A Bag-of-Features Approach for the Classification of Melanomas in Dermoscopy Images: The Role of Color and Texture Descriptors

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

Abstract The identification of melanomas in dermoscopy images is still an up to date challenge. Several Computer Aided-Diagnosis Systems for the early diagnosis of melanomas have been proposed in the last two decades. This chapter presents an approach to diagnose melanomas using Bag-of-features, a classification method based on a local description of the image in small patches. Moreover, a comparison between color and texture descriptors is performed in order to assess their discriminative power. The presented results show that local descriptors allow an accurate representation of dermoscopy images and achieve good classification scores: Sensitivity = 93 % and Specificity = 88 %. Furthermore it shows that color descriptors perform better than texture ones in the detection of melanomas.

Keywords Melanoma diagnosis · Dermoscopy · Bag-of-features · Feature extraction · Feature analysis · Color features · Texture features

Introduction

Dermoscopy is a widely used microscopy technique for the in-vivo observation of skin lesions. A magnification instrument is used to increase the size of the lesion and a liquid (oil, alcohol or water) is placed on top of the lesion prior to the

C. Barata (✉) · M. Ruela · J. S. Marques
Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa, Portugal
e-mail: ana.c.fidalgo.barata@ist.utl.pt

M. Ruela
e-mail: margarida.ruela@ist.utl.pt

J. S. Marques
e-mail: jsm@isr.ist.utl.pt

T. Mendonça
Faculdade de Ciências da Universidade do Porto, Porto, Portugal
e-mail: tmendo@fc.up.pt

J. Scharcanski and M. E. Celebi (eds.), *Computer Vision Techniques for the Diagnosis of Skin Cancer*, Series in BioEngineering, DOI: 10.1007/978-3-642-39608-3_3,

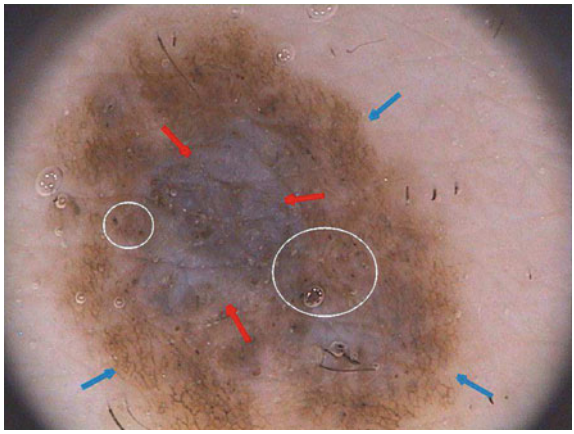

Fig. 1 Melanoma with specific dermoscopic features: blue-whitish veil (*red arrows*); pigment network (*blue arrows*), dots and globules (*white circles*)

observation to eliminate surface reflection. This step makes the cornified layer of the skin translucent, allowing a better visualization of several pigmented structures located within the epidermis, dermis and dermoepidermal junction [3]. Several magnification instruments are currently used by dermatologists: dermatoscope, stereomicroscope or a digital acquisition system. The later allows the attainment of dermoscopy images, that can be processed and analyzed by a Computer Aided-Diagnosis (CAD) system.

The diagnosis of pigmented skin lesions using dermoscopy is based on medical algorithms: ABDC rule [47], 7-point checklist [2] and Menzies' method [29]. All of these methods have in common a set of dermoscopic criteria that can be divided in two groups. The first group is called global features and allows a preliminary and quick categorization of a skin lesion. Global features are a set of patterns (reticular, cobblestone, globular, parallel, etc) that can be found in different pigmented skin lesions. The other group of dermoscopic criteria are the local features (pigment network, dots and globules, streaks, pigmentation related structures, vascular pattern, etc). These features are sometimes called the letters of the dermoscopic alphabet since they are the cues that allow a final diagnosis of the lesion (melanoma or not) [3]. Figure 1 illustrates some of the local dermoscopic features.

Both global and local features play an important role in the diagnosis of melanomas. Some dermatologists perform an analysis of skin lesions using as reference only the global dermoscopic features. This global evaluation method is called pattern analysis and has received some attention in the skin research area, such as [1, 36, 40], which try to reproduce the medical analysis. The published works focus on the identification of the different patterns but do not perform a diagnosis of a skin lesion. However, it is undeniable that local features are the backbone of the common medical algorithms since ABCD rule, 7-point checklist and Menzies' method use these features and their properties (shape or number of colors) to score a skin lesion, thus diagnosing it as melanoma or not. There are several studies which focus on detecting one or more of these dermoscopic criteria, such as pigment network

[6, 37], irregular coloration [10, 31, 45], irregular streaks [35] and granularity [46]. However, as far as the authors know only one study combines a set of detectors and the 7-point checklist algorithm in a CAD system to perform a diagnosis using dermoscopy images [15]. Two reviews on state-of-the art methods can be found in [11, 25].

Most of the CAD systems found in literature use a different procedure, following a pattern recognition approach to classify dermoscopy images [9, 18, 21, 34]. These works have successfully exploited a global representation of the lesion using features inspired by the ABCD rule (color, shape, texture and symmetry). Most of the extracted features are able to perform a good description of the lesion regarding its shape and global color distribution. However, localized texture and color patterns associated to differential structures (e.g., pigment network, dots, streaks or blue-whitish veil) might be missed since a global analysis is being performed. To overcome this situation, this chapter describes a different approach for the analysis of dermoscopy images. Since experts usually try to characterize local structures in the image, the described strategy will try to mimic this behavior and represent the image by a set of local features, each of them associated to a small region in the image. The local features used describe the texture and color of each region and a comparative study between the two types of descriptive features is performed, in order to assess their degree of discrimination.

Bag-of-Features

The description of an image with local features have been successfully used in several complex image analysis problems, such as scene recognition and object-class classification [22, 23, 27, 42, 44, 50]. The used approach is called Bag-of-Features (BoF) [42, 44] and it is inspired by the bag-of-words (BoW) [5], which is a well known text retrieval algorithm. The procedure used by BoW to model documents evolves in three different steps. The first step consists of parsing the documents of the dataset into words, i.e., dividing the documents to smaller components. Images can also be sampled into smaller regions (patches). Two sampling strategies are commonly used in BoF: sparse and dense sampling. Sparse sampling is performed by detecting a set of informative keypoints (e.g., corners) and their respective support regions (square patches). This detection can be done using one or more of the several detectors proposed in literature (e.g. Difference of Gaussian [28] or Harris-Laplace [30]). A comparative study between the six most popular keypoint detectors can be found in [22]. For dense sampling it is assumed that each keypoint is the node of a regular grid defined in the image domain. The patches associated with the keypoints are extracted by sampling uniformly over the grid. Both sampling methods have been used in different works and a comparison between the two strategies was performed by van de Sande et al. [38]. Their experimental results showed that dense sampling outperformed sparse sampling. The BoF approach proposed in this work uses dense

sampling to extract the patches from a given image and only patches whose area is more than 50 % inside the lesion are considered.

The second step in BoW document analysis is to represent each word by its stem. The equivalent for the image analysis case is to represent each patch by a feature vector $x_i \in \mathbb{R}^n$. Different features can be used to locally describe the patches. This chapter focus on two specific kinds of features, color and texture, which will be addressed in section "Local Features". N square patches are extracted from each image I on the dataset. Therefore, a family of local features will be associated with I as follows

$$F = \{x_1, \ldots, x_N\}, \quad x_i \in \mathbb{R}^n. \tag{1}$$

The last step of both BoW and BoF corresponds to the training process. In the first, each discriminative word receives an unique label. Very common words, which occur in most documents, are rejected and do not receive a label. This process can be seen as the creation of a dictionary of representative words. Then, each document is analyzed separately and its discriminative words are compared with the ones from the dictionary. From this comparison will result an histogram of the frequency of occurrence of the dictionary words within the document. This histogram will represent the document and will be used to compare different documents and assess their degree of similarity. Reproducing this histogram representation in the BoF case requires some extra effort. First, assuming that there is a dataset of L images, this dataset has associated with it set of all the extracted local features

$$\mathscr{F} = \bigcup_{k=1}^{L} F^{(k)}. \tag{2}$$

In practice, the set $\mathscr{F}$ has many thousands (or even millions) of feature vectors. Therefore, in order to obtain a visual dictionary (analogous to the dictionary of BoW), this set has to be approximated by a collection of prototypes $c_1, \ldots, c_K$, called *visual words*. The *visual words* are obtained using a clustering algorithm (in this work K-means is used). After obtaining a visual dictionary, all feature vectors in the training set are classified in the nearest *visual word* and a label

$$l_i^{(k)} = \arg\min_j \|x_i^{(k)} - c_j\|, \tag{3}$$

which identifies a specific *visual word*, assigned to each feature vector $x_i^{(k)}$. The final step is to characterize each image $I^{(k)}$ by a histogram of visual words frequency

$$h^{(k)}(l) = \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \delta(l_i^{(k)} - l), \tag{4}$$

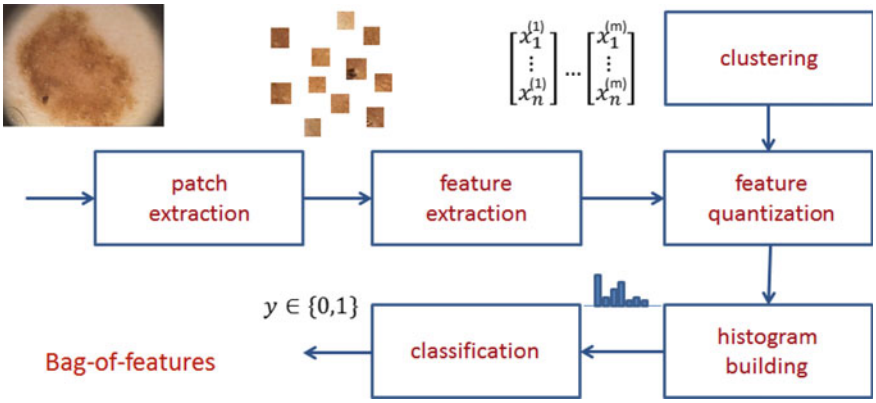

Fig. 2 Block diagram of the BoF classification system

where $\delta(.)$ denotes the Kronecker delta ($\delta(x) = 1$, if $x = 0$; $\delta(x) = 0$, otherwise). As in BoW, this histogram will act as the feature vector that describes the image and the set of L feature vectors obtained this way will be used to train a classifier.

For each new image I to be classified the process is similar to the one described previously. The image is sampled and local features are extracted. Then, the local features are compared with the dictionary of visual words obtained in the training phase and, finally, the histogram of visual words frequency is computed. The image is classified using the computed histogram and the classifier learned using the training set.

All the steps of the BoF strategy described previously are summarized in Fig. 2.

There are several factors that can impact the performance of BoF. Following the blocks sequence on Fig. 2 these factors are: the size of the regular grid (δ) used in the patch extraction step, the type and quantity of extracted features, the size of the dictionary (K), and the classification algorithm used. All these factors are thoroughly analyzed in this work. Several values for δ and K are tried and three classification algorithms with different properties are tested: k-nearest neighbor (kNN) [16], AdaBoost [17] and support vector machines (SVM) [8, 12].

Local Features

The local dermoscopic criteria used by dermatologists to diagnose skin lesions can be represented by two different kinds of image features: texture and color features. Local dermoscopic structures such as pigment network, dots and streaks can be characterized by texture features since these features represent the spatial organization of intensity in an image, allowing the identification of different shapes. Color features describe the color distribution, thus they are able to characterize particular pigmented regions such as blue-whitish veil or regression areas.

In theory both of these features provide a good description of the extracted patches (see Fig. 2) and both play an important role in the final classification. One of the main objectives of this chapter is to determine if the previous hypothesis are correct. This objective is accomplished by assessing the performance of color and texture features separately and by combining both of them. Moreover, since both color and texture features can be extracted using different types of descriptors, a comparison between some of them is also performed. The several texture and color descriptors tested are described in the next sections.

Texture Descriptors

Texture features characterize the intensity of an image. Therefore, it is necessary to convert the original RGB image into a gray level one before extracting texture descriptors. This is done by selecting the color channel with the highest entropy [41].

Texture descriptors can be divided into several categories depending on the methodology used. This chapter focus on three different methods: statistical, signal processing and gradient [33]. In statistical methods, the features are extracted by computing neighbor pixel statistics. A very well known method for computing these statistics is the gray level co-occurrence matrix (GLCM) proposed by Haralick et al. [20]. This matrix stores the relative frequencies of gray level pairs of pixels at a certain relative displacement and can then be used to compute several statistics which will be the elements of the feature vector. The results presented in this chapter are obtained using five of the most common statistics: contrast, correlation, homogeneity, energy and entropy. The performance of these features is directly related with GLCM, since it has been already proved that the performance of a classification system is influenced by the number of gray levels (G) used as well as the way of combining the orientations of the nearest neighbors [13]. Therefore, several values of G and two ways of combining the orientations (average GLCM versus four GLCM) are tested, according to what is proposed in [13].

Signal processing approaches have in common three sequential steps. First, the image $I(x, y)$ is convolved with a bank of N filters, with a certain impulse response $h_i(x, y), \quad i = 1, \ldots, N$. Then, an energy measure of the output $J_i(x, y) \quad i = 1, \ldots, N$, is performed

$$E_i = \sum_x \sum_y |J_i(x, y)|. \tag{5}$$

Finally, the energy content is used to computed statistics that are the components of the feature vector [33]. The two statistics computed in this chapter are the mean μ_i and standard deviation σ_i

$$\mu_i = \frac{E_i}{M}, \quad \sigma_i = \sqrt{\frac{E_i - \mu_i}{M}}, \tag{6}$$

where M is the number of pixels (x, y) in J_i.

Several filter banks can be found on literature [33]. This chapter compares two of the most well known: Laws [26] and Gabor [4] filter masks. The filter masks proposed by Laws [26] have been widely used for texture characterization. These masks can have a dimension 3×3 or 5×5 and result from convolving two of of the five possible 1-D kernels. Each 1-D kernel focus on specific textural characteristics like edges, waves or ripples. In this chapter only three kernels will be used: L, which computes the average grey Level, E that extracts Edges (describe linear structures, such as pigment network) and S that extracts Spots (describe circular structures, such as dots). The 1-D kernel values are the following: $L_3 = [1\ 2\ 1]$, $E_3 = [1\ 0\ -1]$, $S_3 = [1\ -2\ 1]$, $L_5 = [1\ 4\ 6\ 4\ 1]$, $E_5 = [-1\ -2\ 0\ 2\ 1]$ and $S_5 = [-1\ 0\ 2\ 0\ -1]$. All the possible combinations of 1-D kernels are considered, thus the filter bank has a dimension $N = 9$. Since it is not known the dimension of the masks that leads to the best results, both 3×3 and 5×5 filter banks are tested.

Gabor filters have been used for texture classification [4] and edge detection [19]. Therefore, they can be used to characterize dermoscopic structures that have a linear shape (e.g. pigment network or streaks). The impulse response of a Gabor filter is the following

$$h_i(x, y) = e^{\frac{\tilde{x}^2 + \gamma^2 \tilde{y}^2}{2\sigma_G^2}} \cos\left(2\pi \frac{\tilde{x}}{\lambda} + \varphi\right), \tag{7}$$

where γ is an aspect ratio constant. σ_G is the standard deviation, λ is the wavelength, φ is the phase of the filter and $(\tilde{x}, \tilde{y})$ are obtained from rotating (x, y) as follows [19]

$$\tilde{x} = x \cos\theta_i + y \sin\theta_i, \qquad \tilde{y} = -x \sin\theta_i + y \cos\theta_i. \tag{8}$$

The angle amplitude $\theta_i \in [0, \pi]$ determines the orientation of the filter h_i and the step between two consecutive filters is $\frac{\pi}{N_\theta}$, where N_θ is the number of filters in the filter bank [19]. This descriptor depends on several parameters. In this chapter two of them are varied: $\sigma_G \in \{1, 2, \ldots, 5\}$ and $N_\theta \in \{2, 3, \ldots, 10\}$. All the others are kept constant and equal to: $\gamma = 0.5$, $\varphi = 0\ rad$ and $\frac{\tilde{x}}{\lambda} = 0.56$ [19].

Gradient features such as gradient histograms have also been successfully used in several classification problems [14, 28]. In this work, two gradient histograms are used: amplitude and orientation. The image gradient $g(x, y) = [g_1(x, y)\ \ g_2(x, y)]$ is computed using Sobel masks. Then, gradient magnitude and orientation are respectively computed as follows

$$\| g(x, y) \| = \sqrt{g_1(x, y)^2 + g_2(x, y)^2}, \phi(x, y) = \tan^{-1}\left(\frac{g_2(x, y)}{g_1(x, y)}\right). \tag{9}$$

Finally, the histograms of gradient amplitude and orientation are obtained

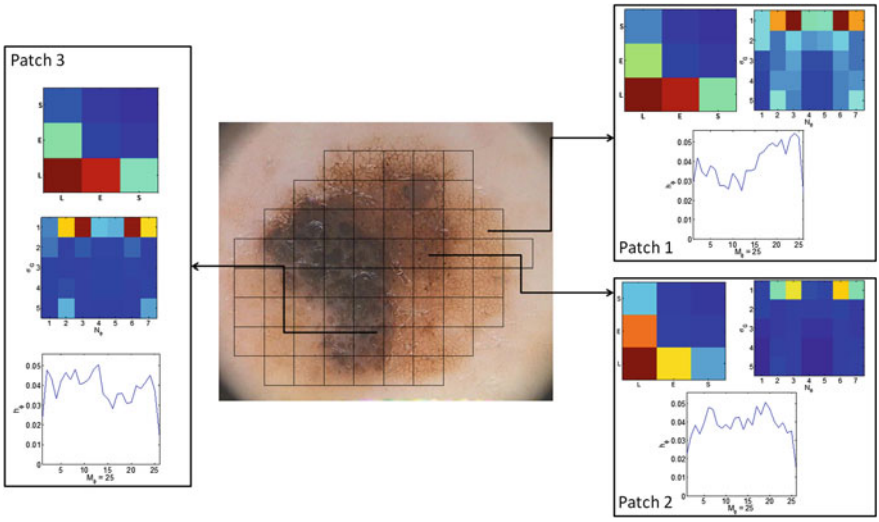

Fig. 3 Texture features for three different 60×60 patches: energy content for Laws (5×5 masks) and Gabor ($N_\theta = 1, 2, \ldots, 7$ and $\sigma_G = 1, 2, \ldots, 5$) filters and histogram of the gradient phase ($M_\phi = 25$)

$$h_a(i) = \frac{1}{N} \sum_x \sum_y b_i(\|g(x, y)\|), \qquad i = 1, \ldots, B_a,$$
$$h_\phi(i) = \frac{1}{N} \sum_x \sum_y \tilde{b}_i(\phi(x, y)), \qquad i = 1, \ldots, B_\phi, \tag{10}$$

where N is the number of pixels inside the patch and B_a, B_ϕ are the number of bins of the magnitude and orientation histograms, respectively. Finally, $b_i(.)$, $\tilde{b}_i(.)$ are the characteristic functions of the ith histogram bin

$$b_i(a) = \begin{cases} 1 \text{ if} a \text{belongs to the} i\text{th amplitude bin} \\ 0 \text{ otherwise} \end{cases}$$
$$\tilde{b}_i(\phi) = \begin{cases} 1 \text{ if} \phi \text{belongs to the} i\text{th orientation bin} \\ 0 \text{ otherwise} \end{cases}. \tag{11}$$

The parameter varied for both gradient features is the number of bins of the histograms ($B_a \in \{15, 25, 35, 45\}$ and $B_\phi \in \{15, 25, 35, 45\}$).

Figure 3 shows some of the extracted texture features for three different patches within the same lesion. The exemplified patches were selected in order to include a specific dermoscopic structure: pigment network (patch 1), dots (patch 2) and globules (patch 3). It is clear that the extracted descriptors (Laws, Gabor and h_ϕ) are different for each patches, which demonstrates that different dermoscopic structures have different textural properties and, therefore, must be described separately.

Color Descriptors

Several color descriptors, such as histograms and mean color, have been used in object and scene recognition problems [39]. The descriptors are usually computed over one or more color spaces like RGB, HSV/I [49], CIE La*b* and L*uv [49] and the biologically inspired opponent color space (Opp) [7]. These six color spaces have different properties, thus they might provide different information for the melanoma classification problem addressed in this chapter. For this reason the six previous color spaces are tested.

For each color space a set of three histograms is computed (one for each of the three color components). For each patch, the histogram associated with the color channel $I_c,\ c \in \{1, 2, 3\}$ is given by

$$h_c(i) = \frac{1}{N} \sum_{x,y} b_c(I_c(x, y))\ i = 1, \ldots, B_c, \tag{12}$$

where N is the number of pixels inside the patch, i is the histogram bin, B_c is the number of bins and $b_c(.)$ is the characteristic function of the ith bin

$$b_c(I_c(x, y)) = \begin{cases} 1 \text{ if} I_c(x, y) \text{belongs to the} i\text{th color bin} \\ 0 \text{ otherwise} \end{cases}. \tag{13}$$

The bins are defined by dividing the color component range into intervals with the same width. For all histograms, the number of intervals $B_c \in \{15, 25, 35, 45\}$ is a tested parameter.

Another common color descriptors are color moments [48]. Color moments result from assuming that the distribution of color in an image can be seen as a probability distribution. Since probability distributions are usually characterized by a set of unique moments, they can be used as color features. The general definition of the $1st$ order color moment for the color channel $I_c,\ c \in \{1, 2, 3\}$ is the following

$$M_c^1 = \frac{\sum_x^N I_c(x)}{N}, \tag{14}$$

where N is the number of pixels inside the patch. Higher order (p) color moments are defined by [48]

$$M_c^p = \left(\frac{\sum_x^N (I_c(x) - M_c^1)^p}{N} \right)^{\frac{1}{p}}. \tag{15}$$

The first three order moments are used in this chapter. These moments correspond to: mean (M^1), standard deviation (M^2) and skewness (M^3). Therefore, each patch will be represented by a total of nine color moments (three for each color component).

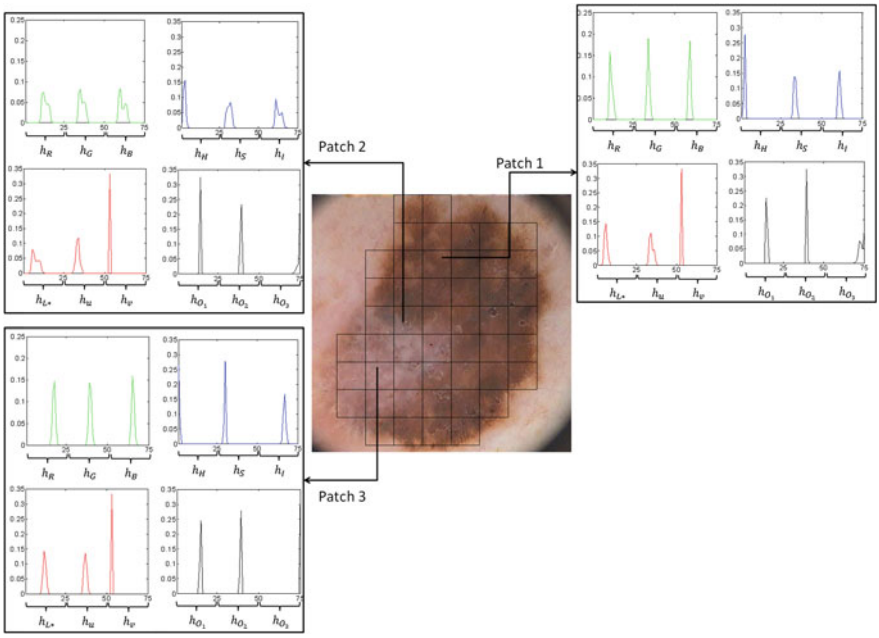

Fig. 4 Color histograms ($M_c = 25$) for three different 60×60 patches: h_{RGB} (*green*), h_{HSI} (*blue*), h_{L^*uv} (*red*) and h_{Opp} (*black*)

Figure 4 exemplifies some of the extracted color histograms (RGB, HSI, L*uv and Opp) for three different patches. Each patch was selected in order to represent a different color section of the lesion: patch 1 was extracted from the brown region, patch 2 was extracted from the brown-white transition region and patch 3 was extracted from the white region. As in the texture features case, the three feature vectors are different, thus each color region is characterized differently.

Experimental Results

The proposed method was evaluated with a dataset of 176 dermoscopy images (25 melanomas and 151 nevi). These images were taken during clinical exams performed at Hospital Pedro Hispano, Matosinhos, with a digital acquisition system that allows a magnification of $20\times$. Images are stored in BMP and $JPEG$ formats and their average resolution is 573×765. Each image was classified by an experienced dermatologist as *melanoma* or *non-melanoma* (ground truth label).

The evaluation metrics used are the Sensitivity (SE) and Specificity (SP). These two measures are combined in a cost function (C)

$$C = \frac{c_{10}(1 - SE) + c_{01}(1 - SP)}{c_{10} + c_{01}}, \tag{16}$$

where c_{10} is the cost of an incorrectly classified melanoma and c_{01} is the cost of an incorrectly classified non-melanoma. This cost function represents the trade-off between SE and SP. In this chapter it is assumed that an incorrect classification of a melanoma is worse and, therefore, the classification error costs are defined as: $c_{10} = 1.5c_{01}$ and $c_{01} = 1$. The selected classifiers are those which achieve the lowest values of C.

Since the dataset is small, the different possibilities are tested using a stratified 10-fold cross validation method. Both classes were evenly distributed by the ten folds. To decrease the impact of class unbalance, local features associated with each melanoma in the different training sets were repeated. Gaussian noise ($w \sim N(0, \sigma_n^2)$, with $\sigma_n = 0.0001$) was added to each repeated local features to prevent exact matching between feature vectors on the training set.

BoF depends on several parameters. The local feature extraction process relies on the size of the patches (δ) while the classification process depends both on the size of the codebook (K) and the classification algorithm used. The best δ is searched in the set $\{20, 40, \ldots, 100\}$ and K in the set $\{100, 200, 300\}$. Each one of the tested classification algorithms depends on several parameters as well. In the kNN case, the parameters tested are the number of neighbors ($k \in \{5, 7, \ldots, 25\}$) and the distance d used to compare the feature vectors. For two vectors $\mathbf{x}$ and $\mathbf{y}$ these distances are computed as follows

- **Euclidean**

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| \tag{17}$$

- **Histogram Intersection**

$$d(\mathbf{x}, \mathbf{y}) = \sum_i min(x_i, y_i) \tag{18}$$

- **Kullback-Leibler**

$$d(\mathbf{x}, \mathbf{y}) = \sum_i \log\left(\frac{y_i}{x_i}\right) y_i \tag{19}$$

For AdaBoost, the parameter tested is the number of weak classifiers ($W \in [2, 300]$). SVM is tested using the BoF default kernel χ^2, defined for two vectors $\mathbf{x}$ and $\mathbf{y}$ as

$$Kernel_{\chi^2} = e^{-\rho d_{\chi^2}(\mathbf{x},\mathbf{y})} \tag{20}$$

where width parameter ρ is searched in the set $\{2^{-9}, 2^{-8}, \ldots, 2^8, 2^9\}$ and

$$d_{\chi^2}(\mathbf{x}, \mathbf{y}) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i} \tag{21}$$

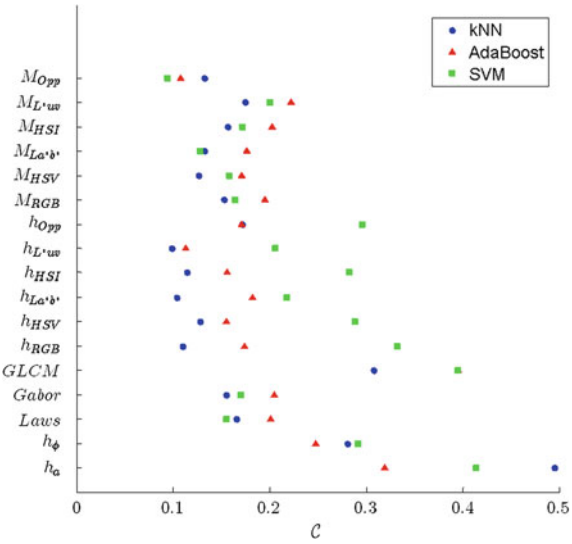

Fig. 5 Best cost results (x axis) for single features using kNN (*blue circle*), AdaBoost (*red triangle*) and SVM (*green square*)

The optimal parameters and results are computed using a total of more than 4,90,000 possible combinations.

Single Features Comparison for the Best Configuration of Classifiers

Figure 5 shows the cost results obtained for the different local features and classifiers. These results show that good classification scores can be achieved using single features (e.g., $C = 0.094$ for M_{Opp}, $C = 0.099$ for h_{L^*uv} and $C = 0.104$ for $h_{La^*b^*}$).

For texture features, the two best descriptors are the signal processing ones. Both Laws and Gabor texture descriptors achieve promising results: $C = 0.155$ for Gabor and $C = 0.166$ for Laws. GLCM and gradient features achieve worse results, which suggests that filter descriptors provide more discriminative information regarding local dermoscopic features.

Color space histograms outperform the corresponding color moments for kNN and AdaBoost classifiers. In the case of SVM, color moments perform much better than their corresponding histograms and the best single descriptor classification result is achieved in this case: M_{Opp}, $C = 0.094$, that corresponds to $SE = 94\,\%$ and $SP = 88\,\%$. The best results achieved with color features outperform those obtained with texture features, which suggest that the former are more discriminative.

Table 1 shows the performance measures and the best configurations for some of the best texture and color features. Good results are achieved both with kNN

Table 1 Classification results and configurations for best texture and color features

Features	SE (%)	SP (%)	C	Feature parameters	Classifier parameters
Laws	100	61	0.155	$\delta = 40\ 5 \times 5$	$K = 100$, SVM: $\rho = 2^{-5}$
Gabor	98	64	0.155	$\delta = 80$, $\sigma_G = 4$ $N_\theta = 5$	$K = 100$, kNN: $k = 19$, Histogram intersection
h_{L^*uv}	100	75	0.099	$\delta = 80$, $B_c = 15$	$K = 300$, kNN: $k = 13$, Kullback-Leibler
$h_{La^*b^*}$	93	85	0.104	$\delta = 80$, $B_c = 25$	$K = 300$, kNN: $k = 19$, Kullback-Leibler
M_{Opp}	93	88	0.094	$\delta = 40$	$K = 100$, SVM: $\rho = 2^{-4}$

and SVM and, for some features, both classifiers lead to very similar results (see Fig. 5). However, kNN appears to achieve the best overall results. It is interesting to notice that for Gabor, h_{L^*uv} and $h_{La^*b^*}$ the best comparative distance are the statistical ones: Kullback-Leibler or Histogram Intersection. This occurrence is also noted for the other tested features and can be explained by the fact that the actual features provided for training and classification are the histograms of *visual words* frequencies, i.e., distributions.

Fusion of Color and Texture Features

Combining different descriptors of the same class may improve the results. To test this hypothesis the two best texture descriptors (Gabor and Laws) and the two types of color descriptors (moments and histograms) were combined. The fusion strategy used is early fusion where the feature vectors are concatenated into a single one [43]. Figure 6 shows the results achieved for each pair of descriptors using kNN and SVM (similar performance is achieved with AdaBoost). The results for the pairs color moments/histograms are identified with the respective color space, whilst the pair Gabor/Laws is labeled Tex. Finally, the performance achieved with each descriptor of the pair is also represented in the graphics (red asterisks). These results show that the fusion of descriptors can either improve the overall results (more evident on the SVM case) or improve the results when compared with the worst descriptor of the pair. Moreover, the fusion of descriptors approximate the results achieved with color and texture descriptors by significantly improving the last ones.

Table 2 shows the best results achieved for each pair as well as the configurations that led to those results. It is interesting to notice that the best global results are achieved with RGB despite the drawbacks of this color space. Nonetheless, this result is worse than the one achieved with the best single descriptor (see Table 1).

Combining different classes of descriptors is an usual approach in BoF and other pattern recognition methods. In this chapter, the best color and texture descrip-

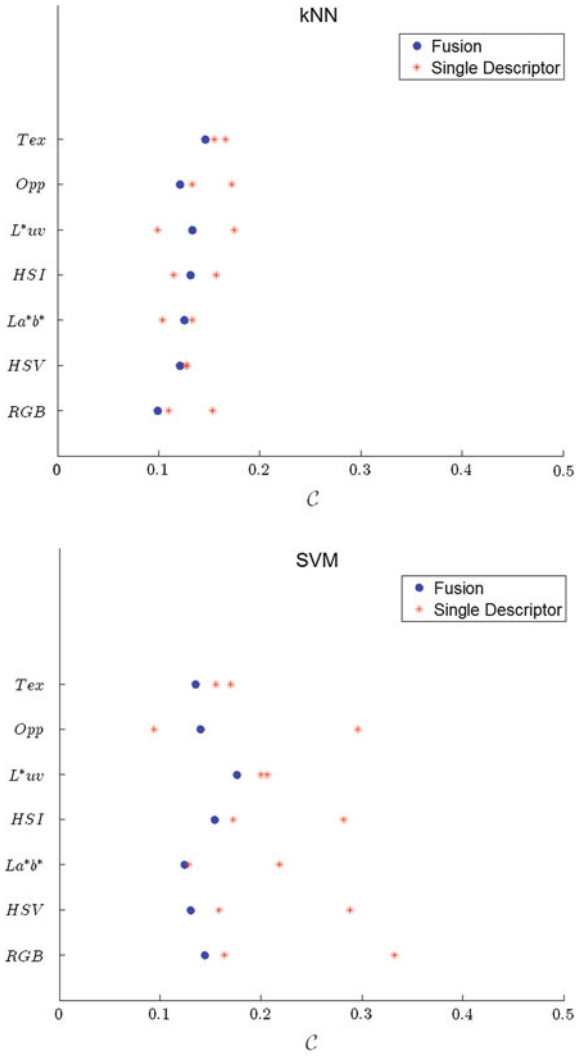

Fig. 6 Early fusion cost results (x axis) for kNN (*top*) and SVM (*bottom*). The performance of each descriptor of the pair is also shown (*red asterisk*)

tors were combined. Due to the different properties of the feature sets and to the large dimension of the feature vector that would result from an early fusion, a late fusion strategy was applied in this case [43]. In this method, the final decision is made by combining the outputs of different classifiers (in this case one classifier for color and other for texture descriptors). The SVM classifier trained using the M_{Opp} (see Table 1) descriptor was combined with two classifiers that achieved similar classification results: kNN ($C = 0.145$) and SVM ($C = 0.135$), both trained using the early fusion of Laws and Gabor descriptors (see Fig. 6). Late fusion strategies use the posteriori probability of each class and one of several possible rules [24] to make

Table 2 Classification results and configurations for descriptors fusions

Fusion	SE (%)	SP (%)	C	Feature parameters	Classifier parameters
Tex	91	79	0.135	$\delta = 40, 5 \times 5$ $\sigma_G = 1$ $N_\theta = 5$	K = 100, SVM: $\rho = 2^{-3}$
Opp	96	77	0.121	$\delta = 20$, $B_c = 25$	$K = 100$, kNN: $k = 15$, Histogram intersection
L^*uv	93	75	0.141	$\delta = 20$, $B_c = 15$	$K = 200$, kNN: $k = 9$, Kullback-Leibler
HSI	100	67	0.131	$\delta = 60$, $B_c = 25$	$K = 300$, kNN: $k = 23$, Histogram intersection
La^*b^*	89	86	0.124	$\delta = 80$, $B_c = 25$	$K = 100$, SVM: $\rho = 2^{-3}$
HSV	98	73	0.121	$\delta = 60$, $B_c = 45$	$K = 300$, kNN: $k = 11$, Histogram intersection
RGB	100	75	0.099	$\delta = 80$, $B_c = 45$	$K = 200$, kNN: $k = 15$, Histogram intersection

the final decision. For SVM, the probabilities are computed using the Platt's method [32] while for kNN, the posterior probabilities are computed as follows:

$$P(w|x) = \frac{k_w}{k}, \tag{22}$$

where w represents the class that can be either 0 or 1, x is a pattern to be classified and k_w is the number of patterns amongst the total number of neighbors k that belong to class w. The combination is computed using the Sum-Rule [24].

The best fusion result was achieved using the kNN classifier trained with texture descriptors combined with the M_{Opp}: $C = 0.097$, SE $= 96\,\%$, SP $= 82\,\%$. Since the color descriptor performs much better than the texture descriptors (Gabor and Laws), it is understandable that the performance of the fusion is slightly inferior to the color descriptor alone.

Visual Words Analysis

The analysis of the identified *visual words* can provide important information regarding the most distinctive dermoscopic structures. Therefore, a simple study of the *visual words* that lead to the best results is performed in this chapter.

This study is done using the information of local color features M_{Opp}, using the best configuration for these features (see Table 1). The average *visual words* histograms for melanomas and non-melanomas were computed using all the histograms of the dataset. Figure 7 shows the obtained average histograms. These two

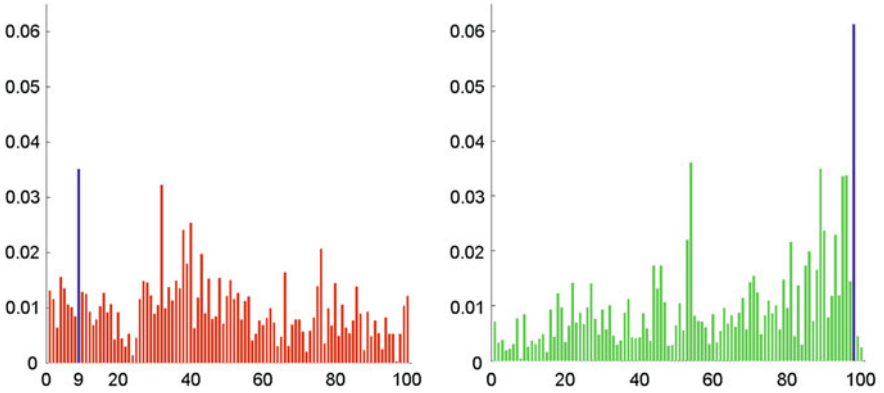

Fig. 7 Average visual words histograms for melanoma (*red*) and non-melanoma (*green*), obtained using the local color features M_{Opp}. Most frequent *visual word* is highlighted (*blue*)

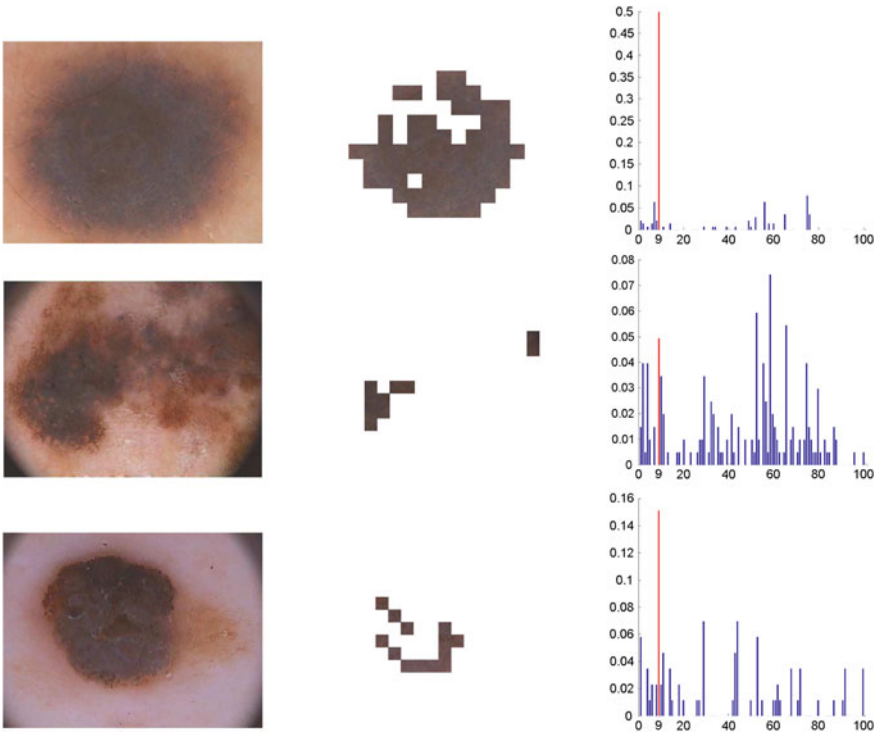

Fig. 8 Example of the most selected color visual word in melanomas: melanomas (*left*); 9th bin visual word (*mid*); visual words histograms, with the 9th bin visual word highlighted (*right*)

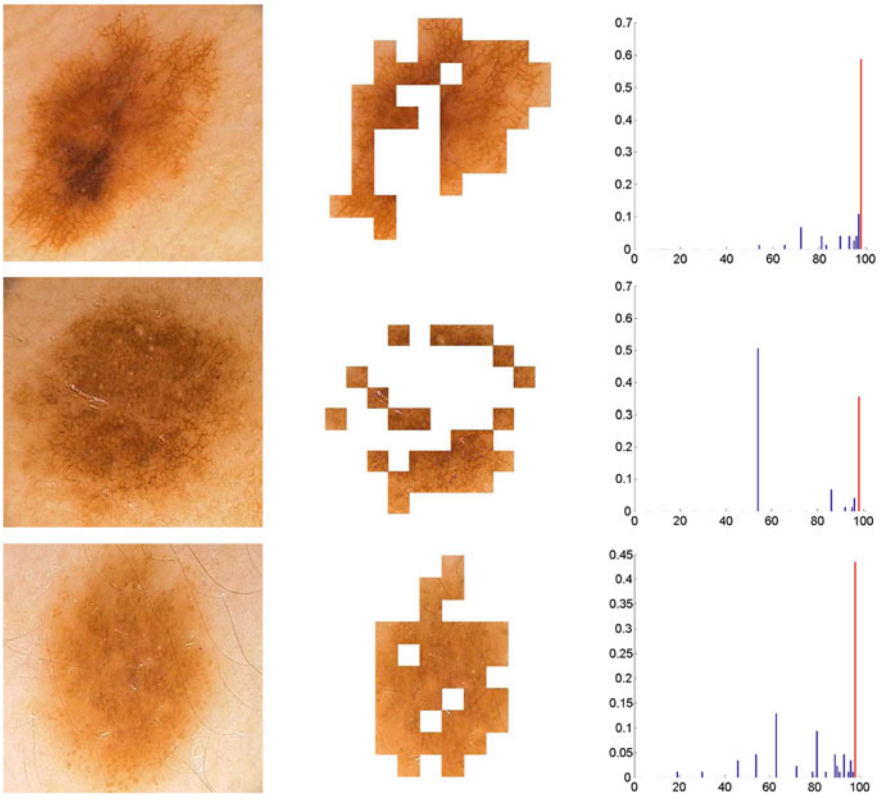

Fig. 9 Example of the most selected color visual word in non-melanomas: non-melanomas (*left*); 98th bin visual word (*mid*); visual words histograms, with the 98th bin visual word highlighted (*right*)

histograms are significantly different and it is interesting to notice that there are some *visual words* which are more common than others.

The next step is to select the most frequent *visual word* for melanomas (see Fig. 6, highlighted 9th bin) and assess the patches associated with this word. Figure 8 shows three melanomas from the dataset, the patches associated with the *visual word* and the corresponding histograms. Although each lesion is described by a different histogram, the same *visual word* (9th bin) is present in all of them and is one of the most frequent. The detected patches are extracted from a blue-gray region (blue-whitish veil), which is one of the atypical pigmentations associated with melanoma [3].

A similar analysis can be performed using the most frequent *visual word* of non-melanomas (see Fig. 7). As before, the patches identified as being this specific visual word were extracted from examples of benign lesions (see Fig. 9). This *visual word* corresponds to a healthy light brown region in all the exemplified lesions. Moreover,

this *visual word* is associated with a considerable number of patches, which suggests that these lesions have a more or less uniform pigmentation. Experts usually associate these two evidences (light brown and uniform pigmentation) with benign lesions [3]. It is interesting to notice that, as in the previous analysis, the observations are consistent with the medical knowledge.

The *visual words* analysis performed is simple. However, the results are interesting and in accordance to what is expected to observe in both melanoma and non-melanoma lesions. This suggests that BoF can be used to classify melanomas and to identify specific dermoscopic features and patterns [3] by associating them to *visual words*. Future work should focus on this task, which can be seen as a multiple object recognition problem.

Conclusions

This chapter investigates the applicability of local color and texture features to the melanoma classification problem. Several factors associated with the performance of BoF were tested, namely the type of descriptors used and the classification algorithm.

The results show that individually color descriptors perform better than texture descriptors and that good classification results can be achieved using kNN (SE = 93 %, SP = 85 % with $h_{La^*b^*}$ and SE = 100 %, SP = 75 % with h_{L*uv}) and SVM (SE = 93 %, SP = 88 % with M_{Opp}). The fusion of color and texture descriptors also achieved good results, with a score of SE = 96 %, SP = 82 % for the combination of Opp moments with Gabor and Laws texture descriptors.

A simple analysis of the *visual words* showed that the dictionary obtained using BoF has potential to be used as a detector/identifier for specific dermoscopic features and patterns. Future work will rely on testing this hypothesis in order to develop a more medical oriented system. Moreover, sparse sampling methods should be tested in order to compare their performances with that of the dense sampling used in this chapter. Finally, high-level descriptors should be tested as well.

Acknowledgments The authors thank to Dr. Jorge Rozeira for providing the dermoscopy images. This work was supported by Fundação Ciência e Tecnologia in the scope of the grant SFRH/BD/84658/2012 and projects PTDC/SAU-BEB/103471/2008 and PEst-OE/EEI/LA0009/2011.

References

1. Abbas, Q., Celebi, M.E., Serrano, C., García, I.F., Ma, G.: Pattern classification of dermoscopy images: a perceptually uniform model. Pattern Recogn. **46**, 86–97 (2013)
2. Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., Delfino, M.: 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. **134**, 1563–1570 (1998)

3. Argenziano, G., Soyer, H., De Giorgi, V., Carli, P., Delfino, M., Ferrari, A., Hofmann-Wellenhof, R., Massi, D., Mazocchetti, G., Scalvenzi, M., Wolf, I.: Interactive atlas of dermoscopy. Edra Medical Publishing and New Media, Milan (2000). http://www.dermoscopy.org/atlas/
4. Arivazhagana, S., Ganesanb, L., Priyala, S.P.: Texture classification using gabor wavelets based rotation invariant features. Pattern Recogn. Lett. **27**, 1976–1982 (2006)
5. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, New York (1999)
6. Barata, C., Marques, J.S., Rozeira, J.: A system for the detection of pigment network in dermoscopy images using directional filters. IEEE Trans. Biomed. Eng. **59**(10), 2744–2754 (2012)
7. Bratkova, M., Boulos, S., Shirley, P.: oRGB: A pratical opponent color space for computer graphics. IEEE Comput. Graphics Appl. **29**, 42–55 (2009)
8. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**, 121–167 (1998)
9. Celebi, M.E., Kingravi, H.E., Uddin, B., Iyatomi, H., Aslandogan, Y., Stoecker, W.V., Moss, R.: A methodological approach to the classification of dermoscopy images. Comput. Med. Imag. Graphics **31**(6), 362–373 (2007)
10. Celebi, M.E., Iyatomi, H., Stoecker, W., Moss, R.H., Rabinovitz, H., Soyer, H.P.: Automatic detection of blue-white veil and related structures in dermoscopy images. Comput. Med. Imag. Graph. **32**(8), 670–677 (2008)
11. Celebi, M.E., Stoecker, W.V., Moss, R.H.: Advances in skin cancer image analysis. Comput. Med. Imag. Graph. **35**, 83–84 (2011)
12. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011). Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm
13. Clausi, D.: An analysis of co-occurrence texture statistics as a function of grey level quantization. Can. J. Remote Sens. **28**(1), 45–62 (2002)
14. Dala, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and. Pattern Recogn. **1**, 886–893 (2005)
15. Di Leo, G., Paolillo, A., Sommella, P., Fabbrocini, G.: Automatic diagnosis of melanoma: a software system based on the 7-point check-list. In: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, pp. 1818–1823 (2010)
16. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (1999)
17. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**, 119–139 (1997)
18. Ganster, H., Pinz, A., Wildling, E., Binder, M., Kittler, H.: Automated melanoma recognition. IEEE Trans. Med. Imag. **20**(3), 233–239 (2001)
19. Grigorescu, C., Petkov, N., Westenberg, M.A.: Contour detection based on nonclassical receptive field inhibition. IEEE Trans. Image Process. **12**, 729–739 (2003)
20. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3**, 610–621 (1973)
21. Iyatomi, H., Oka, H., Celebi, M.E., Hashimoto, M., Hagiwara, M., Tanaka, M., Ogawa, K.: An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. Comput. Med. Imag. Graphics **32**(7), 566–579 (2008)
22. Jiang, Y.G., Ngo, C.W., Yang, J.: Towards optimal bag-of-features for object categorization and semantic video retrieval. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval, pp. 494–501 (2007)
23. Khan, F.S., van de Weijer, J., Vanrell, M.: Top-down color attention for object recognition. In: Proceedings of the IEEE 12th International Conference on Computer Vision, pp. 979–986 (2009)
24. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 226–239 (1998)
25. Korotkov, K., Garcia, R.: Computerized analysis of pigment skin lesions: a review. Artif. Intell. Med. **56**, 69–90 (2012)

26. Laws, K.I.: Rapid texture identification. In: Proceedings of SPIE Conference on Image Processing for Missile Guidance (1980)
27. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2169–2178 (2006)
28. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
29. Menzies, S., Ingvar, C., Crotty, K., McCarthy, W.H.: Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. Arch. Dermatol. **132**, 1178–1182 (1996)
30. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. Int. J. Comput. Vis. **60**(1), 63–86 (2004)
31. Pellacani, G., Grana, C., Cucchiara, R., Seidenari, S.: Automated extraction of dark areas in surface microscopy melanocytic lesion images. Dermatology **208**(1), 21–26
32. Platt, J.: Probabilities for sv machines. In: Advances in Large Margin Classifiers, pp. 61–74. MIT Press, Cambridge (2000)
33. Randen, T., Husoy, J.H.: Filtering for texture classification: A comparative study. IEEE Trans. Pattern Anal. Mach. Intell. **21**, 291–310 (1999)
34. Rubegni, P., Cevenini, G., Burroni, M., Perotti, R., Dell'Eva, G., Sbano, P., Miracco, C.: Automated diagnosis of pigmented skin lesions. Int. J. Cancer **101**(6), 576–580 (2002)
35. Sadeghi, M., Lee, T.K., McLean, D., Lui, H., Atkins, M.S.: Detection and analysis of irregular streaks in dermoscopic images of skin lesions. IEEE Trans. Med. Imag. **32**(5), 849–861. doi:10.1109/TMI.2013.2239307
36. Sadeghi, M., Lee, T.K., McLean, D., Lui, H., Atkins, M.S.: Global pattern analysis and classification of dermoscopic images using textons. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 168–173 (2012)
37. Sadeghi, M., Razmara, M., Wighton, P., Lee, T.K., Atkins, M.S.: A novel method for detection of pigment network in dermoscopic images using graphs. Comput. Med. Imag. Graph. **35**(2), 137–143 (2011)
38. Van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: A comparison of color features for visual concept classification. In: Proceedings of the 2008 international conference on Content-based image and video retrieval (2008)
39. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1582–1593 (2010)
40. Serrano, C., Acha, B.: Pattern analysis of dermoscopic images based on markov random fields. Pattern Recogn. **42**, 1052–1057 (2009)
41. Silveira, M., Nascimento, J.C., Marques, J.S., Marçal, A.R.S., Mendonça, T., Yamauchi, S., Maeda, J.: Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. IEEE J. Sel. Top. Sign. Process. **3**, 35–45 (2009)
42. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: Proceedings of the 9th IEEE International Conference on Computer Vision, pp. 1470–1477 (2003)
43. Snoek, C.G.M.: Early versus late fusion in semantic video analysis. In: In ACM Multimedia, pp. 399–402 (2005)
44. Squire, D.M., Mller, W., Mller, H., Raki, J.: Content-based query of image databases, inspirations from text retrieval: inverted files, frequency-based weights and relevance feedback. Pattern Recogn. Lett. **21**(13–14), 143–149 (1999)
45. Stoecker, W.V., Gupta, K., Stanley, R.J., et al.: Detection of asymmetric blotches in dermoscopy images of malignant melanomas using relative color. Skin Res. Technol. **11**(3), 179–184 (2005)
46. Stoecker, W.V., Wronkiewicz, M., Chowdhury, R., Stanley, R., Xu, J., Bangert, A., Shrestha, B., Calcara, D., Rabinovitz, H., Oliviero, M., Ahmed, F., Perry, L., Drugge, R.: Detection of granularity in dermoscopy images of malignant melanoma using color and texture features. Comput. Med. Imag. Graph. **35**(2), 144–147 (2011)

47. Stolz, W., Riemann, A., Cognetta, A.B.: ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. Eur. J. Dermatol. **4**, 521–527 (1994)
48. Stricker, M., Orengo, M.: Similarity of color images. In: Proceedings SPIE, vol. 2420, pp. 381–392 (1995)
49. Tkalcic, M., Tasicl, J.F.: Colour spaces: perceptual, historical and applicational background. In: Proceedings of the IEEE Region 8 EUROCON 2003. Computer as a Tool, vol. 1, pp. 304–308 (2003)
50. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: An in-depth study. Tech. Rep. 5737, Institut National De Recherche en Informatique et en Automatique (2005)