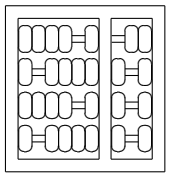


INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Points of Interest and Visual Dictionary for Retina Pathology Detection

Anderson Rocha *Tiago Carvalho*
Siome Goldenstein *Jacques Wainer*

Technical Report - IC-11-07 - Relatório Técnico

March - 2011 - Março

Points of Interest and Visual Dictionary for Retina Pathology Detection

Anderson Rocha* Tiago Carvalho Siome Goldenstein Jacques Wainer

Abstract

Diabetic retinopathy (DR) is a diabetes development that affects the retina's blood flow. The effect of DR is the weakening of retina's vessels, resulting on anything from small hemorrhages to the growth of new blood vessels. If left untreated, DR eventually lead to blindness, and, in fact, this is the leading cause of blindness in persons in the age range of 20 to 74 years in developed countries. One of the most successful means for fighting DR is early diagnosing through the analysis of *ocular-fundus* images of the human retina. In this paper, we present a new approach to detect retina-related pathologies from ocular-fundus images.

Our work is intended for an automatic triage scenario, where patients whose retina is considered not-normal by the system will see a specialist. This implies that automatic screening needs an evaluation criteria that rewards low false negative rates, i.e., we should avoid images incorrectly classified as normal as much as possible.

Our solution constructs a visual dictionary of the desired pathology' important features and classifies whether an *ocular-fundus* image is normal or a DR candidate. We evaluate the methodology on hard exudates, deep hemorrhages, and microaneurysms, test different parameter configurations, and demonstrate the robustness and reliability of the approach performing cross-data-set validation (using both our own and other publicly available data-sets).

1 Introduction

Diabetes Mellitus (DM) is a systemic, chronic, and life-threatening disease that, according to World Health Organization (WHO) [1] affects 135 million people worldwide, and may affect as many as 300 million by the year 2030 [2].

DM can cause micro and macro vascular changes. Diabetic Retinopathy (DR) is mainly the result of microvascular retinal changes triggered by diabetes. If not treated in time, DR can lead to the complete loss of sight. Recent reports account that about 25 thousand people with diabetes go blind every year in the US due to diabetic retinopathy [3]. According to Abràmoff et al. [3], in the US and Europe, DR is the major cause of blindness for the

*The authors are with the *Reasoning for Complex Data* (RECOD) Lab.,Institute of Computing, University of Campinas (UNICAMP), Campinas, Brazil. **Contacting author**: Anderson Rocha (anderson.rocha@ic.unicamp.br).

economically active population[1] and, according to a 2002 report, it is estimated that DR is responsible for 5% of all the world's blindness cases [4].

Early diagnostic, and treatment [5], of DR may prevent blindness, so systematic screening (by specialists) of diabetic patients is a cost-effective procedure to reduce DR-caused blindness [6].

This paper solution is based on dealing separately with different DR related anomalies, in this case, hard exudates, deep hemorrhages, and microaneurysms. The approach constructs a visual dictionary to represent important features of each anomaly, an approach inspired in the current computer vision literature. We validate our approach with a series of experiments on publicly available data, and compare the results against the state-of-the-art. Different than the related literature, we do cross-data-set validation, showing the robustness and reliability of introduced method.

The results described in this is paper are part of a larger project to develop, and deploy, automatic screening systems for DR in the Brazilian Universal Health System. The project is a concerned effort of two universities, and it is the result of a collaboration between several computer scientists, engineers, and ophthalmologists. Broadly speaking, we look for a reliable method to deploy systematic screening program in developing regions, where there may be a lack of on-site specialized doctors compared to the ever-growing number of patients [7].

With automatic screening, only patients with a tagged ocular-fundus retina image will see a specialist. The system does not give a diagnosis, since it is against the law for a non-medical doctor to do so in many places – this approach is an attempt to make the best of each doctor's time, and also allows for the remote image analysis and diagnosis on only the subset of the population most likely to have a pathology. This model of care requires a low false negative rate — the percentage of sick people that never get to see the specialist.

We start the paper with Section 2, where we briefly describe the main medical aspects of DR. Section 3 presents state-of-the-art achievements regarding DR diagnosing. Section 4 introduces our method, based on Points of Interest and Visual Dictionary for detecting pathologies on ocular-fundus images. Section 5 reports our experiments to compare our approach to the literature. Finally, Section 6 wraps up the paper and sheds light onto several considerations regarding directions for future work.

2 Diabetic Retinopathy into Context

The US National Eye Institute (NEI) categorize the progression of diabetic retinopathy into four stages [8, 9]:

1. **Mild Nonproliferative Retinopathy.** The first stage of DR, where the retina's tiny blood vessels present the first occurrences of localized balloon-like swelling (microaneurysms).

[1] In some developing countries (e.g., Brazil) this is not yet the case but it is rapidly going in this direction [4].

2. **Moderate Nonproliferative Retinopathy.** The second stage, where there is blockage of some of the retina's nourishing vessels.

3. **Severe Nonproliferative Retinopathy**. In this stage, there is blockage of several blood vessels, seriously compromising the retina's blood supply. Due to the insufficient nourishment, these areas send signals to the body to grow new blood vessels.

4. **Proliferative Retinopathy.** It is the most advanced DR stage in which the signals sent by the retina for nourishment trigger the growth of new blood vessels. These new vessels are fragile and highly prone to leaking which ultimately result in severe vision loss and blindness.

Figure 1 depicts an ocular-fundus image with the retina's main regions highlighted.

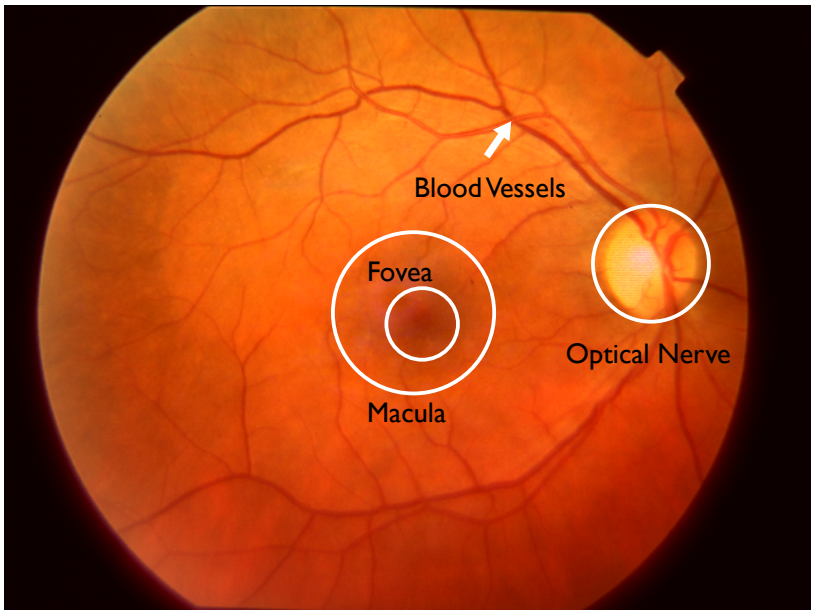


Figure 1: Ocular-fundus image and the retina's main regions.

2.1 DR-related Anomalies

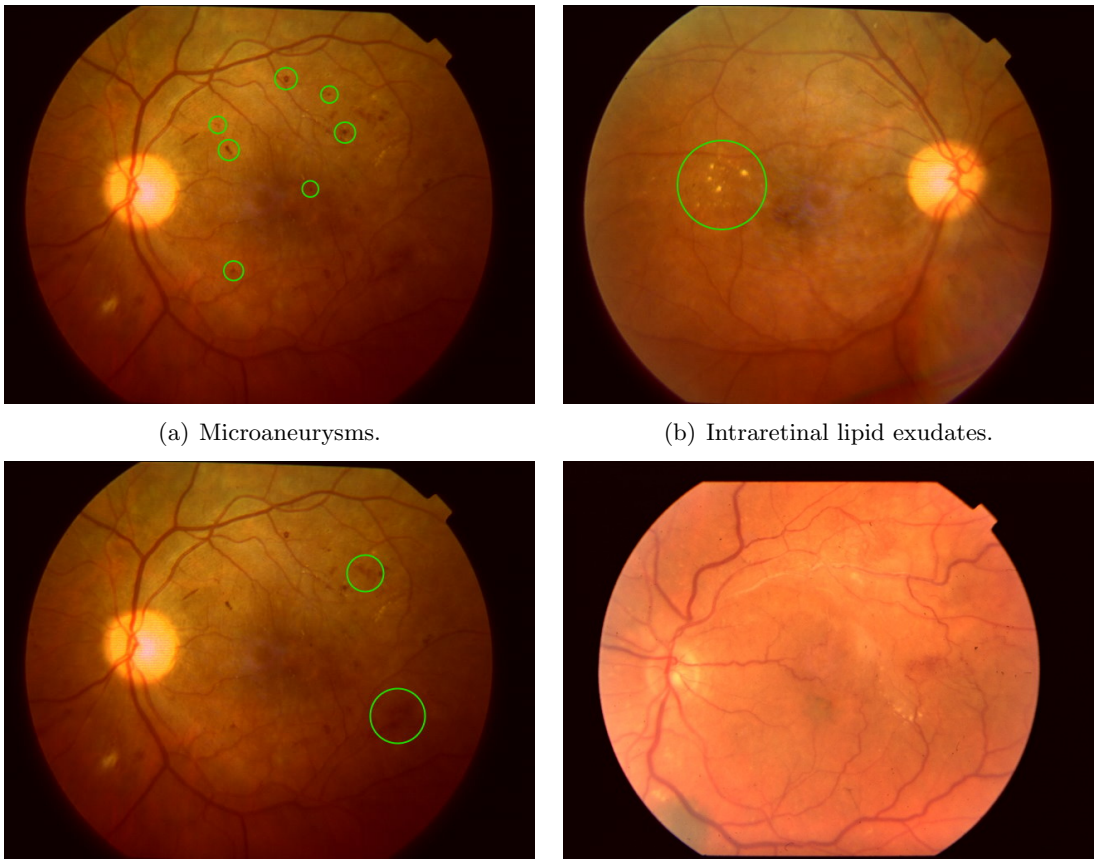
In the following, we discuss some of the DR-related anomalies which may be detected with the proper analysis of ocular-fundus images [10]:

- **Microaneurysms** are focal dilatations of retinal capillaries. Despite multiple layers of basements membrane, they are susceptible to the accumulation of water and lipid in the retina. (Figure 2(a)). They have an appearance similar to red dots in these images.

- **Intra-retinal lipid exudates (hard exudates)** reflect the breakdown of the blood-retinal barrier. DR triggers several physiological processes in the retina leading to the total incompetence of the blood-retinal barrier. Failures in the blood-retinal barrier enable access of fluid rich in lipids and proteins to the parenchyma, causing retinal edema and exudation (Figure 2(b)).

- **Hemorrhages.** As microaneurysm's or capillary's walls weaken, ruptures may occur and cause an intraretinal hemorrhage (Figure 2(c)).

- **Neovascularization** is the growth of new blood vessels, triggered in the retina when some of its regions are about to collapse due to the blocking of their nourishing vessels. These newly-formed vessels are fragile and grow out-of-control on the inner surface of the retina, eventually causing vision impairing (Figure 2(d)).



(a) Microaneurysms.

(b) Intraretinal lipid exudates.

(c) Hemorrhages.

(d) Neovacularization.

Figure 2: Examples of anomalies caused by diabetic retinopathy.

3 Diabetic Retinopathy Diagnosing — State-of-the-Art

In this section, we present an overview of the main contributions for automatic and semi-automatic detection of anomalies related to diabetic retinopathy in the literature. Most of the state-of-the-art approaches focus on single anomalies and, in most of these approaches, their extensions to other pathologies or a more general case are not straightforward or even possible. Only recently, there were the first approaches focused on the general *disease*×*non-disease* problem.

3.1 Detection of blood vessels

One of the classic approaches for the identification of the diabetic retinopathy in the literature is the detection and analysis of retina's blood vessels [11–14].

For instance, Chaudhuri et al. [12] used bidimensional matching filters to detect blood vessels in retinoscopy ocular-fundus images. Acharya et al. [11] employed several traditional image processing algorithms to isolate blood vessels in ocular-fundus images. After the identification and extraction of the blood vessels (Figure 3), the final decision is either handed to human specialists for further analysis or need further methods for issuing an automatic diagnosing.

3.2 Detection of Hard Exudates

Exudates are one of the most common diabetic retinopathy anomalies and has attracted a lot of attention from researchers. In retinoscopy images, exudates are characterized by yellowish regions with varying size and brightness. The size of these regions varies according to the stage of the patient's disease. Figure 4 depicts two ocular-fundus images — with and without exudates.

Exudates can lead to the loss of sight when present in the macula's center region. Its detection is important because its presence is highly correlated to the presence of other DR anomalies.

Welfer et al. [15] introduced an exudate detection method based on Mathematical Morphology (MM) operators for ocular-fundus images in LUV color space. Welfer et al's method uses traditional MM operators [16–18] such as morphological opening and closing, minima and maxima detection, and even watershed transforms.

Sopharak et al. [19] presented an approach for the same problem considering elementary image processing algorithms such as filtering and contrast enhancement [20]. The authors assume the conjecture that pixels close to exudate lesions are directly separable from normal pixels using only their intensity. The authors also use MM operators for local analysis of small pixel variations [19].

In another related work, Sopharak et al. [21] proposed a new exudate detection approach using fuzzy clustering and data analysis algorithms over a set of features carefully chosen by ophthalmologists: (1) pixel intensities after pre-processing operations; (2) standard deviations of pixel intensities; (3) hue; (4) number of border pixels and others. The authors used a set of mathematical morphology operations in the final stage to refine the results.

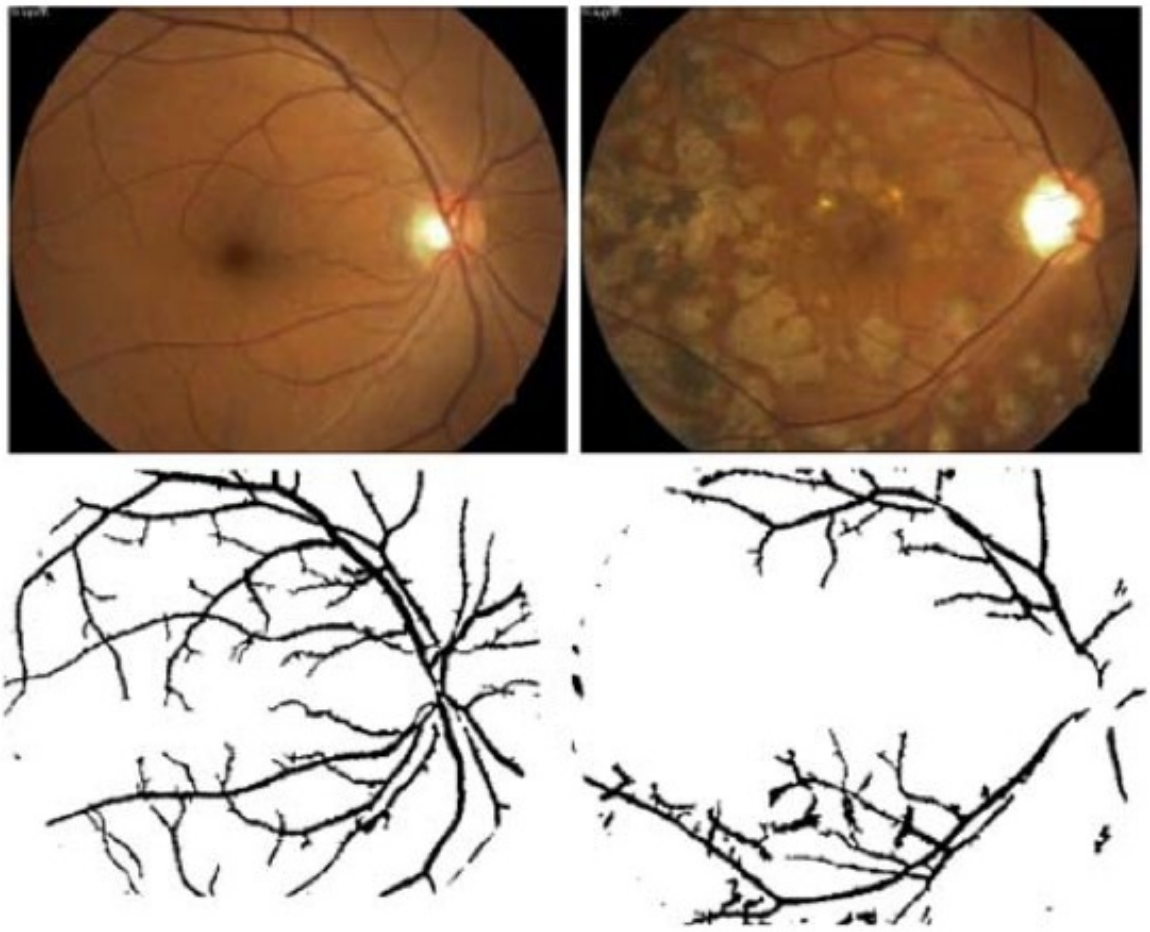


Figure 3: Blood vessels detection and highlighting in retinoscopy ocular-fundus images. A normal retina (left) and a retina under proliferative retinopathy (right). Images from [11].

Garcia et al. [22] introduced a different solution for exudate detection using classifiers and machine learning techniques. The authors used some image processing algorithms to find and isolate candidate regions of exudates. After detection, the authors characterize such regions using features such as average and standard deviation of RGB values within and outside the selected regions among other features. In the end, the method classifies the candidate regions into exudate or non-exudate ones using machine learning classifiers fed with the extracted features.

Fleming et al. [23] showed a method to detect exudates using mathematical morphology operators, multi-scale decomposition and analysis of probability maps. The authors obtain the exudate candidate regions after decomposing the input image into five scales. For each scale, the authors use dynamic thresholding [23] to eliminate candidate regions with brightness due to retina's reflectance. In the later stages of the method, the authors use a region

(a) Normal Retina. (b) Retina with Exudates.

Figure 4: A normal retina *vs.* a retina with *exudates*.

growing technique to determine the candidate regions. The proposed method characterizes each region using their pixels' normalized intensities, standard deviation, number of pixels, normalized border's gradient among others. The last stage of the method uses a *Support Vector Machine* (SVM) [24] classifier fed with the collected features to point out whether or not a given candidate region is an exudate one.

Sanchez et al. [25] introduced a method for detecting hard exudates using a set of pre-processing algorithms for enhancing image's overall luminosity and contrast. The method uses dynamic thresholding and mixture of statistical models in order to separate normal regions from regions with exudates. In the end, the authors use a post-processing sequence of edge-based operations to refine the results. The authors report 100% sensitivity for 90% specificity. Notwithstanding, the authors have used only 106 non-public images from which 26 were used for training and the remaining 80 were used for testing (40 with exudates and 40 normals). Another problem with this method is that it is better suited for images with a high quantity of exudates. The lower the amount of exudates in the images the worse the classification.

Dupas et al. [26] presented a pixel-based hard exudate detection method in ocular-fundus images. The authors' approach achieves 92.8% sensitivity while no specificity is reported. The authors validated the approach using only 30 images from which 15 contained exudates.

Sometimes, the classification results for an anomaly-based detection system can be highly improved when taking into consideration contextual information about the patients. Sanchez et al. [27] showed how to boost exudate detection results from 72% sensitivity and 70% specificity to 91% sensitivity when considering patients contextual information. The authors have used 144 images (69 with exudates and 75 normals) divided into two 72-image groups.

Table 1 summarizes the proposed methods in the literature for exudate detection. *Sensitivity* measures the proportion of actual positives which are correctly identified as such. In this case, the percentage of ocular-fundus images identified as having exudates. *Speci-*

ficity measures the proportion of negatives which are correctly identified. In our case, the percentage of normal ocular-fundus images which are correctly identified as not having exudates [24]. As we discussed, if the automated detection system is being used for triage, then *specificity* must be close to 100%, and sensitivity should be as high as possible.

Table 1: Exudate detection methods (Sensitivity *vs.* Specificity).

Technique	**Sensitivity**	**Specificity**	**# of images in the data set**	**Method Used**
Sopharak et al. [19]	80%	99.5%	60	Mathematical Morphology operators
Yun et al. [28]	80%	99.5%	124	Mathematical Morphology operators and Neural Networks
Welfer et al. [15]	70.5%	98.8%	89	Mathematical Morphology operators
Wang et al. [29]	100%	70%	154	Brightness enhancement and statistical classifiers
Garcia et al. [22]	88%	84%	117	Neural Networks and Support Vector Machines over image patches
Sopharak et al. [21]	87.3%	99.3%	60	Mathematical Morphology operators and fuzzy clustering data analysis
Fleming et al. [21]	95%	86.6%	13,219	Mathematical Morphology operators and multi-scale decomposition
Sanchez et al. [25]	100%	90%	106	Dynamic thresholding and mixture of statistical models combined with image pre- and post-processing operators
Dupas et al. [26]	92.8%	—	30	Pixel-based classification
Sanchez et al. [27]	91.0%	70%	144	Common image processing associated with patient's contextual information

3.3 Detection of Microaneurysms

Microaneurysms are present since the early stages of diabetic retinopathy [30]. Jalli et al. [31] analyzed the appearance of microaneurysms using the fluorescein angiography technique. Jelinek et al. [32] introduced an automatic approach to detect microaneurysms using optometrics features in retinoscopy ocular-fundus images. The authors reported 85% of sensitivity and 90% specificity.

3.4 Detection of Hemorrhages

Hemorrhages are a diabetic retinopathy anomaly normally present in the later stages of the disease. The higher the amount of retina's hemorrhage, the higher is the progression of the damages to the retina. There are two general approaches in the literature for hemorrhages detection from ocular-fundus images:

- Detection of blood vessels;
- Detection of blood vessels with hemorrhages.

Sinthanayothin et al. [33] introduced a method for hemorrhages detection based on a custom-tailored operator called *moat operator* and traditional image processing algorithms such as pixel's intensity normalization, thresholding, and region growing. The authors reported 77% sensitivity and 89% specificity.

Zhang e Chutatape [34] presented a *top-down* & *bottom-up* technique to detect hemorrhages aimed to be robust to illumination changes. The authors use contrast enhancement algorithms with region growing approaches to delineate candidate regions. Later on, the method groups the image pixels through fuzzy clustering and classifies the groups into lesion *vs.* non-lesion *vs.* exudates *vs.* hemorrhages regions using hierarchical Support Vector Machines.

4 Visual Dictionaries for DR Detection

The main problem with single anomaly-oriented solutions so far has been how hard it is to extend these approaches to more general problems. Most of the times, exudate detection custom-tailored approaches trained with one data set will not work in another data set, allowing some doubt regarding the progress of DR detection. Furthermore, it is hard to compare the existing approaches in the literature, as most of them use private data sets for training and testing, and rarely attempt cross-data-set validation. In this paper, we address both of these problems.

We introduce a new method for single pathology detection that is easy to extend, testing it on the detection of hard exudates, deep hemorrhages, and microaneurysms. Also, we show experimentally how the method allows training in one data set and testing across different data sets while keeping a very high specificity and a high sensitivity rates.

Our detection methodology uses retinoscopy ocular-fundus images and the concept of visual dictionaries [35–37]. This relative new paradigm in the computer vision literature uses a set of highly extensible feature representation, and achieves high detection/classification results without pre- or post-processing in the analyzed images.

Visual dictionaries constitute a robust representation approach where each image is treated as a collection of regions. In this representation, the only information we care about is the appearance of each region [37].

Our objective when creating a visual dictionary is to learn, from a training set of examples, the generative model [38] that selects the d more representative regions for a given problem. The number of selected regions, d, must be large enough to distinguish relevant changes in the images, but not so large as to distinguish irrelevant variations such as noise [39, 40]. These regions create a d-dimensional [41] Hilbert space $\mathcal{H}$, in which each region is now represented by a visual word [42].

Given a visual dictionary, we represent an image according to the visual words it contains. We map the original input image regions ϕ to a Hilbert space $\mathcal{H}$ represented by the calculated visual words. One of the main challenges we face in this new scenario is to create a representative dictionary that captures all the nuances of a given problem.

4.1 Overview of the Approach

Recently, several researchers have been using features around locally invariant interest points. Although originally developed for large baseline correspondence applications, there are some attempts for image retrieval and classification [43–50].

The principle behind most of these approaches is that one can represent every image in a collection using a large number of points of interest (PoI). It is possible to calculate a local descriptor around each PoI, and store them in an indexing data structure [51, 52].

The hypothesis of this approach is that the PoIs convey more information than the other points in the image. Therefore, PoIs can be robustly estimated, even if the image suffers distortions – the major criterion of quality for a PoI algorithm is repeatability [51].

After locating them, each PoI is described by the analysis of a small patch around it. The literature has shown that local descriptors computed around points of interest are more robust to represent image's nuances than global descriptors [43–49]. However, this representative power comes with an advantage and a drawback. When searching for a specific target, this discriminative power is extremely important. Notwithstanding, when searching for complex categories, it is a problem since the ability to generalize becomes paramount. Therefore, as these solutions are often designed for exact matching, they do not translate directly into good results for image classification.

A possible solution to this problem is the technique of visual dictionaries which considers the high-dimensional descriptor space and split it into multiple regions. Usually, one uses a non-supervised learning technique (e.g., clustering) for this task in order to find the most discriminative points of interest. Each region of PoIs, becomes a visual "word" of a "dictionary".

After the creation of the vocabulary, one summarizes each image of the collection analyzing each of its PoIs and assigning them the closest word in the dictionary. In the end, each image is represented by a set of visual words [36, 51, 52].

With this simple idea, the biggest challenge is to design a good dictionary. The creation of the dictionary requires the quantization of the description space which can be done using clustering approaches or, as we discuss in this paper, by bringing in the specialists to "select" the important words.

4.2 Local Features

To represent the visual content of a given image, we find a set of points of interest in such images and characterize their surrounding regions. It is desired to choose scale-invariant interest points in order to achieve a representation robust to some possible image transformations. To do so, we can use several different approaches.

We have found out that *Speeded-Up Robust Features* (SURF) [53] performs best. We also have experiments evaluating the performance of the *Scale-Invariant Features Transform* (SIFT) [43]. Both of these methods achieve high repeatability and distinctiveness.

4.2.1 Scale-Invariant Features Transform (SIFT)

SIFT algorithm is one of the most robust methods under translations, scale and rotation transformations [43].

The SIFT algorithm has four major steps.

1. **Scale-space extrema detection**: in this stage, the algorithm searches for candidate points invariant to scale changes. For that, the method looks for stable features in a scale-space composed of differences of Gaussians (DoGs) of progressively larger standard deviations [43].

2. **Feature point localization**: Scale-space extrema detection produces many keypoint candidates stable and some unstable ones. In this stage, the algorithm performs a detailed fit to the nearby data for accurate location, scale, and ratio of principal curvatures. This stage allows the elimination of low contrast points (sensitive to noise) as well as poorly localized points along an edge. The remaining points are called points of interest (PoI).

3. **Orientation assignment**: this stage assigns each PoI one or more orientations based on local image gradient directions. This is important to achieve invariance to rotation as the PoI descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.

4. **PoI characterization**: in order to achieve invariance to rotation, this stage selects a patch around each PoI and rotates such patch towards the most frequent direction of the gradient. In addition to invariance to rotations, the method also uses several normalizations and thresholds to become more robust to illumination changes.

4.2.2 Speeded-Up Robust Features (SURF)

SURF algorithm is based on the Hessian matrix, but uses a very basic approximation, just as DoG is a very basic Laplacian-based detector. The descriptor, on the other hand, uses a distribution of Haar-wavelet responses within the interest point's neighborhood.

The SURF algorithm has four major stages:

1. **Feature point detection**: this stage uses an Hessian detector approximation based on low-pass box filters (Haar filters) [54]. In addition, the method uses integral images [55] to speed up the involved operations.

2. **Feature point localization**: rather than using a different measure for selecting the location and the scale (as done by the Hessian-Laplace detector [56]), SURF uses the determinant of the Hessian for both location and scale. Given a point $x = (x, y)$ in an image I, the Hessian matrix $H(x, \sigma)$ in x at scale σ is defined as follows

$$H(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}, \tag{1}$$

where $L_{xx}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order derivative with the image I in point $\mathbf{x}$.

Usually, scale-spaces are implemented as image pyramids. The images are repeatedly smoothed with a Gaussian and subsequently sub-sampled in order to achieve a higher level of the pyramid. Given that SURF uses box filters and integral images, the method does not have to iteratively apply the same filter to the output of a previously filtered layer. Instead, the method applies the filters of any size at exactly the same speed directly on the original image [53].

To localize the interest points in the image across different scales, the method performs non-maximum suppression in a $3\times3\times3$ neighborhood. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space.

3. **Orientation assignment**: to assign an orientation to each interest point, SURF calculates the Haar-wavelet responses in x and y directions using a circular neighborhood of radius $6s$ around the interest point, where $s = \sigma$ is the scale at which the interest point was detected. In each scale s, the method also calculates the wavelet responses using integral images for fast filtering. The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window covering an angle of π. The longest such vector lends its orientation to the interest point [53].

4. **PoI characterization**: In this stage, SURF creates a square region centered around the interest point, and oriented along the orientation selected in the previous stage. The region is split up regularly into smaller 4×4 square sub-regions. For each sub-region, the method computes a few simple features at 5×5 regularly-spaced sample points.

4.3 Visual Vocabulary

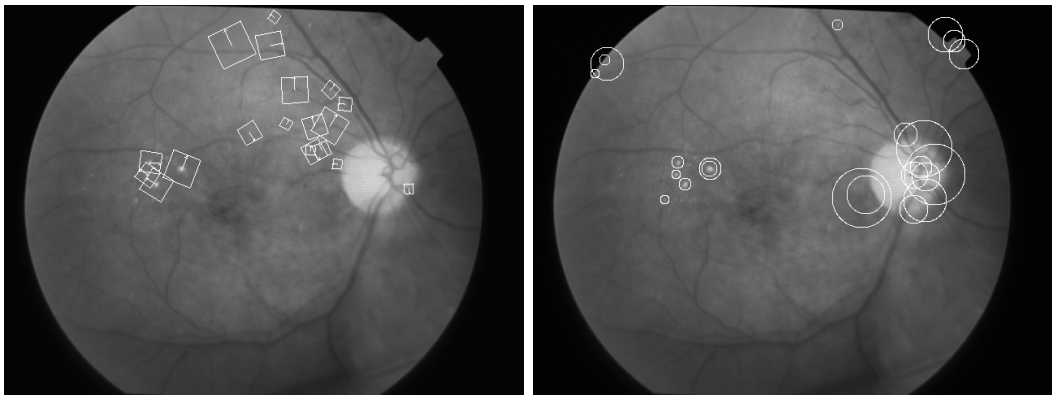
As we have discussed in Section 4.1, SURF and SIFT are good low-level representative feature detectors. However, this distinctiveness power comes with a price: as these solutions are often designed for exact matching, they do not translate directly into good results for image classification in broad or even constrained domains.

In our case, these approaches are not well suited for direct use when classifying exudate and normal retinoscopy ocular-fundus images, for instance. To preserve the distinctiveness power of such descriptors while increasing their generalization, we use the concept of visual vocabularies.

In the construction of a visual vocabulary, each region of PoIs becomes a visual "word" of a "dictionary". In the following, we consider the problem of exudate detection for the sake of explanation. The approach we introduce in this paper is general enough to detect other DR-related anomalies as we show in Section 5. To solve the problem of detecting exudate in ocular-fundus images, we select and create a database of training examples comprising training positive images with exudates and negative images considered normal by specialists. In this training stage, we perform the localization of the interest points in

all available images using either SIFT or SURF. Note that we do not perform any pre-processing on the images.

Each image in the training generates a series of points of interest as Figure 5 depicts. The figure shows SIFT PoIs in an ocular-fundus image in contrast to SURF PoIs.



(a) Points of Interest – SIFT. (b) Points of Interest – SURF.

Figure 5: Points of Interest – SIFT *vs.* SURF.

After finding the PoIs, we need to create the dictionary or codebook representing distinctive regions of images with exudates as well as images tagged as normal by specialists. For that, we need to choose the size (number of words) k of the dictionary.

To create the dictionary, we can perform clustering such as k-means for finding representative centers for the cloud of PoIs. Another option is to bring specialists to the loop and ask them to roughly mark regions in the ocular-fundus images with normal and exudate regions and, later on, to select PoIs within such tagged regions as representative words of the codebook.

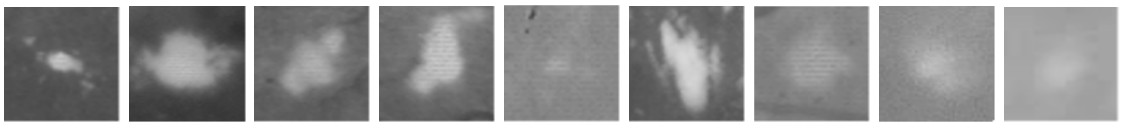
Figure 6 depicts an example of a set of visual words representing exudates and retina's normal structures.

4.4 Training and Classification

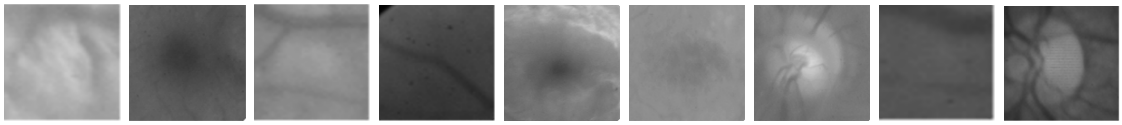
From the training images, we create the visual dictionary using one of the following approaches: random selection, clustering, and coarse or fine selection of descriptive regions.

In the coarse selection approach, we consider images manually graded by retina specialists and roughly mark candidate regions in normal images as well as images with exudates. Then we select only feature points lying within such regions. The number of normal and exudate regions are the same.

In the fine selection approach, after calculating the feature points over manually graded images by retina specialists, we perform a fine selection of feature points lying within a marked region. The difference between coarse and fine selection is that in the coarse approach there is a coarse marking around a candidate region while in the fine selection a feature point is selected if and only if it lies 100% over an exudate region or a region with



(a) Visual Words for *Exudate* Regions.



(b) Visual Words for *Normal* Regions.

Figure 6: Visual words example.

no exudates at all. The selection process is performed only in the training stage of the proposed approach.

After creating the dictionary, we perform the assignment of each of the training images' PoIs to the closest visual word of the dictionary. This step is known as *quantization*. In the end of the quantization process, we are left with a set of feature vectors representing the histogram of the selected visual words for each image.

To perform the final classification procedure, we select a two-class machine learning classifier such as *Support Vector Machine* (SVM) [24]. For training the classifier, we feed it with feature vectors calculated using the training images containing positive (e.g., images containing exudates) and negative (normal images) examples.

Figure 7 depicts the sequence of steps of the proposed approach (training and testing).

4.5 Approach's Extension

One of the most important advantages of the approach discussed in this paper is that it does not need to perform any pre- or post-processing stages when classifying the images. Thus, we can use this very same approach to solve other diabetic retinopathy detection problems such as: detection of deep hemorrhages and microaneurysms as we show in Section 5.

5 Experiments and Validation

In this section, we present the experiments we have performed to validate the technique we introduce in this paper.

We designed the experiments in three rounds:

- **Round #1**. In this round, we discuss important parameters regarding the introduced technique such as the number of words representing the classification dictionary, the most appropriate region descriptor (e.g., SIFT *vs.* SURF) and the policy for selecting the representative words such as cluster-, random- or user-steered (coarse or fine region selection). We perform the experiments in this stage using the DR1 data set (c.f., Sec. 5.1).

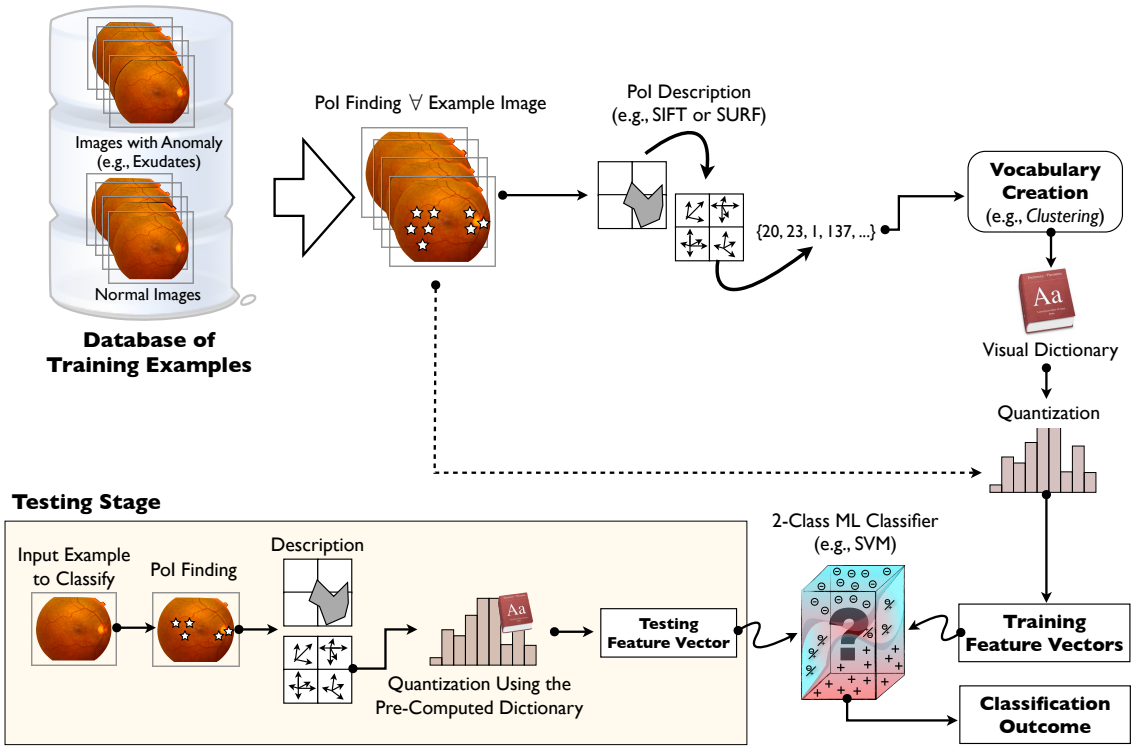


Figure 7: Sequence of steps for classifying DR anomalies (e.g., exudates) from retinoscopy ocular-fundus images.

- **Round #2**. In this round, we evaluate the cross-data-set effect on the our approach. For that, we train the proposed approach in one data set and test in another one. We perform experiments and report cross training and testing results for two very common data sets in the literature: *RetiDB* (c.f., Sec. 5.1) and *Messidor* (c.f., Sec. 5.1).

- **Round #3**. In this round, we show results for our approach for detecting deep hemorrhages and microaneurysms in ocular-fundus images. It is important to emphasize that we do not use any kind of pre- or post-processing when performing these new tasks. For these experiments, we use the DR1 data set (c.f., Sec. 5.1).

The results of these experiments are showed as ROC curves. The ROC curve is generated by retraining the SVM with different values for the cost C parameters for the two classes, positive and negative. All the reported results consider the average of 5-fold cross-validation procedures.

5.1 Data sets

In this paper, we perform experiments on three different data sets.

- **DR1**: this is a data set from the Ophthalmology Department of Federal University of São Paulo (Unifesp), collected across several months. The data set comprises 2,307 images from which 687 are normal. Of the anormal retinas, 264 images contain exudates, 79 images have microaneurysms, and 170 images have deep hemorrhage (the remaining anomalies in this data set are not relevant for this paper). All the images in this data set were manually annotated by three medical specialists. The average resolution of the images is 640×480 pixels. To allow researchers to perform fair comparisons with respect to the approach we discuss in this paper, we will make this data set public available[2].

- **RetiDB**[3]: is a public database for benchmarking diabetic retinopathy detection from ocular-fundus images. The data set comprises 22 normal and 71 images with exudates. The images resolution id $1,500 \times 1,152$-pixels.

- **Messidor**[4]: is a public data set of ocular-fundus images, in three resolutions ($1,440 \times 960$, $2,240 \times 1,488$, or $2,304 \times 1,536$ pixels). The data set comprises 547 normal and 226 DR images.

5.2 Experiments – Round #1

In this section, we discuss important aspects of our technique for exudate detection from ocular-fundus images such as: the number of visual words representing the reference dictionary, the most appropriate region descriptor (e.g., SIFT *vs.* SURF) and the policy for selecting the representative words such as cluster-, random- or user-steered.

5.2.1 SIFT *vs.* SURF and the size of the dictionary

Figure 8 shows results for the first experiment. In this experiment, we show the effectiveness of using SIFT and SURF descriptors as the basis for generating the visual descriptive words. We also show the number of necessary words for creating the reference dictionary in order to analyze new unseen ocular-fundus images.

With this experiment, we conclude that the SURF descriptor is more effective for representing the visual descriptive words. Another interesting conclusion is regarding the number of necessary words for the reference dictionary. Some of the image and object categorization literature [36,57,58] argue that larger dictionaries provide better classification results. This might be the case for complex categories as well as for large multi-class scenarios, but it was not the case in the two-class classification problem we deal with in this paper.

Contrary to what has been reported in the literature for image and object categorization, a very large dictionary does not help us any more than a dictionary of only 100 representative words. We obtain the best exudate detection results considering the SURF descriptor and 100 visual words. In this case, we get 80% sensitivity while paying the price of just 10% of false negatives (90% specificity). In this experiment, the visual words are randomly selected.

[2]Please visit `http://www.ic.unicamp.br/~rocha/pub/ communications.html`

[3]Kindly provided by the RetiDB program partners (see `http://www2.it.lut.fi/project/imageret/diaretdb1/`)

[4]Kindly provided by the Messidor program partners (see `http://messidor.crihan.fr`)

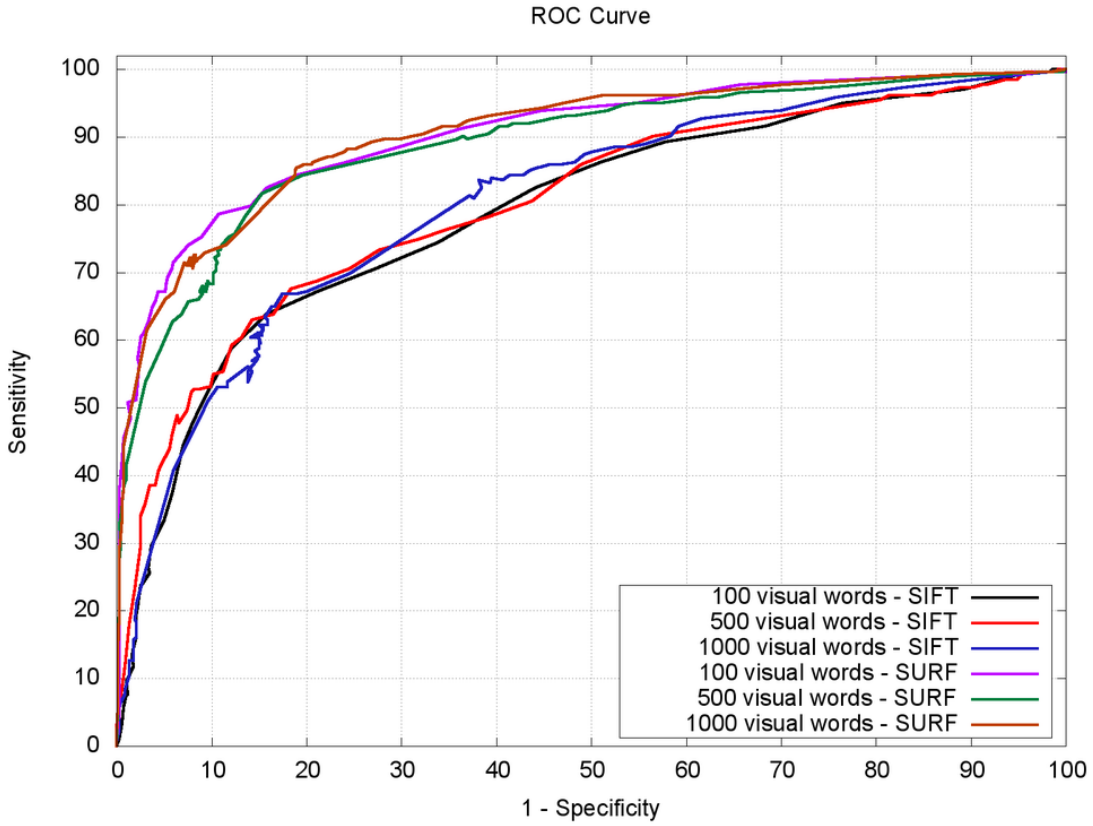


Figure 8: SIFT *vs.* SURF descriptor for creating the visual dictionary.

Figure 9 shows that when using SURF, 100 visual words is the most interesting dictionary size for distinguishing normal ocular-fundus images from images containing exudates. It is worth noting, however, that this approach is effective even for just 25 visual words. In this case, the method achieves 70% sensitivity for 90% specificity.

5.2.2 Clustering *vs.* Non-Clustering *vs.* User-Selection

In restricted domains such as the one we have in this paper, it is possible to bring human specialists to the classification loop and take advantage of their experience to increase the discriminability of the reference dictionary.

In this section, we assess the exudate *vs.* normal images classification problem when considering a random selection of 100 words, a clustering technique with 100 words, and two user-steered approaches: coarse selection and fine selection.

Figure 10 shows the results for the dictionary creation considering random selection of visual words, clustering, and specialists' coarse and fine selection. Not surprisingly, random selection yields the worst results, while the input from the specialist yields the best results.

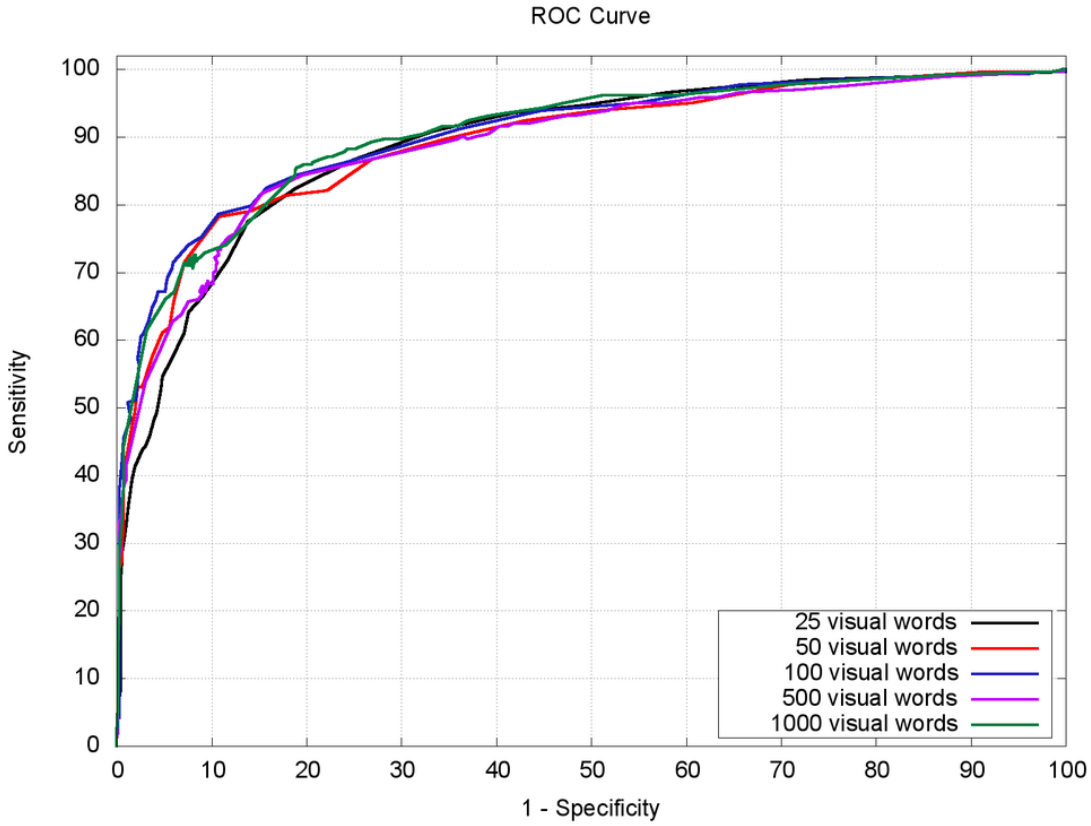


Figure 9: Different dictionary sizes with SURF descriptor.

When considering the fine region selection performed by human specialists, the results are comparable with the clustering selection. However, with the user-steered approach, the specialists have more control over the classification performance without the need for re-clustering if any change needs to take place. For instance, with 2% of false negatives (98% specificity), the method already achieves $\cong$ 64% sensitivity. With 1% of false negatives (99% specificity), it achieves $\cong$ 55% sensitivity which is impressive for a classification system which does not use any kind of pre- or post-processing techniques. For a scenario with 10% of false negatives (90% specificity), the introduced solution achieves 82% sensitivity with the specialists' fine region selection.

Another advantage about bringing specialists to the classification loop lies on the fact that they can adapt a classification system to their needs (to be in accordance to the specifics of a given retinoscopy, for instance). Usually, the specialist needs to select only a few regions. For instance, for 100 visual words, the specialist needs to mark only 50 regions within normal images and 50 regions within images with exudates.

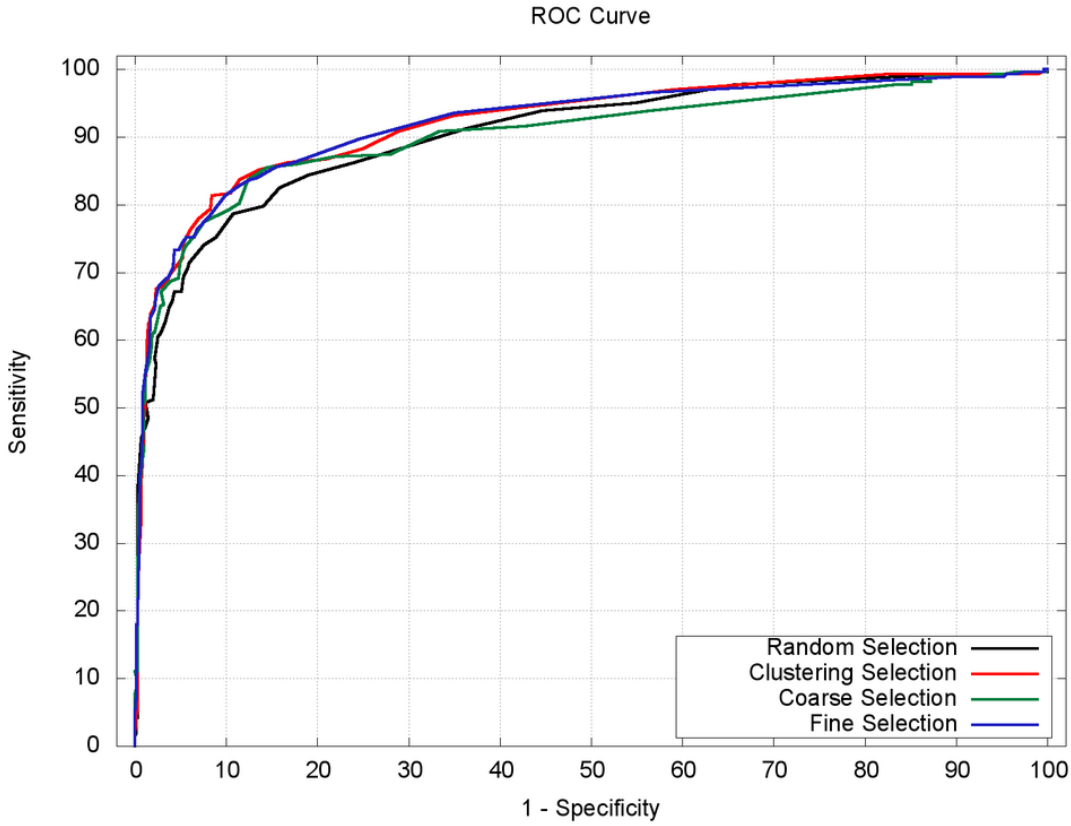


Figure 10: Visual dictionary creation – Clustering *vs.* Random Selection *vs.* User-Steered (Coarse/Fine).

5.3 Experiments – Round #2

In this round, we evaluate the cross-data-set effect on our approach. For that, we train our approach in one data set and test in another one. We perform experiments and report cross training and testing results for two common data sets in the literature: *RetiDB* (c.f., Sec. 5.1) and *Messidor* (c.f., Sec. 5.1). Whenever possible, we compare the results with state-of-the-art ones obtained with the same data sets.

The experiments in this section are an important step for setting initial standards when comparing exudate detectors. To our knowledge, the researchers so far have not used cross training and testing with this important two public data sets. Most of the results are only showed with non public data sets making any kind of comparison unfair. To complicate even more the comparisons, most of the experiments in the literature only validate the proposed solutions with very few images, usually less than 50.

Figure 11 shows results for training our approach with DR1 data set and testing the learned training parameters over two public data sets: RetiDB and Messidor. For compar-

ison purposes, Welfer et al. [15] was one of a very few researchers to provide results using RetiDB data set. The authors report ≅ 70% specificity for a common operational scenario. In comparison, our approach trained with DR1 and tested over RetiDB yields ≅ 50% sensitivity for 100% specificity. The sensitivity increases if we allow more false negatives. For 90% specificity (10% false negatives instead of zero), the approach results in 73% sensitivity. For Messidor data set, the approach trained with DR1 yields ≅ 60% sensitivity for ≅ 98% specificity (2% false negatives).

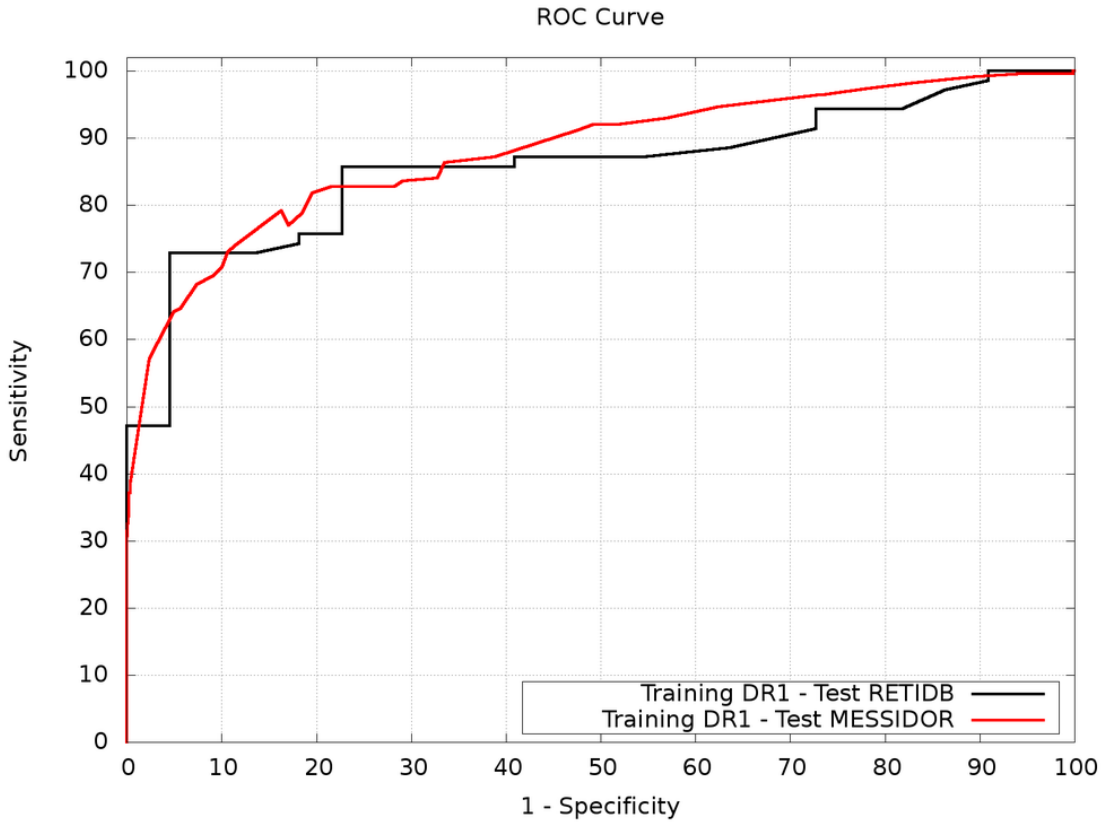


Figure 11: Cross training-testing results. Training with DR1 data set and testing over RetiDB and Messidor data sets. The training is performed using user-steered fine selection.

Figure 12 shows the results for training the approach with Messidor data set and testing with DR1 and RetiDB. For 90% specificity (10% false negatives), the approach achieves ≅ 76% sensitivity considering RetiDB data set. Similar results are obtaining when testing with DR1 data set.

Figure 13 presents results when training the proposed approach with RetiDB data set. For 90% specificity (10% false negatives), the approach yields ≅ 50% sensitivity for Messidor while resulting ≅ 60% for DR1 data set.

Any additional comparison with the state-of-the-art approaches presented in Section 3

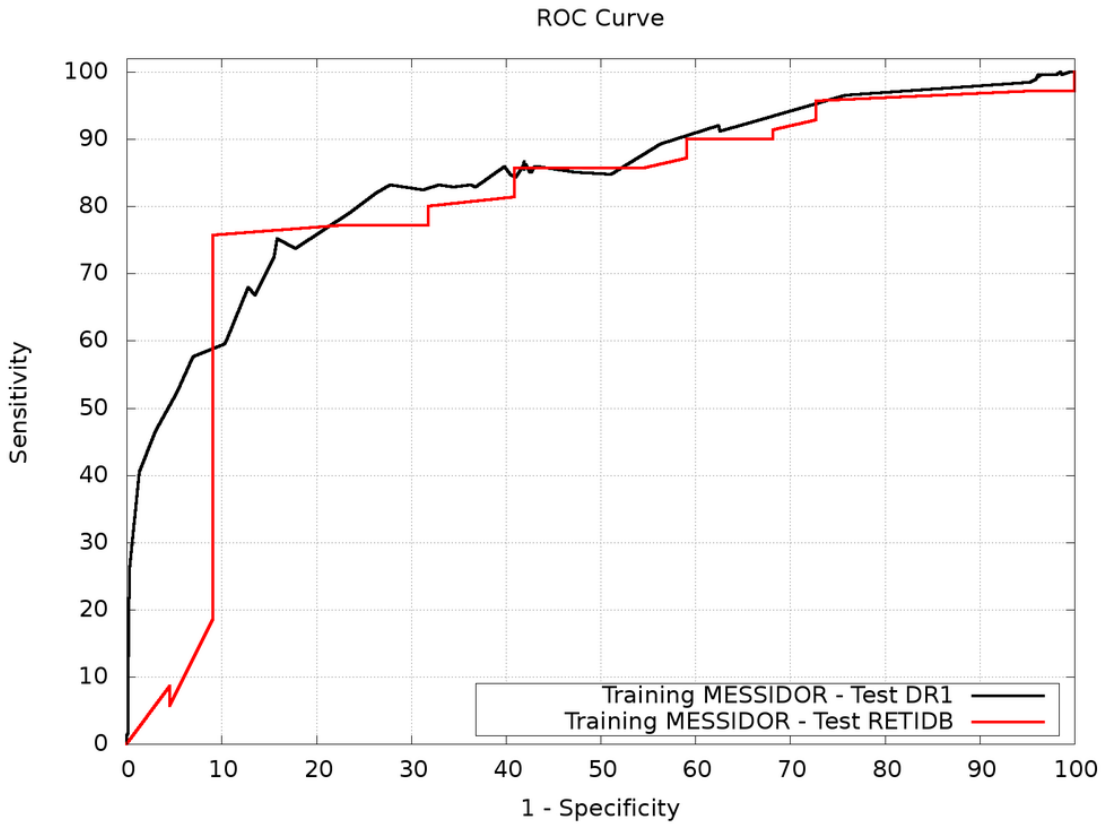


Figure 12: Cross training-testing results. Training with Messidor data set and testing over DR1 and RetiDB data sets. The training is performed using user-steered fine selection.

would be unfair given that the authors do not use the same research protocol (validation data sets, training/testing partitions etc.). In addition, the approaches we showed in Section 3 are custom-tailored for exudate detection and, to our knowledge, none of them has a straightforward extension for detecting other DR-related anomalies.

5.4 Experiments – Round #3

Most of the state-of-the-art approaches dealing with the analysis of ocular-fundus images are custom-tailored to focus on just one anomaly and, most of the times, their extension to the general case is not straightforward or even possible.

In this section, we show the results for detecting two additional DR anomalies, where the only change is the training set of positive examples: to detect deep hemorrhages, we train with deep hemorrhages images, to detect microaneurysms we train with microaneurysms images. As before, all experiments use a 5-fold cross-validation procedure.

Figure 14 brings the detection results for deep hemorrhages and microaneurysms using

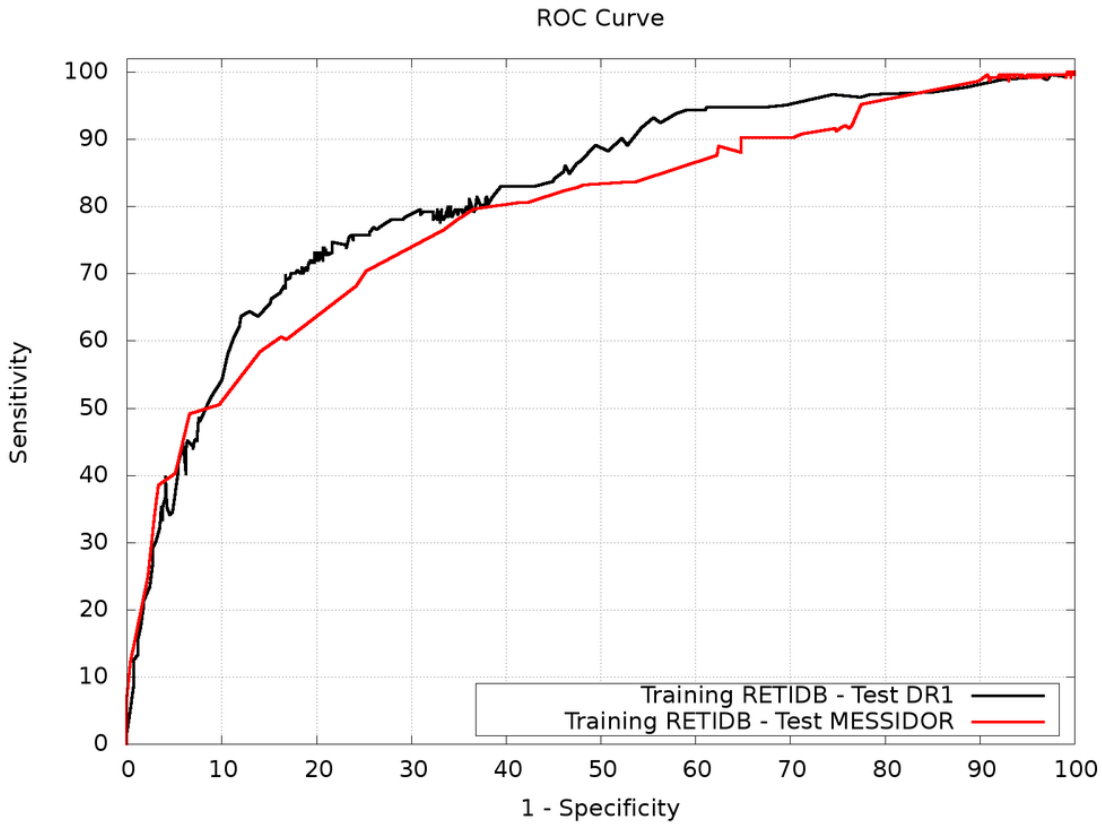
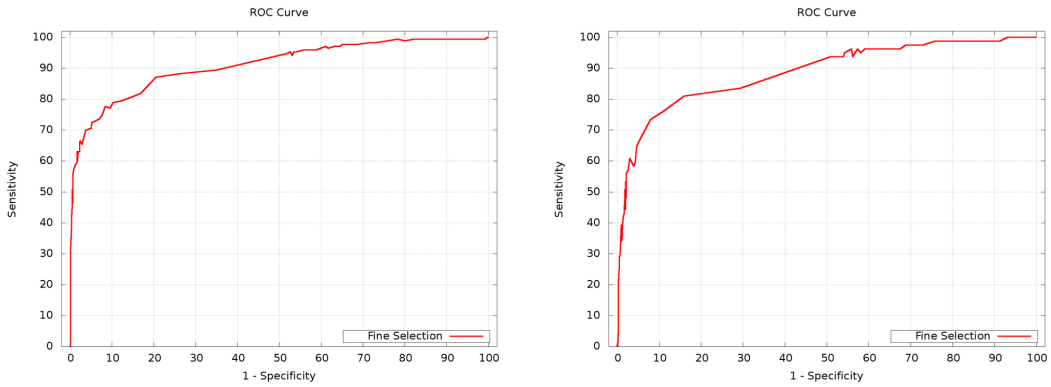


Figure 13: Cross training-testing results. Training with RetiDB data set and testing over DR1 and Messidor data sets. The training is performed using user-steered fine selection.

the proposed visual dictionary approach. There is no pre- or post-processing of any kind when performing these experiments. Considering $\cong$ 95% specificity, the approach detects deep hemorrhages with 60% sensitivity. When detecting microaneurysms, for 100% specificity (no false negatives), the approach yields $\cong$ 45% sensitivity using fine selection of 100 visual words in both cases.

As many ocular-fundus images present more than one anomaly at the same time, an approach with straightforward extension as the one we present in this paper is paramount. We can easily devise a classification system which stands on several simpler detectors such as exudate, hard hemorrhages and microaneurysms detectors. For such a detector, we can apply each of the simple detectors and, if any of them spots an anomaly, we tag this ocular-fundus image for further analysis.



(a) Deep hemorrhages detection.

(b) Microaneurysms detection.

Figure 14: Detection results for Hemorrhages and Microaneurysms using our approach based on a visual dictionary. Experiments performed with 2-fold cross validation over DR1 data set.

6 Conclusions

One of the most successful means for fighting DR in developed and developing countries continues to be its early diagnosis through the analysis of ocular-fundus images of the human retina.

In this direction, in this paper we have used a visual dictionary approach for detecting single pathologies in retinoscopy ocular-fundus images. To detect a specific DR-related anomaly, all it takes is to use the proper training set, as we have demonstrated with experiments in hard exudates, deep hemorrhages, and microaneurysms.

Visual dictionary is an elegant method to learn and represent important features of a given anomaly, and allows us to classify whether an ocular-fundus image is normal or a DR candidate.

We have performed experiments on a series of public data sets and showed the efficacy of our contribution. Using just as few as 100 visual words we are already able to achieve good classification results. To create the visual dictionary representing an ocular-fundus image with a given anomaly and its normal counterparts, we evaluate four possibilities: random selection and clustering for a complete automated solution and also coarse and fine selection when we bring human specialists to the classification loop. For coarse and fine selection, we have showed that a specialist just need to mark a few regions on images to provide the system with the basis for classification. This procedure is performed just once during the training stage.

Our best results are achieved using SURF for the detection and description of the points of interest and a visual dictionary of 100 words, selected by a specialist. However, both the coarse user selection and the fully automated clustering approach for the selection of the 100 visual words are viable alternatives.

Our future work include combining simple detectors specialized in the detection of simple

anomalies such as: exudate, microaneurysms, deep hemorrhages and other detectors to create a final high-level detector for normal *vs.* non-normal ocular-fundus images. We believe the combination of such simple detectors will provide high classification results representing a step forward in computer aided diagnosing using ocular-fundus retinoscopy images.

References

[1] World Health Organization (WHO), "Diabetes programme," Online, August 2010, `http://www.who.int/diabetes/en`.

[2] K. W. Tobin, E. Chaum, V. P. Govindasamy, and T. P. Karnowski, "Detection of anatomic structures in human retinal imagery," *IEEE Transactions on Medical Imaging*, vol. 26, no. 12, pp. 1729–1739, 2007.

[3] M. D. Abràmoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. van Ginneken, "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes," *Diabetes Care*, vol. 31, no. 2, pp. 193–198, 2008.

[4] S. R. S. ao, M. R. K. H. Mitsuhiro, and R. B. Jr., "Visual impairment and blindness: an overview of prevalence and causes in Brazil," *Anais da Academia Brasileira de Ciências*, vol. 81, pp. 539–549, 2009.

[5] Q. Mohamed, M. C. Gillies, and T. Y. Wong, "Management of diabetic retinopathy: A systematic review," *JAMA*, vol. 298, no. 8, pp. 902–916, 2007.

[6] M. James, D. A. Turner, D. M. Broadbent, J. Vora, and S. P. Harding, "Cost effectiveness analysis of screening for sight threatening diabetic eye disease," *BMJ*, vol. 320, p. 1627, 2000.

[7] T. J. de Carvalho, "Aplicação de técnicas de visão computacional e aprendizado de máquina para a detecção de exsudatos duros em imagens de fundo de olho," Master's thesis, Universidade Estadual de Campinas, 2010.

[8] National Eye Institute (NEI), "Facts about diabetic retinopathy," Online, September 2010, `http://www.nei.nih.gov/health/diabetic`.

[9] ——, "Diabetic retinopathy: What you should know," National Institutes of Health – National Eye Institute, Tech. Report NEI 03-2171, September 2003.

[10] Z. M. da Silva Correa and R. E. Jr, "Aspectos patológicos da retinopatia diabética," *Revista de Oftalmologia de São Paulo*, vol. 68, no. 3, pp. 410–414, 2005.

[11] U. R. Acharya, C. M. Lim, E. Y. K. Ng, C. Chee, and T. Tamura, "Computer-based detection of diabetes retinopathy stages using digital fundus images," *Journal of Engineering in Medicine*, vol. 223, pp. 545–553, 2009.

[12] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions Medical Imaging*, vol. 8, pp. 263–269, 1989.

[13] J. Hayashi, T. Kunieda, J. Cole, R. Soga, Y. Hatanaka, M. Lu, T. Hara, and F. Fujita, "A development of computer-aided diagnosis system using fundus images," in *Intl. Conference on Virtual Systems and MultiMedia*, 2001, pp. 429–438.

[14] K. H. Englmeier, K. Schmid, C. Hildebrand, S. Bichler, M. Maurino, and T. Bek, "Early detection of diabetes retinopathy by new algorithms for automatic recognition of vascular changes," *European Journal of Medical Research*, vol. 9, pp. 473–488, 2004.

[15] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images," *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 228–235, 2010.

[16] P. G. B. Jahne, H. Haussecker, Ed., *Handbook of Computer Vision and Applications*. Academic Press, London, 1999.

[17] E. Dougherty and R. Lotufo, "Hands-on morphological image processing," in *SPIE PRESS*, 2003.

[18] J. B. T. M. Roerdink and A. Meijster, "The watershed transform: definitions, algorithms and parallelization strategies," *Fundamenta Informaticae*, vol. 41, no. 1-2, pp. 187–228, 2000.

[19] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," *Computerized Medical Imaging and Graphics*, vol. 32, p. 8, 2008.

[20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall; 2 edition, 2002.

[21] A. Sopharak, B. Uyyanonvara, and S. Barman, "Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy c-means clustering," *Sensors*, vol. 9, no. 3, pp. 2148–2161, 2009.

[22] M. Garcia, C. I. Sanchez, M. I. Lopez, and R. H. Daniel Abasolo, "Neural network based detection of hard exudates in retinal images," *Computer Methods and Programs in Biomedicine*, vol. 93, pp. 9–19, 2009.

[23] A. D. Fleming, S. Philip, K. A. Goatman, G. J. Williams, J. A. Olson, and P. F. Sharp, "Automated detection of exudates for diabetic retinopathy screening," *Physics in Medicine and Biology*, vol. 52, no. 24, pp. 7385–7396, 2007.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Springer, 2006.

[25] C. I. Sanchez, M. Garcia, A. Mayo, M. I. Lopez, and R. Hornero, "Retinal image analysis based on mixture models to detect hard exudates," *Medical Image Analysis*, vol. 13, pp. 650–658, 2009.

[26] B. Dupas, T. Walter, A. Erginay, R. Ordonez, N. Deb-Joardar, P. Gain, J.-C. Klein, and P. Massin, "Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages and exudates, and of a computer-assisted diagnostic system for grading diabetic retinopathy," *Diabets & Metabolism*, vol. 36, pp. 213–220, 2010.

[27] C. I. Sanchez, M. Niemeijer, M. S. A. S. Schulten, M. Abramoff, and B. van Ginneken, "Improving hard exudate detection in retinal images through a combination of local and contextual information," in *IEEE Intl. Symposium on Biomedical Imaging*, 2010, pp. 5–8.

[28] W. L. Yun, A. Rajendra, Y. V. Venkatesh, C. Chee, L. Min, and E. Y. K. Ng, "Identification of different stages of diabetic retinopathy using retinal optical images," *Intl. Journal on Information Sciences*, vol. 178, pp. 106–121, 2008.

[29] H. Wang, W. Hsu, K. G. Goh, and M. L. Lee, "An effective approach to detect lesions in color retinal images," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2000.

[30] O. Faust, R. Acharya, E. Y. K. Ng, K.-H. Ng, and J. S. Suri, "Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review," *Journal of Medical Systems*, vol. In press, 2010.

[31] P. Y. Jalli, T. J. Hellstedt, and I. J. Immonen, "Early versus late staining of microaneurysms in fluorescein angiography," *Retina*, vol. 17, pp. 211–215, 1997.

[32] H. J. Jelinek, M. J. Cree, D. Worsley, A. Luckie, and P. Nixon, "An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice," *Clinical and Experimental Optometry*, vol. 89, pp. 299–305, 2006.

[33] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, and H. L. Cook, "Automated detection of diabetic retinopathy on digital fundus image," *Diabetic Medicine*, vol. 19, pp. 105–112, 2002.

[34] X. Zhang and O. Chutatape, "Top-down and bottom-up strategies in lesion detection of background diabetic retinopathy," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2005, pp. 422–428.

[35] A. Pinz, *Object Categorization*. Now Publishers, 2005.

[36] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *IEEE Intl. Conference on Computer Vision*, 2003, pp. 1470–1477.

[37] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *IEEE Intl. Conference on Computer Vision*, 2005, pp. 1800–1807.

[38] I. Ulusoy and C. M. Bishop, "Generative versus discriminative methods for object recognition," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005.

[39] F.-F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 524–531.

[40] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on Statistical Learning in Computer Vision*, 2004.

[41] M. Hazewinkel, Ed., *Encyclopaedia of Mathematics*. Springer, 1995.

[42] E. Valle and M. Cord, "Advanced techniques in CBIR: Local descriptors, visual dictionaries and bags of features," *Tutorials of the Brazilian Symposium on Computer Graphics and Image Processing*, vol. 0, pp. 72–78, 2009.

[43] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, February 2004.

[44] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. 506–513.

[45] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Intl. Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, January 2004.

[46] E. N. Mortensen, H. Deng, and L. G. Shapiro, "A SIFT descriptor with global context," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2005, pp. 184–190.

[47] M. F. Demirci, A. Shokoufandesh, Y. Keselman, L. Bretzner, and S. Dickinson, "Object recognition as many-to-many feature matching," *Intl. Journal of Computer Vision*, vol. 69, no. 2, pp. 203–222, 2006.

[48] J. Stoettinger, A. Hanbury, N. Sebe, and T. Gevers, "Do colour interest points improve image retrieval," in *ICIP*, 2007, pp. 169–172.

[49] R. Datta, D. Joshi, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–77, April 2008.

[50] J. A. Junior, R. Torres, and S. Goldenstein., "SIFT applied to CBIR." *Revista de Sist. de Inf. da Fac. Salesiana Maria Auxiliadora*, vol. 4, pp. 41–48, 2009.

[51] E. Valle, "Local-descriptor matching for image identification systems," PhD Thesis, Universit de Cergy-Pontoise, Cergy-Pontoise, France, 2008.

[52] E. Valle, M. Cord, and S. Philipp-Foliguet, "High-dimensional descriptor indexing for large multimedia databases," in *ACM Intl. Conference on Information and Knowledge Management*, 2008, pp. 739–748.

[53] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *European Conference on Computer Vision*, 2006, pp. 1–14.

[54] R. Gonzalez and R. Woods, *Digital Image Processing*, 3rd ed. Prentice-Hall, 2007.

[55] P. Viola and M. Jones, "Robust real-time face detection," *Intl. Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, February 2004.

[56] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *IEEE Intl. Conference on Computer Vision*, 2001, pp. 525–531.

[57] F. Wolf, T. Poggio, , and P. Sinha, "Human document classification using bags of words," MIT, Tech. Rep., 2006.

[58] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2169–2178.