Machine Learning and Pattern Classification in Identification of Indigenous Retinal Pathology

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

Abstract—**Diabetic retinopathy (DR) is a complication of diabetes, which if untreated leads to blindness. DR early diagnosis and treatment improve outcomes. Automated assessment of single lesions associated with DR has been investigated for sometime. To improve on classification, especially across different ethnic groups, we present an approach using points-of-interest and visual dictionary that contains important features required to identify retinal pathology. Variation in images of the human retina with respect to differences in pigmentation and presence of diverse lesions can be analyzed without the necessity of preprocessing and utilizing different training sets to account for ethnic differences for instance.**

I. INTRODUCTION

INDIGENOUS populations such as the Australian Aborigine, all have an increased incidence of diabetes compared to the Caucasian population resident in these countries [1]. To optimize screening, detection and treatment, mobile screening combined with automated classification of disease can be used [2].

Automated assessment of pre-proliferative diabetic retinopathy has been possible for some time using fluorescein-labeled images [3]. Results for color fundus analysis identifying microaneurysms, exudates and cotton-wool spots as well as proliferative retinopathy have only been reported more recently [4-6].

To optimize automated processing of color images one has to consider intra-image variation such as light diffusion, pathology, variation in fundus reflectivity and fundus thickness and inter-image variation (being the result of using different cameras, illumination, acquisition angle and retinal pigmentation). Several methods are available including grey world normalization, histogram equalization and histogram specification [7]. Color normalization also increases the discrimination in almost all of the color features [8].

A. Rocha, T. Carvalho, S. Goldenstein, and J. Wainer are with the Reasoning for Complex Data Lab., Institute of Computing, Univ. of Campinas (Unicamp), Campinas, SP, Brazil. Contact (phone/fax: +55 19 35215854; e-mail: anderson.rocha@ic.unicamp.br).

H.F. Jelinek is with the Centre for Research in Complex Systems and the School of Community Health, Albury, NSW, Australia (phone: +61 2 60519219; fax: +61 2 60519219; e-mail: hjelinek@csu.edu.au).

This paper proposes a process that does not require preprocessing but deals with image differences of the retinal fundus directly. The approach constructs a visual dictionary to represent important features that characterize the pathology of interest and uses pattern recognition tools to classify the retinal images into disease and non-disease.

II. METHODS

A. Diabetic Retinopathy Images

For creating the visual words and training the detectors (two-class classifiers), 672 non-DR images, 261 images with bright lesions and 246 images with red lesions from the Ophthalmology Dept., Federal University of São Paulo, Brazil were used. The images came from patients with different racial background and were manually graded by specialists. The images ranged from 640×640 to 1,581×1,113 pixels in resolution. All of these images were used for creating the dictionaries and for training the two-class classifiers (normal *vs.* bright lesions and normal *vs.* red lesions). The aboriginal test images were obtained from the Albury Eye Clinic using a Topcon camera at 1,200×1,200-pixel resolution. Retinal images contained either no pathology, red lesions, or bright lesions.

B. Detection of Features

Every image in a collection is represented using a large number of points of interest (PoI) [9] and a local descriptor around each PoI using the Speeded-Up Robust Features (SURF) approach is calculated [10]. This is then stored in an indexing data structure. PoIs are robustly estimated as they convey more information than other points in the image. A PoI shows repeatability after several image transformations across different scales.

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) and integral images [11] to speed up the operations.

(2) **Feature point localization**: The determinant of the Hessian for both location and scale is used. 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

$$H(x, \sigma) \quad \begin{bmatrix} Lxx(x, \sigma) & Lxy(x, \sigma) \\ Lxy(x, \sigma) & Lyy(x, \sigma) \end{bmatrix} \quad (1)$$

where $Lxx(x, \sigma)$ is the convolution of the Gaussian second order derivative with the image I in point x.

The scale-spaces are implemented as image pyramids by repeatedly smoothing the images with a Gaussian and subsequently sub-sampling to achieve a higher level of the pyramid. To localize the PoIs in the image across different scales, the method performs non-maximum suppression in a $3 \times 3 \times 3$ neighborhood. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space.

(3) **Orientation assignment**: SURF calculates the Haar-wavelet responses in x and y directions using a circular neighborhood of radius $6s$ around the interest point, where s σ is the scale at which the interest point was detected. Fast filtering, in each scale s, the method calculates wavelet responses using integral images. The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window covering an angle of π. The point of interest gets the orientation from the longest vector.

(4) **PoI characterization**: SURF creates a square region centered on the PoI, and oriented along the orientation selected in Step (3). The region is split up regularly into smaller 4×4 square sub-regions. For each sub-region, the method computes some simple features at 5×5 regularly spaced sample points.

Basically, Stages (1) and (2) gives the points of interest (PoIs) while Stage (3) assigns the orientation of each PoI and Stage (4) performs the description of each PoI.

C. Visual Vocabulary

The creation of the dictionary is outlined in Fig. 1. SURF is a good low-level representative feature detector with several applications in computer vision. However, SURF-based approaches are often designed to provide exact matching and they do not translate directly into good results for image classification in broad or even constrained domains. Therefore, we use the concept of visual vocabularies [12] to capture the high-distinctiveness of PoIs while using such discrimination for image classification.

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 discuss in this paper is general enough to detect other DR-related anomalies as we show in Section IV and V.

To solve the problem of detecting bright lesions 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 SURF. In this work, we do not perform any preprocessing on the images.
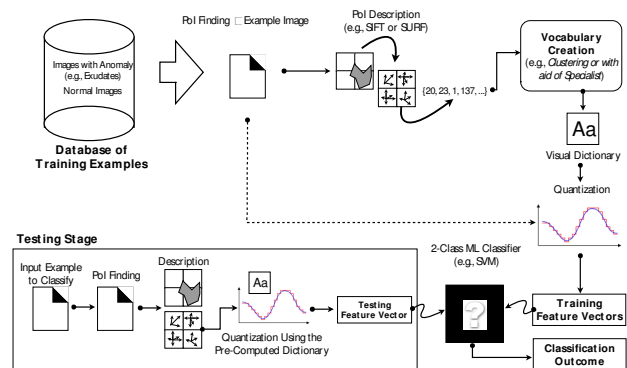

Fig. 1. General pipeline of the proposed approach.

Each image in the training generates a series of points of interest. After finding the PoIs, the dictionary or codebook was created, which represents distinctive regions of the images with bright lesions as well as images tagged as normal by specialists. Our objective when creating a visual dictionary is to learn, from a training set of examples, a model that selects the more representative regions for our problem. The size of our dictionary must be large enough to distinguish relevant changes in the images and disregard irrelevant features.

To create the "dictionary", we need to choose its size k, or number of representative words. During training, the specialists select regions of interest in the analyzed images and creating masks for candidate regions more likely to contain the DR anomaly of interest. The points of interest are then considered in these more likely DR regions. To create the dictionary, we can then perform clustering such as k-means for finding representative centers for the cloud of PoIs or simply pick PoIs within the specialist marked regions as we indeed do in this paper.

A good dictionary is the one that captures properties of the DR anomaly of interest as well as properties of normal images. Therefore, we create the dictionary using 50 "words" representing the DR anomaly of interest (e.g., bright lesion) and 50 "words" representing normal retinas (non-DR) [12]. Note that in the training stage we used common images and not specific indigenous images.

### D. Training and Classification

The visual dictionary was created from the training images, using the fine selection of candidate regions in normal and abnormal images, which the specialists marked. The selection process is only performed in the training stage. Once the dictionary was created, each of PoIs from the training set was assigned to the closest visual word of the dictionary. This step is known as *quantization*. At the end of the quantization process, a set of feature vectors representing the histogram of

the selected visual words for each image are obtained. The final classification procedure was performed using the Support Vector Machine (SVM) algorithm. The classifier was trained by adding the feature vectors calculated from the training images containing positive (e.g., images containing pathology) and negative (normal images) examples. The parameters for the SVM were settled during training using SVM grid search.

To analyze multiple pathologies, we can create different dictionaries (normal *vs.* bright lesions, normal *vs.* red lesions) and train different two-class classifiers that can be combined later using state-of-the-art machine learning classifier fusion.

III. RESULTS AND DISCUSSION

In this section, we present the results for our classification techniques based on points of interest and visual words. Fig. 2 shows typical signatures for normal images *vs.* bright lesions.
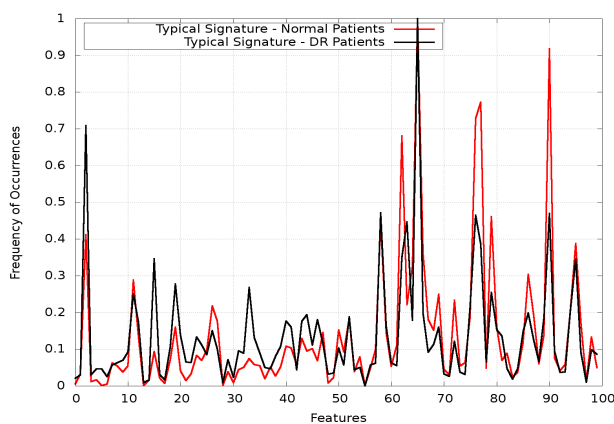

Fig. 2. Typical signatures for normal patients *vs.* patients with exudates. DR patients have higher frequencies for abnormal words (positions 1-50) while the contrary should happen for positions 51-100.

The plot depicts 100 "words" and their frequency in the training set. For this research we chose to use 100 words as this number was found to be most efficient and effective for identification of lesions. Increasing the number of words normally gives higher responses and also increases the computational time. Thus positions 1-50 represent anomaly-based words (e.g., exudates in this case) while positions 51-100 represent words for normal regions (non-DR). For positions 1-50, it is expected that the abnormal words dominate the normal words while the contrary should happen for positions 51-100.

The typical signatures for bright and red lesions are calculated based on the training images. Given an unseen aboriginal image for testing, the process consists of calculating the points of interest in this image, and mapping such points to the proper visual dictionary (normal and with anomalies).

Fig. 3 depicts two aboriginal images for bright lesions detection/classification.

Fig. 3 (top) shows an example of a good characterization of an aboriginal image in which traces of bright lesions can clearly be seen. This is due to the dominance of the tested signatures with respected to the typical signature of a DR patient with bright lesions (higher frequencies in positions 1-50).
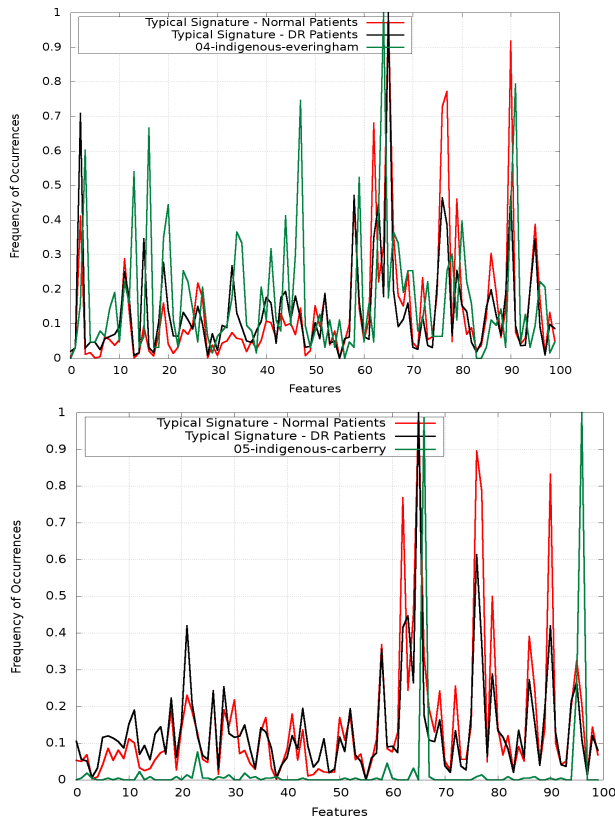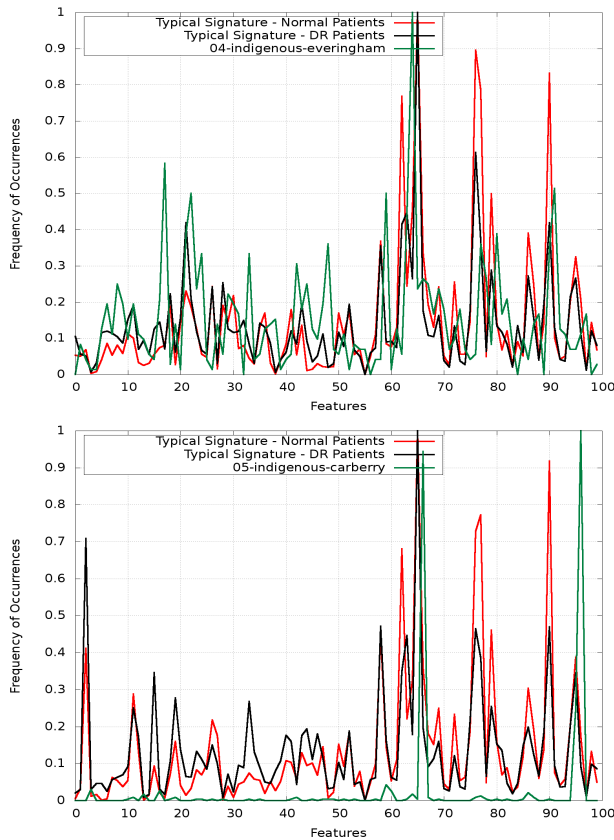

Fig. 3. Two examples of visual words characterization for DR aborigine bright lesions detection. Best characterization results (top) against worse characterization results (bottom).

On the other hand, Fig. 3 (bottom) shows a bad characterization in which it is not possible to differentiate between a DR candidate for bright lesion since the signature found in the tested image does not dominate the typical signature for DR or normal patients. Fig. 4 presents similar results for red lesions. Fig. 4 (top) indicates that the tested image is a DR candidate for red lesion since its signature for positions 1-50 dominates the typical red lesions signature while for positions 51-100 the typical normal signature dominates. Fig. 4 (bottom) depicts an example in which the red lesions characterization is not as good and no clear conclusion can be drawn.

This kind of analysis is very important to visualize what is happening in the feature level characterization of the analyzed images, but difficult to automate. Therefore, use is made of a typical machine learning classifier to learn such behaviors and

test unknown images. For this we selected an operational point in which false negatives, for instance, were penalized.

Fig. 4. Two examples of visual words characterization for DR aborigine red lesions detection. Best characterization results (top) against worse characterization results (bottom).

The classification then achieved promising results, correctly classifying about 80% of the aboriginal images from a training set of non-aboriginal images. The two-class SVM classifier with a radial basis kernel was used with class weighting during training to balance differences in the number of examples for normal, bright, and red-lesion images, and at the ROC operational point of 90% sensitivity and 80% specificity.

This is an important result as it indicates not only that the method works on indigenous images but also that cross training has not reduced the accuracy of the procedure. Having a common approach which is robust to cross-training can be an important milestone in DR pathology detection research.

IV. Conclusion

The use of the visual dictionary is a robust method to learn and represent important features of a given anomaly even in the presence of noise and differences in color background of the retinal fundus in different populations.

Ocular-fundus images were classified as normal or DR candidates using a cross-training methodology that was robust against differences in retinal fundus color and therefore easy to implement worldwide. In addition, the unified proposed approach allows us to develop different detectors under the same simple underlying characterization procedure allowing more than one lesion to be identified at a time therefore yielding better differentiation in terms of DR disease progression. As a future direction, we aim at investigating machine learning feature and classifier fusion techniques in order to combine different anomaly detectors toward more discriminative DR *vs.* Non-DR classifiers.

References

[1] P. Mitchell, S. Foran, T. Y. Wong, B. Chua, I. patel, and E. Ojaimi, *Guidelines for the Management of Diabetic Retinopathy*. Canberra: NHMRC, 2008.

[2] H. F. Jelinek, D. Cornforth, M. Cree, J. Cesar, R.M., J. J. G. Leandro, J. V. B. Soares, and P. Mitchell, "Automated characterisation of diabetic retinopathy using mathematical morphology: a pilot study for community health," in *NSW Primary Health Care Research and Evaluation Conference*, Sydney, 2003, pp. 48.

[3] M. J. Cree, J. A. Olson, K. McHardy, P. Sharp, and J. Forrester, "A fully automated comparative microaneurysm digital detection system.," *Eye,* vol. 11, pp. 622-628, 1997.

[4] A. L. Karperien, H. F. Jelinek, J. J. G. Leandro, J. V. B. Soares, J. Cesar, R.M., and A. Luckie, "Automated detection of proliferative retinopathy in clinical practice," *Clinical Ophthalmology,* vol. 2, pp. 109-122, 2008.

[5] H. Wang, W. Hsu, K. G. Goh, and M. L. Lee, "An effective approach to detect lesions in colour retinal images," in *IEEE Int. Conf. in Computer Vision and Pattern Recognition*, 2000, pp. 181-187.

[6] L. Streeter and M. J. Cree, "Microaneurysm detection in colour fundus images," in *Image and Vision Computing*, 2003, pp. 280-284.

[7] K. A. Goatman, M. J. Cree, J. A. Olson, P. F. Sharp, and J. V. Forrester, "Automated measurement of microaneurysm turnover," *Investigative Ophthalmology and Visual Science,* vol. 44, pp. 5335-5341, 2003.

[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 *Workshop on Digital Image Computing*, Brisbane, Australia, 2005, pp. 163-169.

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

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

[11] P. Viola and M. Jones, "Robust real-time face detection," *Intl. Journa of Computer Vision,* vol. 52, pp. 137–154, 2004.
[12] A. Rocha, T. Carvalho, S. Goldenstein, and J. Wainer. "Points of Interest and Visual Dictionary for Retina Pathology Detection," Technical Report IC-11-07, Institute of Computing, Univ. of Campinas, Campinas, Brazil, 2011.