Classification of Melanoma Lesions Using Sparse Coded Features and Random Forests
Mojdeh Rastgoo, Guillaume Lemaître, Olivier Morel, Johan Massich, Rafael Garcia, Fabrice Mériaudeau, Frank Marzani, Désiré Sidibé

To cite this version:
Mojdeh Rastgoo, Guillaume Lemaître, Olivier Morel, Johan Massich, Rafael Garcia, et al.. Classification of Melanoma Lesions Using Sparse Coded Features and Random Forests. SPIE Medical Imaging, Feb 2016, San Diego, United States. hal-01250955

HAL Id: hal-01250955
https://u-bourgogne.hal.science/hal-01250955
Submitted on 5 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Classification of Melanoma Lesions Using Sparse Coded Features and Random Forests

Mojdeh Rastgoo\textsuperscript{a,b} and Guillaume Lemaître\textsuperscript{a,b} and Olivier Morel\textsuperscript{a} and Johan Massich\textsuperscript{a} and Rafael García\textsuperscript{b} and Fabrice Mériaudeau\textsuperscript{a} and Frank Marzani\textsuperscript{a} and Désiré Sidibé\textsuperscript{a}

\textsuperscript{a}LE2I UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, 12 rue de la Fonderie, 71200 Le Creusot, France; \\
\textsuperscript{b}ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain

ABSTRACT

Malignant melanoma is the most dangerous type of skin cancer, yet it is the most treatable kind of cancer, conditioned by its early diagnosis which is a challenging task for clinicians and dermatologists. In this regard, CAD systems based on machine learning and image processing techniques are developed to differentiate melanoma lesions from benign and dysplastic nevi using dermoscopic images. Generally, these frameworks are composed of sequential processes: pre-processing, segmentation, and classification. This architecture faces mainly two challenges: (i) each process is complex with the need to tune a set of parameters, and is specific to a given dataset; (ii) the performance of each process depends on the previous one, and the errors are accumulated throughout the framework. In this paper, we propose a framework for melanoma classification based on sparse coding which does not rely on any pre-processing or lesion segmentation. Our framework uses Random Forests classifier and sparse representation of three features: SIFT, Hue and Opponent angle histograms, and RGB intensities. The experiments are carried out on the public PH\textsuperscript{2} dataset using a 10-fold cross-validation. The results show that SIFT sparse-coded feature achieves the highest performance with sensitivity and specificity of 100\% and 90.3\% respectively, with a dictionary size of 800 atoms and a sparsity level of 2. Furthermore, the descriptor based on RGB intensities achieves similar results with sensitivity and specificity of 100\% and 71.3\%, respectively for a smaller dictionary size of 100 atoms. In conclusion, dictionary learning techniques encode strong structures of dermoscopic images and provide discriminant descriptors.

Keywords: Melanoma, Classification, Sparse coding, Random forests, Dermoscopy

1. INTRODUCTION

Despite the fact that malignant melanoma only accounts for nearly 2\% of all skin cancer cases, the vast majority of skin cancer deaths are due to malignant melanoma. According to the World Health Organisation, the incidence of this pathology has increased up to 132,000 diagnosed cases of melanoma\textsuperscript{[4]} during the past decades. The American Cancer Society estimated that during 2014, 76,100 new cases of melanoma would be diagnosed, leading to death in 9710 cases\textsuperscript{[3]} At advanced stage, melanoma is incurable and the patients should go through surgery, possibly immunotherapy, chemotherapy, and/or radiation therapy. If diagnosed at its early stage, however, melanoma is the most treatable kind of cancer\textsuperscript{[2,3]} Indeed the patient survival rate has well increased, in the past decades, thanks to early diagnosis and treatment of melanoma in its early stages.

A well established criteria for early stage melanoma prognosis is the “ABCDE” rule\textsuperscript{[3]} illustrated in Fig.1. This criteria is meant for a human reader to visually inspect an image of a skin lesion and characterize this lesion based on the following visual cues:

(A) Asymmetry of the lesion,

(B) irregularity of the lesion Borders,
Figure 1: **ABCD**-lexicon comparison of two lesions. The top row corresponds to melanoma while the bottom row is a benign lesion.

- (a) melanoma
- (b) **Asymmetric**
- (c) **irregular Borders**
- (d) **variated Colors**
- (e) **Diameter**

- (f) benign lesion
- (g) **non-Asymmetric**
- (h) **regular Borders**
- (i) **unique Color**
- (j) **Diameter**

(C) presence of variegated **Colors** within the lesion,

(D) lesion size, evaluating if the **Diameter** of the lesion is greater than 6 mm,

(E) evaluation of the lesion **Evolution** over time, through regular screening.

Despite the aforesaid positive impact of these methodologies, visual inspection of medical images for diagnosis is prone to errors. Manning *et al.* state that the diagnosis error rates are comparable to the error rates found in any other human visual inspection task. Thus, double readings and Computer-Aided Diagnosis (CAD) systems are placed to aid dermatologists and clinicians to reduce this weakness. Diverse CAD systems are proposed which take advantage of image processing and Machine Learning to mimic the criteria defined by the “ABCD” rule. A prevalent schema for such CAD systems, consists of: (i) **image pre-processing** to facilitate any subsequent task; (ii) **segmentation** to determine the location and region properties of the lesion (likewise the reading of the image by a human grader); (iii) **feature extraction** to characterize the lesion; and further, (iv) **classification** of these features to drive the decision. Towards diagnosis, accurate delineations are required to describe the lesions with highly cognitive criteria, such as asymmetry or boundary irregularities. For the clinicians, finding this delineation is intrinsic to the visual inspection. For the case of CAD systems, however, the delineation is subject to the process of segmentation, which remains an open research topic.

In this paper, we propose a more general framework which does not rely upon pre-processing and segmentation of the lesions and is based on sparse coded features and Random Forests (RF) classifier to detect melanoma in dermoscopic images. The rest of this paper is organized as follows: an overview of related work is presented in Sect. 2. The proposed method is discuss in Sect. 3 while the experiment and obtained results are presented in Sect. 4 and 5 respectively. Finally, conclusions are drawn in Sect. 6.

## 2. RELATED WORK

In the past decade, numerous approaches have been proposed for automated recognition of melanoma lesions. The developed methods are commonly based on clinical or dermoscopy modality. As previously mentioned, these methods follow the usual classification framework of computer vision and consists of four common steps. Korotkov *et al.* summarize these methods and their properties. Unfortunately, a fair comparison among the state-of-the-art presented methods is not possible due to lack of a benchmark and common datasets. Nevertheless recently a public dataset (PH2) has been resealed for research purposes, thanks to Mendoncca *et al.* Section 4 presents a detailed description of this dataset.
Table 1: Summary of the proposed classification methods using PH dataset.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Segmentation</th>
<th>features</th>
<th>Classification</th>
<th>Balancing</th>
<th>Validation</th>
<th>Best performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruela et al.</td>
<td>✓</td>
<td>Shape, FD&lt;sup&gt;1&lt;/sup&gt;</td>
<td>AdB</td>
<td>ROS</td>
<td>OvA&lt;sup&gt;1&lt;/sup&gt;</td>
<td>92</td>
</tr>
<tr>
<td>Ruela et al.</td>
<td>✓</td>
<td>Color statistics</td>
<td>k-NN, AdB</td>
<td>ROS</td>
<td>OvA</td>
<td>96</td>
</tr>
<tr>
<td>Barata et al.</td>
<td>✓</td>
<td>Opponent histogram gradients</td>
<td>AdB, SVM</td>
<td>BoW</td>
<td>ROS 10-fold&lt;sup&gt;2&lt;/sup&gt;</td>
<td>100</td>
</tr>
<tr>
<td>Barata et al.</td>
<td>✓</td>
<td>Opponent histogram</td>
<td>k-NN</td>
<td>BoW</td>
<td>ROS 10-fold&lt;sup&gt;2&lt;/sup&gt;</td>
<td>98</td>
</tr>
<tr>
<td>Abuzaghleh et al.</td>
<td>✓</td>
<td>FFT2, DCT2</td>
<td>SVM</td>
<td>-</td>
<td>-</td>
<td>97.7</td>
</tr>
<tr>
<td>Rastgoo et al.</td>
<td>✓</td>
<td>Shape, color statistics opponent angle</td>
<td>-</td>
<td>RF&lt;sup&gt;3&lt;/sup&gt;</td>
<td>DOS</td>
<td>94</td>
</tr>
</tbody>
</table>

<sup>1</sup> Fourier descriptor (FD), One versus all (OvA).  
<sup>2</sup> Completed Local Binary Pattern (CLBP), Gray-Level Co-occurrence Matrix (GLCM), Histogram of Gradients (HoG), Learner combination (LC).

Subsequently, we emphasize on the most recent methods which are evaluated using this dataset. Table 1 summarizes these methods. Barata et al.<sup>10</sup> and Ruela et al.<sup>13,14</sup> used different subsets of PH dataset in their works.<sup>10</sup> Ruela et al.<sup>13,14</sup> compared the role of shape and colors to detect melanoma vs. benign and dysplastic lesions using AdaBoost (AdB) classifier. Concerning the same problem, Barata et al.<sup>11</sup> proposed to use Bag-of-Words (BoW) representation of colors and gradient features and compared different classifiers, such as AdB, kernel Support Vector Machine (SVM), and k-nearest-neighbor (NN). The configuration based on BoW representation and k-NN classifier lead to the best results with Sensitivity (SE) and Specificity (SP) of 100% and 75%, respectively. The authors later used a similar scheme (BoW representation, k-NN classifier, and histogram of opponent color space histogram) to compare the effects of manual and automatic segmentation in the classification process. The results indicate that manual segmentation outperformed the automatic segmentation (SE and SP of 98% and 86%, respectively). Feature space, random over-sampling (ROS) was used in all the aforementioned methods to tackle the imbalance problem.

Abuzaghleh et al.<sup>15</sup> also used PH dataset in which they proposed an automated recognition system based on color and shape features such as 2-D Fast Fourier Transform features (FFT2), Discrete Cosine Transform features (DCT2), size and complexity features. The authors proposed two classification approaches: (i) multi-class and (ii) two-level classification, both using SVM classifier. In the latter approach, the first level classifies normal and abnormal lesions while the second level works only with the abnormally classified lesions to differentiate melanoma and dysplastic nevi.

In our previous work, we compared the effects of various colors, shape and texture features using ensemble approaches.<sup>16</sup> The features were extracted from the segmented area and data space over-sampling (DOS) was used instead of ROS.<sup>16</sup> The best results with SE and SP of 94% and 92%, respectively, was achieved using RF ensemble and combination of color and texture features.

### 3. METHODOLOGY

As mentioned in the Sect. 1, the proposed framework, in comparison to previous methods (see Table 1), do not rely on pre-processing and segmentation and focus only on feature detection, extraction, and classification. Figure 2 illustrates our proposed framework.

#### 3.1 Feature extraction

##### 3.1.1 Low-level features

In clinical environment, the prognosis of early stage melanoma relies upon visual cues, represented by a set of rules such as “ABCDE”. One of the main characteristic is variegated colors and difference of color representation between melanoma and benign lesions. Textural difference, irregular, and sharp borders are another valuable and distinguishable characteristics. In this study, three low-level features in line with the previous discriminative criteria are used. Color variation criteria is described through two descriptors: opponent color space angle and hue histograms ($C_1$) and R,G, and B intensities ($C_2$). In addition the texture, gradient and irregular borders
is characterized via Scale-Invariant Feature Transform (SIFT) descriptor. These descriptors are presented into
details in the remainder of this section. Note that these features are locally extracted by partitioning the
dermoscopic images into patches.

Dense Scale-Invariant Feature Transform (SIFT) descriptors are used to encode the local gradient infor-
mation using an histogram-based representation\cite{18} SIFT descriptors are extracted over a regular grid such
that each grid point is fixed at the center of each image patch. Typically, a region around the center
is divided into $4 \times 4$ sub-regions from which a 8-bins histogram of the gradient orientations weighted by
the gradient magnitudes is computed. Finally, these histograms are concatenated to build the final de-
scriptor with a final size of 128 dimensions. The SIFT implementation used in this work is provided by
Vedaldi et al\cite{19}.

The opponent color space angle and hue histogram ($C1$) have been first proposed by Van de Weijer and
Schmidt as local color features\cite{20}. These descriptors are robust to photometric variations (i.e., shadow,
shading, specularities, and light source changes) as well as geometrical variations (i.e., viewpoint, zoom,
and object orientation). The hue ($H_C$) and angle ($\theta_C$) of opponent color space ($O_{1,2,3}$) are formulated as
shown in Eq.2 and Eq.3, respectively. The opponent color space transformation is defined as in Eq.1.

$$
\begin{bmatrix}
O_1 \\
O_2 \\
O_3 
\end{bmatrix} =
\begin{bmatrix}
\frac{(R - G)}{\sqrt{2}} \\
\frac{(R + G - 2B)}{\sqrt{6}} \\
\frac{(R + G + B)}{\sqrt{3}} 
\end{bmatrix},
$$

\hspace{1cm} (1)

$$
H_C = \arctan\left(\frac{\sqrt{3}(R - G)}{R + G - 2B}\right),
$$

\hspace{1cm} (2)

$$
\theta_C^d = \arctan\left(\frac{\sqrt{3}(R_d' - G_d')}{R_d' + G_d' - 2B_d'}\right),
$$

\hspace{1cm} (3)

where $d$ denotes the spatial coordinates of $(x,y)$ and $R_d', G_d', B_d'$ denote the first order derivatives of $RGB$
with respect to the coordinates.

This color descriptor is built by taking a 42 bins histogram for the opponent angle $\theta_C^d$ and the hue channel
$H_C$, for a final descriptor size of 84 dimensions.

The color intensities ($C2$) represent the color information in its simplest form, their intensities. This de-
scriptor concatenates the color intensities $R$, $G$ and $B$ to create the feature descriptor.

3.1.2 High-level features

High-level descriptor is computed using sparse coding techniques. Sparse signal representation has become very
popular in the past decades and lead to state-of-the-art results in various applications such as face recognition\cite{21}.
image denoising, image inpainting, and image classification. The main goal of sparse modeling is to efficiently represent the images as a linear combination of a few typical patterns, called atoms, selected from a dictionary. Here, we intend to use sparse representation of the low-level extracted features for melanoma classification. Sparse coding consists of three main steps: (i) dictionary learning, (ii) low-level features projection, and (iii) feature pooling, as illustrated in Fig. 3.

**Sparse approximation** Given a dictionary $D \in \mathbb{R}^{n \times K}$ composed of $K$ atoms and an original signal $y \in \mathbb{R}^{n}$ (i.e., one feature vector), the sparse approximation corresponds to find the sparsest vector $x \in \mathbb{R}^{K}$ such that:

$$\arg \min_{x} \|y - Dx\|_{2} \quad \text{s.t.} \quad \|x\|_{0} \leq \lambda ,$$

where $\lambda$ is a specified sparsity level.

Solving the above optimization problem is an NP-hard problem. However, approximate solutions are obtained using greedy algorithms such as Matching Pursuit or Orthogonal Matching Pursuit (OMP). We used the batch-OMP variant which offers a more efficient algorithm than the standard OMP for our specific problem.

**Dictionary learning** As stated previously, the sparse approximation is computed given a specific dictionary $D$, which involves a learning stage from a set of training data. This dictionary is learned using K-SVD which is a generalized version of K-means clustering and uses Singular Value Decomposition (SVD). The dictionary is built, by iteratively solving the optimization problem of Eq. 5, by alternatively computing the sparse approximation of $X$ and the dictionary $D$.

$$\arg \min_{D, X} \|Y - DX\|_{2} \quad \text{s.t.} \quad \|x_{i}\|_{1} \leq \lambda ,$$

where $Y$ is a training set of low-level descriptors, $X$ is the associated sparse coded matrix (i.e., set of high-level descriptors) with a sparsity level $\lambda$, and $D$ is the dictionary with $K$ atoms.

Given $D$, $X$ is computed using the batch-OMP algorithm, while given $X$, $D$ is sequentially updated, one atom at a time using SVD.

**Low-level features projection** Once the dictionary is learned, each set of low-level features $F_{I} \in \mathbb{R}^{n \times p}$ extracted from $p$ patches in an image is encoded using the dictionary $D$, solving the optimization problem presented in Eq. 4 such that $F_{I} \approx DX_{I}$.

**Feature pooling** The sparse coded matrix $X_{I}$ is max-pooled to build a final descriptor $f$ characterizing the given image, such that:

$$f_{i} = \max_{j} (|X_{I}(i, j)|), \forall i = 1, \cdots, K .$$

3.2 Feature classification

3.2.1 Over-sampling from imbalanced dataset

Similarly to Barata et al. the imbalanced issue of the dataset is tackled by over-sampling the samples of the minority class. New samples are generated to get a balanced set by randomly repeating original samples of the minority class with an additional Gaussian noise $N(0, 0.0001)$. 
3.2.2 Random Forests (RF)

The classification is performed using a RF classifier. RF is an ensemble of decision trees which generalizes the classification process by applying two types of randomization: at the tree level, each tree is fed by a bootstrap made of \( S' \) samples built from the original data of size \( S \) such that \( S = S' \), and at the node level, a subset of feature dimensions \( m \) is randomly selected from the original dimension \( M \) such that \( m = \sqrt{M} \). The trees in RF are grown to their maximum length without any pruning. Each tree in the ensemble casts a unit vote in the final prediction and the final prediction is based on combination of all the votes. RF is used with 1000 un-pruned trees using gini criterion and the original feature dimension.

4. EXPERIMENTS

The experiments are conducted on the public PH\textsuperscript{2} dataset. This dataset has been acquired at Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal with Tuebinger Mole Analyzer system with a magnification of 20\texttimes. The 8-bits RGB color dermoscopic images are obtained under the same conditions with a resolution of 768 px \( \times \) 560 px. This dataset contains 200 dermoscopic images divided into 80 benign, 80 dysplastic and 40 melanoma lesions. The lesions are segmented and their histological diagnosis are provided as ground-truth.

In our experiments seven images are discarded due to artefacts such as hair oculusions. Thus, they are conducted on a subset of the dataset consisting of 39 melanoma, 78 benign, and 76 dysplastic lesions. The patch size used to extract the feature is 10 px \( \times \) 10 px. The three low-level features are sparsely encoded considering three sparsity levels \( \lambda = \{2, 4, 8\} \) and different number of atoms \( K = \{100, 200, \ldots , 1000\} \). The classification is performed in a 10-fold cross-validation model in which 80% of the data is used for training and 20% for testing.

5. RESULTS

The results of the experiment are shown in terms of Sensitivity and Specificity in Table 2. The highest classification rate in respect of each feature type and each sparsity level are highlighted in different shades of gray from dark to light. The results show that \( C1 \) and SIFT sparse coded features perform better with sparsity levels of 2 and 4, respectively, while \( C2 \) performs better with a sparsity level of 8. In general, larger dictionary sizes lead

\[ \text{arg min}_x \|y - Dx\|_2 \quad \text{s.t.} \|x\|_1 \leq \lambda \]

\[ F = \begin{bmatrix} \vdots & \vdots & \vdots \end{bmatrix}_{n \times p} \rightarrow X = \begin{bmatrix} \vdots & \vdots & \vdots \end{bmatrix}_{n \times K} \rightarrow f = \begin{bmatrix} \vdots \end{bmatrix}_K \]

Figure 3: Sparse coded representation for training and testing samples.
Table 2: The obtained results with different number of atoms and sparsity levels. The first, second, and third highest results for each sparsity level are highlighted in different shades of gray from dark to light color, respectively.

<table>
<thead>
<tr>
<th>Features</th>
<th>Color 1</th>
<th>Color 2</th>
<th>SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>SP</td>
<td>SE</td>
</tr>
<tr>
<td>Sparsity level</td>
<td>2</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Dictionary size</td>
<td>100</td>
<td>65.6</td>
<td>49.6</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>50.1</td>
<td>59.3</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>59.8</td>
<td>65.4</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>67</td>
<td>78.6</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>78.7</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>98.6</td>
<td>82.5</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>92.8</td>
<td>89.9</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>92.9</td>
<td>81.4</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>90</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>100</td>
<td>86.8</td>
</tr>
</tbody>
</table>

Figure 4: The 12 highest result achieved by RF classifier and sparse representation of SIFT, color1, and color2 features with different sparsity levels are illustrated in blue and black as sensitivity and specificity, respectively, while their dictionary size is represented in red. A comparison of the dictionary sizes is also presented in middle of the graph, which contains five levels with maximum dictionary sizes of 1000, 800, 600, 400, 200 for each level respectively.

To better classification performance, independently to the feature type and the sparsity level. More precisely, dictionaries with more than 600 atoms are preferable. Figure 4 illustrate the 12 best results. Although, SIFT and C1 sparse coded features achieve the best classification performance in comparison with C2, it can be noted that C2 features represent the dermoscopic images in their simplest form and create comparable results.
6. CONCLUSION AND FUTURE WORK

In this work, we proposed a novel classification framework of melanoma lesions, based on sparse representation of the low-level features. Our framework does not rely on the primary steps of pre-processing and segmentation of the lesions and provide more general algorithm to solve this problem. We proposed to use a well-known color descriptor based on hue and angle histograms of opponent color space as well as SIFT as a texture descriptor. We also consider to represent the images in their simplest form and consider the second color descriptor as R,G and B intensities. An extensive comparison based on different dictionary sizes and several sparsity levels were carried out on the PH2 dataset. The results highlighted the advantage of the proposed method where an RF classifier and a sparse representation of SIFT features with a dictionary size of 800 and a sparsity level of 2 achieved the highest performance (SE and SP of 100% and 90.3%, respectively). In general, the obtained results outperform the state of the art. As avenues for future research, a comparison of sparse learned dictionary with Bag-of-Word models can be performed.

REFERENCES

1. “World health organization, skin cancer.”