Zernike moments. As will be demonstrated later, HG and HL outperform the conventional HOG and HOL, respectively.

2.5.1. Histogram of Gradients (HG)

Lowe (2004) has first used a variant of the histogram of oriented gradients (HOG) as a descriptor with orientation alignment. Dalal and Triggs (2005) later proposed the HOG descriptor for pedestrian detection. This has then been widely applied in object detection and image recognition applications. The rational behind the HOG descriptor is that local features in an image can be described by the distribution of intensity gradients and edge directions. In this work, however, we ignore the orientation of edges and texture and propose to use the histogram of gradients for each selected patch. The idea is basically to compute the gradient magnitude $m(x, y)$ for each patch as

$$\partial_x I = I(x+1, y) - I(x-1, y) \tag{5}$$

$$\partial_y I = I(x, y+1) - I(x, y-1) \tag{6}$$

$$m(x, y) = \sqrt{(\partial_x I)^2 + (\partial_y I)^2} \tag{7}$$

Recall that a skin image is divided into a number of patches. Each patch can be described by a HG accordingly where the boundaries used to calculate the histograms are determined by the minimum and maximum values of all the gradient magnitudes of the image. Then, each HG is L2-normalized. We emphasize here that the gradient orientation is neglected because it does not bring any useful information on dermoscopic and standard skin images. Indeed, regardless of the lesion type and class, its gradient orientation cannot be discriminative as illustrated by Fig. 6 and Fig. 7 (see third column), respectively. As a result, the use of the gradient orientation may reduce the inter class dissimilarity and, hence, deteriorate the classification performance.

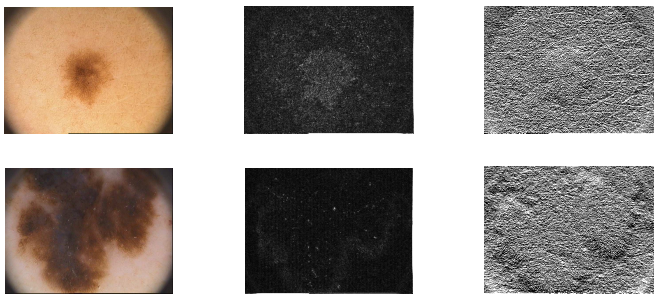

Figure 6: Gradient magnitude and orientation for normal and abnormal dermoscopic images. First row: Normal images. Second row: Abnormal images. First column: Original image. Second column: Gradient magnitude. Third column: Gradient orientation.

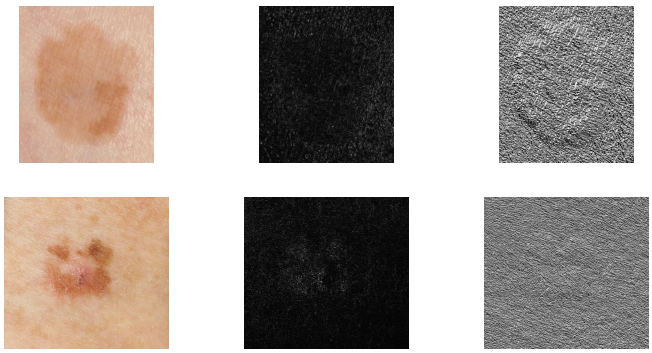

Figure 7: Gradient magnitude and orientation for normal and abnormal standard skin images. First row: Normal images. Second row: Abnormal images. First column: Original image. Second column: Gradient magnitude. Third column: Gradient orientation.

On the other hand, the gradient magnitude can discriminate normal lesions from abnormal ones in the sense that skin cancer lesions are less textured than normal lesions (see the second column of Fig. 6 and Fig. 7). In Fig. 8, the HOG and HG of a block of melanoma and non melanoma lesion are illustrated. The corresponding Chi-square measures between the histograms of melanoma and non melanoma blocks are depicted in Table 1 for both HG and HOG. As can be seen, the dissimilarity with HG is more significant than that with HOG.

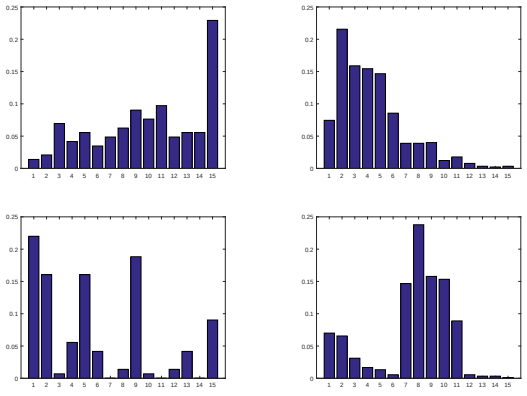

Figure 8: HOG and HG for normal and abnormal dermoscopic images. First row: Normal image. Second row: Abnormal image. First column: HOG. Second column: HG.

Table 1: Difference of HOG and HG between a skin cancer lesion and another normal lesion in terms of the Chi-square measure.

Histogram	*HOG*	*HG*
Chi-square	0.34	0.45

2.5.2. Histogram of Lines (HL)

The Histogram of Oriented Lines (HOL) was initially adopted in (Jia et al., 2014) for palm-print identification. It basically uses the same approach as the HOG but computes the directed lines in the image instead of gradients. Practically, the summation of image pixels over a certain set of lines can be obtained via the finite Radon transform (FRAT) (Do and Vetterli, 2003). Indeed, in a textured image, a line can be detected by the FRAT from the neighborhood of each pixel in a small local area. However, because the FRAT treats the input image as a periodic signal, Huang et al. (2008) proposed a Modified Finite Radon transform (MFRAT) to highlight line patterns effectively through a filtering-like process. The authors of (Jia et al., 2014) used the MFRAT image to generate a HOL descriptor for palm-prints by adopting the same steps of the HOG descriptor. One beneficial property of the HOL is that it has good invariance to changes of illumination than the HOG. Moreover, the HOL is robust against transformations because slight translations and rotations lead to only small histogram value changes (Jia et al., 2014).

The MFRAT of a real function f(x,y) on the finite grid Z_p^2 centered at (x_0, y_0) is defined as

$$r(L_k(x_0, y_0)) = \sum_{x,y \epsilon L_k} f(x, y) \tag{8}$$

where $f(x, y)$ is the pixel value located in (x, y) and L_k denotes the set of points that make up a line on the lattice Z_p^2, which means that:

$$L_k(x_0, y_0) = \{(x, y) : y = k(x - x_0) + y_0, (x, y) \epsilon Z_p^2\} \tag{9}$$

where (x_0, y_0) denotes the centre point of the lattice Z_p^2 and k is the corresponding slope of L_k. If the centre of lattice Z_p^2 moves over an image pixel by pixel (or pixels by pixels), the energies and directions of all pixels are calculated. Two new images, direction and magnitude images respectively can be created, if the values of all pixels for an image $I(x, y)$ of size $n \times m$ are replaced by their directions and energies. In this work, we compute the energy function for each pixel at (x_0, y_0) as

$$E(x_0, y_0) = \min(r(L_k(x_0, y_0)) \tag{10}$$

The equation below show how the new energy images can be created (Huang et al., 2008):

$$F_E = \begin{vmatrix} E(1,1) & E(1,2) & ... & E(1,m) \\ E(2,1) & E(2,2) & ... & E(2,m) \\ ... & ... & ... & ... \\ E(n,1) & E(n,2) & ... & E(n,m) \end{vmatrix}, \quad (11)$$

where F_E is the energy matrix. The same steps for computing the histogram of gradients will be used here for computing the histogram of lines, denoted by HL. The HL extracted from each patch is adopted as a feature vector instead of HOL. Note that the line orientation is neglected because it does not bring any useful information on dermoscopic and standard skin images as discussed earlier in section 2.5.1. Obviously, as can be seen in Fig. 9 the dissimilarity with HL is better than that with HOL for normal and abnormal images.

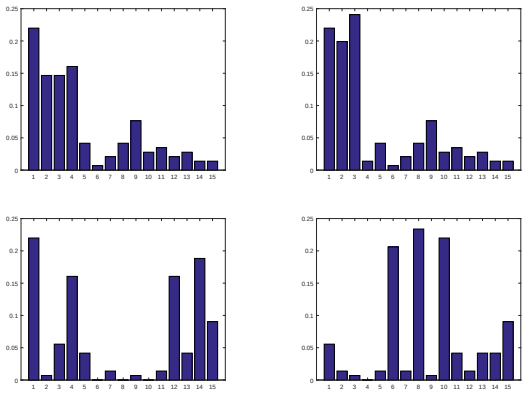

Figure 9: HOL and HL for normal and abnormal dermoscopic images. First row: Normal image. Second row: Abnormal image. First column: HOL. Second column: HL.

Furthermore, the analysis of the two previous lesions using the Chi-Square measure has been included in Table 2. As shown, the dissimilarity with HL is more significant than that with HOL.

Table 2: Difference of HOL and HL between a skin cancer lesion and another normal lesion in terms of the Chi-square measure.

Histogram	*HOL*	*HL*
Chi-square	0.31	0.55

2.5.3. Color vector angle (CVA)

Vector angle is effective in evaluating color contrast and can be used for extracting color features. Color Vector Angles (CVA) have been widely used in edge detection, image hashing and image retrieval (Tang et al., 2014). CVA is used for identifying color edges because it is insensitive to variations in intensity while being sensitive to differences in hue and saturation (Dony and Wesolkowski, 1999). In this work, CVA is adopted in extracting color features from dermoscopic and standard skin images. As angle calculation needs two colors, a reference color is generated from $P_{ref} = [R_{ref}, G_{ref}, B_{ref}]^T$, where R_{ref}, G_{ref} and B_{ref} are the means of the red, green and blue components of all pixels. Therefore, for each pixel, the color vector angle between its RGB vector and P_{ref} is calculated. The advantage of color vector angle is attributed to its sensitivity to hue differences. Let $P_1 = [R_1, G_1, B_1]^T$ and $P_2 = [R_2, G_2, B_2]^T$ be the vectors of two colors. The angle Δ between two pixels P_1 and P_2 can be computed by

$$\Delta = \arcsin\left(\left(1 - \frac{\left(P_1^T P_2\right)^2}{P_1^T P_1 P_2^T P_2}\right)^{\frac{1}{2}}\right) \quad (12)$$

where arcsin(.) is the operation arcsin. Once a matrix of angles Δ is extracted from an image patch, a histogram is constructed using a certain number of bins. This constitutes the feature vector corresponding to the CVA.

2.5.4. Zernike moments

Zernike moments have been used in shape-based image retrieval and edge detection as they are essentially retrieved from geometric moments by replacing the conventional transform kernel with orthogonal Zernike polynomials (Khotanzad and Hong, 1990). In general, orthogonal moments are better than other types of moments in terms of information redundancy and image representation and they have important advantages such as the rotational invariance, robustness to noise and expression efficiency (Zhao et al., 2013). In fact, because Zernike polynomials are orthogonal to each other, Zernike moments can represent the properties of an

image with no redundancy or overlap of information. Therefore, a skin lesion can be comprehensively described by the Zernike moments. Zernike Moments (ZM) with different orders have been implemented in this work according to the following equation. The best performance is obtained when using the 3^{rd} order moment, and therefore it has been used here. Zernike Moments are defined as:

$$Z_{\alpha,\beta} = \frac{\alpha+1}{\pi} \sum_{(p,\theta)\epsilon\ unit\ disk} \sum I(p,\theta)\, V^{*}_{\alpha,\beta}(p,\theta) \quad (13)$$

where $V_{\alpha,\beta}(p,\theta)$ is a Zernike polynomial of order α and repetition β and the symbol $*$ denotes the complex conjugate. The form of these polynomials is given as (Zhao et al., 2013)

$$V_{\alpha,\beta}(p,\theta) = R_{\alpha,\beta}(p)\, e^{j\beta\theta} \quad (14)$$

where α is either a positive integer or zero and β is subject to the constraint that $\alpha - |\beta|$ is even, and $0 \leq |\beta| \leq \alpha$, and $R_{\alpha,\beta}(p)$ are the radial polynomials (Zhao et al., 2013). Table 3 lists for each order α the corresponding number of moments which were experimentally tested. Five Zernike moments at the third order are extracted from each color plane of a patch in the RGB color representation. The respective set of features consists of the L2-normalized fifteen Zernike moments (e.g. five from each color channel).

Table 3: List of Zernike moments up to order three

α	Moments	No. of moments
1	$Z_{1,1}$	1
2	$Z_{2,0}, Z_{2,2}$	2
3	$Z_{3,1}, Z_{3,3}$	2

2.6. Codebook generation

At the training stage, a large number of feature vectors, where each is extracted from a patch of a training image, are used to generate a codebook. This will then serve as a dictionary to represent each test image as a histogram since each patch corresponds to a codeword. Three codebooks have been generated for each type of features, i.e. HG, HL, and CVA+ZM. The first codebook is for HG features, the second one is for the HL features, and the third codebook is generated using a concatenation of CVA and the 3^{rd} order Zernike Moments (ZM). The K-means method is adopted in this work for generating the centroids (codewords) in each codebook. The idea of codebook generation is that each feature vector extracted from a patch of any training image is used to form a codebook via K-means with a certain number of centroids (also called codewords).

2.7. Building histograms and concatenation

Once the codebooks are generated, each patch of an image is assigned a label corresponding to the closest centroid according to the Euclidean distance. For each codebook, the image is described by the frequency of labels in the form of a histogram. As mentioned earlier, this applies to the three types of features. Therefore, the final descriptor F_d is a concatenation of three histograms as

$$F_d = [C_{HG}, C_{HL}, C_{CVA+ZM}] \quad (15)$$

where C_{HG}, C_{HL}, and C_{CVA+ZM} represent the histogram of gradients, the histogram of lines, and the histogram of CVA and ZM corresponding to their respective features, accordingly.

2.8. Classification

Given a number of descriptors, F_d, obtained from medically annotated images, a binary classifier is trained on two classes, i.e., melanoma and non melanoma.

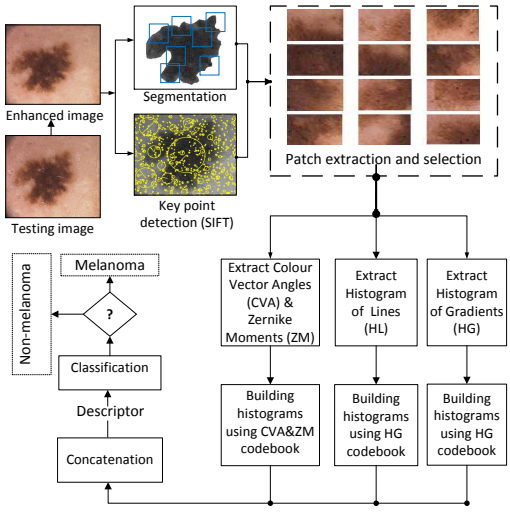

Figure 10: The testing phase of the proposed system

Once trained, the classifier can be used for melanoma skin cancer diagnosis. Once the codebooks are generated and the classifier is trained, the testing phase follows the same steps as shown in Fig. 10. When a test image is presented to the system, it is first subjected to the pre-processing stage, segmentation, path extraction and selection, and feature extraction prior to the obtention of the descriptor via the constructed codebooks. The descriptor is then presented to the trained classifier in order to detect the presence of melanoma skin cancer.

3. Experimental results

In our experiments, five-fold cross-validation is carried out for training and testing the proposed system. A dataset of 200 medically annotated dermoscopic images (40 melanomas and 160 non-melanomas) with a truth table obtained from the database of the Hospital Pedro Hispano (HPH) (Mendonc et al., 2015) has been used here for classification purposes. The system has also been tested on the Dermofit database where 236 standard images were used (76 melanomas and 160 non-melanomas). This has been obtained from the University of Edinburgh (Ballerini et al., 2013). Unless otherwise stated, the setting used in our experiments for the process of feature and descriptor extraction is listed in Table 4.

Table 4: Parameters setting of the system

Features	Parameter	Value
CVA	codebook size	100
	patch size	30×30
	bins per patch	15
ZM	codebook size	100
	patch size	30×30
	moments per patch	15
HG	codebook size	100
	patch size	120×120
	bins per patch	15
HL	codebook size	100
	patch size	120×120
	bins per patch	15

It is worth mentioning that only the patches whose area covers more than 50% of the lesion were selected for feature extraction. The measures used in our experiments consist of the False Positive rate (FP), the False Negative rate (FN), Sensitivity (SE), Specificity (SP), and Accuracy (Acc).

3.1. System analysis

In the first set of experiments, the combination of color moments and color histograms, as proposed in (Barata et al., 2014a,b), have been assessed in the bag of features approach as opposed to our combined CVA+ZM features. This is to illustrate the contribution of the proposed color features and Zernike moments. To this end, three different classifiers were used, namely the SVM, AdaBoost, and an Artificial Neural Network (ANN) with 120 neurons in the hidden layer and the Radial Basis Function (RBF) as an activation function. The results on the dermoscopy HPH dataset are depicted in Tables 5 and 6.

Table 5: Comparisons of color features

	$\mathbf{C_{RGB+M}}$				$\mathbf{C_{Opp+M}}$			
	FN	FP	SE%	SP%	FN	FP	SE%	SP%
SVM	0.21	0.35	78.75	65.33	0.16	0.31	84.29	69.21
AdaBoost	0.14	0.39	86.26	60.67	0.11	0.35	89.36	64.62
ANN (RBF)	0.11	0.28	89.45	71.89	0.09	0.29	91.39	70.51

Table 6: Comparisons of color features

	$\mathbf{C_{CVA+ZM}}$			
	FN	FP	SE%	SP%
SVM	0.15	0.32	85	68.39
AdaBoost	0.09	0.33	90.96	67.29
ANN (RBF)	0.06	0.14	93.73	86.29

As can be seen, the proposed CVA+ZM features appear more powerful than the conventional combination of color histograms (using the Opponent color model) and color moments (using the first, second and third order statistical moments). Next, the performance of the Histogram of Gradients (HG) and the Histogram of Lines (HL) is evaluated in comparison to the HOG and HOL, respectively on the same dataset. This is to verify our claim on the negative affect of texture and edge orientation in the bag of features approach. Note that the orientation of the gradient was used in exiting techniques such as (Barata et al., 2014a,b; Alfed et al., 2016). The classification results are shown in Tables 7 and 8, respectively.

Table 7: Comparisons of HOG and HG

	C_{HOG}				C_{HG}			
	FN	FP	SE%	SP%	FN	FP	SE%	SP%
SVM	0.15	0.30	84.93	70.17	0.15	0.23	85.05	76.56
AdaBoost	0.08	0.34	91.99	66.29	0.09	0.27	91.34	73.29
ANN (RBF)	0.06	0.16	93.84	83.83	0.05	0.15	94.65	84.6

Table 8: Comparisons of HOL and HL

	C_{HOL}				C_{HL}			
	FN	FP	SE%	SP%	FN	FP	SE%	SP%
SVM	0.15	0.26	84.65	74.17	0.12	0.26	88.26	74.24
AdaBoost	0.08	0.31	92.29	68.56	0.07	0.33	93.15	67.24
ANN (RBF)	0.04	0.19	95.65	80.67	0.05	0.16	95.15	84.11

It is clear that the orientation of gradients and lines in HOG and HOL, respectively, decreases the performance when compared to HG and HL. This can also be noticed in Table 9 where the HOG was combined with HOL and compared to the other combination of HG and HL. Interestingly, one can observe that the combination of HG and HL enhances the accuracy.

Table 9: Combination of texture features

	$[C_{HOG}, C_{HOL}]$				$[C_{HG}, C_{HL}]$			
	FN	FP	SE%	SP%	FN	FP	SE%	SP%
SVM	0.1	0.22	90.43	77.56	0.08	0.23	92.41	77
AdaBoost	0.07	0.18	93.22	82.1	0.04	0.18	95.6	82.43
ANN (RBF)	0.04	0.1	95.76	90.14	0.02	0.08	97.5	91.51

To gauge the performance of the complete system, the previously assessed features, i.e., CVA+ZM, HG and HL are combined as proposed in (15). The corresponding results are depicted in Table 10 where a combination of conventional texture and color features is also considered.

Table 10: Combinations of texture and color features

	$[C_{HOG}, C_{HOL}, C_{Opp+M}]$				$[C_{HG}, C_{HL}, C_{CVA+ZM}]$			
	FN	FP	SE%	SP%	FN	FP	SE%	SP%
SVM	0.12	0.28	88.12	72.39	0.02	0.06	97.6	93.96
AdaBoost	0.06	0.37	93.53	62.83	0.01	0.04	98.8	96.18
ANN (RBF)	0.07	0.15	93.19	85.5	0.01	0.02	99.41	98.18

Obviously, the proposed combination of features significantly outperforms the competing one. This is adopted in the rest of the paper for further analysis and comparisons with state-of-the-art systems.

3.2. Comparison with existing systems

For comparison purposes, the authors have reproduced a number of melanoma skin cancer detection systems that used the bag of features approach, in particular (Barata et al., 2014a), (Barata et al., 2014b), (Alfed et al., 2015) and (Alfed et al., 2016). Furthermore, since the HPH dataset has been widely used in the literature with the same settings, we have collected the results of other related systems as reported in their original papers on 200 dermoscopy images. The reported techniques are listed in Table 11.

Table 11: List of acronyms used for describing existing techniques and strategies

Existing techniques	Acronym
Pigment Network and Bank of directional Filters	**PNBF** (Barata et al., 2012)
Sparse key-point sampling using Harris Laplace	**HLaplace** (Barata et al., 2013)
Color Moments	**CM** (Barata et al., 2014b)
Local color-based features	**LCB** (Barata et al., 2014a)
Color Constancy (shades of gray)	**CC** (Barata et al., 2015)
Scale Adaptive Local Binary Patterns	**LBP – based** (Riaz et al., 2014)
Combination of shape, color and texture features	**SCT** (Abuzaghleh et al., 2015)
Mean color vector-based symmetry features	**MCV** (Ruela et al., 2015)
Pigment network-based Standard deviations	**PNStds** (Alfed et al., 2015)
Detection of Pigment Networks	**DPN** (Eltayef et al., 2017)

Table 12 lists the results of the aforementioned techniques on the HPH dataset (Mendonc et al., 2015). Note that local features normally yield better performance than global features as illustrated in (Barata et al., 2014a). It is also worth mentioning that the combination of color and texture plays a major role in melanoma skin lesion diagnosis. This can be justified by the low performance of the systems that used color features only, in particular (Barata et al., 2015) and (Barata et al., 2014b) or, on the other hand, the systems that used texture features only such as in (Barata et al., 2012) and (Riaz et al., 2014). Obviously, the performance has been improved when color, shape and texture features are combined such as in (Abuzaghleh et al., 2015) or even by combining color and texture features such as the system presented in this work.

Table 12: Performance of existing techniques on the Dermoscopy HPH image database

Technique	FN	FP	SE%	SP%	Acc%
PNBF (Barata et al., 2012)	0.09	0.18	91.1	82.1	86.6
HLaplace (Barata et al., 2013)	0.02	0.14	98	86	92
CM (Barata et al., 2014b)	0.07	0.12	93	88	90.5
LCB (Barata et al., 2014a)	**0**	0.25	**100**	75	87.5
CC (Barata et al., 2015)	0.07	0.24	92.5	76.3	84.4
LBP – based (Riaz et al., 2014)	0.16	0.06	84	94	89
SCT (Abuzaghleh et al., 2015)	**0**	0.08	**100**	91.5	95.75
MCV (Ruela et al., 2015)	0.04	0.17	96	83	89.5
PNStds (Alfed et al., 2015)	0.05	0.08	95.45	92.33	93.89
$[\mathbf{C_{HOG}}, \mathbf{C_{Opp+M}}]$ (Alfed et al., 2016)	0.09	0.15	91	85	88
DPN (Eltayef et al., 2017)	0.08	0.05	92.3	95.0	90.0
This work	0.01	**0.02**	99.41	**98.18**	**98.79**

Overall, the proposed system delivers the third lowest score on the HPH dataset in terms of FN, slightly outperformed by SCT and LCB that seem to detect all test cancer images. However, these two systems perform poorly in terms of FP where a good number of false skin cancer detections were noticed. As a consequence, the proposed system appears more reliable in terms of the overall accuracy, beating the closest competitor by 3% approximately. Finally, 256 standard images have been used from the Dermofit dataset, namely 76 of malignant melanoma and 180 non-melanoma images. The non-melanoma images have been selected from 9 different types of lesions as depicted in Table 13.

Table 13: Categories of lesions and number of selected images used in the Dermofit database

	Dermofit database	
Lesion Type	Total images	Selected images
Actinic Keratosis	45	20
Basal Cell Carcinoma	239	20
Melanocytic Nevus (mole)	231	20
Seborrhoeic Keratosis	257	20
Squamous Cell Carcinoma	88	20
Intraepithelial Carcinoma	78	20
Pyogenic Granuloma	24	20
Haemangioma	97	20
Dermatofibroma	65	20
Malignant Melanoma	76	76

In Table 14, the performance is displayed for the proposed system as well as the competing techniques that use a bag of features approach.

Table 14: Performance of existing techniques using standard images from the Dermofit database

	FN	FP	SE%	SP%	Acc%
LCB (Barata et al., 2014a)	**0.04**	0.28	95.59	72.02	88.27
CM (Barata et al., 2014b)	0.05	0.30	94.62	70.49	87.11
PNStds (Alfed et al., 2015)	0.08	0.27	92.23	73.28	85.94
$[C_{HOG}, C_{Opp+M}]$ (Alfed et al., 2016)	0.19	0.29	81.32	70.86	78.13
This work	**0.04**	**0.15**	**96.04**	**84.78**	**92.96**

For the Dermofit standard database, the proposed system clearly has the upper hand in every aspect and outperforms the competing techniques by more than 4.5% of overall accuracy. It is also worth noting that standard images appear less efficient than dermoscopic images for melanoma skin cancer diagnosis as can be seen from the drop in performance. This can be justified by the fact that dermoscopic images are characterised by more details and texture, hence more discriminative features, than standard images.

4. Conclusion

In this paper, a melanoma skin cancer detection system has been presented. The system relies on a bag of features approach using multiple codebooks where new color and textural features are proposed for describing skin cancer lesions efficiently, namely, the histogram of gradients (HG), the histogram of lines (HL), the 3^{rd} order Zernike moments, and color vector angles. It has been claimed and demonstrated that the orientation information in the conventional histogram of oriented gradients (HOG) and the histogram of oriented lines (HOL) reduces the inter class dissimilarity and consequently decreases the discriminative nature of the respective features in skin cancer detection. The histogram of gradients and the histogram of lines are used separately to create two respective codebooks whereas the color vector angles and the 3rd order Zernike moments are combined for the creation of a third codebook. The automated skin cancer detection system has been assessed on two different datasets, namely the HPH dermoscopy database and the Dermofit standard database. Results have shown that the proposed system can reach 98.79% of overall accuracy on HPH and 92.96% on the standard Dermofit. Compared with related state-of-the-art techniques, the system outperforms its closest competitor by, approximately, 3% on HPH and 4% on Dermofit.

References

Abuzaghleh, O., Barkana, B. D., and Faezipour, M. (2015). Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention. *IEEE Journal of Translational Engineering in Health and Medicine*, 3:1–12.

Alfed, N., Khelifi, F., and Bouridane, A. (2016). Improving a bag of words approach for skin cancer detection in dermoscopic images. In *Proc. IEEE International Conference on Decision and Information Technologies (CoDIT)*, pages 228–232, Saint Julian's, Malta.

Alfed, N., Khelifi, F., Bouridane, A., and Seker, H. (2015). Pigment network-based skin cancer detection. In *Proc. IEEE International Conference on Engineering in Medicine and Biology Society (EMBS)*, pages 7214–7217, Milan, Italy.

Argenziano, G., Fabbrocini, G., Carli, P., Giorgi, V. D., Sammarco, E., and Delfino, M. (1998). 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. *Archives of Dermatology*, 134(12):1563–1570.

Ballerini, L., Fisher, R. B., Aldridge, B., and Rees, J. (2013). A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. In *Color Medical Image Analysis*, pages 63–86. Springer.

Barata, C., Celebi, M. E., and Marques, J. S. (2015). Improving dermoscopy image classification using color constancy. *IEEE Journal of Biomedical and Health Informatics*, 19(3):1146–1152.

Barata, C., Marques, J. S., and Rozeira, J. (2012). A system for the detection of pigment network in dermoscopy images using directional filters. *IEEE Transactions on Biomedical Engineering*, 59(10):2744–2754.

Barata, C., Marques, J. S., and Rozeira, J. (2013). The role of keypoint sampling on the classification of melanomas in dermoscopy images using bag-of-features. In *Proc. Springer Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, pages 715–723, Madeira, Portugal.

Barata, C., Ruela, M., Francisco, M., Mendonca, T., and Marques, J. S. (2014a). Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Systems Journal*, 8(3):965–979.

Barata, C., Ruela, M., Mendonca, T., and Marques, J. S. (2014b). 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*, pages 49–69. Springer.

Chakravorty, R., Liang, S., Abedini, M., and Garnavi, R. (2016). Dermatologist-like feature extraction from skin lesion for improved asymmetry classification in ph2 database. In *Proc. IEEE International Conference on Engineering in Medicine and Biology Society (EMBC)*, pages 3855–3858, USA.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, San Diego, USA.

Do, M. N. and Vetterli, M. (2003). The finite ridgelet transform for image representation. *IEEE Transactions on Image Processing*, 12(1):16–28.

Dony, R. and Wesolkowski, S. (1999). Edge detection on color images using rgb vector angles. In *Proc. IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, volume 2, pages 687–692, Edmonton, Canada.

Eltayef, K., Li, Y., and Liu, X. (2017). Detection of pigment networks in dermoscopy images. In *Proc. IOP International Conference on Communication, Image and Signal Processing (CCISP)*, volume 787, page 012034, Dubai.

Giotis, I., Land, N. M. S., Biehl, M., Jonkman, M. F., and Petkov, N. (2015). MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Systems with Applications*, 42:6578–6585.

Huang, D., Jia, W., and Zhang, D. (2008). Palmprint verification based on principal lines. *Pattern Recognition*, 41(4):1316–1328.

Jia, W., Hu, R. X., Lei, Y. K., Zhao, Y., and Gui, J. (2014). Histogram of oriented lines for palmprint recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(3):385–395.

Khotanzad, A. and Hong, Y. H. (1990). Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):489–497.

Kiani, K. and Sharafat, A. R. (2011). E-shaver: an improved DullRazor ® for digitally removing dark and light-colored hairs in dermoscopic images. *Computers in Biology and Medicine*, 41(3):139–145.

Korotkov, K. and Garcia, R. (2012). Computerized analysis of pigmented skin lesions: a review. *Artificial Intelligence in Medicine*, 56(2):69–90.

Kruk, M., Swiderski, B., Osowski, S., Kurek, J., Slowinska, M., and Walecka, I. (2015). Melanoma recognition using extended set of descriptors and classifiers. *EURASIP Journal on Image and Video Processing*, 2015(1):1–10.

Lee, H. and Chen, Y. P. P. (2015). Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42:5356–5365.

Lee, T., Ng, V., Gallagher, R., Coldman, A., and McLean, D. (1997). Dullrazor®: a software approach to hair removal from images. *Computers in Biology and Medicine*, 27(6):533–543.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157, Kerkyra, Greece.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Mendonc, T., Ferreira, P. M., Marcal, A. R. S., Barata, C., Rocha, J. S. M. J., and Rozeira, J. (2015). *PH2: A Public Database for the Analysis of Dermoscopic Images*. PhD thesis, Universidade do Porto.

Nachbar, F., Stolz, W., Merkle, T., Cognetta, A. B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., and Plewig, G. (1994). The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, 30(4):551–559.

Nock, R. and Nielsen, F. (2004). Statistical region merging. *IEEE Transactions on pattern analysis and machine intelligence*, 26(11):1452–1458.

Oliveira, R. B., Marranghello, N., Pereira, A. S., and Tavares, J. M. (2016). A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Systems with Applications*, 61:53–63.

Riaz, F., Hassan, A., Javed, M. Y., and Coimbra, M. T. (2014). Detecting melanoma in dermoscopy images using scale adaptive local binary patterns. In *Proc. IEEE Interna-*

tional Conference on Engineering in Medicine and Biology Society (EMBS), pages 6758–6761, Chicago, USA.

Ruela, M., Barata, C., Marques, J. S., and Rozeira, J. (2015). A system for the detection of melanomas in dermoscopy images using shape and symmetry features. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, pages 1–11.

She, Z., Duller, A., Liu, Y., and Fish, P. J. (2006). Simulation and analysis of optical skin lesion images. *Skin Research and Technology*, 12(2):133–144.

Silveira, M., Nascimento, J. C., Marques, J. S., Marcal, A. R. S., Mendonca, T., Yamauchi, S., Maeda, J., and Rozeira, J. (2009). Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):35–45.

Tanaka, T., Yamada, R., Tanaka, M., Shimizu, K., and Oka, H. (2004). A study on the image diagnosis of melanoma. In *Proc. IEEE International Conference on Engineering in Medicine and Biology Society (EMBS)*, volume 1, pages 1597–1600, San Francisco, Ca, USA.

Tang, Z., Dai, Y., Zhang, X., Huang, L., and Yang, F. (2014). Robust image hashing via colour vector angles and discrete wavelet transform. *The Institution of Engineering and Technology (IET), Image Processing*, 8(3):142–149.

Tobias, O. J. and Seara, R. (2002). Image segmentation by histogram thresholding using fuzzy sets. *IEEE Transactions on Image Processing*, 11(12):1457–1465.

Vasconcelos, M., Rosado, L., and Ferreira, M. (2015). A new color assessment methodology using cluster-based features for skin lesion analysis. In *Proc. IEEE International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 373–378, Opatija, Croatia.

Zhao, Y., Wang, S., Zhang, X., and Yao, H. (2013). Robust hashing for image authentication using zernike moments and local features. *IEEE Transactions on Information Forensics and Security*, 8(1):55–63.

Zhu, S. C. and Yuille, A. (1996). Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900.