Points of Interest and Visual Dictionaries for Automatic Retinal Lesion Detection

Anderson Rocha*, *Member, IEEE*, Tiago Carvalho, Herbert F. Jelinek, *Member, IEEE*, Siome Goldenstein, *Senior Member, IEEE*, and Jacques Wainer

Abstract—In this paper, we present an algorithm to detect the presence of diabetic retinopathy (DR)-related lesions from fundus images based on a common analytical approach that is capable of identifying both red and bright lesions without requiring specific pre- or postprocessing. Our solution constructs a visual word dictionary representing points of interest (PoIs) located within regions marked by specialists that contain lesions associated with DR and classifies the fundus images based on the presence or absence of these PoIs as normal or DR-related pathology. The novelty of our approach is in locating DR lesions in the optic fundus images using visual words that combines feature information contained within the images in a framework easily extendible to different types of retinal lesions or pathologies and builds a specific projection space for each class of interest (e.g., white lesions such as exudates or normal regions) instead of a common dictionary for all classes. The visual words dictionary was applied to classifying bright and red lesions with classical cross validation and cross dataset validation to indicate the robustness of this approach. We obtained an area under the curve (AUC) of 95.3% for white lesion detection and an AUC of 93.3% for red lesion detection using fivefold cross validation and our own data consisting of 687 images of normal retinae, 245 images with bright lesions, 191 with red lesions, and 109 with signs of both bright and red lesions. For cross dataset analysis, the visual dictionary also achieves compelling results using our images as the training set and the RetiDB and Messidor images as test sets. In this case, the image classification resulted in an AUC of 88.1% when classifying the RetiDB dataset and in an AUC of 89.3% when classifying the Messidor dataset, both cases for bright lesion detection. The results indicate the potential for training with different acquisition images under different setup conditions with a high accuracy of referral based on the presence of either red or bright lesions or both. The robustness of the visual dictionary against image quality (blurring), resolution, and retinal background, makes it a strong candidate for DR screening of large, diverse communities with varying cameras and settings and levels of expertise for image capture.

Index Terms—Diabetic retinopathy (DR), diabetes automated screening, hard exudate detection, hemorrhage detection, microaneurysm detection, red and bright lesion classification, visual dictionaries.

*A. Rocha is with the Institute of Computing, University of Campinas (Unicamp), Campinas 13083-852, Brazil (e-mail anderson.rocha@ic.unicamp.br).

T. Carvalho, S. Goldenstein, and J. Wainer are with the Institute of Computing, University of Campinas (Unicamp), Campinas 13083-852, Brazil (e-mail tiagojc@gmail.com; siome@ic.unicamp.br; wainer@ic.unicamp.br).

H. F. Jelinek is with the Centre for Research in Complex Systems, Charles Sturt University, N.S.W. 2830, Australia, and also with the Australian School of Advanced Medicine, Macquarie University, N.S.W. 2113, Australia (e-mail hjelinek@csu.edu.au).

I. INTRODUCTION

DIABETIC retinopathy (DR) is the result of microvascular retinal changes triggered by diabetes that can lead to a complete loss of sight if not treated in a timely manner. Recent reports have shown that approximately 25 000 people with diabetes go blind every year in the U.S. due to DR [1]. According to [1], in the U.S. and Europe, DR is the major cause of blindness for the economically active population and, according to [2], it is estimated that DR is responsible for 5% of all the world's blindness cases.

Early diagnosis of DR and treatment [3] can prevent blindness, and therefore, systematic screening (by specialists) of diabetic patients is a cost-effective health care practice [4]. However, due to the large number of people that require screening and annual reviews, an automated and accurate screening tool is a useful adjunct in diabetes clinics. Currently, several highly accurate programs exist for automated detection of specific DR-related lesions [5]–[7]. These algorithms require different pre- and postprocessing steps of the retinal images depending on the lesion of interest as well as corrections for resolution and color normalization to account for images with different fields of view and ethnicity [8].

The presence of microaneurysms and dot hemorrhages (red lesions) and/or hard exudates (bright lesions) are indicative of early stage DR. Fig. 1(a) depicts one example with the main retinal regions highlighted. Microaneurysms are focal dilations of retinal capillaries [see Fig. 1(b)] and have an appearance similar to red dots in these images. Intraretinal lipid exudates (hard exudates) are caused by the breakdown of the blood-retinal barrier, which leads to fluid rich in lipids and proteins to leave the parenchyma, causing retinal edema and exudation [see Fig. 1(c)]. Finally, dot hemorrhages are similar to microaneurysms but slightly larger and are found where capillary walls weaken. These may rupture causing intraretinal hemorrhages [see Fig. 1(d)].

Automated bright lesion detection has resulted in highly accurate classification and has been discussed recently by [9]–[11]. For automated red lesion detection similar high sensitivity and specificity has been reported. The Iowa retinopathy online challenge published results by five research groups using different algorithms for pre- and postprocessing and detection of microaneurysms [12]. More recent results reported by Niemeijer and colleagues including Giancardo *et al.* [5] and the group led by Antal *et al.* [6] have shown a further improvement.

Addressing the detection of multiple DR-related lesions, Abràmoff and coworkers combined several different lesion-specific detectors into a single automatic detection program and

suggested that a single algorithm that is able to identify multiple lesions is required [9]. Bright and red lesions have different image characteristics and, therefore, require different pre- and postprocessing making automatic multilesion detection difficult in practice. A program that allows for multiple lesions to be detected without requiring different processing of the images is an important step forward in the development of automatic retinal fundus assessment programs. Detecting the presence of a specific lesion or multiple different lesions offers a solution to improve the accuracy of detecting DR and correctly classifying images for the presence of either specific single lesions or combinations thereof. Computer-based feature detection has the advantage of being able to utilize characteristics of images that do not necessarily correspond to DR-related lesion characteristics but rather use features such as colorization of a region of interest (RoI) [13] or discontinuities in the image such as texture, color, or boundaries [14].

In clinical practice, machine vision learning algorithms have been applied for screening of DR in the southern United States using content-based image retrieval for automated diagnosis [15]. This system achieves a sensitivity of 94.8% and a specificity of 78.7% with an area under the curve (AUC) of 95.3% but requires extensive pre- and postprocessing to differentiate between different stages of DR progression and type of lesions. Web-based applications of machine learning tools have also been investigated to enhance the uptake of current automated technologies [16].

This paper introduces an algorithm that addresses several of the aforementioned issues. The solution is based on a unified feature detection and analysis framework that is capable of identifying different DR-related lesions such as hard exudates and microaneurysms by correlating image specific features with the presence of specific lesions without the necessity of pre- and postprocessing of images. The approach utilizes experts to indicate RoIs containing specific lesions and then identifies points of interest (PoIs) within these regions that become "words" within a visual dictionary. Our approach extends previous work as it is independent of the image resolution, color space representation, and amenable to detect different lesions other than just microaneurysms. Finally, our approach does not rely on any *ad hoc* parameter (e.g., Gaussian standard deviations) nor does it assume any prespecific size of the lesions such as required by some of the current microaneurysm detectors in the literature.

The visual word dictionary is based on measuring metrics at the image level and does not follow a "decision-support tool" approach, where the system is used to aid diagnosis, as it is found in most of the current methods. The decision-support approach requires pinpointing the location of each lesion to allow the specialist to evaluate the image for diagnosis. In this case, metrics based on the accuracy of detecting each type of lesion is more relevant. Using a visual words dictionary as the basis for a classification algorithm was inspired by the computer vision and image processing literature [17]–[21], where visual dictionaries and PoIs are used as a basis for several applications such as image retrieval and classification.

Section II presents state-of-the-art achievements for automatic DR-related lesion detection. Section III introduces our

method. Section IV reports the experimental setup and the classification results for different retinal datasets. Finally, Section V concludes this paper and discusses some considerations regarding directions for future work.

II. DETECTION OF DIABETIC RETINOPATHY LESION

A number of recent reviews on automatic retinal analysis with emphasis on DR lesion detection have described some of the major research results in the area. Abràmoff *et al.* [22] is a more general review in retinal image processing, while [23] and [24] are more specific to DR.

Most of the techniques used for automatic DR lesion detection are based on specific segmentation or feature extraction techniques developed for identifying *each* specific lesion. These techniques have been achieving increasing accuracy rates but normally a technique developed for the detection of one kind of lesion cannot be directly used to detect another kind. Fleming and colleagues recently reported combining red and bright lesion detection by adapting their microaneurysm detector to identify bright lesions but requiring separate processing of the images depending on whether bright or red lesions detection was required [11].

Detection of bright or red lesions also relies heavily on color [25] and thus some color normalization is needed [5], [8]. Detecting microaneurysms requires additional processing to remove retinal blood vessels and the optic disk. Thus, detection and removal of these structures is a necessary initial step when single-lesion detection is required. Many of the automatic detectors also use feature size (in terms of pixels) as a criterion [26] and, thus, require preprocessing due to confounding by image resolution of the absolute pixel size. For these algorithms to function, training and test sets usually have to have identical resolution for the algorithm to function optimally. An interesting example is [27] that presents a microaneurysm detector trained on different images for detection in an optometry clinic.

Single-lesion detectors may also be sensitive to fundus images with white reflection that is present in retinae of the young [27]. We refer the reader to the recent reviews [23] and [24] for more details on these techniques.

Including RoI for detection of lesions was reported by Agurto *et al.* [28] and [26]. Agurto *et al.* used 120 regions of 40×40 pixels representative of four types of lesions commonly associated with DR (i.e., microaneurysms, exudates, neovascularization, and hemorrhages). The regions were characterized using texture descriptors at multiple scales (e.g., cumulative distribution functions of the amplitude, frequency magnitude, and relative instantaneous frequency angle). Similar to Agurto *et al.*'s approach the method described in this paper also uses region characterization and performs multiscale analysis of the images. Our work uses representative regions in the image to identify lesions and creates a dictionary to capture common properties among such regions [26]. The main differences to Agurto *et al.*'s work is how the regions are characterized and that our method does not specify a specific size for any of the RoIs.

III. Visual Dictionaries for DR Detection

In this paper, we present a solution to classify DR-related lesions based on the concept of selecting features around locally invariant interest points and visual dictionaries of images. This paradigm uses a set of highly extensible feature representations, and characterizes red and bright lesions in the optic fundus images using visual words that incorporates information provided by specialists into a framework easily extendible to different types of retinal abnormalities.

Two fundamental differences to previous work is that: 1) the method builds a specific projection space for each class of interest (e.g., exudates or normal regions) instead of using a common dictionary for all classes; and 2) PoIs are only considered during training if they fall within the regions in the fundus image highlighted by the specialists rather than including all PoIs within the fundus images when training.

Visual dictionaries constitute a robust representation approach as each image is treated as a collection of regions. In this representation, the only important information is the appearance of each region [18]. Dictionary "words" are not in the space of images (or patches of images), but at the space of feature description of the "PoIs" in the image, with the PoIs located within the RoIs.

The objective when creating a visual dictionary is to learn, from a training set of examples, the optimal representative PoIs for a given problem. The dictionary must be large enough to distinguish relevant differences between images but not include irrelevant variations [29], [30]. Given a visual dictionary, an image can be represented according to the visual words it contains.

We can summarize the approach that we introduce in this paper for DR-related lesion detection into two phases: training and detection. The training phase consists of learning the overall behavior of the lesions of interest and what makes the images with lesions different to normal images. The detection phase consists of using the learned knowledge in an automatic fashion for testing unknown images. Algorithm 1 presents the main training steps, while Algorithm 2 shows the detection procedure. The next sections present details about each step of the approach described in the two algorithms.

Algorithm 1 DR-Pathology training procedure.

Input: a collection of image examples containing the DR-related pathology of interest, X_p;

Input: a collection of control images (normal images with no pathology), X_c.

1: experts mark regions of interest (RoIs) in X_p images;
2: $\mathcal{P}_p \leftarrow$ points of interest in X_p constrained by the located RoIs;
3: $\mathcal{P}_c \leftarrow$ points of interest in X_c constrained by the retina's region;
4: $\mathcal{D}_p \leftarrow$ a representative dictionary for the set of points of interest in $\mathcal{P}_p$ via K-Means clustering;
5: $\mathcal{D}_c \leftarrow$ a representative dictionary for the set of points of interest in $\mathcal{P}_c$ via K-Means clustering;
6: $\mathcal{D} \leftarrow \mathcal{D}_p \cup \mathcal{D}_c$;
7: **for** each image I_i in $X_p \cup X_c$ **do**
8:     project its points of interest onto $\mathcal{D}$ creating a signature vector fv_i for each image;
9: **end for**
10: **train** an SVM classifier using all signature vectors fv_i generating the classification model $\mathcal{M}_{\text{SVM}}$ able to differentiate normal images from the DR-pathology of interest.

Algorithm 2 DR-Pathology detection procedure.

Input: the test image I_j
Input: the pre-computed projection dictionary $\mathcal{D}$
Input: the pre-computed training model $\mathcal{M}_{\text{SVM}}$

1: calculate a bounding box of I_j to limit the analysis to its retina's region. This is performed by means of Hough Transform [31];
2: $\mathcal{P}_j \leftarrow$ points of interest in I_j constrained by the bounding box in previous step;
3: project the points of interest $\mathcal{P}_j$ onto the pre-computed dictionary $\mathcal{D}$ creating signature fv_j;
4: classify fv_j as normal or DR-pathology by means of the learned model $\mathcal{M}_{\text{SVM}}$

A. Background

Identifying features around locally invariant interest points was originally developed for large baseline correspondence applications and has recently been a focus of intense research for other applications. A few attempts have been made to apply this model for image retrieval and classification [19], [33]

The PoIs approach is based on the hypothesis that PoIs within an image convey more information than other points and that every image in a collection can be represented using a large number of PoIs. It is then possible to calculate a local descriptor around each PoI, and store these local descriptors in an indexing data structure [34]. Local descriptors computed around PoIs are more robust in identifying subtle changes within images compared to global descriptors [19]–[21]. The advantage of using PoIs is that they can be robustly estimated, even if the image suffers distortions as the major criterion of quality for a PoI algorithm is repeatability [34]. Robustness of this kind might be interesting in retinal imagery since small changes can occur during different equipment setup, motion blur, and small geometric distortions.

Although the discriminative power of the PoIs is an extremely important attribute and advantageous to use when searching for a specific target, they are less than optimal when searching for complex categories as the ability to generalize becomes paramount. Therefore, alternative techniques combined with PoIs are required. A solution is to apply the concept of visual dictionaries, which considers the high-dimensional descriptor spaces associated with the PoIs and finds the ones that best represent the whole set of PoIs by means of a nonsupervised learning technique (e.g., clustering).

The biggest challenge for developing an accurate algorithm is to design a good visual dictionary. The creation of the dictionary requires the quantization of the description space that is achieved using clustering approaches and including the information provided by the expert.

B. Identification of Points of Interest

The protocol we devised and followed for our tests are as follows. To detect bright or red lesions, the specialists mark RoIs within the retinal images are considered as good representatives of bright or red lesions. For normal/control images, the entire retinal region represented in the image can be considered a RoI. On average 2–5 RoIs for each training image with a given
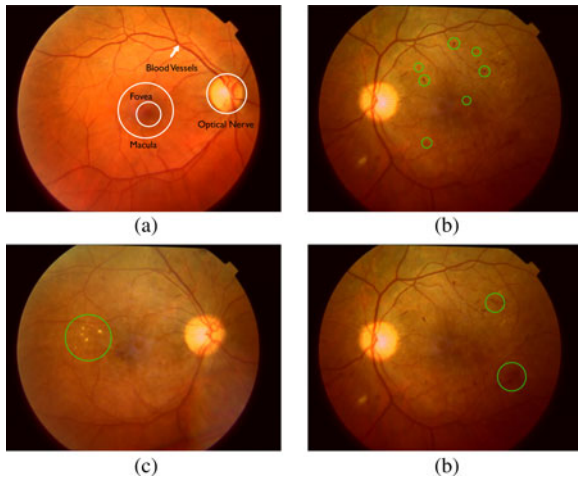
Fig. 1. Retina's main regions as well as examples of DR-related pathologies. (a) Retina's main regions. (b) Microaneurysms. (c) Intraretinal lipid exudates. (d) Hemorrhages.

Fig. 2. PoIs located by SURF (white circles) and the RoIs markings performed by a medical specialist (yellow dashed polygons).

DR-related lesion are marked by the specialists. The training stage then locates PoIs within all images. The parameters are a good tradeoff between classification accuracy and computation efficiency for finding good representative PoIs.

PoIs are found at discontinuities within the image, being either textural or containing some other boundary condition. It is desirable to choose scale-invariant PoIs in order to achieve a representation that is robust to possible image transformations (e.g., rotations, scale, and partial occlusions). There are several options for finding and characterizing PoIs such as *Speeded-Up Robust Features* (SURF) [35] and *Scale-Invariant Features Transform* (SIFT) [32]. Both methods achieve high repeatability and distinctiveness. Empirically, we determined that the results using SURF were somewhat better than the ones using SIFT; therefore, we present only results with SURF.

Using SURF, each image in the training set generates a series of PoIs as illustrated in Fig. 2. All of these PoIs are then filtered and only PoIs within the RoIs marked by the specialist are kept for further processing. In Fig. 2, only the points lying within the dashed RoIs are kept. To cover the most important characteristics within an image, a reasonable number of PoIs need to be used. We have found that, normally, 1000 PoIs per image is optimal within 3–5 RoIs for the images with lesions. Once the PoIs in an image are found, their neighborhoods are characterized by means of a local descriptor.

In brief, SURF has two steps: detection of the PoIs and description of these points. The detection is based on multiscale decomposition of the image using the convolution of Gaussian second-order derivatives within the image, at different scales. The PoI characterization considers a square region of radius $20s$ centered around the PoI, where s is the scale on which the point was detected. This square region is then split into regular smaller 4×4 square subregions and, for each subregion, the method computes a number of features (sums and differences) at 5×5 regularly spaced sample points.

The result of the SURF is a set of PoIs each one represented in n-dimensions. In this paper, we use each PoI with $n = 128$
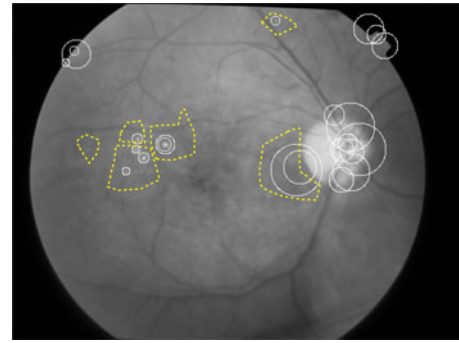
dimensions. For more details about SURF, we refer the reader to [35].

C. Visual Dictionary

A database of training examples comprising positive images (i.e., images with bright or red lesions present) and negative control images (i.e., considered normal by specialists) is created.

As described in Section III-A, SURF is a good low-level representative feature detector. To preserve the discriminatory power of such descriptors found by SURF while increasing their generalization, the concept of visual dictionaries [30] can be introduced.

After finding the PoIs, a dictionary representing distinctive features of images with a specific lesion of interest as well as images tagged as normal by specialists is created. An important parameter for classification is the number of words k in this dictionary. A k too low groups together too many PoIs into the same visual word and looses the ability to distinguish important information. A k too high memorizes the details of the PoIs in the training set and looses its ability to generalize. To achieve this, the clustering task is performed using the k-means algorithm [36] and all PoIs in normal images are clustered into $k/2$ groups, as are the PoIs in the RoI containing bright or red lesions. This results in a dictionary with k words. Applying k-means clustering at this stage ensures that the visual dictionary has the same number of normal and DR-related lesion words. However, it is not a strict requirement to have the same number of words for control and DR-related lesions. We have experimental results showing that, for some lesions, other proportions may be better. However, in this paper, we focused on a 50%–50% proportion. We empirically found the dictionary sizes considering a tradeoff between classification accuracy and computational efficiency. However, we can employ automatic approaches such as [37] for this task.

After creating the visual word dictionary, each of the remaining PoIs within the RoIs in the training images are assigned to the closest visual word of the dictionary using a process called projection or *quantization* [17], [34]. At the end of the quantization process, each image is represented by a *signature* or histogram of the visual words it contains, which is the input to a machine learning classifier in the last stage. Formally, the
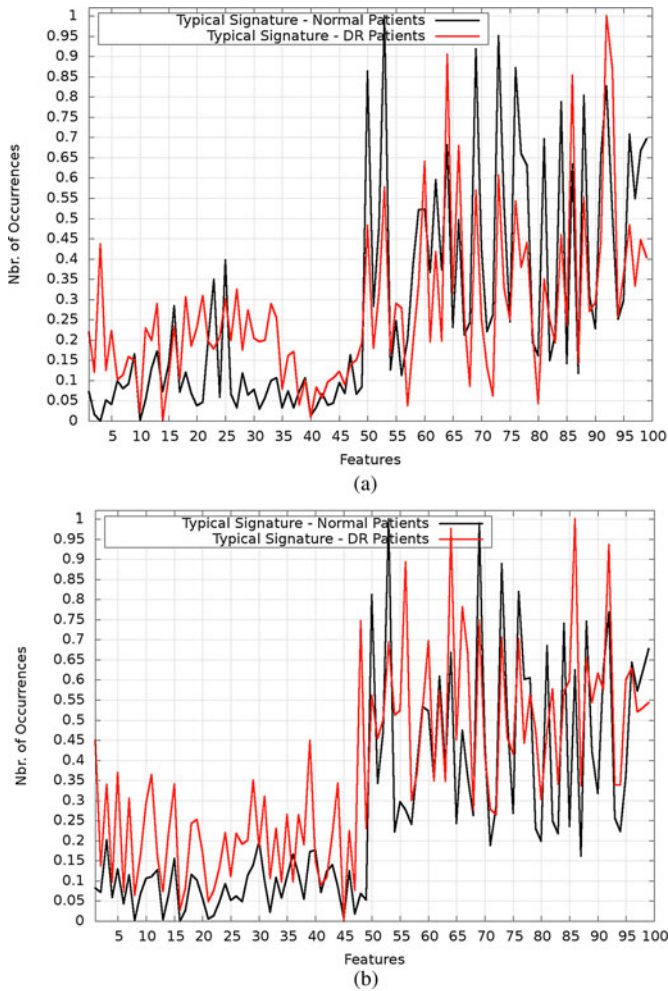
Fig. 3.   Typical signatures for images showing bright and red lesions calculated based on averaging out all the signatures of the training set considering the average of a fivefold cross-validation procedure for DR1 dataset (c.f., Section IV-A. (a) Bright lesions. (b) Red lesions.).

quantization performs a hard assignment (each PoI is attributed to the closest visual word in the dictionary) and the aggregation function is the sum [34] (once a PoI is matched to a visual word the corresponding entry in the signature is increased by one).

Fig. 3 depicts two examples of typical signatures found for images with and without lesions and provides an indication of the discriminative power of the proposed approach. Each plot shows a typical signature for normal images versus images with either bright or red lesions. The typical signature for images showing bright and red lesions is calculated based on averaging out all the signatures of a training set (DR1 in this case, see Section IV-A).

Specifically, each plot depicts 100 "visual words" and their frequency in the training set. Positions 1–50 represent lesion-based visual words (e.g., bright or red lesions in this case) while positions 51–100 represent words for the normal regions. For positions 1–50, it is expected that the visual words associated with lesions dominate (i.e., have a higher frequency of occurrence) the normal visual words while the opposite is the case for positions 51–100.

To perform the final classification, a two-class machine learning classifier such as a *support vector machine* (SVM) was selected [36]. For training the classifier, the signature vectors of the training images containing examples of images with a given lesion and normal/control images are fed into the classifier. To test a new image, its PoIs are located and projected onto the precomputed dictionary to create its signature vector, which forms the input to the trained classifier. However, some PoIs may lie outside the retinal boundary of the image. Therefore, before the quantization, the PoIs lying within the retinal boundary are identified by calculating a binary mask of the image using a pixel thresholding procedure and then calculating the Hough transform [31] of the resulting binary mask. Other possible algorithms could be used but the Hough transform is simple and fast. The rationale for its use is that the binary mask of a retinal image can contain several image portions not related to the retina. The Hough transform then finds the best circle that encompasses the retina. Fig. 4 shows an example of a retinal image, its binarization, and the resulting Hough transform delimited retinal portion.

Fig. 5 summarizes the sequence of steps of the proposed approach (training and detection) and illustrates what we discussed earlier in Algorithms 1 and 2.

IV. EXPERIMENTS AND VALIDATION

In this section, we present the experimental methods and results that we used to validate the approach. The experiments are divided into two parts.

1) *Part #1*. Experiments for identifying the optimal number of words representing the classification dictionary using the DR1 dataset described shortly. Here, we use a cross-validation protocol dividing the DR1 dataset into fivefolds. The SVM parameters are tunned to the training examples (four out of the fivefolds) and that parameterization is used on the remaining (not used) fold. As this refers to a five-fold cross-validation protocol, each time the testing fold changes, the parameters are recalculated on the appropriate training sets.

2) *Part #2*. Experiments using cross dataset classification based on the visual word dictionary approach. Here, we use the a cross-database validation protocol in which we train on DR1 images and test on RetiDB and Messidor. The SVM is tunned on all examples in the DR1.

No pre- or postprocessing was undertaken prior to training and classification. All experiments use SURF with 128 dimensions for finding the PoIs and SVM as the final classification method. The results are reported as ROC curves. In both experiment sets, the SVM parameters are found using the standard LibSVM's grid search fine tuning algorithm [38].

A. Experiments—Part #1

The first part of our experiments was conducted using the DR1 dataset from the Ophthalmology Department of the Federal University of São Paulo, collected during 2010. The images are clinical images from all retinal patients of the department, including a diverse range of patients.
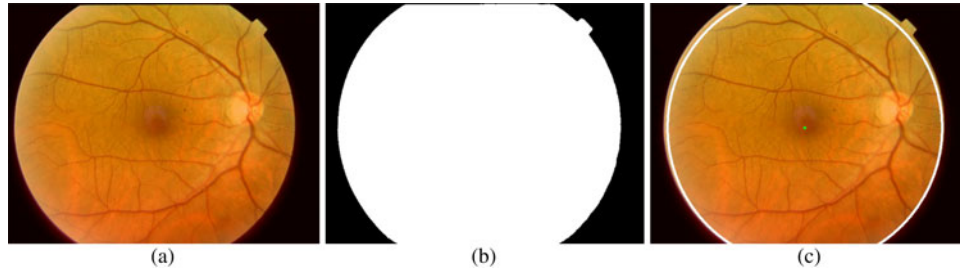
Fig. 4. Constraining the analysis of normal retina image using Hough transform. (a) Original image. (b) Its binarization. (c) Result of a Hough transform to delimit the retina's region.
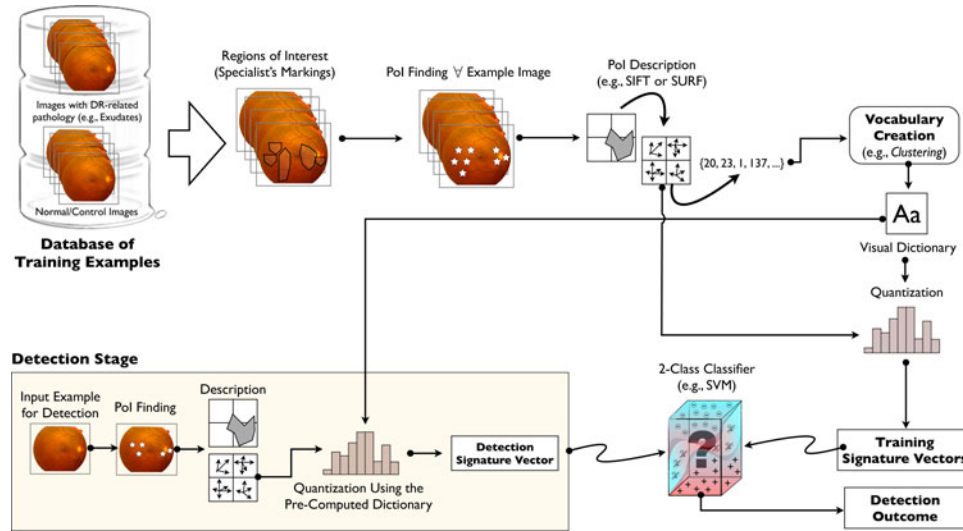

Fig. 5. Sequence of steps for classifying DR lesions (e.g., bright or red lesions) from fundus images.

The DR1 dataset comprises 1014 images with an average resolution of 640×480 pixels, (687 are normal retinae, 245 images contain bright lesions, 191 contain red lesions, and 109 contain signs of both bright and red lesions). All of the images in the DR1 dataset were manually annotated for DR-related lesion (presence/absence) by three medical specialists. For each image, the specialist tags the image as having or not having bright or red lesions. All of the images in which the three agree were kept in the final dataset. The RoIs were marked on some of the images with red, bright, or both lesions by a single ophthalmologist with five years of experience. The expert draws a boundary surrounding the lesion(s) and thus creates a RoI in the digital training images (see Fig. 2). The images were captured using a TRC50X (Topcon Inc., Tokyo, Japan) mydriatic camera with maximum resolution of one megapixel and a field of view (fov) of 45°. This dataset is publicly available[1] and adds to currently existing images [39], [40].

In the first part of the experiments, we explored the importance of how the number of words considered to characterize each lesion in the creation of the dictionary influences the outcome of the detector. We also investigated how changing the discriminatory power of the SURF PoIs, that is the space surrounding the PoIs, influenced the classification outcome and

determined empirically that the characterization of PoIs containing 128 orientations is more effective than using 64 orientations. A possible reason for this can be that discontinuities associated with lesions have a higher number of subtle characteristics that require more orientations to be captured. The results of training the classifier using the DR1 dataset are based on the average of a fivefold cross-validation procedure (five independent evaluations of the full algorithm using at each time onefold as test and the remaining four as training).

Fig. 6 shows the ROC classification results for images with bright and red lesions with the best and worst dictionary sizes. For bright lesions, the best-performing dictionary size consisted of 500 words (250 bright lesions-based words plus 250 normal-based words) with a corresponding AUC = 95.3%. For reference, at 90% sensitivity, the system achieves 87% specificity. The worst performing dictionary size was 50. For red lesions, the differences among the different dictionary sizes are not as significant, but nevertheless, the best performing dictionary has 200 words (100 for lesion-based words and 100 for normal), with a corresponding AUC = 93.3%. For reference, at 90% sensitivity, the system achieves 83% specificity.

In the case of bright lesion classification, there is a large difference between the characterizing PoIs for bright lesion and nonlesion structures and a 500-word, being a sparse large dictionary, is able to best capture the characteristics associated with
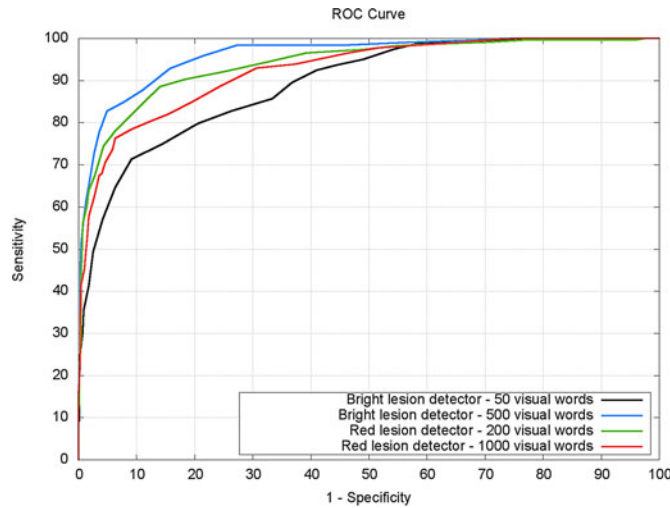
[1]http://www.recod.ic.unicamp.br/site/asdr/

Fig. 6.   Classification results for images showing bright and red lesions using visual dictionaries. Experiments were performed with fivefold cross validation using the DR1 dataset. The two ROC curves for red and bright lesions represent the best and worst choices for the dictionary size. AUC = 95.3%. for bright lesions, AUC = 93.3% for red lesions.
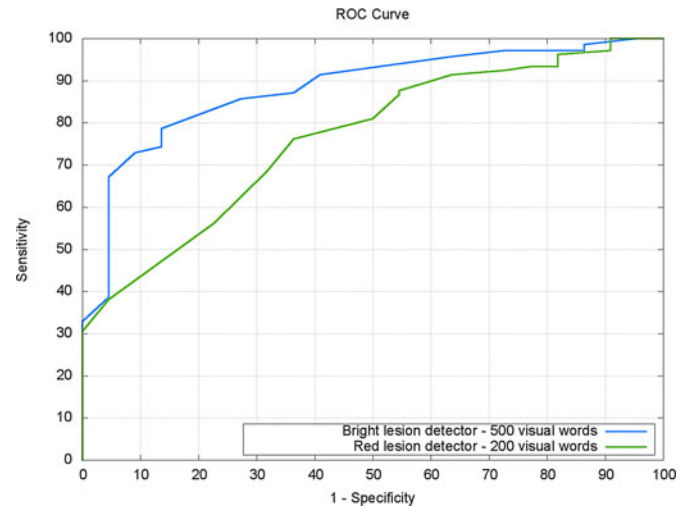

Fig. 7.   Cross training-testing classification results for images showing bright and red lesion classification. Training with the DR1 dataset and testing over the RetiDB dataset. AUC = 88.1% for bright lesions and AUC = 76.4% for red lesions.

white lesions. On the other hand, for red lesions and some nonlesions, the PoIs are more difficult to differentiate, which means a small and dense (200-word) dictionary is more appropriate since the visual words quantified is the main factor that distinguishes among these.

B. Experiments—Part #2

The second part of our experiments consisted of cross dataset classification. The experiments were conducted using the DR1 dataset as the training data and the RetiDB and Messidor datasets as the testing datasets. The dictionary size was fixed based on the outcome of experiments in Part #1.

The RetiDB[2] dataset contains 22 normal images, 71 images tagged as having bright lesions, and 106 tagged as having red lesions. The images are in 1500×1152-pixel resolution. An image can be normal or contain either bright, red, or both lesions.

The Messidor[3] dataset contains 1200 images with three resolutions (1440×960, 2240×1488, or 2304×1536 pixels). The images are tagged by specialists according to the following.
1) *Risk of macular edema*: whether or not the image contains hard exudates. It comprises three groups, $G_{0,1,2}$. G_0 contains images with no visible exudates and $G_{1,2}$ contain images with exudates. For our experiments, we used all the images in these three groups. Groups G_1 and G_2 together contain 226 images.
2) *Retinopathy grade*: grading was performed with respect to the presence of red lesions. The grading resulted in four groups ($Risk_{1,2,3}$ and normal images). We considered the normal images and the $Risk_2$ and $Risk_3$ images only, which contain examples of, at least, five microaneurysms or at least one hemorrhage spot. The $Risk_2$ and $Risk_3$ groups contain 247 and 254 images, respectively.

[2]http://www2.it.lut.fi/project/imageret/diaretdb1/
[3]http://messidor.crihan.fr

Currently, our approach still does not classify images with less than five spots of MAs. Therefore, we did not include $Risk_1$ in this paper.

The normal group contains 546 images with no lesions.

During training, the parameters of the SVM were automatically determined using the standard LibSVM's grid search fine tuning algorithm [38] on the DR1 images only. The best SVM parameters were SVM regularizer $C = 32.0$ and the kernel RBF parameter $\gamma = 0.0078125$ for the bright lesion classifier. For the red lesion classifier, $C = 8.0$ and $\gamma = 0.03125$.

Fig. 7 shows the cross dataset classification results for images with bright and red lesions for RetiBD. For bright lesion detection, our approach used a dictionary with 500 words and resulted in an AUC = 88.1% with images from the RetiDB dataset. For comparison purposes, [41] provides results using the RetiDB dataset for training and classification. The authors reported a 70%/99% sensitivity/specificity result. For this level of sensitivity, our approach results in 85% specificity. However, in our case the training data does not include any images from RetiDB. For red lesions, our method achieved an AUC = 76.4%. For reference, at 90% sensitivity, the cross dataset method yielded a 60% specificity.

Fig. 8 depicts the classification results for Messidor images showing bright lesions, and three results for red lesions: the result for $Risk_2$ and $Risk_3$ separated, and the union of both sets. For bright lesions, AUC = 89.3%, and for a 90% sensitivity the result was a 64% specificity. For red lesions, AUC = 86.2%, 63.3% and 72.2% for $Risk_3$ alone, $Risk_2$ alone, and both combined.

The results for red lesions show a negative characteristic of our method—it is somewhat sensitive to the number of red lesions, and images with a low number of red lesions are not detected as such. We are currently exploring two alternatives to explain this limitation, and correct it. The first explanation is that the PoI detector in SURF is not detecting most of the
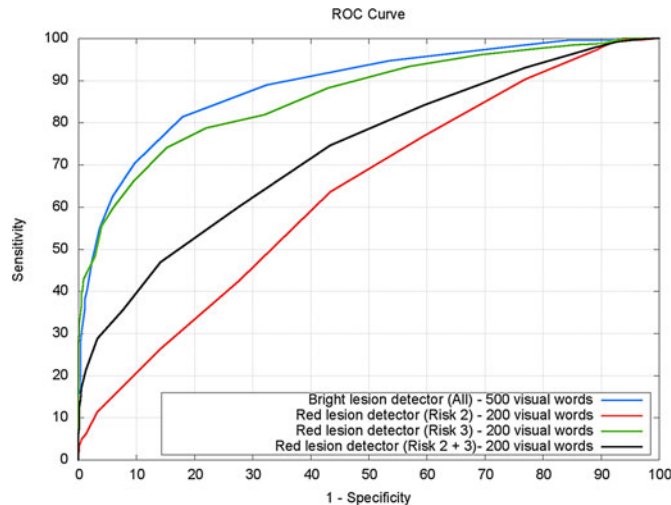
Fig. 8. Cross dataset training-testing classification results for images showing bright and red lesions classification. Training with the DR1 dataset and testing over the Messidor dataset. AUC = 89.3% for bright lesion and AUC = 86.2%, 63.3%, and 72.2% for $Risk_3$ alone, $Risk_2$ alone, and both combined.

points relevant to detection of the red lesion (microaneurysms, MAs). Indeed, the PoI in SURF uses discontinuities to mark a point as interesting, and sometimes MAs are not too different to their surroundings—and that is why they are so hard to detect. If this explanation is correct, then there are two alternatives to improve the results for a low number of MAs: either preprocessing the image so that MAs become more discontinuous from their surroundings, and thus more likely to be detected by SURF, or dropping SURF's point detection itself and performing a more dense sample of the retina, following the method used by Li and Perona for general scene categorization [29].

The alternative explanation is that the points are being detected, but they are too few that their influence on the image signature is too small to be detected by the learning method (SVM). A solution to the problem is to alter the quantization step. The current quantization is a hard assignment with sum aggregation. There are alternatives of soft assignment (each PoI is attributed to all visual words inversely proportional to its distance) and the aggregation using the max function (each entry remembers only the largest PoI attribution).

V. Conclusion

Automated screening algorithms need to be accurate in detecting lesions, identifying the type of lesions, be easily applied across diverse populations and the use of different image capturing equipment. Locating precisely a specific lesion necessitates extensive pre- and postprocessing of images. In addition, similar acquisition conditions are often needed for good performance. This is not often possible in large screening programs and processing centers, where local screening programs often use different technologies, have varying expertise in image capture and consist of diverse ethnic groups. Therefore, an automated method needs to be robust against these conditions. The approach described in this paper addresses these issues by being able to identify the presence of specific lesions.

The objective of this paper was to describe a solution for detecting different lesions (a task that normally require different preprocessing depending on the lesion) without deploying any pre- or postprocessing operations. We introduced a visual word dictionary-based approach that applies PoIs and visual dictionaries for classifying bright or red lesions found in optic fundus images.

The main requirement of the method was to use a training set where ophthalmic specialists have marked RoIs within the retinal image that contain the lesion. The approach automatically calculates PoIs that are representative and highly distinctive of such regions and, at the same time, are scale-invariant and robust to some image transformations. The method builds a powerful visual dictionary upon the PoIs. After the projection of the PoIs of a test image onto the dictionary (projection) space, the method uses a machine learning classifier to point out the classification of such an image.

The novelty of our approach is in the characterization of lesions using visual words that incorporate information such as texture and boundary discontinuities with the specialists' knowledge in a framework easily extendible to different types of DR-related lesions. The approach also extends the current state-of-the-art dictionary-based classification techniques (e.g., [17]–[21]) by building a specific projection space for each lesion of interest (e.g., normal, bright, or red lesions) instead of a common dictionary for all types of lesions. This has the advantage of correctly capturing the subtleties of the lesion of interest and avoiding areas of the image dominating the calculus where no lesions are present as is the case for the majority of images. Finally, the incorporation of RoIs as defined by the specialists, which then define the boundaries within which the PoIs are located into the dictionary calculus is also innovative. Current state-of-the-art dictionary-based classification techniques tend to find PoIs anywhere in the training images and not within specific RoIs.

We validated the proposed approach with a series of experiments on publicly available datasets of retinal images. Our work adds to previous reports using cross dataset training, which is an important area of research and an important component in clinical practice where classifiers cannot always be trained on an identical image set (same camera, resolution, operator, FOV) that needs to be classified.

The best results were achieved using SURF, for the detection and description of the PoIs and a visual dictionary with 500 visual words for images showing bright lesions and 200 visual words for images showing red lesions. The approach achieved results comparable to those in the literature. However, comparing our results for classification of bright or red lesions directly is difficult as our classifier classifies the images as having signs of specific lesions and accuracy is not determined by identifying the correct location.

By focusing on the goal of image-based metrics, as opposed to DR-related lesion detection per se, the method is able to learn what images containing the lesion "look like" (based on the quantity and distribution of the visual words in the image) instead of detecting the lesion in the image itself.

The visual dictionary is an elegant method to learn and represent important features of a specific lesion, and allows classifying whether or not an image contains these lesions with a unified approach. This approach is independent of the image resolution, color space representation, and does not assume any specific size of the lesions. Furthermore, the approach is robust across differences in the appearance of the lesions and the retina across different ethnicities. In [14], we report the result of applying this technique on Australian aboriginal cases, with better results than any other approach published.

Cross-dataset learning—training of one dataset to test the performance on an entirely different dataset—is the ultimate test to expose the weaknesses and robustness of a method. As we report here, it showed us that there is still some room for improvement and adjustment on the deployment details of our approach. We strongly believe the community as a whole should adopt this validation procedure from now on.

A. Future Work

This paper has shown how to construct detectors for two of the most common lesions in DR. The work continues in three different directions. As mentioned, the red lesion detector is still dependent on the number of spot signs in the images. We are currently exploring the alternatives discussed in Section IV-B—pre-processing the image, dense SURF, and soft quantization with max aggregation [34].

On the other hand, it is possible that even a combination of techniques will not be enough to detect a very small number of microaneurysms. However, retinae with fewer than three microaneurysms have a higher disappearance rate compared to retinae with four or more aneurysms. This suggests that the identification of five microaneurysms can be seen as the stage where a referral is necessary [42], [43]. Our results are in line with current clinical practice as the Wisconsin work was based on seven-field ETDRS criteria and the U.K. prospective study on four retinal fields while current automated screening uses one or two retinal views (foveal field and optic disk centered). Microaneurysms outside of the foveal field are not immediately sight threatening and only 50% of microaneurysms were within this region reported by Klein *et al.* [44]. A plausible conclusion is that, for screening purposes, detection of less than five microaneurysms is not that useful.[4]

The second direction of future research involves the development of an automatic DR screening system, that can detect the presence of many abnormalities related to DR. Thus, we are developing detectors for other lesions (cotton wool spots, drusen, superficial hemorrhage, neovascularization) and defining a way to combine the output of each of these detectors into a single decision on whether or not the patient should be referred to a specialist and the level of disease progression. The development of other detectors follows in part the steps laid down by the two detectors described herein—detecting PoIs, mapping them to visual words and learning a classifier to distinguish the different histograms. However, not all forms of lesions are equally
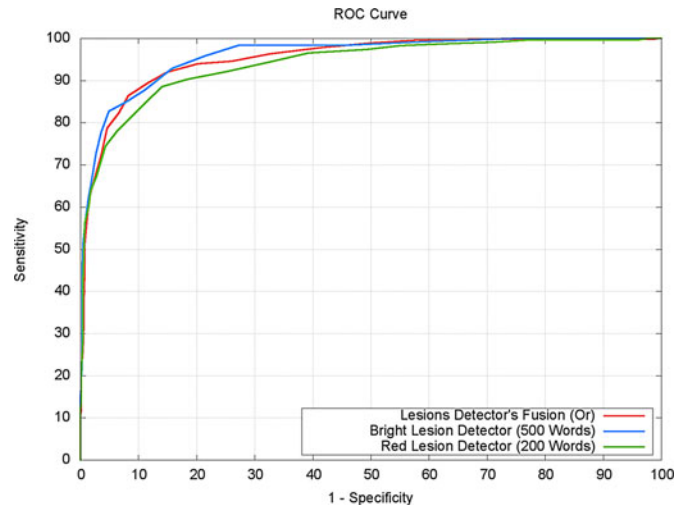

Fig. 9. Results with combining the two detectors with an OR. The average result of a fivefold cross validation on DR1 alone. The images are only the ones in DR1 where either bright or red lesions were detected. AUC = 95.2%.

amenable to such approach—and some of them may require some changes to the process. For example, neovascularization suffers the problem of not having highly discontinuous points to be detected by SURF, and thus we are following the dense traditional SURF approach in such cases.

The third direction of future research is how to combine the results of each of the detectors. In principle, the combination is simple if any detector identifies a lesion, the patient should be referred to the specialist. This is what we call an OR-rule. Fig. 9 presents the result of using the OR rule for red and bright lesion fusion. The figure shows the average of a fivefold cross-validations using the DR1 data. The combination of detectors has a promising AUC of 95.2. But we can go beyond the OR-rule. The SVM classifiers used in the detectors may return not just the class to which the image belongs to, but some measure of distance of that image to the boundary region. Thus, we can use this information to combine the results in other ways. For example, if all identify that the image is normal but with a low certainty of that determination (this may be the case if there are dust particles captured in the image during photography), it might be prudent to refer the patient to the specialist. We are currently exploring some of these alternatives to combine the output of each detector.

REFERENCES

[1] 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]Personal Communication, Allan Luckie, Albury Eye Clinic, Australia.

[2] S. R. Salomão, M. R. K. H. Mitsuhiro and R. Belfort, Jr., "Visual impairment and blindness: An overview of prevalence and causes in Brazil," *Anals Braz. Acad. Sci.*, vol. 81, pp. 539–549, 2009.

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

[4] 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," *Br. Med. J.*, vol. 320, pp. 1627–1631, 2000.

[5] L. Giancardo, F. Meriaudeau, T. Karnowski, Y. Li, K. Tobin, and E. Chaum, "Microaneurysm detection with radon transform-based classification on retina images," in *Proc. Intl. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 5939–5942.

[6] B. Antal, I. Lazar, A. Hajdu, Z. Torok, A. Csutak, and T. Peto, "Evaluation of the grading performance of an ensemble-based microaneurysm detector," in *Proc. Intl. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 5943–5946.

[7] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1223–1232, 2006.

[8] M. J. Cree, E. Gamble, and D. J. Cornforth, "Colour normalisation to reduce inter-patient and intra-patient variability in microaneurysm detection in colour retinal images," in *Proc. Workshop Digital Image Comput.*, 2005, pp. 163–168.

[9] M. Niemeijer, B. van Ginneken, S. R. Russell, and M. S. A. Suttorp-Schulten, M. D. Abràmoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 5, pp. 2260–2267, 2007.

[10] 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," *Phys. Med. Biol.*, vol. 52, no. 24, pp. 7385–7396, 2007.

[11] A. D. Fleming, K. A. Goatman, S. Philip, G. J. Williams, G. J. Prescott, G. S. Scotland, P. McNamee, G. P. Leese, W. N. Wykes, P. F. Sharp, and J. A. Olson, "The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy," *Br. J. Ophthalmol.*, vol. 94, no. 6, pp. 706–711, 2010.

[12] M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, L. Qin, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M. D. Abràmoff, "Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 185–195, Jan. 2010.

[13] H. F. Jelinek, K. Al-Saedi, and L. Backlund, "Computer assisted top-down assessment of diabetic retinopathy," in *Proc. World Congr. Biophys. Biomed. Eng.*, 2009, pp. 127–130.

[14] H. F. Jelinek, A. Rocha, T. Carvalho, S. Goldenstein, and J. Wainer, "Machine learning and pattern classification in identification of indigenous retinal pathology," in *Proc. Intl. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 5951–5954.

[15] Y. Li, T. P. Karnowski, K. W. Tobin, L. Giancardo, S. Morris, S. E. Sparrow, S. Garg, K. Fox, and E. Chaum, "A health insurance portability and accountability act-compliant ocular telehealth network for the remote diagnosis and management of diabetic retinopathy," *Telemedicine e-Health*, vol. 17, no. 8, pp. 627–634, 2011.

[16] A. Ruggeri, E. Poletti, D. Fioroin, and L. Tramontan, "From laboratory to clinic: The development of web-based tools for the estimation of retinal diagnostic parameters," in *Proc. Intl. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 3379–3382.

[17] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *Proc. IEEE Intl. Conf. Comput. Vision*, 2003, pp. 1470–1477.

[18] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Proc. IEEE Intl. Conf. Comput. Vision*, 2005, pp. 1800–1807.

[19] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Intl. J. Comput. Vision*, vol. 60, no. 1, pp. 63–86, Jan. 2004.

[20] E. N. Mortensen, H. Deng, and L. G. Shapiro, "A SIFT descriptor with global context," in *Proc. IEEE Intl. Conf. Comput. Vision Pattern Recogn.*, 2005, pp. 184–190.

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

[22] M. D. Abràmoff, M. K. Garvin and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010.

[23] R. Winder, P. Morrow, I. McRitchie, J. Bailie, and P. Hart, "Algorithms for digital image processing in diabetic retinopathy," *Computerized Med. Imag. Graph.*, vol. 33, pp. 608–622, 2009.

[24] 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," *J. Med. Syst.*, vol. 36, no. 1, pp. 145–157, Feb. 2012.

[25] H. Wang, W. Hsu, K. G. Goh, and M. L. Lee, "An effective approach to detect lesions in color retinal images," in *IEEE Intl. Conf. Comput. Vision Pattern Recogn.*, vol. 2, pp. 181–186, 2000.

[26] B. Zhang, K. Karray, L. Zhang, and J. You, "Microaneurism (MA) detection via sparse representation classifier with MA and non-MA dictionary learning," in *Proc. Intl. Conf. Pattern Recogn.*, 2010, pp. 277–280.

[27] H. F. 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," *Clin. Exp. Optometry*, vol. 89, pp. 299–305, 2006.

[28] C. Agurto, V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, S. Russel, and M. Abràmoff, P. Soliz, "Multiscale AM-FM methods for diabetic retinopathy lesion detection," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 502–512, Feb. 2010.

[29] F.-F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Intl. Conf. Comput. Vision Pattern Recogn.*, vol. 2, pp. 524–531, Jun. 2005.

[30] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Workshop Stat. Learn. Comput. Vision*, 2004, pp. 1–8.

[31] P. Hough and B. Powell, "A method for faster analysis of bubble chamber photographs," *Il Nuovo Cimento*, vol. 18, pp. 1184–1191, 1960.

[32] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Feb. 2004.

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

[34] E. Valle, "Local-descriptor matching for image identification systems," Ph.D. dissertation, École Doctorale Sciences et Ingénierie, Universit de Cergy-Pontoise, Cergy-Pontoise, France, 2008.

[35] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Comput. Vision Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[36] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Berlin, Germany: Springer, 2006.

[37] J. P. Papa and A. Rocha, "Image categorization through optimum path forest and visual words," in *Proc. Intl. Conf. Image Process.*, 2011, pp. 3586–3589.

[38] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines*, 2001, software at [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm

[39] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med. Image Anal.*, vol. 16, no. 1, pp. 216–226, 2012. Available: http://www.ncbi.nlm.nih.gov/pubmed/21865074

[40] C. I. Sánchez, M. Niemeijer, A. V. Dumitrescu, M. S. A. Suttorp-Schulten, M. D. Abràmoff, and B. van Ginneken, "Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data," *Invest. Ophthalmol. Vis. Sci.*, vol. 52, no. 7, pp. 4866–4871, Jul. 2011.

[41] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images," *Comput. Med. Imag. Graph.*, vol. 34, no. 3, pp. 228–235, 2010.

[42] A. K. Sjølie, R. Klein, M. Porta, T. Orchard, J. Fuller, H. H. Parving, R. Bilous, S. Aldington, and N. Chaturvedi, "Retinal microaneurysm count predicts progression and regression of diabetic retinopathy. Post-hoc results from the direct programme," *Diabetic Med.*, vol. 28, no. 3, pp. 345–351, 2011.

[43] E. M. Kohner, I. M. Stratton, S. J. Aldington, R. C. Turner, and D. R. M. for the UK Prospective Diabetes Study (UKPDS) Group", "Microaneurysms in the development of diabetic retinopathy (UKPDS 42)," *Diabetilogia*, vol. 42, pp. 1107–1112, 1999.

[44] R. Klein, S. M. Meuer, S. E. Moss, and B. E. Klein, "Retinal microaneurysm counts and 10-year progression to diabetic retinopathy," *Arch. Ophthalmol.*, vol. 113, pp. 1386–1391, 1985.