Discrimination of retinal images containing bright lesions using sparse coded features and SVM

Désiré Sidibé, Ibrahim Sadek, Fabrice Mériaudeau

HAL Id: hal-01145892
https://hal.archives-ouvertes.fr/hal-01145892

Submitted on 27 Apr 2015

Discrimination of retinal images containing bright lesions using sparse coded features and SVM

Désiré Sidibé*, Ibrahim Sadek, Fabrice Mériaudeau

Université de Bourgogne - LE2I, CNRS, UMR 6306, 12 rue de la fonderie, 71200 Le Creusot, France

Abstract

Diabetic Retinopathy (DR) is a chronic progressive disease of the retinal microvasculature which is among the major causes of vision loss in the world. The diagnosis of DR is based on the detection of retinal lesions such as microaneurysms, exudates and drusen in retinal images acquired by a fundus camera. However, bright lesions such as exudates and drusen share similar appearances while being signs of different diseases. Therefore, discriminating between different types of lesions is of interest for improving screening performances. In this paper, we propose to use sparse coding techniques for retinal images classification. In particular, we are interested in discriminating between retinal images containing either exudates or drusen, and normal images free of lesions. Extensive experiments show that dictionary learning techniques can capture strong structures of retinal images and produce discriminant descriptors for classification. In particular, using a linear SVM with the obtained sparse coded features, the proposed method achieves

---

*Corresponding author

Email addresses: `dro-desire.sidibe@u-bourgogne.fr` (Désiré Sidibé),
`ibrahimsadek87@gmail.com` (Ibrahim Sadek), `fmeriau@u-bourgogne.fr`
(Fabrice Mériaudeau)

superior performance as compared with the popular Bag-of-Visual-Word approach for image classification. Experiments with a dataset of 828 retinal images collected from various sources show that the proposed approach provides excellent discrimination results for normal, drusen and exudates images. It achieves a sensitivity and a specificity of 96.50% and 97.70% for the normal class; 99.10% and 100% for the drusen class; and 97.40% and 98.20% for the exudates class with a medium size dictionary of 100 atoms.

Key words: Diabetic retinopathy, Exudates, Drusen, Sparse coding, Classification.

---

1. Introduction

Diabetic Retinopathy (DR) is a chronic progressive disease of the retinal microvasculature that has became a major cause of vision loss in the world. Due to the burden of diabetes over the past decades, the prevalence of DR is expected to grow exponentially and affect over 300 millions people worldwide by 2025 [1, 2]. Despite the high risk factor, it has been established that early detection and timely treatment can reduce the development of severe vision loss in 60% of cases [3]. However, DR does not exhibit any distinctive symptoms which the patient can easily perceive until a severe stage is reached. It is therefore important to develop Computer Aided Diagnosis (CAD) systems that can help clinicians in tasks such as large scale population screening and early diagnosis. The benefit of such a CAD system is to quickly pre-screen patients and orient those at risk to an ophtalmologist, reducing time and cost of a visit.

The diagnosis of DR is based on the detection of lesions or other vascular

2

abnormalities in retinal images acquired by fundus cameras. Lesions of interest include microaneurysms, cotton wool spots, exudates, macular edema, and hemorrhages [4]. Many algorithms have been developed for DR detection over the past two decades, following a general pipeline: image quality verification, vessel segmentation, lesion detection, and automatic classification with a machine learning technique [5, 6].

However, most of the existing algorithms are designed to detect a specific type of lesions in retinal images, though different lesions might be present in a single image indicating the stage of the disease. Morever, the lesions can appear very similar while being signs for different diseases. For instance, DR is caracterized, at early stages, by red lesions (microaneurysms) and bright lesions (exudates) which appear as small white or yellowish deposits with sharp edges in retinal images. On the other hand, Age Related Macular Degeneration (ARMD) which causes a gradual loss of central vision, is mainly caracterized by variable sized yellowish deposits in the retina called drusens. Therefore, discriminating between different types of lesions is of great interest for decision making in an automatic CAD system.

In this paper, we propose an automatic classification of retinal images that discriminates between images containing different bright lesions. The method is based on sparse signal representation techniques that can capture strong structures of retinal images and produce discriminant descriptors for classification. In particular, the proposed methodology does not require any pre-processing of the images such as blood vessels segmentation or optic disc removal, and achieves higher performance than other feature representation techniques such as bag-of-visual-words approaches. This method can be used

as an initial step for guessing the type of lesions a retinal image contains and then specific detection algorithms can be used for each type of lesion, increasing the performance of each specific detector by reducing the false positives.

The rest of the paper is organized as follows. We briefly review existing approaches for retinal images classification and discrimination in Section 2. Section 3 gives a detailed presentation of the proposed approach. In Section 4, different experiments are performed and the results are analyzed showing the good performances of our approach. Finally, concluding remarks are given in Section 5.

2. Related Works

Several approaches have been proposed in the literature for the detection of retinal features and lesions. The methods usually involve three main steps [5, 6, 7]. First, a pre-processing step is applied in order to compensate for the great variability between and within retinal images. The green channel is usually considered the most preferable choice, because it provides a maximum contrast between different retinal lesions and structures. The second step detects candidate regions that may correspond to retinal lesions and extracts features, while the last step classifies the candidate regions as retinal lesions or not.

Many of the methods proposed in the literature deal with the detection of a specific type of lesion, exudates or drusen for instance. A method to detect hard exudates from color fundus images was introduced by Garcia *et al.* [8]. The proposed method comprises a pre-prossessing step for luminos-

ity and contrast normalization, candidate regions extraction based on local adaptive histogram thresholding, and classification using a neural network classifier. Note that prior to classification, the optic disk is removed since its characteristics are similar to that of exudate lesions. The method was evaluated on a dataset of 117 images, and achieved a sensitivity of 100%, a specificity of 77.78%, and a mean accuracy of 91.04%. An automatic method for identification of hard exudates in retinal color fundus images was proposed by Osareh *et al.* [9]. In this method, after image normalization via histogram specification and local contrast enhancement, the images are segmented using Fuzzy C-means (FCM) clustering algorithm. Finally, a three layers perceptron neural network is used for classification. Using 75 color images for training and 67 for testing, the method achieves a sensitivity of 95%, and a specificity of 88.9% for the image based classification task. Giancardo *et al.* [10] proposed a method for the diagnosis of diabetic macular edema (DME) based on exudates detection. The method uses a SVM classifier trained with a single feature vector per image obtained through color and wavelet analysis. The authors also provided a new publicly available dataset (HEI-MED) consisting of 169 patients from various ethnic groups and different levels of DME, and the method achieves an AUC (area under curve) of 88%. A detection and classification approach of DME severity is introduced by Deepak and Sivaswamy [11]. The detection is based on a supervised learning of characteristics of normal images. Thus, any deviation from these normal characteristics is considered an indication of abnormality. The characteristics are computed from color and Radon features from different orientations. The method achieves a sensitivity of 100% and a specificity

between 74% and 90% on a dataset of 644 images from publicly available datasets. Recently, Ali *et al.* [12] proposed an exudate detection method using a statistical atlas. The method first build a statistical atlas using a set of normal color retinal images. The atlas captures the chromatic distribution of the retinal images for a given ethnic group. Then, any test image from the same group is warped to the atlas coordinates frame and a simple distance map between the atlas and the test image suppresses anatomical structures such as optic disk, vessels and macula, and gives a good segmentation of the lesions. This method has the advantage of avoiding complex pre-processing of the images.

Age Related Macular Degeneration (ARMD), which causes a gradual loss of central vision, is mainly caracterized by drusens. Hijazi *et al.* [13] proposed a method for drusen detection in retinal images based on angular and circular decomposition. The output of the decomposing step is a set of trees representing the images, and a weighted frequent sub-tree mining approach is used to determine the most frequent sub-trees. The weighted frequent sub-trees are then employed as features to train a classifier. Two classifiers are used for final classification, SVM and Naive Bayes. A total of 258 images from two different publicly dataset were used and the system achieved an accuracy of 100% and 95% with SVM and Naive Bayes respectively. Akram *et al.* [14] proposed an algorithm to automatically segment drusens in fundus images for ARMD diagnosis. The images are first pre-processed using morphological operations and contrast enhanced by applying an adaptive histogram equalization. Then, Gabor filters are used to select candidate regions from which color and texture features are extracted. Finally, a SVM is used for

classification. The method achieves a sensitivity, a specificity and an accuracy of 95%, 98.4% and 97% respectively, using the STARE dataset of 400 images. Zheng *et al.* [15] introduced a system that uses a two-step procedure for drusen detection. First, a pixel-wise classification is performed using color features and SVM to predict whether a pixel corresponds to drusen or not. Then, a group-wise classification is exploited to remove false positive components from the pixel-wise classification. The system is validated by comparing its output with manually segmented drusen on a pixel by pixel basis. The system achieves an accuracy between 80% and 86% depending on the dataset used. In [16], Garnier *et al.* proposed an ARMD detection method from fundus images using a multiresolution texture analysis. The texture is analyzed at several scales using a wavelet decomposition in order to identify all the relevant texture patterns, and an image is finally described with the textural pattern distributions of the wavelet coefficients images obtained at each level of decomposition. The method achieves an accuracy of 93.3% on a small dataset of 45 images.

As mentioned in [17], a common limitation of current algorithms for the detection and classification of bright lesions in retinal images is the need for complex and empirical pre- and post-processing steps depending on the lesion of interest. These include image size and illumination normalization [18], the segmentation and removal of blood vessels [19, 20], and the detection and removal of optic disc [21].

To avoid these pre- and post-processing steps, Rocha *et al.* [22] proposed a framework based on bags-of-visual-words (BoVW) representation which can be used for the detection of different lesions. This BoVW approach first

7

extracts low-level features from the retinal images and build a visual vocabulary, or codebook, by quantization of the low-level features. The visual vocabulary is then employed to code each image and the obtained representation in terms of visual words in the dictionary is used to train a classifier that can distinguish different retinal lesions. The method has shown promising results for detection of both hard exudates and microaneurysms [22], and has been the basis of future work addressing the task of retinal images discrimination [17]. In particular, the system achieved an AUC of 95.3 for white lesion detection and 93.3 for red lesion detection using a dataset of 1232 images.

Few methods were proposed for the task of discriminating between retinal images containing different types of bright lesions. Niemeijer *et al.* [23] proposed a method that can automatically detect bright lesions in retinal images and can differentiate among exudates, cotton wool spots, and drusen. The method first detects potential lesions pixels in the images using a set of 14 digital filters and a k-nearest neighbour (k-NN) classifier. These pixels are grouped into probable lesions regions from which features such as contrast, shape and size characteristics, and proximity to vessels are extracted. Finally, a linear discriminant analysis classifier is trained to classify bright lesions into exudates, cotton-wool or drusen. The system achieves a sensitivity of 95% and a specificity of 88% using a validation set of 300 images. Grinsven *et al.* [24] proposed an algorithm to automatically retrieve and classify images with bright lesions, namely drusen and exudates, using the BoVW approach. The algorithm initially partions the image into a fixed number of square patches and extracts a set of features from each patch. The features set

includes color histograms, histogram of oriented gradients (HoG) and local binary patterns (LBP). A visual dictionary is created using these features and classification is performed with a weighted nearest neighbor classifier. The method achieves an AUC of 0.9 for the classification task with a dataset of 415 images. Pires *et al.* [25] have also used the BoVW approach in order to identify images with bright lesions such as hard exudates, cotton wool spots, and drusen, in addition to images with red lesions like hemorrhages and microaneurysms. They created a visual vocabulary and trained a SVM detector for each type of lesion. Then, the results of the individual detectors are combined into a meta-classification step that indicates whether or not a patient should be referred to an ophtalmologist for further review. The system achieves an AUC score of 93.4 using a dataset of 1077 images. A similar approach is adopted by Sadek et al. [26] where authors create a dictionary for different features (LBP, HoG and SURF) and finally concatenate the obtained histograms representation into a unique feature vector. This representation is then used with a SVM classifier to distinguish between retinal images containing drusen or exudates. Experiments with a dataset of 430 images show a mean classification accuray of about 97.2%. A different approach for retinal images discrimination is proposed by Ujjwal *et al.* [27] based on visual saliency framework. They used a visual saliency detection method [28] to extract potential locations of abnormalities in retinal images, and regions with lesions are detected based on their saliency values and LBP features using a k-NN classifier. The method is reported to achieve a correct discrimination rate of 96.41% between respecicely 171 drusen and 217 hard-exudates images.

Although the previously described methods reported high correct bright lesions classification and discrimination rates, they still require some pre-processing step or some manual annotation of the images. Indeed, the BoVW approach [24] needs a prior knowledge about the location of the optic disk and the macula, while the method described in [25] requires manual annotation of unhealthy regions. The correct classification rate of the method of Ujjwal *et al.* [27] depends on the output of the visual saliency detector, and the method in [26] requires the removal of blood vessels. In this paper, we propose an automatic classification and discrimination method for retinal images that require no prior knowledge such as the locations of some retinal structures (optic disc or macula), and that can learn discriminant features from the images directly based on the sparse coding principle.

3. Methodology

This section describes the proposed method for the discrimination of retinal images containing bright lesions based on sparse coding techniques. Sparse signal representations are becoming increasingly popular and lead to state-of-the-art results in various applications such as face recognition [29], image denoising and impainting [30], and image classification [31]. The main reason being the intrinsic sparse nature of images representations when using fixed bases such as DCT or wavelets [32, 33]. In addition, the basis vectors can be learned from the data itself and be constrained to produce a sparse representation.

3.1. A brief introduction to sparse coding

The goal of sparse modeling is to find efficient representations of signals (or images viewed as two-dimensional signals) as a linear combination of a few typical patterns, called atoms, selected from a dictionary. Given a dictionary matrix $\mathbf{D} \in \mathbb{R}^{n \times K}$ that contains K atoms as column vectors $\mathbf{d}_j \in \mathbb{R}^n$, $j = 1, \ldots K$, the sparse coding problem of a signal $\mathbf{y} \in \mathbb{R}^n$ can be stated as finding the sparsest vector $\mathbf{x} \in \mathbb{R}^K$ such that $\mathbf{y} \simeq \mathbf{D}\mathbf{x}$. The problem is therefore to solve the following optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon, \tag{1}$$

where ϵ is the reconstruction error of the signal $\mathbf{y}$ using the dictionary $\mathbf{D}$ and the sparse code $\mathbf{x}$.

Alternatively, one can also solve the following optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \text{ subject to } \|\mathbf{x}\|_0 \leq \lambda, \tag{2}$$

where λ is a specified sparsity level.

The vector $\mathbf{x} \in \mathbb{R}^K$ contains the representation coefficients of the signal $\mathbf{y}$ w.r.t. the dictionary $\mathbf{D}$. As opposed to other representations such as PCA, in sparse coding one is looking for the sparsest vector $\mathbf{x}$, i.e. the vector with the smallest number of nonzero coefficients. In the above formulation, the l^0-norm (which is actually a pseudo-norm) counts the nonzeros entries of a vector.

Since exactly solving the above optimization problem is a NP hard problem [32], approximate solutions are obtained using greedy algorithms such as *matching pursuit* (MP) [34] or *orthogonal matching pursuit* (OMP) [35]. A

second class of methods relies on relaxation and replace the l^0-norm with an l^1-norm making the optimization problem a convex one that can be **solved** efficiently. Such methods are called *basis pursuit* (BP) [36].

A key issue for practical applications is the choice of the dictionary **D**. One can use a pre-determined and fixed dictionary as is the case with wavelets, curvelets, or steerable filters transforms. But, the pre-specified dictionary do not necessary well describe a given signal at hand. Hence, it is more appealing to learn the dictionary from a given set of training data. The K-SVD algorithm [37] is a standard unsupervised dictionary learning algorithm that generalizes the well known K-means clustering algorithm.

Given a set of training signals $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]$, where each $\mathbf{y}_i \in \mathbb{R}^n$, K-SVD jointly finds a dictionary $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K]$ and an associated sparse codes matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]$, $\mathbf{x}_i \in \mathbb{R}^K$, by solving the following problem:

$$\min_{\mathbf{D}, \mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2 \text{ subject to } \forall i, \ \|\mathbf{x}_i\|_1 \leq \lambda. \tag{3}$$

The K-SVD algorithms iteratively solves this optimization problem by alternating between computing the sparse code matrix $\mathbf{X}$ and the dictionary $\mathbf{D}$. Given $\mathbf{D}$, the sparse coding problem is solved using any pursuit algorithm (MP, OMP or BP). Given the codes $\mathbf{X}$, the dictionary is sequentially updated, i.e. one atom at a time, using singular value decomposition (SVD). This provides a very effective unsupervised dictionary learning technique. After learning the dictionary, each given signal $\mathbf{y}_i \in \mathbb{R}^n$ is represented by the corresponding feature vector (or sparse code) $\mathbf{x}_i \in \mathbb{R}^K$. This feature can then be used for recognition or classification tasks.

3.2. Image classification with sparse coded features

The first step in the process of using sparse representation for image classification is to learn a dictionary that captures strong and discriminative structures of the images. We use the K-SVD algorithm for dictionary learning [37].

3.2.1. Features extraction and representation

The dictionary is learnt from low-level features extracted from the training data. Several works have shown that applying sparse coding to local parts or descriptors of the images can capture higher-level features compared to raw image patches [38, 39]. In our work, we divide the image into square patches of fixed size and extract low-level features from each of the patches. Since the images used in the experiments are from different sources, as explained in Section 4, we resize them to a fixed resolution of 512×512 using bilinear interpolation, then we extract the following types of low-level features:

- **Color features:** color features are extracted in image patches of size 8×8. For an image size of 512×512, one thus have 4096 feature vectors extracted. Each of the feature vector is the concatenation of color histograms from normalized r, g and b components from RGB color space, and from h, s, v, Cb and Cr from HSV anb YCbCr color spaces respectively. We use histograms of size 8 in our experiments, so that, in total, every feature vector is of dimension 64.

- **SIFT:** dense SIFT features are extracted over a regular grid on the image. The grid points correspond to the corners of the square patches,

so we use a grid step size of 8 pixels. For every grid point, a SIFT descriptor of dimension 128 is extracted.

- **HOG:** histogram of oriented gradients (HOG) are extracted over a regular grid on the image, with a grid step size of 8 pixels. For every grid point, a descriptor of dimension 31 is extracted.

- **LBP:** local binary patterns (LBP) are extracted over a regular grid on the image, with a grid step size of 8 pixels. For every grid point, we extract a feature vector of size 58.

Color features capture the local color appearance of image regions and are good indicator of the presence of bright lesions which appear as small white or yellowish deposits in the retina. LBP features capture the local texture of image regions, while SIFT and HOG features are based on the local distribution of gradient information. Thus, they capture the local shape of the regions. Note that except for color features, all other features are extracted from the green channel of the fundus image. For each feature type, the features extracted from all training images are put as columns of a matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]$, where each $\mathbf{y}_i \in \mathbb{R}^n$ is a feature vector and m is the total number of local patches from the training images. We use the K-SVD algorithm, explained in Section 3.1, to learn a dictionary $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K]$, where K is the size of the dictionary.

Every image I can then be coded using the learnt dictionary $\mathbf{D}$. First, a set of low-level features is extracted from the image to form a feature matrix $\mathbf{F}_I \in \mathbb{R}^{n \times p}$, where p is the number of local patches. Given the dictionary $\mathbf{D}$, the sparse codes matrix $\mathbf{X}_I \in \mathbb{R}^{K \times p}$ is found that solves the optimization
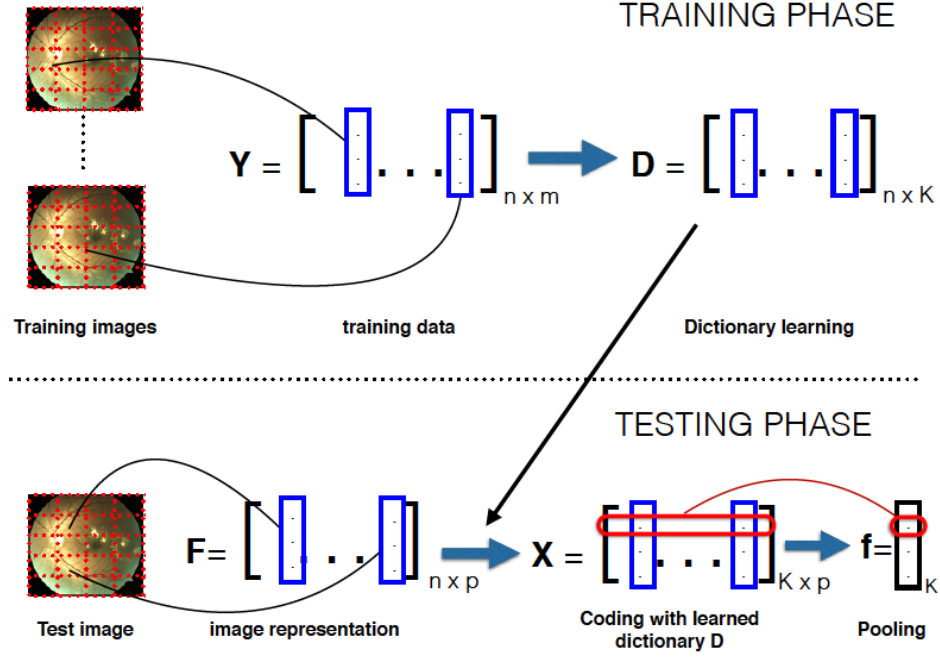
Figure 1: Flowchart of the feature extraction and coding method.

problem in Equation 1, so that $\mathbf{F}_I \simeq \mathbf{DX}_I$. Note that each column of $\mathbf{X}_I$ is a sparse vector that represent the corresponding feature vector from the image I, i.e. the corresponding column in $\mathbf{F}_I$. Finally, the set of sparse codes is combined into a single and global descriptor as follows [31]:

$$\mathbf{f}_i = \max_j(|\mathbf{X}_I(i,j)|), \quad \forall i = 1, \ldots, K. \tag{4}$$

Fig. 1 shows the overall flowchart of the proposed feature extraction method based on sparse coding.

3.2.2. Image classification

After dictionary learning, every image is sparse coded using the obtained dictionary and represented as a single feature vector as explained in Sec-

tion 3.2.1. These feature vectors are then used as input for training a SVM classifier. We used linear SVM as they can be implemented with linear complexity in training and testing phases [40].

3.2.3. Relation with the BoVW model

The sparse coding technique described above is closely related to the common bag-of-visual-worlds (BoVW) approach. In the BoVW model [41], a visual dictionary, or codebook, is created by quantization of the low-level features extracted from the training images. A traditional unsupervised method for computing the dictionary is to apply K-means clustering technique to the set of low-level features, and use the k centroids as codewords, i.e. elements of the visual dictionary. A new image is then represented by a vector indexed by the codewords: for each element d_i of the dictionary, ones counts the number of low-level features in the image whose closest codeword is d_i.

The visual dictionary in the BoVW model corresponds to the one obtained via sparse coding (Section 3.1). However, a key difference lies in the fact that the BoVW approach creates the dictionary from quantization of the low-level features directly and assigns each feature to the nearest codeword. Sparse coding techniques, on the contrary, encode each low-level feature as a weighted sum of dictionary elements, optimizing a tradeoff between reconstruction error and the l_1 norm of the reconstruction weights. This non-linear encoding scheme provides better representation of the images, as we will see in the experiments (Section 4).

4. Experiments and Results

In this section, we describe the different experiments performed to evaluate our method, and present the results obtained in discriminating between retinal images containing bright lesions, namely exudates and drusen, and normal retinal images which contain no lesions. In particular, we compare different low-level features and compare our proposed method using sparse coding with BoVW approaches used in [26, 24].

4.1. Dataset

We use in this study a dataset of 828 images with 452 normal images, 85 images with drusen and 291 images containing exudates. These images are sampled from six public datasets and one private database, listed in Table 1. The idea of using images from different datasets is to evaluate the robustness of the proposed method to images acquisition variabilty. Indeed, the images from the several datasets were collected in different environments with different cameras. The partition of the images into the three classes is given in Table 1, and some images are shown in Fig. 2. Note the variability in the images appearances, different illumination and contrast.

In all our experiments, we have used the KSVD-box[1] for sparse modeling, and the Vlfeat library[2] for features extraction.

---

[1]http://www.cs.technion.ac.il/∼ronrubin/software.html

[2]http://www.vlfeat.org

Dataset	Normal	Drusen	Exudate
ORNL[3]	36	61	20
HEI-MED[4]	20	-	26
STARE[5]	-	24	-
HRF[6]	15	-	-
DRIDB[7]	10	-	27
DRIVE[8]	20	-	-
MESSIDOR[9]	351	-	218
Combined	**452**	**85**	**291**

Table 1: Distribution of images from public datasets used in the experiments.

Figure 2: Example of images in the dataset. From left to right: a normal image, an image with drusen and an image with exudates.

[3]private dataset kindly provided by T. Karnowski from ORNL.

[4]see (`http://vibot.u-bourgogne.fr/luca/heimed.php`)

[5]see (`http://www.ces.clemson.edu/~ahoover/stare/`)

[6]see (`http://www5.cs.fau.de/research/data/fundus-images/`)

[7]see (`http://www.fer.unizg.hr/ipg/resources/image_database`)

[8]see (`http://www.isi.uu.nl/Research/Databases/DRIVE/`)

[9]see (`http://messidor.crihan.fr`)

4.2. Retinal images classification

In this section, we evaluate the performances of the proposed retinal images discrimination method. We perform different experiments, and use 10-fold cross validation. The dataset is split into 10 subsets, and the dictionaries are learnt with 9 of the subsets while classification is performed with the remaining subset. This process is repeated 10 times ensuring that each image is used both for training and testing. Thus, the method is tested with 10 different training and testing sets and the average classification performance is reported. The performance is evaluated in terms on classification accuracy, sensitivity and specificity. Since, we have three classes (normal, exudates and drusen) we compute the classification performance scores for each invidual class in a *one-vs-rest* manner. For example, for the drusen class we consider all drusen images as positive samples and all normal and exudates images as negative samples.

In the experiments, we vary the size of the dictionary from 10 to 1000, and we set the sparsity level (the parameter λ in Eq. 3) to 3 as this value gives the best classification results.

4.2.1. Normal-vs-rest

In this experiment, we test the performance of the proposed method in correctly identifying a normal image, i.e. a retinal image free of lesions. All 452 normal images in the dataset are used as positive examples and all drusen and exudates images are used as negative ones. The results obtained with the proposed sparse coding approach using SIFT features are given in Table 2. As can be seen, the classification performance increases with the size of the dictionary. For a medium size of 100, the method achieves an accuracy of

97.50%, a sensitivity of 96.50% and a specificity of 97.70%. For a larger dictionary of size 300, the method achieves near perfect classification results despite the high variability in the images which are selected from various sources. The method achieves an accuracy of 99.80%, a sensitivity of 100% and a specificity of 99.70%. This shows that the proposed method is able to correctly discriminate a retinal image containg no lesions from images with bright lesions.

Dictionary size	Accuracy	Sensitivity	Specificity
10	83.80 (±5.29)	56.50 (±9.83)	90.70 (±4.06)
20	88.20 (±4.24)	77.20 (±10.32)	91.00 (±4.27)
30	90.30 (±4.27)	79.20 (±17.81)	92.90 (±4.63)
50	93.70 (±3.71)	82.40 (±15.33)	96.60 (±3.17)
100	97.50 (±2.84)	96.50 (±5.76)	97.70 (±3.50)
200	99.40 (±0.97)	98.50 (±3.17)	99.7 (±0.95)
300	99.80 (±0.63)	100 (±0)	99.70 (±0.95)
500	99.60 (±0.84)	100 (±0)	99.5 (±1.08)
1000	99.80 (±0.63)	100 (±0)	99.7 (±0.95)

Table 2: Normal-vs-rest classification results with sparse coded features and SVM classifier. The results indicate the mean measure and standard deviation obtained with 10-fold cross validation.

4.2.2. Drusen-vs-rest

This second experiment aims at evaluating the performance of the proposed method in correctly identifying an image containing drusen. All 85 drusen images of the dataset are used as positive examples and all normal and exudates images are used as negative ones. The results obtained with the

proposed sparse coding approach using SIFT features are given in Table 3. As in the previous experiment, the classification performances increase as the dictionary size increases. However, for the identification of drusen images, near perfect results are obtained with a dictionary of medium size 100 with an accuracy of 99.80%, a sensitivity of 99.10% and a specificity of 100%. For a dictionary of size 300 or above, all drusen images are correctly identified and no other image is misclassified. The proposed method achieves an accuracy, a sensitivity and a specificity of 100%.

Dictionary size	Accuracy	Sensitivity	Specificity
10	91..90 (±3.38)	81.20 (±12.59)	95 (±2.79)
20	96.30 (±3.33)	87.90 (±15.44)	98.60 (±1.90)
30	96.60 (±2.41)	87.90 (±5.67)	98.60 (±2.37)
50	98.20 (±1.93)	92.40 (±9.57)	99.40 (±1.90)
100	99.80 (±0.63)	99.10 (±2.85)	100 (±0)
200	99.80 (±0.63)	99.20 (±2.53)	100 (±0)
300	100 (±0)	100 (±0)	100 (±0)
500	100 (±0)	100 (±0)	100 (±0)
1000	100 (±0)	100 (±0)	100 (±0)

Table 3: Drusen-vs-rest classification results with sparse coded features and SVM classifier. The results indicate the mean measure and standard deviation obtained with 10-fold cross validation.

4.2.3. Exudates-vs-rest

n this last experiment, we evaluate the performance of the proposed method in correctly identifying an image containing exudates. All 291 exudates images of the dataset are used as positive examples and all normal

and drusen images are used as negative ones. The results obtained with the proposed sparse coding approach using SIFT features are given in Table 4. As can be seen, the proposed method performs well in this experiment and is able to correctly identify retinal images containing exudates from other images. In particular, the method achieves an accuracy, a sensitivity and a specificity of 99.80%, 100% and 100% respectively, with a dictionary of size 300.

Dictionary size	Accuracy	Sensitivity	Specificity
10	88.30 ($\pm$6.73)	91.80 ($\pm$9.83)	83.4 ($\pm$11.37)
20	91.50 ($\pm$3.66)	92.10 ($\pm$10.32)	90.90 ($\pm$7.06)
30	93.30 ($\pm$3.68)	94.10 ($\pm$17.81)	91.50 ($\pm$9.18)
50	95.70 ($\pm$3.06)	97.70 ($\pm$15.33)	92.40 ($\pm$6.72)
100	97.0 ($\pm$2.54)	97.40 ($\pm$5.76)	98.20 ($\pm$3.05)
200	99.60 ($\pm$0.84)	100 ($\pm$0)	99 ($\pm$2.11)
300	99.80 ($\pm$0.63)	100 ($\pm$0)	100 ($\pm$0)
500	99.80 ($\pm$0.63)	100 ($\pm$0)	100 ($\pm$0)
1000	99.80 ($\pm$0.63)	100 ($\pm$0)	100 ($\pm$0)

Table 4: Exudates-vs-rest classification results with sparse coded features and SVM classifier. The results indicate the mean measure and standard deviation obtained with 10-fold cross validation.

A comparison of the classification results obtained for each one of the three classes is shown in Figures 3, 4 and 5. We can observe from these figures that, the performances obtained for the *normal* class (normal-vs-rest experiment) are slightly lower than those obtained for the *drusen* class (drusen-vs-rest experiment) and the *exudates* class (exudates-vs-rest exper-

iment). For example with a dictionary of size 100, the acurracy values are 97.5%, 99.8% and 97% for the *normal, drusen* and *exudates* classes respectively. The sensitivity values are 96.5%, 99.10% and 97.4% for the three classes, and the specificity values are respectively 97.7%, 100% and 98.2%. For larger dictionaries, size above 500 atoms, the method achieves perfect or close to perfect results for all three classes.
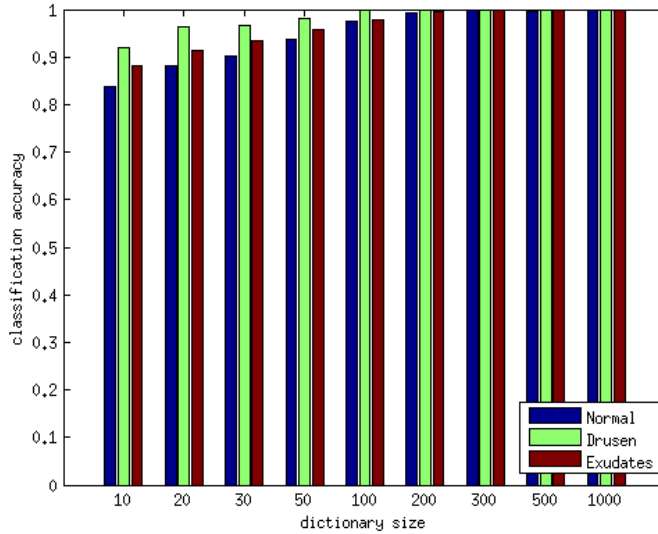

Figure 3: Classification accuracy with SIFT sparse coded features and SVM classifier.

4.3. Comparison of the different features

In this section, we evaluate the performances of the different feature types introduced in Section 3.2.1, namely SIFT, LBP, HOG and COLOR features. Again, we evaluate the performances of the classification method using different types of features for each individual class in a *one-vs-rest* manner. We use for this comparison a dictionary of size 100 (the main observations are
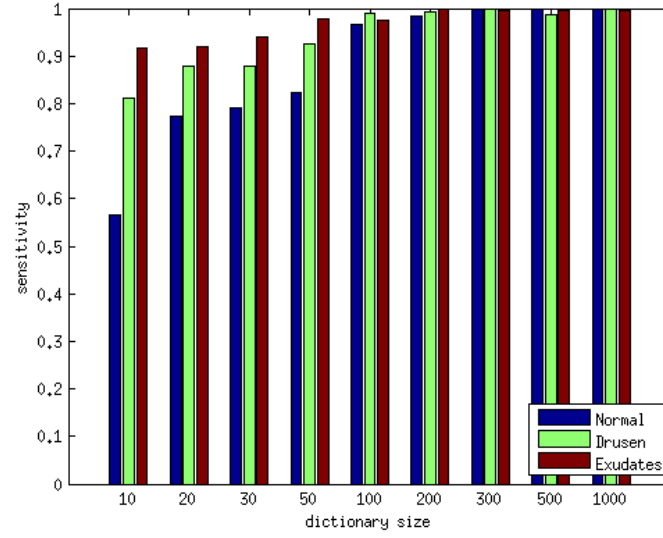
Figure 4: Sensitivity with SIFT sparse coded features and SVM classifier.
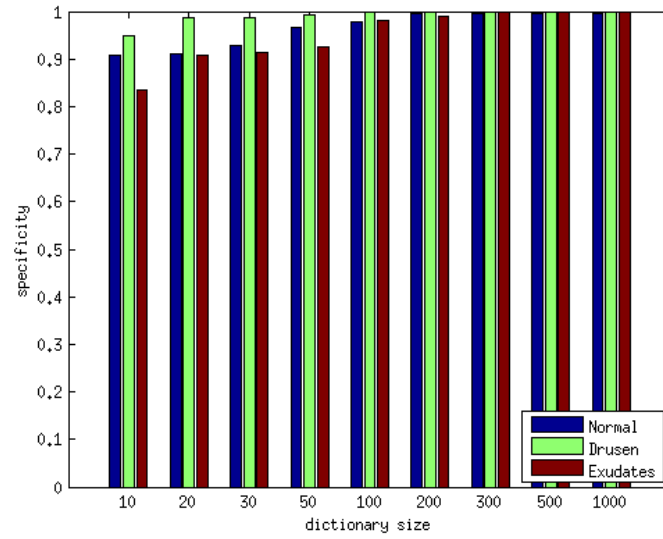

Figure 5: Specificity with SIFT sparse coded features and SVM classifier.

24

similar for varying dictionary sizes).

The results for the experiments are given in Table 5 and in Figures 6, 7 and 8. We can see from these results that sparse coding applied to SIFT features outperforms other feature types. The classification accuracy obtained with SIFT features is higher than the ones obtained with other features whose performances are comparable. We obtain an accuracy of 97.5%, 92.6%, 88.3% and 92.3% with SIFT, COLOR, LBP and HOG features respectively, for the *normal* class. For the *drusen* class, COLOR features achieve very good classification results but slightly below the results obtained with SIFT features. However, for the *exudates* class, the difference is more significantly in favor of SIFT features. For this class, using SIFT features we obtain an accuracy of 97.7%, a sensitivity of 97.4% and a specificity of 98.2%, while using COLOR features these figures are respectively 93.8%, 95.4% and 91.7%.

From Figures 6, 7 and 8, we can conclude that SIFT features perform best among all features for all three classes. In particular for the difficult class of normal images, SIFT achieves a sensitivity of 96.5%, while the second best feature COLOR achieves a sensitivity of only 80.6%. We can also see that all features provide slightly better results for the *drusen* and *exudates* classes in comparison with the *normal* class.

The results in Table 5 also show the standard deviation of the different performance measures for each feature type. As can be seen, the results obtained with SIFT features are more stable than results with other features as shown by lower standard deviation values.

		Feature type			
		COLOR	LBP	HOG	SIFT
Normal	Acc	92.60 (±5.42)	88.30 (±3.77)	92.30 (±3.13)	97.50 (±2.84)
	Sens	80.60 (±14.61)	75.10 (±16.35)	80.90(±17.88)	96.50 (±5.76)
	Spec	96.20 (±4.02)	92.10 (±4.63)	95.00 (±2.00)	97.7 (±3.50)
Drusen	Acc	98.00 (±2.36)	96.90 (±2.38)	95.90 (±2.56)	99.80 (±0.63)
	Sens	95.60 (±10.63)	91.60 (±9.26)	94.30 (±6.34)	99.10 (±2.85)
	Spec	98.20 (±1.99)	98.20 (±2.10)	97.00 (±2.58)	100 (±0)
Exudates	Acc	93.80 (±3.91)	88.30 (±2.87)	93.50 (±2.27)	97.70 (±2.54)
	Sens	95.40 (±2.59)	90.30 (±5.44)	93.50 (±3.63)	97.40 (±3.75)
	Spec	91.70 (±9.04)	86.40 (±7.66)	93.40 (±7.50)	98.20 (±3.05)

Table 5: Comparison of different features. For each class, we give the classification accuracy (Acc), the sensitivity (Sens) and the specificity (Spec). The results indicate the mean value and standard deviation obtained with 10-fold cross validation.

4.3.1. Combination of features

In this section, we propose to combine different feature types to improve the classification performances. Based on the results of previous section, where we observed that SIFT features give the best classification performances, we tested the combination of SIFT features with other features. More precisely, we tested the following combinations: COLOR+SIFT, LBP+SIFT, HOG+SIFT and LBP+HOG+COLOR.

The results in Figures 9, 10 and 11 show that combining SIFT features with other features types slightly improve the classification results. In particular, we observe a slight improvement when using SIFT combined with COLOR features. The classification accuracy increases from 97.50% with SIFT to 98.90% COLOR+SIFT for the *normal* class, and from 97.7% to
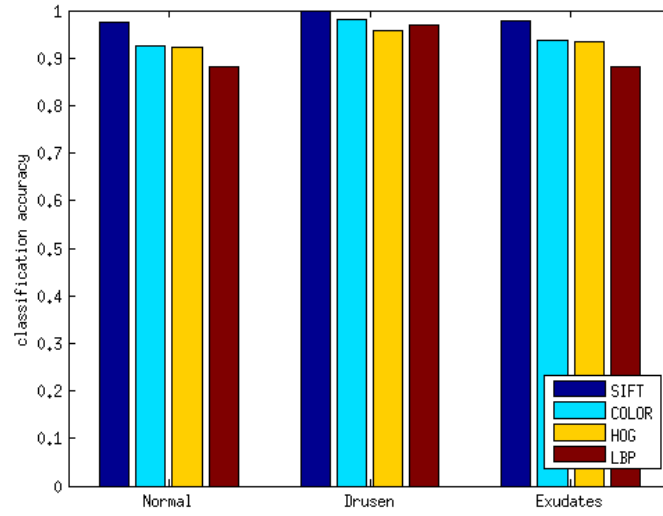
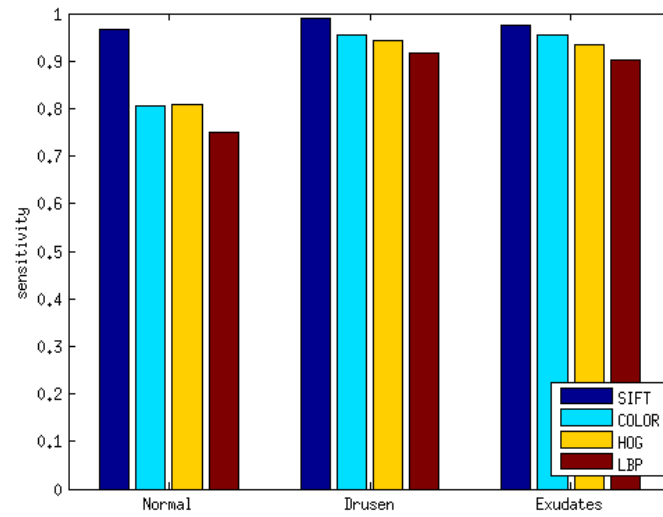Figure 6: Comparison of different features: classification accuracy.

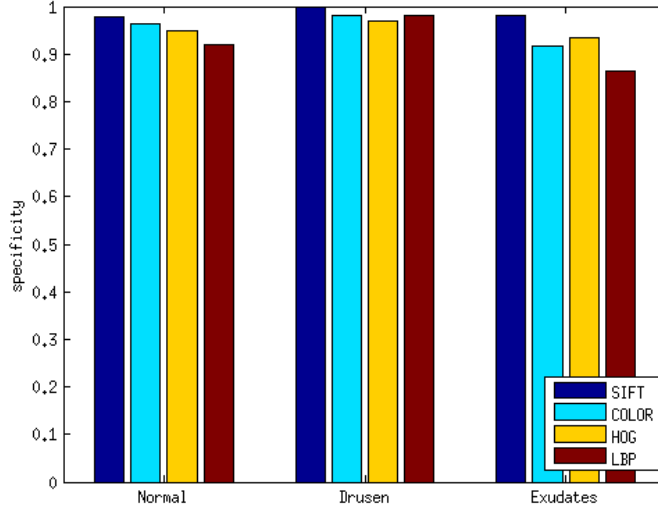Figure 7: Comparison of different features: sensitivity.

Figure 8: Comparison of different features: specificity.

99.8% for the *exudates* class. For the *drusen* class, the performances are the same when using a dictionary of size 100. The combinations LBP+SIFT and HOG+SIFT give similar results compared with SIFT features alone. It is also interesting to observe that the combination LBP+HOG+COLOR performs the worse achieving lower performance than SIFT alone. This is particularly true for the sensitivity measure (Figure 10) as LBP+HOG+COLOR gives a sensitivity of 80.4% for the normal class while SIFT achieves a sensitivity of 96.5% and COLOR+SIFT a sensitivity of 98.9%.

The good results obtained when combining COLOR and SIFT features can be explained by the fact that SIFT features are extracted from the green channel of the color fundus image, thus ignoring the chromatic information. Thus adding color information to SIFT features, the descriptor captures more information. However, as seen in the results, combining LBP, HOG and

28

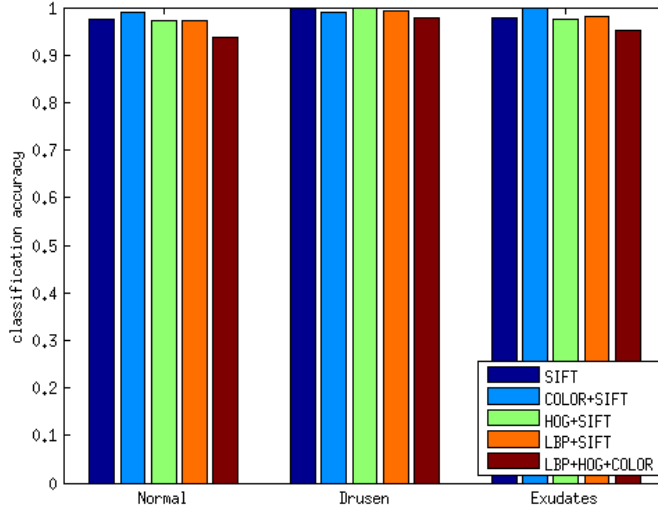COLOR features does not provide better performances compared with SIFT.

Figure 9: Features combination: classification accuracy.

4.4. Comparison with bag-of-visual-words methods

In this section we compare the proposed method based on sparse coding with the bag-of-visual-words (BoVW) approach. The BoVW approach is used by Grinsven *et al.* [24] and by Sadek et al. [26] to create a dictionary and represent an image by a histogram of visual words occurrences.

Table 6 shows the results of the comparison for the *normal* class. We can see that both methods perform well for this class with the same accuracy of 93.7% for a small dictionary of size 50, and a slightly better accuracy for sparse coding for larger dictionaries. With a dictionary of size 100, the sparse coding approach and the BoVW method give an accuracy of 97.50% and 95.30% respectively. However, there is a significant difference when
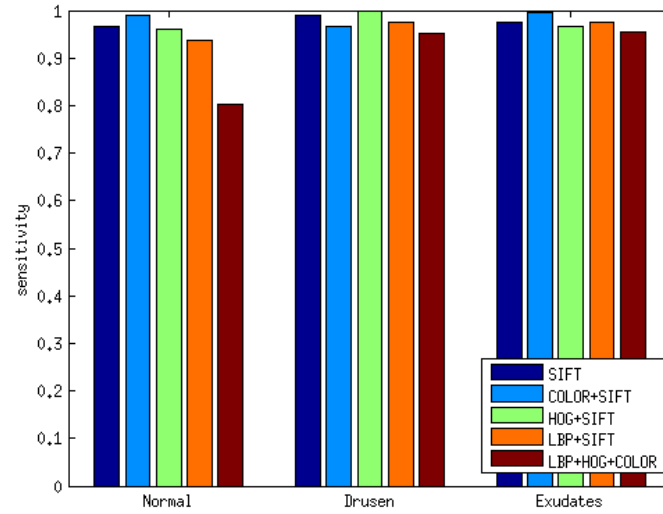
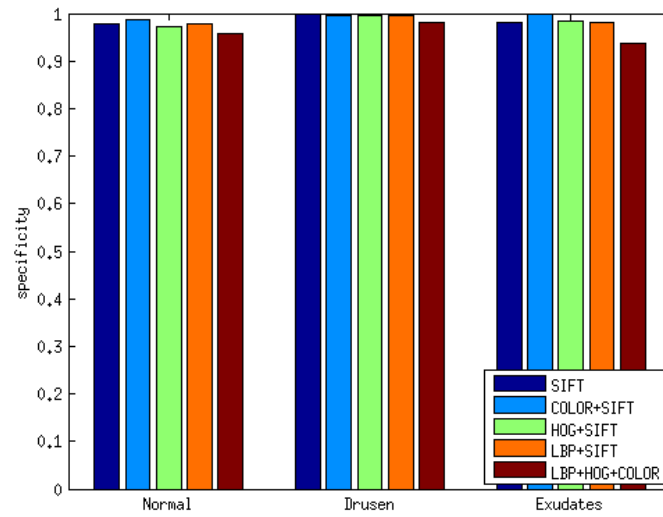29

Figure 10: Features combination: sensitivity.

Figure 11: Features combination: specificity.

		Dictionary size			
		50	100	500	1000
Proposed method	Acc	93.70 (±3.71)	97.50 (±2.84)	99.40 (±0.97)	99.80 (±0.63)
	Sens	92.40 (±5.33)	96.50 (±5.76)	98.50 (±3.17)	100 (±0)
	Spec	96.60 (±3.17)	97.70 (±3.50)	99.70 (±0.95)	99.70 (±0.95)
Bag-of-Words	Acc	93.70 (±2.58)	95.30 (±2.06)	97.20 (±2.04)	97.70 (±2.06)
	Sens	90.20 (±8.11)	87.30 (±12.59)	92.50 (±6.57)	92.20 (±12.04)
	Spec	94.60 (±3.50)	96.60 (±3.50)	98.20 (±1.55)	98.80 (±1.55)

Table 6: Comparison between the proposed method based on sparse coding and the BoVW approach for the *Normal* class, using SIFT features.

considering the sensitivity and specificity measures. For example, with a dictionary of size 100, the BoVW approach achieves a sensitivity of 87.30% while the sparse coding approach achieves a sensitivity of 96.50%. With a larger dictionary, 500 atoms, these two figures are respectively 92.5% and 98.5%. Thus, we can see that the sensitivity of BoW, for the normal class, reaches a plateau around 92% when the sensitivity value of sparse coding keep increasing to reach 100% with 1000 atoms. The same observations apply to the other classes and the results are shown in Tables 7 and 8. It is important to note that the classification results obtained with the proposed sparse coding approach are more stable than those obtained with the BoVW approach with respect to the training set used. This can be seen in the standard deviation values in Table 6, 7 and 8.

		Dictionary size			
		50	100	500	1000
Proposed method	Acc	98.20 (±1.93)	99.80 (±0.63)	99.80 (±0.63)	100 (±0)
	Sens	92.40 (±9.57)	99.10 (±2.85)	98.60 (±4.43)	100 (±0)
	Spec	99.40 (±1.90)	100 (±0)	100 (±0)	100 (±0)
Bag-of-Words	Acc	97.90 (±3.84)	98.80 (±1.03)	99.10 (±1.66)	99.80 (±0.63)
	Sens	90.40 (±8.42)	96.70 (±5.33)	98.20 (±5.69)	100 (±0)
	Spec	98.80 (±2.10)	99.10 (±1.45)	99.40 (±1.26)	99.70 (±0.95)

Table 7: Comparison between the proposed method based on sparse coding and the BoVW approach for the *Drusen* class, using SIFT features.

		Dictionary size			
		50	100	500	1000
Proposed method	Acc	95.70 (±3.06)	97.60 (±2.54)	99.80 (±0.63)	99.80 (±0.63)
	Sens	97.70 (±4.06)	97.40 (±3.75)	99.70 (±0.95)	99.70 (±0.95)
	Spec	92.40 (±6.72)	98.20 (±3.05)	100 (±0)	100 (±0)
Bag-of-Words	Acc	94.80 (±2.78)	95.70 (±2.45)	98.00 (±2.36)	97.90 (±2.18)
	Sens	94.00(±4.45)	96.00 (±3.20)	98.10 (±3.21)	97.80(±2.53)
	Spec	92.60 (±3.86)	94.90 (±5.32)	97.60 (±3.13)	97.90 (±3.41)

Table 8: Comparison between the proposed method based on sparse coding and the BoVW approach for the *Exudates* class, using SIFT features.

5. Conclusion

In this paper, we have proposed an automatic classification of retinal images that discriminates between images containing different bright lesions, namely drusen and exudates. The method is based on sparse signal representation techniques that capture strong structures of retinal images and

produce discriminant descriptors for classification. In particular, the proposed methodology does not require any pre-processing of the images such as blood vessels segmentation or optic disc removal, and achieves higher performence than other feature extraction schemes such as bag-of-visual-words approaches. We obtained excellent classification results with a dataset of 828 retinal images from various sources, showing the robustness of the approach to images acquisition variabilty such varying illumination, contrast and resolution. In particular, using a dictionary of size 100 and a linear SVM with sparse coded features, the method achieves a sensitivity and a specificity of 96.50% and 97.70% for the normal class; 99.10% and 100% for the drusen class; and 97.40% and 98.20% for the exudates class.

The proposed framework can easily be extended to include other type of lesions such microaneurysms which are subtle and difficult to distinguish. Another future work direction would be the identification of the precise location of the lesions as well as their size and quantity in order to estimate the degree of severity of the disease.

References

[1] Fong D.S., Aiello L.P., Ferris F.L., and Klein P., "Diabetic retinopathy", *Diabetes Care*, vol. 27 (10), pp. 2540-53, 2004.

[2] Wild, S., Roglic, G., Green, A., Sicree, R. and King, H., "Global prevalence of diabetes estimates for the year 2000 and projections for 2030", *Diabetes Care*, vol. 27(5), pp.1047-1053, 2004.

[3] Kertes P.J., and Johnson T.M., *Evidence-based eye care*, Lippincott William & Wilkins, 2007.

[4] Phillips P.J., "Visible manifestations of diabetic retinopathy", *Medecine Today*, vol. 5(5), pp. 83.

[5] Abramoff M.D, Garvin M.K., and Sonka M., "Retinal image analysis: a review", *IEEE Review Biomed. Eng.*, vol. 3, pp. 169-208, 2010.

[6] Usman Akram,M., Khalid, S., Tariq, A., Khan, S. A., and Azam, F., "Detection and classification of retinal lesions for grading of diabetic retinopathy", *Computers in Biology and Biomedicine*, vol. 45, pp. 161-171, 2014.

[7] Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J.P., al-Diri, B., Cheung, C.Y., Wong, D., Abrmoff, M., Lim, G., Kumar, D., Burlina, P., Bressler, N.M., Jelinek, H. F., Meriaudeau, F., Quellec, G., MacGillivray, T., and Dhillon, B., "Validation retinal fundus image analysis algorithms: issues and proposal", in *Investigative Ophthalmology & Visual Science*, vol. 54(5), pp. 3546-3569, 2013.

[8] Garcia, M., Sanchez, C., Lopez, M., Abasolo, D., and Hornero, R., "Neural network based detection of hard exudates in retinal images", *Computer Methods and Programs in Biomedecine*, vol. 93(1), pp. 9-19, 2009.

[9] Osareh, A., Mirmehdi, M., Thomas, B., and Markham, R., "Automated identification of diabetic retinal exudates in digital colour images", *Britich Journal of Ophtalmology*, vol. 87(10), pp. 1220-1223, 2003.

[10] Giancardo, L., Meriaudeau, F., Karnowski, T., Li, Y., Garg, S., Tobin Jr, K., and Chaum, E., "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets", *Medical Image Analysis*, vol. 16(1), pp. 216-226, 2012.

[11] Deepak, K.S., and Sivaswamy, J., "Automatic assessment of macular edema from color retinal images", *IEEE Tran. on Medical Imaging*, vol. 31(3), pp. 766-776, 2012.

[12] Ali, S., Sidibé, D., Adal, K.M., Giancardo, L., Chaum, E., Karnowski, T., and Meriaudeau, F., "Statistical atlas based exudate segmentation", *Computerized Medical Imaging and Graphics*, vol. 37(5-6), pp. 358-368, 2013.

[13] Hijazi, M., Coenen, F., and Zheng, Y., "Image classification for age-related macular degeneration screening using hierarchical image decompositions and graph mining", in *Proc. of ECML/PKDD, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pp. 65-80, 2011.

[14] Akram, M.U., Mujtaba, S., and Tariq, A., "Automated drusen segmentation in fundus images for diagnosing age related macular degeneration", in *Proc. of ICECCO, International Conference on Electronics Computer and Computation*, pp. 17-20, 2013.

[15] Zheng, Y., Vanderbeek, B., Daniel, E., Stambolian, D., Maguire, M., Brainard, D., and Gee, J., "An automated drusen detection system for classifying age-related macular degeneration with color fundus pho-

tographs", in *IEEE 10th International Symposium on Biomedical Imaging*, pp. 1448-1451, 2013.

[16] Garnier, M., Hurtut, T., Ben Tahar, H., and Cheriet, F., "Automatic multiresolution age-related macular degeneration detection from fundus images", in *Proc. SPIE Medical Imaging; Computer-Aided Diagnosis*, 2014.

[17] Pires, R., Jelinek, H.F., Wainer, J., Valle, E., and Rocha, A., "Advancing Bag-of-Visual-Words Representation for Lesion Classification in Retinal Images," in *PLoS One*, vol. 9(6), e96814, 2014.

[18] Foracchia, M., Grisan, E., and Ruggeri, A., "Luminosity and contrast-normalization in retinal images", *Medical Image Analysis*, vol. 9(3), pp. 179-190, 2005.

[19] Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., and Barman, S.A., "Blood vessel segmentation methodologies in retinal images-a survey", *Computer Methods and Programs in Biomedicine*, vol. 108(1), pp. 407-433, 2012.

[20] Moghimirad, E., Rezatofighi, S. H., and Soltanian-Zadeh, H., "Retinal vessel segmentation using a multi-scale medialness function", *Computers in Biology and Biomedicine*, vol. 42(1), pp. 50-60, 2012.

[21] Muramatsu, C., Nakagawa, T., Sawada, A., Hatanaka, Y., Hara, T., Yamamoto, T., and Fujita, H., "Automated segmentation of optic disc region on retinal fundus photographs: Comparison of contour modeling

and pixel classification methods", *Computer Methods and Programs in Biomedicine*, vol. 101(1), pp. 23-32, 2011.

[22] Rocha, A., carvalho, T., Jelinek, H.F., Goldenstein, S., and Wainer, J., "Points of interest and visual dictionaries for automatic retinal lesion detection", *IEEE Trans. on Biomedical Engineering*, vol. 59, pp. 2244-2253, 2012.

[23] Neimeijer, M., van Ginneken, B., Russell, S., Suttorp-Schulten, M., and Abramoff, M., "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis", *Investigate Ophtalmology & Visual Science*, vol. 48(5), pp. 2260-2267, 2007.

[24] van Grinsven, M., Chakravarty, A., Sivaswamy, J., Theelen, T., van Ginneken, B., and Sanchez, C., "A bag of words approach for discriminating between retinal images containing exudates or drusen", in *IEEE 10th International Symposium on Biomedical Imaging*, pp. 1444-1447, 2013.

[25] Pires, R., Jelinek, H., Wainer, J., Goldenstein, S., Valle, E., and Rocha, A., "Assessing the need for referral in automatic diabetic retinopathy detection," in *IEEE Trans. on Biomedical Engineering*, vol. 60, pp. 3391-3398, 2013.

[26] Sadek, I. Sidibé, D. and Meriaudeau, F., "Automatic discrimination of color retinal images using the bag of words approach," in *SPIE Medical Imaging*, 94141J-94141J-8, 2015.

[27] Ujjwal, Deepak, K.S., Chakravarty, A., and Sivaswamy, J., "Visual saliency based bright lesion detection and discrimination in retinal images," in *IEEE 10th International Symposium on Biomedical Imaging*, pp. 1436-1439, 2013.

[28] Hou, X., and Zhang, L., "Saliency detection: a spectral residual approach", in *Proc. IEEE International Conference on Computer Vision*, pp. 1-8, 2007.

[29] Wright J., Yang, A.Y., Ganesh, A., Sastry, S.S., "Robust face recognition via sparse representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31(2), pp; 210-227, 2009.

[30] Elad, M., and Aharon, M., "Image denoising via sparse and redundant representations over learned dictionaries", *IEEE Trans. on Image Processing*, vol. 15(12), pp. 3736-3745, 2006.

[31] Yang J., Yu K., Gong Y., and Huang T., "Linear spatial pyramid matching using sparse coding for image classification", in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1794-1801, 2009.

[32] Elad, M., *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[33] Wright J., Ma Y., Mairal J., Sapiro G., and Zisserman A., "Sparse representation for computer vision an pattern recognition", *Proceedings of the IEEE*, vol. 98(6), pp; 1031-1044, 2010.

[34] Mallat, S., and Zhang, Z., "Matching pursuits with time-frequency dictionaries", *IEEE Trans. on Signal Processing*, vol. 41(12), pp. 3397-3415, 1993.

[35] Davis, G., Mallat, S., and Zhang, Z., "Adaptive time-frequency decompositions", *Optical Engineering*, vol. 33(7), pp. 2183-2191, 1994.

[36] Chen, S.S., Donoho, D.L., Saunders, M.A., "Atomic decomposition by basis pursuits", *SIAM Reviews*, vol. 43(1), pp. 129-159, 2001.

[37] Aharon M., Elad M., and Bruckstein A., "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation", *IEEE Trans. on Signal Processing*, vol. 45(11), pp. 4311-4322, 2006.

[38] Kavukcuoglu K., Ranzato M.A., Fergus R., and LeCun Y., "Learning invariant features through topographic filter maps", in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1605-1612, 2009.

[39] Raina R., Battle A., Lee H., Packer B., and Ng A.Y., "Self-taugh learning: Transfer learning from unlabeled data", in *Proc. of International Conferance on Machine Learning*, pp. 759-766, 2007.

[40] Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., and Lin, C.J., "Liblinear: a library for large linear classification", *Journal of Machine Learning Research*, vol. 9, pp. 1871-1874, 2008.

[41] Sivic, J., and Zisserman, A., "Video Google: a text retrieval approach to object matching in videos", in *Proc. IEEE International Conference on Computer Vision*, pp. 1470-1477, 2003.