

## A BAG OF WORDS APPROACH FOR DISCRIMINATING BETWEEN RETINAL IMAGES CONTAINING EXUDATES OR DRUSEN

*M.J.J.P. van Grinsven<sup>\*</sup>, A. Chakravarty<sup>†</sup>, J. Sivaswamy<sup>†</sup>, T. Theelen<sup>\*</sup>, B. van Ginneken<sup>\*</sup>, C.I. Sánchez<sup>\*</sup>*

<sup>\*</sup> Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands

<sup>†</sup> Center for Visual Information Technology, IIIT Hyderabad, Hyderabad, India

### ABSTRACT

Population screening for sight threatening diseases based on fundus imaging is in place or being considered worldwide. Most existing programs are focussed on a specific disease and are based on manual reading of images, though automated image analysis based solutions are being developed. Exudates and drusen are bright lesions which indicate very different diseases, but can appear to be similar. Discriminating between them is of interest to increase screening performance. In this paper, we present a Bag of Words approach which can be used to design a system that can play the dual role of content based retrieval (of images with exudates or drusen) system and a decision support system to address the problem of bright lesion discrimination. The approach consists of a novel partitioning of an image into patches from which color, texture, edge and granulometry based features are extracted to build a dictionary. A bag of Words approach is then employed to help retrieve images matching a query image as well as derive a decision on the type of bright lesion in the given (query) image. This approach has been implemented and tested on a combination of public and local dataset of 415 images. The area under the curve for image classification is 0.90 and retrieved precision is 0.76.

### 1. INTRODUCTION

Diabetic Retinopathy (DR) and Age Related Macular Degeneration (AMD) are among the major causes of irreversible blindness in the world. At the early stages, DR is characterized by red and bright lesions (called exudates) which appear as bright, reflective, white or cream colored lesions in color fundus images. AMD is chiefly characterized by drusen which appear as small yellow or white spots located throughout the fundus but more so in the macular region. Thus both drusen and exudates share much in appearance. This is a potential source of problem for readers or a computer aided diagnostic (CAD) tool based screening scenario designed for DR. CAD solutions are needed to help distinguish between the bright lesions. This can be done by designing a system which i) classifies a given bright lesion image as having exudates or drusen or ii) retrieves similar looking images from the collection of the screening program along with labels given by experts. A system which can do both is attractive as it allows flexibility in the type of end use that can be supported.

We employ a Bag of Words (BoW) approach to the problem and offer a solution which results in both retrieval and classification of a given (query) image with bright lesions.

### 2. BACKGROUND

Literature on discriminating between images with exudates and drusen is limited. Further, a Content-Based Image Retrieval (CBIR) approach is absent. The general strategy is to detect lesion candidates, extract a set of features and use supervised classification. Detection methods such as multi-scale morphological filtering [1] as well as pixel level labelling (as lesion or not a lesion) [2, 3] have been attempted. Statistical features [1, 2, 3] and AM-FM features [4] have been used for classification. Contextual information such as proximity to microaneurysm (MA) is used in [3] to help assess the candidate as being a true lesion.

[1, 2, 3] rely on accurate segmentation of the lesions for image level decision between drusen and exudates. Therefore, these methods can suffer due to lack of existing accurate lesion detection methods. Moreover, automatic detection of MA and dark lesions for contextual information in itself is a challenging task. In [4] a grid based approach is adopted to overcome this problem. However, processing of a large number of image patches increases computational load. Moreover, the grid size remains same across images which may be of different resolution and orientation.

CBIR-based approaches have been explored for detecting and staging diabetic macular edema [5, 6, 7], distinguishing AMD from Stargardts disease [8] as well as distinguishing among proliferative, non-proliferative DR and AMD [9]. All these approaches divide the image into regions and perform a local analysis to index a given image with the regions chosen to be anatomically meaningful [10, 9]. Analysis is based on features ranging from local binary patterns (LBP) [5] to AM-FM [9, 8] which are extracted with or without lesion segmentation. We present a CBIR-based approach for identifying and retrieving images with drusen or exudates.

### 3. METHODS

The pipeline of the proposed system is shown in Figure 1. The input image is divided into patches using the macula and optic disk (OD) as reference coordinates. Next a set of features are extracted from each image patch to help discriminate between normal, exudate and drusen patches and used to build a

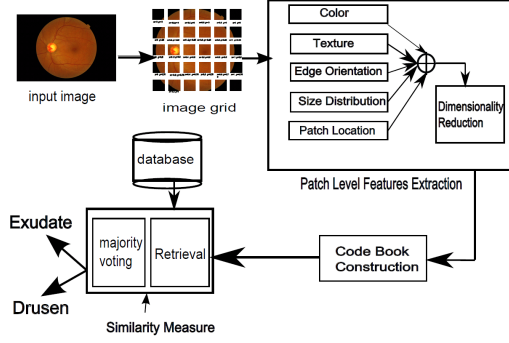


Fig. 1. Pipeline of the Proposed System

dictionary. Given an image, a Bag of Words (BoW) approach is employed to retrieve  $N$  most similar images from the existing dataset and finally an image level decision is made based on the retrieved images.

### 3.1. Patch Extraction

A given image is divided into fixed number of square patches  $P(r)$ , with  $r$  denoting the location with reference to the image origin. If  $r_O$  and  $r_M$  denote the positions (image coordinates) of patches containing OD and macula, it is desirable that they do not change much across the images (views). This condition enables encoding of contextual information using  $r$  and extracting features robust to geometric affine transformations. In our work, we ensured that OD remained at the left of all images and the axis connecting OD and macula remained horizontal by suitably locating them and rotating all images. Further, to ensure the number of patches (30 in our case) remains fixed across images (of different resolution), the size of  $P$  is taken to be 0.6 times the OD-macula distance.

### 3.2. Patch Level Feature Extraction

Features are derived on each patch  $P(r)$  in the image. Since the objective is to retrieve pathologically similar (exudates/drusen) images, derived features must discriminate between normal, exudates and drusen patches. Accordingly, several features based on color, texture, local gradient and morphology were derived. We found color histogram, histogram of Laplacian of Gaussian (LoG), local binary patterns (LBP), histogram of oriented gradients (HoG) and granulometry to be useful. Since a given image can have normal as well as abnormal patches, appropriate features for these two categories were found empirically by carrying out a classification with 5-fold validation. The features found to be suitable for distinguishing normal from abnormal patches were as follows. Histogram features from the RGB, HSV and YCbCr color models: R and B planes (32 bins each); S (64 bins) and V (32 bins); Cb (16 bins). Histogram of Laplacian of Gaussian (8 bins,  $\sigma=12$  of Gaussian), HoG (64 orientation bins); LBP (16 bins) and granulometry (scales 1,3,5,7,9,11).

The features suitable for discriminating a patch with exudates from those with drusen, were as follows. Color histogram features derived from G and B planes (16 bins

each); H, S planes (64 bins each) and V(32); Y (32). Texture features such as HoG and LBP (128 bins each). The two feature sets were individually reduced to 24 and 32 dimensional feature vectors respectively, using Ada-Boost based feature selection. Then they were concatenated to a 56 dimensional feature vector to discriminate between normal, exudate and drusen patches. To handle illumination variations, the features were computed after mean subtraction of the color planes. To account for different size of patches resulting from different resolution images, the histograms were normalized in the range 0,1. All features other than color histograms were computed on the green channel. The position  $r$  was encoded using a 2-D feature vector and concatenated with the above feature vector to obtain a 58-dimensional feature vector per patch.

### 3.3. Codebook construction

To describe the characteristics of each image, we use a BoW approach to encode the information in the image. First, all the patches in the database are categorized into  $k$  clusters using k-means clustering algorithm based on the extracted patch features plus the position features  $r$ . The centroids of the  $k$  clusters are treated as visual words and are used to create a codebook. The codebook consists of  $k$  words describing the image characteristics present in the database. The value for  $k$  was set to 10 which corresponds approximately with the expected number of structures present in the image. Secondly, each image in the database is represented as a histogram of word occurrences (BoW), encoded by assigning each image-patch to one of the  $k$  words in the codebook. To overcome the problem of discriminating between normal and abnormal patches before the codebook construction, all patches are used. Normal patches and abnormal patches are separated into different clusters by the clustering algorithm.

### 3.4. Image retrieval and classification

For a new, unseen query image, patches and features are extracted and the bag of words for that image is constructed. To retrieve images with similar lesions as the query image, a similarity measure is used to compare the query image histogram (BoW) and the database image histograms, providing an image similarity score. Several similarity measures, like squared chord, L1-norm, L2-norm were tested and the one obtaining the best area under the ROC was used. The  $N$  most similar images from the database are retrieved for that query image.

For the classification task, the BoW is used as feature set for supervised classification. Training is done using images from the database with assigned labels. A weighted nearest neighbor approach is used for classification using the similarity metric as distance. The probability  $P(c|i)$  that the query image  $i$  belongs to class  $c$  is calculated as:

$$P(c|i) = \frac{1}{D} \sum_{j=1}^J \frac{1}{d_{ij}}, \quad D = \sum_{n=1}^N \frac{1}{d_{in}} \quad (1)$$

where  $J$  is the number of retrieved images with class  $c$ ,  $N$  the total number of retrieved images and  $d_{ij}$  the distance between the query image  $i$  and the retrieved image  $j$  in feature space.

#### 4. EXPERIMENTS

For this study, we used 188 images containing HE from the public dataset MESSIDOR<sup>1</sup> and 191 images containing drusen from the European Genetic Database (EUGENDA)<sup>2</sup> and 36 images from the public STARE database<sup>3</sup>. Images in the MESSIDOR database were taken at three different clinical sites. All images acquired from the EUGENDA and STARE database were graded as containing drusen only and have no signs of advanced AMD. We constructed 2 datasets, Set A and Set B, which were randomly subdivided into training set and test set. See Table 1 for an overview of the data distribution. In the training set of Set A, we used images

**Table 1.** Data distribution. STA, EUG, MES1, MES2, MES3 stands for STARE, EUGENDA, MESSIDOR (Site 1), MESSIDOR (Site 2) and MESSIDOR (Site 3), respectively.

		STA	EUG	MES1	MES2	MES3
Set A	train	18	95	63	31	18
	test	18	-	-	-	18
Set B	train	-	191	125	63	-
	test	18	-	-	-	18

from all three databases whereas in the training set of Set B we used only images from the EUGENDA database and two clinical sites of MESSIDOR. The test set is the same for set A and B. In this way, we can test how well the method generalizes when the test set contains only images from a different population than the ones included in the training set.

To measure the retrieval performance, the precision, i.e. the ratio between the number of correct retrieved images and the total number of retrieved images  $N$ , against the number  $N$  of retrieved images is calculated. To measure the classification performance, Receiver Operating Characteristic (ROC) analysis is performed.

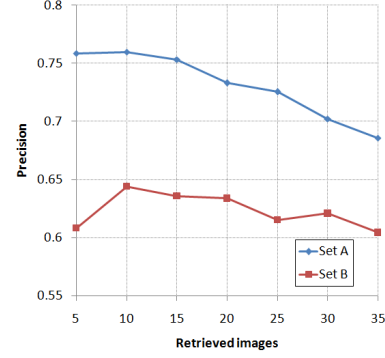
#### 5. RESULTS

For image retrieval, several similarity metrics were examined and the weighted squared chord [11] (SCw) performed best. In Figure 2 two example query images containing exudates 2(a) or drusen 2(b) with the first retrieved images are shown. Figure 3 shows the precision at different number of retrieved images for Set A and Set B. Using more than 10 retrieved images results in a lower retrieval precision. The ROC curves depicted in Figure 4 show the classification performance for Set A and Set B. Area under the ROC (Az) of 0.90 and 0.77 were obtained for Set A and Set B, respectively.

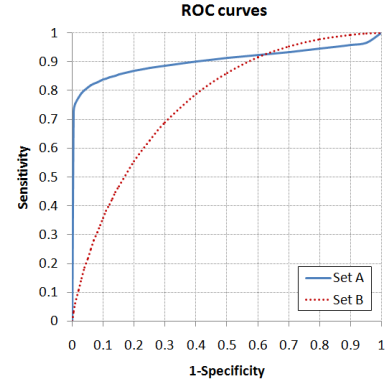
<sup>1</sup>Kindly provided by the Messidor program partners (see <http://messidor.crihan.fr>).

<sup>2</sup>EUGENDA, <http://www.eugenda.org>

<sup>3</sup><http://www.ces.clemson.edu/~ahoover/stare/>



**Fig. 3.** Precision graph for the retrieval task.



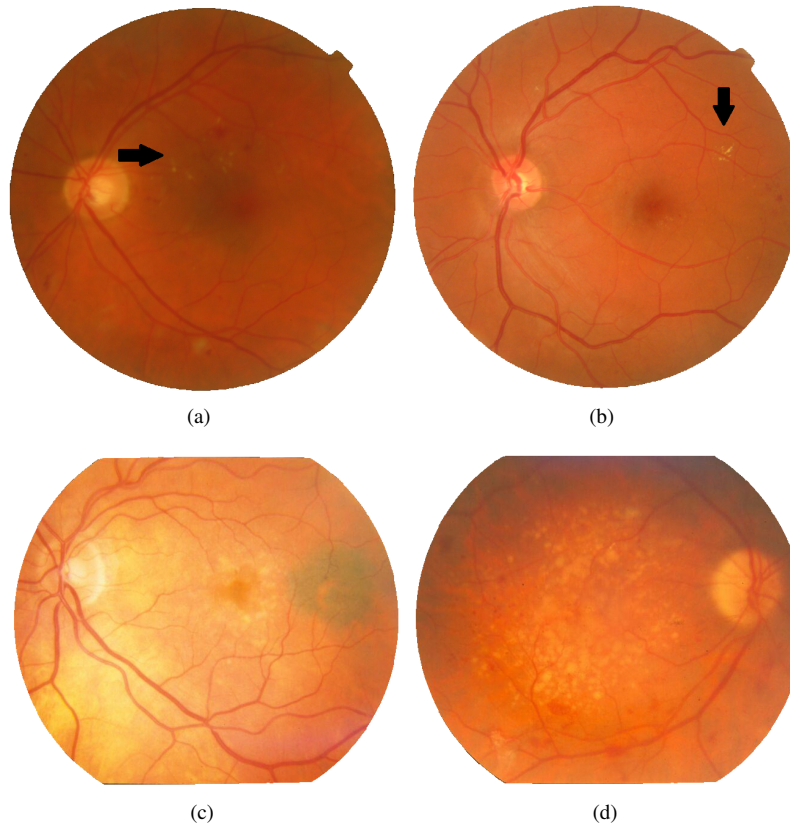
**Fig. 4.** ROC curves for the classification task on set A and B.

#### 6. DISCUSSION AND CONCLUSIONS

In this work, a BoW approach was presented for the retrieval and discrimination of images with bright lesions on retinal images, namely exudates and drusen. In contrast to previously proposed methods [5, 6, 7, 9, 8, 7], the approach focused on the retrieval (and classification) of images containing lesions with similar appearance, independently of the global image appearance and the severity level. To achieve this, a uniform patch sampling strategy based on retinal anatomical landmarks was adopted to reduce the influence of anatomical features on the retrieval task. Additionally, texture and color descriptors invariant to global image illumination and appearance were extracted and selected to improve the discrimination between lesions. Figure 2 shows retrieved images containing lesions with similar appearance than those in the query image, despite the differences in image appearance and/or severity level.

A mean precision of 0.76 and a Az value of 0.90 was obtained for the retrieval and classification tasks, respectively. However, as it can be seen in Figure 3 and Figure 4, the performance decreased when the query images come from a different population than the one represented in the training set. This might be due to differences in the acquisition process or the image quality. In future work, we will investigate this effect on larger datasets.

In DR screening, it is not uncommon to encounter patients with AMD and this can be up to 17% of the patients as expe-



**Fig. 2.** (a) Query image containing exudates (indicated by arrows) and (b) the most similar looking image for (a). (c) Query image containing drusen and (d) the most similar looking image for (c)

rienced in our private database. Many referrals for DR are caused by the presence of AMD. Operating at a high sensitivity, the system will be able to discard about half of the falsely referred cases caused by AMD. In large screening settings, this number will rise to a substantial proportion.

In conclusion, the proposed BoW approach represents a step toward the retrieval and the discrimination of similar eye pathologies in order to improve diagnosis. Specifically, this approach may significantly help DR screening programs by 1) helping inexperienced observers to distinguish between DR related lesions and other visible signs, and 2) increasing the specificity of automatic screening methods by reducing the false detection due to the presence of drusen.

## 7. REFERENCES

- [1] A. D. Fleming et al., "Automated detection of exudates for diabetic retinopathy screening," *PMB*, vol. 52, pp. 7385–7396, August 2007.
- [2] M. Niemeijer et al., "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis," *IOVS*, vol. 48(5), pp. 2260–2267, May 2007.
- [3] C. I. Sanchez et al., "Contextual computer-aided detection: Improving bright lesion detection in retinal images and coronary calcification identification in CT scans," *MIA*, vol. 16, pp. 50–62, 2012.
- [4] C. Agurto et al., "Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images," *IOVS*, vol. 52, no. 8, pp. 5862–5871, Jul 2011.
- [5] K. S. Deepak et al., "Content-based retrieval of retinal images for maculopathy," in *ACM IHI*, 2010, pp. 135–143.
- [6] G. Quellec et al., "Automated assesment of diabetic retinopathy severity using content based image retrieval in multimodal fundus photographs," *IOVS*, vol. 52, no. 11, pp. 1463–1477, 2011.
- [7] K. W. Tobin et al., "A probabilistic framework for content-based diagnosis of retinal disease," in *IEEE EMBS*, 2007, pp. 6744–7.
- [8] S. T. Acton et al., "Content based image retrieval: The foundation for future case-based and evidence-based ophthalmology," in *ICME*, 2008, pp. 541–544.
- [9] E. Chaum et al., "Automated diagnosis of retinopathy by content-based image retrieval," *Retina*, vol. 28, no. 10, pp. 1463–1477, 2008.
- [10] A. Gupta et al., "Content-based retrieval of ophthalmological images," in *ICIP*, 1996, vol. 3, pp. 703–706.
- [11] D. G. Gavin et al., "A statistical approach to evaluating distance metrics and analog assignments for pollen records," *Quaternary Research*, vol. 60, no. 3, pp. 356–367, 2003.