Local Features Applied to Dermoscopy Images: Bag-of-Features versus Sparse Coding

Catarina Barata[1,3]([✉]), Mário A.T. Figueiredo[2,3], M. Emre Celebi[4],
and Jorge S. Marques[1,3]

[1] Institute for Systems and Robotics, Lisbon, Portugal
ana.c.fidalgo.barata@ist.utl.pt
[2] Instituto de Telecomunicações, Lisbon, Portugal
[3] Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[4] Department of Computer Science,
University of Central Arkansas, Conway, AR, USA

Abstract. Feature extraction is a crucial step in any computer aided diagnosis (CAD) system for melanoma diagnosis. Therefore, it is important to select features that are able to efficiently characterize the properties of the different types of lesions. Local features that separately characterize and distinguish different regions of the lesions have been shown to provide good descriptors for these skin lesions. Two powerful methods can be used to obtain local features: bag-of-features (BoF) and sparse coding (SC). Both methods have been applied to dermoscopy with promising results. However, a comparison between the two strategies is lacking. In this work, we fill this gap by developing a framework to compare the two methods in the melanoma diagnosis task. The results show that SC significantly outperforms BoF, achieving sensitivity = 85.5% and specificity = 75.1% versus sensitivity = 81.7% and specificity = 66.5%.

Keywords: Melanoma diagnosis · Computer aided diagnosis · Local features · Bag-of-features · Sparse coding

1 Introduction

Melanoma is one of the most common types of cancer in Europe, North America, and Australia. Due to its rapid growth, it is able to metastasize to other organs, such as lungs, bones, or brain [1]. The diagnosis of skin lesions follows a specific guideline: (i) inspection of the lesion using a magnification device; (ii) assessment of different criteria, such as the ABCD rule [2] or the 7-point checklist [3]; and (iii) scoring the lesion based on the identified criteria. Although the aforementioned medical rules are well established and guarantee an increase in the accuracy of the diagnosis, the evaluation still critically hinges on visual inspection and on the expertise of the dermatologist [1]. This means that the analysis of lesions is a highly subjective and difficult task.

Modern inspection devices are able to acquire images of the lesions, obtained with or without special illumination, dividing the images into two types: dermoscopy and clinical. For the past two decades, research groups have been working on computer aided diagnosis (CAD) systems to diagnose the skin lesions, using either dermoscopy or clinical images [4,5]. These CAD systems can be used as a support tool by dermatologists with any level of expertise, reducing the subjectivity of the diagnosis. This work uses dermoscopy images, thus from this point on we will discuss specific aspects of its automatic analysis.

CAD systems follow three main steps: (i) lesion segmentation; (ii) feature extraction; and (iii) lesion diagnosis [4]. Aside from lesion segmentation, which by itself is a major challenge, there is a significant diversity in the type of features and classifiers used in steps (ii) and (iii). A short list of classifiers that have been applied includes K-nearest neighbor, AdaBoost, support vector machines, and neural networks [5]. The extracted features can be divided into two categories: global and local. The former consists of computing a single vector to describe the entire lesion. This vector can comprise information about the shape and symmetry of the lesion (*e.g.*, area, circularity measure, shape symmetry), color (*e.g.*, RGB or HSV histograms), and texture *e.g.*, gray level co-occurrence Matrix) [6]. Local features allow us to separately characterize different regions of the lesion. This can be seen as an approximation of the analysis performed by dermatologists, since they also assess different regions of the lesions. A simple strategy to compute local features is the bag-of-features (BoF) approach, which has been applied with success in different works (*e.g.*, [7,8]). More recently, a different method has been used to obtain local features: sparse coding (SC) [9,10]. This strategy arises from relaxing the restrictive constraints of the BoF optimization problem, as will be discussed in Sect. 2, and has been shown to be efficient in capturing salient properties of the image in different computer vision problems (*e.g.*, [11,12]).

Both BoF and SC have achieved promising classification results in dermoscopy image analysis. However, a direct comparison between the two types of features has been missing. In this paper, we fill this gap by providing a comparison between the two methods, to assess which one performs the best. The remaining sections of the paper are organized as follows. In Sect. 2, we discuss the formulations of BoF and SC and compare them. In Sect. 3 we describe the experimental framework, and in Sect. 4 we present the results.

2   Local Features - From BoF to Sparse Coding

The BoF method assumes that an image can be represented as a collection of elements of a dictionary of *visual words* (atoms). Assuming that a dictionary D of K elements is known, any image is processed as follows: (i) a set of M patches is extracted and a feature vector $x_m \in \mathbb{R}^D$ is computed for each of them; (ii) the features are matched to the closest dictionary element, as follows

$$\min_{\alpha_m} \|x_m - D\alpha_m\|_2^2$$
$$\textbf{s.t} \quad \alpha_m \in \{0,1\}^K, \ \|\alpha_m\|_0 = 1, \tag{1}$$

where $\|.\|_0$ denotes the ℓ_0 "norm"; (iii) this information is summarized into a histogram of occurrences that counts the number of times each atom was selected.

The constraint used in (1) is very restrictive. In order to deal with this issue one can use the SC formulation. Similarly to BoF, the first step of SC is the extraction of a set of M image patches, followed by the computation of a feature vector to characterize each patch. The following step is to match the feature vectors to atoms of a known dictionary. However, instead of assuming that each vector is only associated with one of the atoms, SC assumes that each vector is a combination of a small number of atoms. This can be formulated as an optimization problem with a regularization based on the ℓ_1 norm

$$\min_{\alpha_m} ||x_m - D\alpha_m||_2^2 + \lambda ||\alpha_m||_1, \tag{2}$$

where $\alpha_m \in \mathbb{R}^K$ is a vector of coefficients and λ is a non-negative parameter specified by the user, which controls the relevance of the regularization term. Using the ℓ_1 norm in the regularization term encourages sparsity of the coefficients, i.e., only a small number of them is non-zero. Additional constraints can be added to the problem, such as setting $\alpha_m \geq 0$ [12].

Aside from the patch representation, another main difference between BoF and SC is the strategy used to estimate the dictionaries. In both cases these are estimated using a training set of N feature vectors $\{x_1, ..., x_N\} \in \mathbb{R}^D$, extracted from the patches of several images. According to the BoF formulation (see (1)) D can be estimated as follows

$$\min_{\alpha_1,...,\alpha_N,D} \sum_{n=1}^{N} ||x_n - D\alpha_n||_2^2$$
$$\textbf{s.t} \quad \alpha_n \in \{0,1\}^K, \ ||\alpha_n||_0 = 1, \ \forall n. \tag{3}$$

This optimization problem can be solved using a clustering algorithm, such as k-means [13].

In the SC formulation, a dictionary of K elements is obtained by solving the following optimization problem

$$\min_{\alpha_1,...,\alpha_N,D} \sum_{n=1}^{N} ||x_n - D\alpha_n||_2^2 + \lambda ||\alpha_n||_1$$
$$\textbf{s.t} \quad ||d_k||_2 \leq 1, \ k = 1, ..., K, \tag{4}$$

where $d_k \in \mathbb{R}^K$ is the k-th column of D. The normalization constraint $\|d_k\|_2 = 1$ is used to avoid trivial solutions for the dictionary, namely having the columns of D growing to infinity and the α coefficients approaching zero.

The estimation of α according to (2) is a convex problem, which can be solved using one of several special-purpose algorithms that have been developed of this problem [14]. On the other hand, the problem (4) is not convex and has been the focus of a considerable amount of recent research. The standard approach to solve (4) is to alternate between estimating the SC coefficients, keeping the dictionary fixed, and updating the dictionary [14]. Formally:

(i) Fix the dictionary D and solve

$$\min_{\alpha_1,...\alpha_N} ||x_n - D\alpha_n||_2^2 + \lambda||\alpha_n||_1. \tag{5}$$

(ii) Fix $\alpha_1,...,\alpha_N$ and solve

$$\min_D \sum_{n=1}^{N} ||x_n - D\alpha_n||_2^2, \ \ \mathbf{s.t.} \ ||d_k||_2 \leq 1, \ k = 1,...,K. \tag{6}$$

These two steps are repeated for a predefined number of iterations or until some convergence criterion is satisfied. An extensive review on different methods to solve these optimizations can be found in [14].

A final difference between the two methods resides in how the final image representation is obtained. As stated at the beginning of this section, BoF represents the image as a histogram of occurrences of atoms. The same approach can not be directly applied to SC, since the vectors α_m select more than one atom, with different weights. Different pooling strategies have been proposed to tackle this issue: *e.g.*, max-pooling or mean of the absolute values of α [11].

3   Experimental Framework

The goal of this paper is to perform a fair comparison between the BoF and SC representations. Therefore, we must maintain common parameters of the methods constant, such as the type and size of patches extracted from the images and the features used to describe them, and adjust only what is specific of each method. In the sequel we present the experimental setting used obtain our results.

(i) **Patch extraction/image sampling:** 16×16 overlapping (step of 8 pixels) patches are extracted from all of the images. Although it is possible to extract patches from the entire image (*e.g.*, [9]), we chose to extract patches only from a bounding box around the lesion. This allows working with the images in their original size (average 560×750), without unbearable computational costs. The area of the image containing the lesion is identified using manual segmentation.

(ii) **Patch features:** Color and texture features are computed for each patch, namely color histograms for the RGB and HSV color spaces, and gradient histograms (amplitude and orientation). All these histograms have 16 bins. The aforementioned fetures are not the ones used in other sparsity-based dermoscopy works [9,10]. Those works use the vectorized patches in either gray level or RGB space, and learn dictionaries to represent that information. Nowadays, learning the dictionaries directly from image patches is a very popular approach [14]. Image patches are not suitable to be tested in the BoF framework, because the size of the resulting feature vectors lead to very high computational costs. Nonetheless, we will use image patches as features in SC, in order to establish a comparison with related works.

(iii) **Dictionary learning:** k-means was used to obtain the BoF dictionaries, while the online dictionary learning method [15] (available in the SPAMS software package[1]) was used to obtain the SC dictionaries. The size of all the dictionaries was chosen in the set $K \in \{2^7, 2^8, 2^9\}$.

(iv) **Pooling:** The final BoF descriptor was obtained using the traditional vector quantization approach (see (1)), followed by histogram building.
In the case of SC, the α vectors of the patches were obtained using the LARS algorithm [16] (available in SPAMS). Two optimization problems were applied on this phase: the traditional one (2) and one obtained by inserting a non-negativity constraint. The combination of all of the α_m vectors of a certain image is performed using two strategies: max-pooling (max) and average absolute pooling (abs), respectively defined as follows:

$$\alpha^j = \max\{|\alpha_1^j|, |\alpha_2^j|, ..., |\alpha_M^j|\}, \tag{7}$$

$$\alpha^j = \frac{1}{M} \sum_{m=1}^{M} |\alpha_m^j|, \tag{8}$$

where α^j is the j-th component of the vector α, M is the number of patches, and α_m^j is the j-th component of the m-th patch vector.

(v) **Classification:** The diagnosis was obtained using a SVM with a radial basis function (RBF) kernel (available in MATLAB 2015b®). A different classifier was trained for each of the possible feature configurations, using a set of dermoscopy images diagnosed by experts. In each of the experiments we tunned the width of the RBF kernel $\rho \in \{2^{-12}, 2^{-5}, ..., 2^{12}\}$ and the penalty term $C \in \{2^{-6}, 2^{-4}, ..., 2^6\}$ given to the soft margin.

4   Experimental Results

4.1   Dataset and Performance Metrics

All of the experiments were carried out on a heterogeneous dataset of 804 images (241 melanomas), selected from the EDRA database [1]. The ground-truth diagnosis was provided by a group of experts.

The different configurations were evaluated in terms of sensitivity (SE), specificity (SP), and a cost score (S) defined as follows

$$S = \frac{c_{10}(1 - SE) + c_{01}(1 - SP)}{c_{10} + c_{01}}, \tag{9}$$

where c_{10} is the cost of an incorrectly classified melanoma (false negative) and c_{01} is the cost of an incorrectly classified non-melanoma (false positive). Since we consider that an incorrect classification of a melanoma is a more serious error, we set $c_{10} = 1.5c_{01}$ and $c_{01} = 1$. The results were obtained using a 10-fold nested cross-validation strategy, where the images were divided into 10-folds, each with

---

[1] http://spams-devel.gforge.inria.fr/.

approximately the same proportion of benign and malignant lesions. One of the folds was kept for testing, while the remaining nine were used for training and parameter selection. This procedure was repeated ten times with a different fold for testing, and the results are the average performance.

4.2   Results

Table 1 shows the comparison between BoF and SC for the different features herein considered. Several conclusions can be drawn from these scores. The first is that *max*-pooling leads to significantly better results than *abs*-pooling. This happens in almost all the features. Moreover, the non-negativity constraint also improves the results (only showed for *max*-pooling). Interestingly, the features used in other dermoscopy works (gray level and RGB patches) achieve worse scores than the other tested color and texture features. Finally, we are able to show that SC outperforms BoF in almost all of the experiments, which suggest that this approach is more efficient.

Table 1. Results for melanoma diagnosis using BoF and SC. In **bold** we highlight the best results.

Features	BoF	SC-abs	SC-max	SC-max & $\alpha \geq 0$
Amplitude histogram	$SE = 64.2\%$	$SE = 68.0\%$	$SE = 69.7\%$	**SE =76.3%**
	$SP = 72.8\%$	$SP = 64.1\%$	$SP = 64.3\%$	**SP =60.4%**
	$S = 0.3236$	$S = 0.3356$	$S = 0.3246$	**S =0.3005**
Orientation histogram	$SE = 66.7\%$	$SE = 75.4\%$	**SE=74.2%**	$SE = 73.8\%$
	$SP = 68.6\%$	$SP = 63.4\%$	**SP =67.9%**	$SP = 67.7\%$
	$S = 0.3254$	$S = 0.2940$	**S =0.2832**	$S = 0.2834$
RGB histogram	$SE = 77.6\%$	$SE = 76.7\%$	$SE = 87.9\%$	**SE =82.9%**
	$SP = 70.3\%$	$SP = 72.3\%$	$SP = 70.0\%$	**SP = 77.3%**
	$S = 0.2532$	$S = 0.2504$	$S = 0.1953$	**S =0.1934**
HSV histogram	$SE = 81.7\%$	$SE = 79.6\%$	$SE = 82.1\%$	**SE =85.5%**
	$SP = 66.5\%$	$SP = 67.3\%$	$SP = 79.0\%$	**SP = 75.1%**
	$S = 0.2438$	$S = 0.2529$	$S = 0.1910$	**S =0.1860**
Gray level patches	-	**SE =68.0%**	$SE = 73.9\%$	$SE = 73.8\%$
		SP = 70.9%	$SP = 52.6\%$	$SP = 54.4\%$
		S =0.3084	$S = 0.3462$	$S = 0.3396$
RGB patches	-	$SE = 65.5\%$	$SE = 72.6\%$	**SE =79.2%**
		$SP = 68.4\%$	$SP = 70.0\%$	**SP = 71.8%**
		$S = 0.3334$	$S = 0.2844$	**S =0.2376**

Table 2 shows the number of images that are correctly and incorrectly classified by BoF and SC, using the best configuration (HSV histogram). These values show that 50% of the images incorrectly classified by BoF are correctly classified by SC. Although the opposite is also true (58 images), it happens in a much

Table 2. Number of images correctly and incorrectly classified by each of the methods using the HSV histograms as patch features.

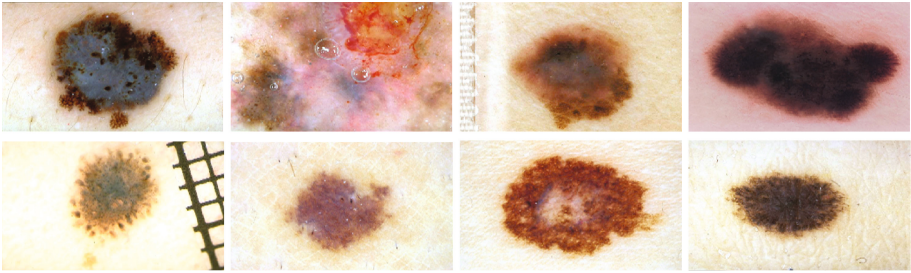|		Sparse coding	
		Correct	Incorrect
BoF	Correct	513	58
	Incorrect	116	117

Fig. 1. Malignant (1st row) and benign (2nd row) lesions, correctly classified by both methods.
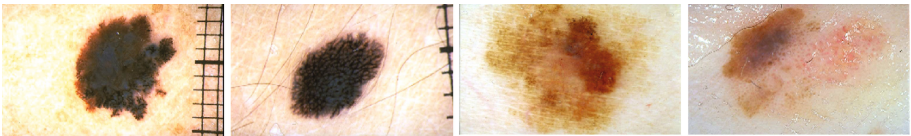

Fig. 2. Malignant (1st and 3rd columns) and benign (2nd and 4th columns) lesions, incorrectly classified by BoF (1st–2nd columns) and SC (3rd–4th columns).

smaller extent. We would like to point out that the scores obtained with SC using a single feature still outperform the best results obtained for this dataset with feature fusion ($SE = 83\%$, $SP = 76\%$) [17].

Figure 1 shows examples of lesions correctly classified by both methods, using their best configurations, while Fig. 2 shows examples of lesions incorrectly classified by one of the methods.

5   Conclusions

In this paper, we have compared bag-of-features and sparse coding in the problem of melanoma diagnosis. A simple framework was used to compare the two methods, where the idea was to keep fixed the common variables and only adjust the key aspects that are specific of each of the methods. This allowed us to perform a fair comparison and show that SC outperforms BoF, obtaining a sensitivity = 85.5% and specificity = 73.4% vs. sensitivity = 81.7% and specificity = 66.5%, for the corresponding best configurations.

References

1. Argenziano, G., Soyer, H.P., De Giorgi, V., Piccolo, D., Carli, P., Delfino, M., Ferrari, A., Hofmann-Wellenhog, V., Massi, D., Mazzocchetti, G., Scalvenzi, M., Wolf, I.H.: Interactive Atlas of Dermoscopy. EDRA Medical Publishing & New Media, Milan (2000)
2. 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)
3. Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., Delfino, E.: 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)
4. Korotkov, K., Garcia, R.: Computerized analysis of pigmented skin lesions: a review. Artif. Intell. Med. **56**(2), 69–90 (2012)
5. Oliveira, R., Papa, J., Pereira, A., Tavares, J.: Computational methods for pigmented skin lesion classification in images: review and future trends. Neural Comput. Appl., 1–24 (2016)
6. Celebi, M., Kingravi, H., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W., Moss, R.: A methodological approach to the classification of dermoscopy images. Comput. Med. Imaging Graph. **31**(6), 362–373 (2007)
7. Barata, C., Ruela, M., Francisco, M., Mendonça, T., Marques, J.: Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Syst. J. **8**(3), 965–979 (2014)
8. Rastgoo, M., Garcia, R., Morel, O., Marzani, F.: Automatic differentiation of melanoma from dysplastic nevi. Comput. Med. Imaging Graph. **43**, 44–52 (2015)
9. Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R.: Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In: Zhou, L., Wang, L., Wang, Q., Shi, Y. (eds.) MLMI 2015. LNCS, vol. 9352, pp. 118–126. Springer, Cham (2015). doi:10.1007/978-3-319-24888-2_15
10. Rastgoo, M., Lemaitre, G., Morel, O., Massich, J., Garcia, R., Meriaudeau, F., Marzani, F., Sidibé, D.: Classification of melanoma lesions using sparse coded features and random forests. In: SPIE Medical Imaging, p. 97850C. International Society for Optics and Photonics (2016)
11. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1794–1801. IEEE (2009)
12. Zhang, C., Liu, J., Tian, Q., Xu, C., Lu, H., Ma, S.: Image classification by non-negative sparse coding, low-rank and sparse decomposition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1673–1680. IEEE (2011)
13. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: Ninth IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 1470–1477. IEEE (2003)
14. Mairal, J., Bach, F., Ponce, J.: Sparse modeling for image and vision processing. Found. Trends Comput. Graph. Vis. **8**(2–3), 85–283 (2014)

15. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. **11**, 19–60 (2010)
16. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Stat. **32**(2), 407–499 (2004)
17. Barata, C., Celebi, M., Marques, J.: Melanoma detection algorithm based on feature fusion. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2653–2656. IEEE (2015)