Classification of SD-OCT Volumes with LBP: Application to DME Detection

Guillaume Lemaître[a,b,*], Mojdeh Rastgoo[a,b,*], Joan Massich[a,*], Fabrice Mériaudeau[a], Désiré Sidibé[a]

[a] *ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain*
[b] *LE2I UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, 12 rue de la Fonderie, 71200 Le Creusot, France*

Abstract

This paper addresses the problem of automatic classification of Spectral Domain OCT (SD-OCT) data for automatic identification of patients with Diabetic Macular Edema (DME) versus normal subjects. Our method is based on Local Binary Patterns (LBP) features to describe the texture of Optical Coherence Tomography (OCT) images and we compare different LBP features extraction approaches to compute a single signature for the whole OCT volume. Experimental results with two datasets of respectively 32 and 30 OCT volumes show that regardless of using low or high level representations, features derived from LBP texture have highly discriminative power.

Moreover, the experiments show that the proposed method achieves better classification performances than other recent published works.

Keywords: Diabetic Macular Edema, Optical Coherence Tomography, DME, OCT, LBP

☆Document source available in GitHub [1]

*Corresponding author

Email addresses: `g.lemaintre58@gmail.com` (Guillaume Lemaître), `mojdeh.rastgoo@gmail.com` (Mojdeh Rastgoo), `joan.massich@u-bourgogne.fr` (Joan Massich)

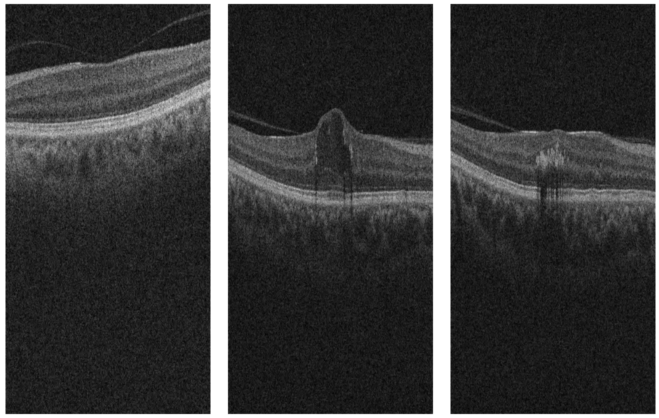
1. Introduction

Eye diseases such as Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME) are the most common causes of irreversible vision loss in individuals with diabetes. Just in United States alone, health care and associated costs related to eye diseases are estimated at almost \$500 M [2]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over 300 M people worldwide by 2025 [3]. Early detection and treatment of DR and DME play a major role to prevent adverse effects such as blindness. Indeed, the detection and diagnosis of retinal diseases are based on the detection of vascular abnormalities or lesions in the retina.

In past decades, Computer Aided Diagnosis systems devoted to ophthalmology, have been developed focusing on the automatic analysis of fundus images [4, 5]. However, the use of fundus photography is limited to the detection of signs which are correlated with retinal thickening such as hard and soft exudates, hemorrhages or micro-aneurysms. Moreover, DME is characterized as an increase in retinal thickness within 1 disk diameter of the fovea center with or without hard exudates and sometimes associated with cysts [6]. Therefore, fundus photography cannot always identify the clinical signs of DME; for example cysts, which are not visible in the retinal surface. In addition, it does not provide any quantitative measurements of retina thickness or information about cross-sectional retinal morphology.

Recently, Optical Coherence Tomography (OCT) has been widely used as a valuable diagnosis tool for DME detection. OCT is based on optical reflectivity and produces cross-sectional and three-dimensional images of the central retina, thus allowing quantitative retinal thickness and structure measurements. The new generation of OCT imaging, namely Spectral Domain OCT (SD-OCT) offers higher resolution and faster image acquisition over conventional time domain OCT. SD-OCT can produce $27,000$ to $40,000$ A-scans/seconds with an axial resolution ranging from 3.5 μm to 6 μm [7]. Figure. 1 shows two B-scan of SD-OCT volumes one for DME patient and one for normal patient. Many

I don't think that's the way to introduce it



(a) Normal (b) DME-cyst (c) DME-exudate

Figure 1: Example of SD-OCT images for normal (a) and DME patients (b)-(c) with cyst and exudate, respectively.

of the previous works on OCT image analysis have focused on the problem of retinal layers segmentation, which is a necessary step for retinal thickness measurements [8, 9]. However, few have addressed the specific problem of DME and its associated features detection from OCT images.

In this research we focus on the latter problem and propose an automatic framework for identification of DME patients versus normal subjects using OCT volumes. The proposed method, which is an extension of our previous work [10], is based on Local Binary Patterns (LBP) features to describe the texture of OCT images and dictionary learning using the Bag-of-Words (BoW) models [11]. We propose to extract 2D and 3D LBP features from OCT images and volumes, respectively. The LBP descriptors are further extracted from the entire sample or local patches within individual samples. In this research beside the comparison of 2D and 3D features, we also compare the effects of common pre-processing steps for OCT data, and different classifiers.

In the following of this paper, first in Sect. 2 a summary of the related studies is presented.

This paper is organized as follows, Section 2 presents a summary of the related studies. The proposed framework is explained in Sect. 3, while the experiments and results are discussed in Sect. 4. Finally, the conclusion and avenue for future directions are drawn in Sect. 5.

2. Related Work

This section reviews the works straightly addressing the problem of classifying OCT volumes as normal or abnormal. A summary can be found in 1.

Srinivasan *et al.* [12] proposed a classification method to distinguish DME, Age-related Macular Degeneration (AMD) and normal SD-OCT volumes. The OCT images are pre-processed by reducing the speckle noise by enhancing the sparsity in a transform-domain and flattening the retinal curvature to reduce the inter-patient variations. Then, Histogram of Oriented Gradients (HOG) are extracted for each slice of a volume and a linear Support Vector Machines (SVM) is used for classification. On a dataset of 45 patients equally subdivided into the three aforementioned classes, this method leads to a correct classification rate of 100%, 100% and 86.67% for normal, DME and AMD patients, respectively.

Venhuizen *et al.* proposed a method for OCT images classification using the BoW models [13]. The method starts with the detection and selection of keypoints in each individual B-scan, by keeping the most salient points corresponding to the top 3% of the vertical gradient values. Then, a texton of size 9×9 pixels is extracted around each keypoint, and Principal Component Analysis (PCA) is applied to reduce the dimension of every texton to get a feature vector of size 9. All extracted feature vectors are used to create a codebook using k-means clustering. Then, each OCT volume is represented in terms of this codebook and is characterized as a histogram that captures the codebook occurrences. These histograms are used as feature vector to train a Random Forest (RF) with a maximum of 100 trees. The method was used to classify OCT volumes between AMD and normal cases and achieved an Area Under the Curve (AUC) of 0.984 with a dataset of 384 OCT volumes.

Liu *et al.* proposed a methodology for detecting macular pathology in OCT images using LBP and gradient information as attributes [14]. The method starts by aligning and flattening the images and creating a 3-level multi-scale spatial pyramid. The edge and LBP histograms are then extracted from each block of every level of the pyramid. All the obtained histograms are concatenated into a global descriptor whose dimensions are reduced using PCA. Finally a SVM is used as classifier. The method achieved good results in detection OCT scan containing different pathology such as DME or AMD, with an AUC of 0.93 using a dataset of 326 OCT scans.

Our later study proposes a standard classification procedure to differentiate between DME and normal SD-OCT volumes [1]The data is pre-processed using Non-Local Means (NL-means) filtering. The volumes are mapped into discrete set of structures namely: local, when these structures correspond to patches; or global, when the structures correspond to volume slices or the whole volume. These structures are described in terms of texture using LBP or LBP from Three Orthogonal Planes (LBP-TOP) and encoded using histogram, PCA or BoW to produce a single feature vector in order to present the volumes to a RF classifeir. This methodology was tested against Venhuizen *et al.* [13] using public and non-public datasets showing an improvement within the results achieving a Sensitivity (SE) of 87.5% and a Specificity (SP) of 75%. The obtained results of this study is listed in Sect. 4.

As stated in previous section, this research is a continue of our previous work, where we intend to evaluate the influence of different pre-processing, BoW representation and various classifiers. Our proposed pipeline with detail description of each step is presented in the following section.

Table 1: The summary of the state of the art methods

Ref	Task			Data size	Pre-processing				Features	Representation	Classifier	Evaluation		
	AMD	DME	Normal		De-noise	Flatten	Aligning	Cropping				SE	SP	AUC
[12]	✓	✓	✓	45	✓	✓		✓	HOG		SVM	86.7%,100%,100%		
[13]	✓		✓	384					texton	BoW, PCA	RF			0.984
[14]	✓	✓	✓	326		✓	✓		Edge, LBP	PCA	SVM			0.93
[10]		✓	✓	32	✓				LBP-LBP-TOP	PCA, BoW, histogram	RF	87.5%	75%	

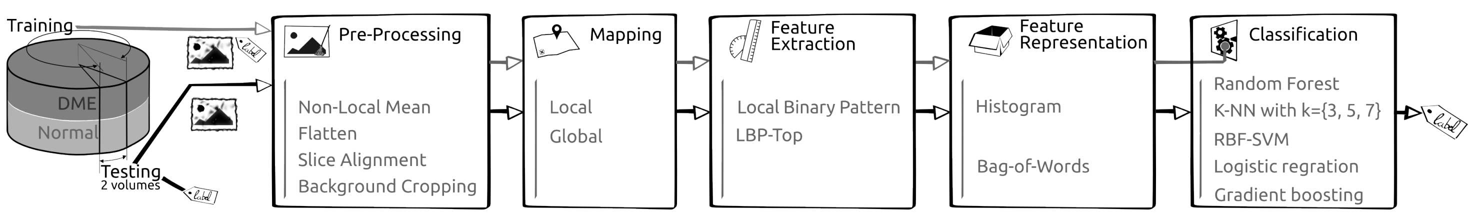


Figure 2: take out bg crop[sik] Machine learning classification basic scheme

3. Materials and Methods

The proposed method, as well as, its experimental set-up for OCT volume classification are outlined in Fig. 2. The methodology is formulated as a standard classification procedure which consists in 5 steps. First, the OCT volumes are pre-processed as presented in details in Sect. 3.1. Then, LBP and LBP-TOP features are extracted, mapped and represented as discussed in depth in Sect. 3.2, Sect. 3.3, and Sect. 3.4, respectively. Finally, the classification step is presented in Sect. 3.5.

3.1. Image pre-processing

This section describes the set of pre-processing techniques which aim at enhancing the OCT volume. The influence of these pre-processing methods and their possible combinations are extensively studied in Sect. **??**.

add the section for the experiment flattening+aligned

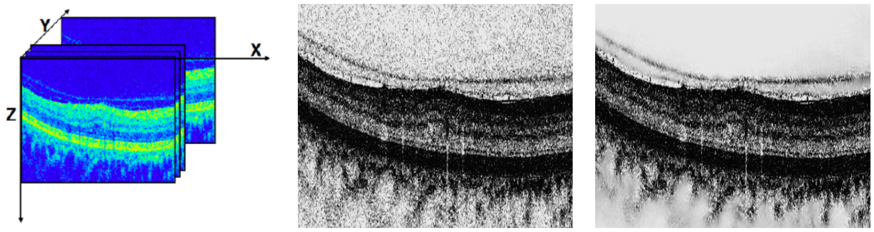
3.1.1. Non-Local Means (NL-means)

OCT images suffer from speckle noise, like other image modalities such as Ultra-Sound (US) [15]. The OCT volumes are enhanced by denoising each B-scan (i.e. each $x - z$ slice) using the NL-means [16], as shown in Fig. 3. NL-means has been successfully applied to US images to reduce speckle noise and outperforms other common denoising methods [17]. NL-means filtering preserve fine structures as well as flat zones, using all the possible self-predictions that the image can provide rather than local or frequency filters such as Gaussian, anisotropic, or Wiener filters [16].

3.1.2. Flattening

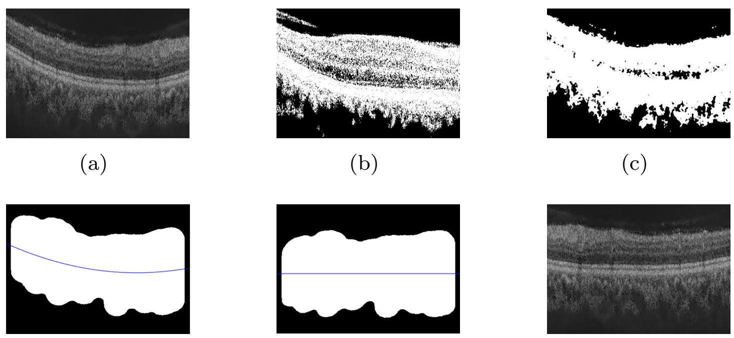
Texture descriptors characterize spatial arrangement of intensities. Therefore, to ensure a consistent characterization of the tissue disposition regardless of the location within the retina, the natural curvature of the retina needs to be taken into account. This can be done in different manners:

- using a descriptor allowing for a wide number of textures, so that each texture orientation accounts for each own descriptor.



(a) (b) (c)

Figure 3: OCT: (a) Organization of the OCT data - (b) Original image - (c) NL-means filtering.



(a) (b) (c)

(d) (e) (f)

Figure 4: Flattening procedure: (a) original image, (b) thresholding, (c) median filter, (d)curve fitting, (e) warping, (f) flatten image.

- using a rotation invariant descriptor.
- by unfolding the curvature of the retina.

This process of unfolding the curvature of the retina is known as image flattening. When flattening, an estimation of the Retinal Pigment Epithelium (RPE) layer is used to modify the volume by imposing that the RPE should be flat. Our implementation modifies the proposal of Liu *et al.* [14], as illustrated in Figure 4. Othsu thresholding is used to segment the retina from the background. A line is fitted to the bottom part of the segmentation hull, since it is assumed to be parallel to the RPE. The image is corrected based on this line.

3.1.3. Slice alignment

Similarly, when using 3D texture, misalignment between the slice introduce error to the texture descriptor. In this case the slices are also aligned based on the segmentation's hull.

3.2. Features extraction

This a generic description. Shall we place it for our images with some picture? (maybe we don't have time)

In this research we chose to extract simple and efficient LBP texture features with regards to each OCT slice and volumes. LBP is a texture descriptor based on the signs of the differences of a central pixel with respect to its neighboring pixels [18]. These differences are encoded in terms of binary patterns as in Eq. (1):

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \ , \qquad s(\cdot) = \begin{cases} 1 & \text{if } (g_p - g_c) \geq 0 \\ 0 & \text{otherwise} \end{cases} \ , \qquad (1)$$

where g_c, g_p are the intensities of the central pixel and a given neighbor pixel, respectively. P is the number of sampling points in the circle of radius R. Figure **??** illustrates the meaning of P and R.

Ojala *et al.* further extend the original LBP formulation to achieve rotation invariance at the expense of limiting the texture description to the notion of circular "uniformity" [18]. Volume encoding is later proposed by Zhao *et al.* by computing LBP descriptors in each orthogonal planes, so called LBP-TOP [19].

In this research we consider uniform and rotation invariant LBP and LBP-TOP features with various sampling points, $\{8, 16, 24\}$, with respect to different radius, $\{1, 2, 3\}$, respectively.

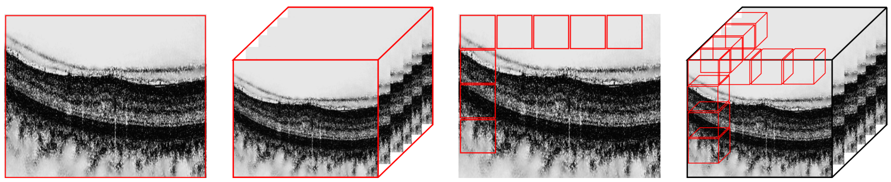
Table. 2 shows the length of uniform rotation invariant histogram (LBP_{hist}) for the used sampling point and radius.

3.3. Mapping

The mapping stage is used to determine a discrete set of elements (or structures) which is used for representing the OCT volume. For this work two map-

Table 2: length of LBP_{hist} for different sampling points and radius ($\{S, R\}$) in LBP descriptor

	Sampling point in a given radius ($\{S, R\}$)		
	$\{8, 1\}$	$\{16, 2\}$	$\{24, 3\}$
LBP_{hist}	10	18	26



(a) (b) (c) (d)

Figure 5: *Global* (a)-(b) and *local* (c)-(d) mapping for LBP and LBP-TOP features (2D B-scane and 3D volume, respectively).

ping strategies are defined: (i) *global* and (ii) *local* mapping.

Global mapping considers to extract the features from the 2D B-scans for LBP and 3D volume for LBP-TOP. Therefore for a volume with d B-scans, the *global*-LBP feature leads to a final descriptor of size $d \times LBP_{hist}$ and *global*-LBP-TOP feature returns the final descriptor of size $3 \times LBP_{hist}$. Here LBP_{hist} refers to the size of uniform and rotation invariant LBP histogram which it's number of bins depends on the number of sampling points in a given neighborhood. The *global* mapping for 2D B-scan end 3D volume is shown in Fig. 5(a) and 5(b).

Local mapping extracts the features from a set of 2D patches for LBP and a set of sub-volumes for LBP-TOP. Considering $(m \times m)$ patch (P) for 2D LBP and $(m \times m \times m)$ sub-volume for LBP-TOP. Using these elements, the *local*-LBP approach provides a final descriptor of size $(LBP_{hist} \times N \times d)$ while *local*-LBP-TOP results in a final descriptor of size $3 \times LBP_{hist} \times N' \times \frac{m}{d}$. Here N and N' are the total number of elements in each B-scan, and 3D volume, respectively. This mapping is illustrated in Fig. 5(c) and 5(d).

For the sake of clarification, the length of final descriptor for *local* and *global* mapping of both LBP and LBP-TOP features with respect to the length of LBP_{hist} are listed in Table. 3.

Table 3: Final length of descriptors of LBP and LBP-TOP features, with respect to different mapping strategies and LBP_{hist} number of bins for various sampling point. Here d is the number of B-scans per volumes and N and N' are the number of patches per B-scane and sub-volumes per volumes, respectively.

Mapping	LBP			LBP-TOP		
	$\{8, 1\}$	$\{16, 2\}$	$\{24, 3\}$	$\{8, 1\}$	$\{16, 2\}$	$\{24, 3\}$
global	$10 \times d$	$18 \times d$	$26 \times d$	3×10	3×18	3×26
local	$10 \times N \times d$	$18 \times N \times d$	$26 \times N \times d$	$3 \times 10 \times N' \times \frac{m}{d}$	$3 \times 18 \times N' \times \frac{m}{d}$	$3 \times 26 \times N' \times \frac{m}{d}$

3.4. Feature representation

Each OCT volume can be described by its texture and we employed two strategies.

Low-level representation The texture descriptor of an OCT volume is defined as the concatenation of the LBP histograms. Therefore, for the LBP-TOP, the feature descriptor is computed through the concatenation of the LBP histograms of the three orthogonal planes and for the LBP, the descriptor is defined either through concatenation of the LBP histograms per each B-scan (*gocal*-mapping), or per each P (*local*-mapping).

High-level representation According to the chosen mapping strategy, the low-level representation can lead to a high dimensional feature space. High-level representation simplifies this high dimensional feature space into a more discriminant lower space. BoW approach is used for this purpose [11]. This model represents the features by creating a visual dictionary or codebook~~“codebook”~~[sik], from the set of low-level features. The set of low-level features are clustered using k-means to create the codebook with k clusters or visual words. After creating the codebook, each of the training example is represented as a histogram of size k. The histogram is obtained by calculating the frequency of occurrences of each of the k words in the extracted features from the training example.

3.5. Classification

Classification corresponds to the mapping of a set of inputs $\mathbf{x}$ into a set of categorical outputs $\mathbf{y}$ using a linear or non-linear function $f(\cdot)$. In supervised learning methods, this function is defined by providing a training set of N samples $\mathbf{x}_{tr}$ with their associated labels $\mathbf{y}_{tr}$. In the remainder of this section, we briefly summarize the supervised classification methods used in the experiments. Details regarding the parameters used in our experiments are provided in Sect. **??**[glm]

k-Nearest Neighbor (NN) is a non-parametric classification method in which an unlabeled feature vector x is assigned to the majority class of its k nearest-neighbors from the training set. To avoid a tie case, the parameter k is set to an odd number.

Logistic Regression (LR) is a linear classifier which uses the logistic function to estimate the probability of x to belong to a particular class c_i [20]. Thus, the posterior probability is expressed as:

$$p(c_i|x) = \frac{1}{1 + \exp(-w^T x)} \tag{2}$$

where w is a vector of the regression parameters to obtain a linear combination of the input feature vector x. The vector w can be inferred by finding the maximum likelihood estimates via optimization methods such as quasi-Newton method [21]. Once the vector w found, an unlabeled feature vector is assigned to the class which maximizes the posterior probability.

Random Forest (RF) is an ensemble of decision trees [22] which generalizes the classification process by applying two types of randomization: at the tree level, each tree is fed by a bootstrap made of S' samples which are built from the original data of size S such that $S = S'$, and at the node level, a subset of feature dimensions m is randomly selected from the original dimension M such that $m << M$. The trees in RF are grown to their maximum length without any pruning. In the testing stage, each tree in the ensemble casts a unit vote in the final prediction and the final prediction is based on combination of all the votes.

Gradient Boosting (GB) is a reformulation of AdaBoost [23] in which the problem of finding an ensemble of real-valued weak learners is tackled as a numerical optimization [24]. A strong learner is built by iteratively finding the best pair of real-valued weak learner function and its corresponding weight which minimizes a given differentiable loss function. Common choice for weak learners is decision stumps or regression trees while the loss function is generally an exponential or logarithmic loss [25], minimized via gradient descent or quadratic approximation.

Support Vector Machines (SVM) is a sparse kernel classification method which aims at finding the best linear hyperplane which separates two classes by maximizing the margin between them [26]. SVM becomes a non-linear classifier by using the kernel trick [27] which consists in replacing each inner product by a non-linear kernel function such as Radial Basis Function (RBF) or polynomial kernels.

4. Experiments and Validation

add repository reference somewhere

To evaluate the effects and influence of the different blocks composing our framework, an experimentation suit has been designed to test different configuration parameters, which are evaluated using different datasets (see Table. 4). The rest of this section details aspects of the experimentation and the design decisions that are consistent across all the experimentation, while subsections report different technicalities.

Unless stated otherwise, all the experiments are run using our own dataset (SERI) alone. Only for the sake of comparison some experiments are re-run on the Duke public dataset using our optimal configurations. SERI and Duke dataset details are reported in section 4.1 and section 4.2 respectively.

For all the experiments, LBP and LBP-TOP features are extracted for different sampling points of 8, 16, and 24 for radius of 1, 2, and 3, respectively. As previously mentioned, two different mapping strategies, *local* and *global*, are used, where for *local* mapping, we consider a (7×7) patch (P) for 2D LBP and $(7 \times 7 \times 7)$ sub-volume for LBP-TOP.

All the experiments are evaluated using Leave-One-Patient Out Cross-Validation (LOPO-CV) strategy. In this validation, at each round a pair DME-normal volume is selected for testing while the rest of the volumes are used for training. The use of this method implies that no variance in terms of SE and SP can be reported. However, and despite this limitation, LOPO-CV has been employed due to the small size of the dataset.

The obtained results of all the experiments except experiment #2 is represented in terms of SE and SP, which are statistics driven from the confusion matrix (see Fig. 6) as stated in eq. (3). The SE evaluates the performance of the classifier with respect to the positive class, while the SP evaluate it's performance with respect to negative class.

$$SE = \frac{TP}{TP + FN} \qquad SP = \frac{TN}{TN + FP} \tag{3}$$

Some experimentation is complemented using Accuracy (ACC) and F1-score (F1). Accuracy is used to have a overall sense of classifier performance, and F1 is used to see the trade off between SE and precision. Equation. 4 shows the formulation of these two measurements.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad F1 = \frac{2TP}{2TP + FP + FN} \tag{4}$$

Experimentation details can be found in Sect. 4.3 to Sect. 4.6 and summarized in Table. 4. In general terms, all the experiments have been carried out using SERI dataset while *Experiment #1 (Sect. 4.3)* has been complemented using Duke dataset for comparison purposes. This *Experiment #1 section 4.3* takes from the experimentation reported in [10] to evaluate the effects of differ-

		Actual	
		A+	A-
Predicted	P+	True Positive (TP)	False Positive (FP)
	P-	False Negative (FN)	True Negative (TN)

Figure 6: Confusion matrix with truly and falsely positive detected samples (TP, FP) in the first row, from left to right and the falsely and truly negative detected samples (FN, TN) in the second row, from left to right.

ent feature representationsand compares the results to those obtained by Venhuizen*et al.* [13]. *Experiment #2 (Sect. 4.4)* studies the effect of the codebook size in order to find the optimal number of words for our application when using BoW. *Experiment #3 (Sect. 4.5)* studies the effect different pre-processing and classifiers.

are not we evaluating feature extraction?

Table 4: The outline and summary of the performed experiments.

	Dataset	Pre-processing	Features	Mapping	Representation	Classification	Validation	Evaluation
Common:	SERI	NL-means	LBP,LBP-TOP $S = \{8, 16, 24\}$ $R = \{1, 2, 3\}$				LOPO-CV	
Experiment#1: Goal: Evaluation of features, mapping and representation	+ Duke	∼	∼	*gocal* *local*	BoW Histogram	RF	∼	SE, SP, [13]
Experiment#2: Goal: Finding the optimum number of words	∼	+ F + F+A	∼	*global* *local*	BoW	LR	∼	ACC, F1
Experiment#3: Goal: Evaluation of different pre-processing for high-level features	∼	+F +F+A	∼	*global* *local*	BoW	3-NN RF SVM GB	∼	SE, SP
Experiment#4: Goal: Evaluation of different pre-processing for low-level features	∼	+F +F+A	∼	*global*	Histogram	3-NN RF SVM GB	∼	SE, SP

4.1. SERI-Dataset

This data was acquired by Singapore Eye Research Institute (SERI), using CIRRUS TM (Carl Zeiss Meditec, Inc., Dublin, CA) SD-OCT device. The datasets consist of 32 OCT volumes (16 DME and 16 normal cases). Each volume contains 128 B-sane with dimension of 512×1024 pixels. All SD-OCT images are read and assessed by trained graders and identifies as normal or DME cases based on evaluation of retinal thickening, hard exudates, intraretinal cystoid space formation and subretinal fluid.

4.2. Duke-Dataset

This data published by Srinivasan *et al.* [12] was acquired in Institutional Review Board-approved protocols using Spectralis SD-OCT (Heidelberg Engineering Inc., Heidelberg, Germany) imaging at Duke University, Harvard University and the University of Michigan. This datasets consist of 45 OCT volumes (15 AMD, 15 DME and 15 normal). In this study we only consider a subset of the original data containing 15 DME and 15 normal OCT volumes.

4.3. Experiment #1

For the completeness of this article, this experiment replicates some of the experiments reported in [10]. The experiment evaluates the effects of different feature extraction, feature representation,... [sik] using the SERI and Duke datasets.

For this experiment, the volumes are pre-processed using NL-means. LBP and LBP-TOP descriptors are extracted using the default configuration. Local and global mapping are used. Volumes are represented using both low-level and high-level representation. For concordance with [10], when using BoW the size of the coodebook is fixed to 32 words. Finally, the volumes are classified using RF classifier with 100 un-pruned trees.

Results are listed in Table. 5, while the two best performing configurations are compared to Venhuizen *et al.* [13] in Table. 6.

Result description[sik]

Table 5: Experient #1 - Obtained results of classification using SERI and Duke datasets.

Features	SERI dataste						Duke dataset					
	8^{riu2}		16^{riu2}		24^{riu2}		8^{riu2}		16^{riu2}		24^{riu2}	
	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP
global-LBP-TOP	56.2	62.5	**87.5**	**75.0**	68.7	68.7	80.0	93.3	73.3	86.6	73.3	86.6
local-LBP	**75.0**	**87.5**	81.2	75.0	68.7	62.5	**80.0**	**86.6**	**86.7**	**100**	93.3	86.6
local-LBP-TOP	62.5	68.7	56.2	37.5	37.5	43.7	80.0	86.6	86.6	86.6	60.0	80.0

Table 6: Experment #1 - Comparing the proposed method by [13] on SERI and Duke datasets.

Data sets	SERI		Duke	
	SE	SP	SE	SP
Venhuizen *et al.* [13]	61.5	58.8	71.4	68.7
{*local*-LBP},8^{riu}	**75.0**	**87.5**	**86.6**	**100.0**
{*global*-LBP-TOP},16^{riu}	**75.0**	**87.5**	80.0	86.6

4.4. Experiment #2

In order to determine the optimal size of the codebook when using BoW, this experiment evaluates several codebook sizes on SERI dataset.

For this experiment, several pre-processing strategies are evaluated: (i) NL-means, (ii) a combination of NL-means and flattening; (iii) a combination of NL-means, flattening and aligning. LBP and LBP-TOP descriptors are extracted using the default configuration. Volumes are represented using the high-level representation BoW using $k \in \{10, 20, 30, \cdots, 100, 200, \cdots, 500, 1000\}$. Finally, the volumes are classified using LR.

The usual construction of the codebook consists of clustering the samples in the feature space using k-means. However, this operation is rather computationally expensive and convergence of the k-means algorithm for all codebook sizes is not granted. Nonetheless, **(author?)** pointed out that randomly generated coodebooks can be used at expenses of accuracy. Since the goal is to assess the best coodebook size not its final performance, for this experiment, the construction of the coodebook has been carried out using k-means++ algorithm (reference is missing)[sik], which is usually used as a k-means initialization algorithm.

Figure. 7 shows the ACC and F1 score graphs obtained for some of the

Figure 7:

configurations[1]. The results obtained when using the optimum number of words for each configuration, are listed in Table 7.

Table 7: Experiment #2 - Optimum number of words for each configuration as a result of LR Classification, for high-level representation of *global* and *local*-LBP, and *local*-LBP-TOP features with different pre-processing. The pre-processing includes: NF, F, F+A, and F+A+C.

Features	Pre-processing	8^{riu2}			16^{riu2}			24^{riu2}		
		ACC%	F1%	W#	ACC%	F1%	W#	ACC%	F1%	W#
global-LBP										
	NF	81.2	78.5	500	62.5	58.06	80	62.5	62.5	80
	F	71.9	71	400	68.7	66.7	300	68.7	66.7	300
	F+A	71.9	71	500	71.9	71	200	75	68.7	500
local-LBP										
	NF	75	75	70	65.6	64.5	90	62.5	60	30
	F	75	73.3	30	71.8	61	70	62.5	62.5	100
	F+A	75	69	40	71.9	71	200	68.7	66.7	10
local-LBP-TOP										
	NF	68.7	68.7	400	75	75	500	71.9	71	60
	F	68.7	68.7	300	68.7	66.7	50	75	76.5	80
	F+A	75	73.3	100	75	73.3	90	75	69	70

4.5. Experiment #3

This experiment is performed for high-level features on SERI dataset. Using the optimum number of words which are obtained from the previous experiment, the low-level feature sets with regards to different pre-processing configurations are re-represented using BoW and k-means clustering approach and are classified using different classifiers such as k-NN, RF, GB, and SVM. The k-means

[1]Full set of scores can be found at the github repository in [1]

algorithm is initialized using k-means++ method and is performed with 5 iteration for each codebook. The RF and GB classifier are trained using 100 un-pruned trees, while SVM classifier is trained with RBF kernel. The regularization and soft-margin parameters of this classifier are chosen with grid-search method. Finally the k-NN classifier is trained by considering the 3 nearest neighbor. Table 8 shows the obtained results from this experiment.

Table 8: Experiment #3 - k-NN and SVM classification with BoW for the *global* and *local* LBP and *local* LBP-TOP features with different pre-processing. The optimum number of words were selected based on the previous experiment.

Features	Pre-processing	k-NN						SVM					
		8^{riu2}		16^{riu2}		24^{riu2}		8^{riu2}		16^{riu2}		24^{riu2}	
		SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%
global-LBP													
	NF	43.7	93.7	43.7	87.5	43.7	62.5	68.7	87.5	62.5	62.5	50.0	56.2
	F	43.7	56.2	50.0	75.0	62.5	56.2	56.2	56.2	56.2	75.0	56.2	68.7
	FA	56.2	62.5	43.7	81.2	68.7	56.2	56.2	68.7	68.7	68.7	56.2	75.0
local-LBP													
	NF	**75.0**	**87.5**	50.0	68.7	43.7	43.7	**75.0**	**93.7**	50.0	75.0	56.2	56.2
	F	56.2	56.2	50.0	50.0	50.0	43.7	**81.2**	**93.7**	68.7	68.7	68.7	75.0
	FA	56.2	43.7	50.0	75.0	50.0	62.5	**75.0**	**93.7**	75.0	68.7	68.7	68.7
local-LBP-TOP													
	NF	56.2	75.0	56.2	75.0	62.5	56.2	**81.2**	**87.5**	**75.0**	**100**	56.2	75.0
	F	62.5	43.7	37.5	68.7	43.7	62.5	**81.2**	**81.2**	75.0	68.7	81.2	68.7
	F+A	56.2	56.2	68.7	50.0	43.7	62.5	62.5	75.0	68.7	75.0	62.5	81.2

Features	Pre-processing	RF						GB					
		8^{riu2}		16^{riu2}		24^{riu2}		8^{riu2}		16^{riu2}		24^{riu2}	
		SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%
global-LBP													
	NF	**68.7**	**93.7**	43.7	62.5	50.0	68.7	56.2	50.0	37.5	31.2	50.0	43.7
	F	56.2	50.0	56.2	75.0	50.0	75.0	50.0	56.2	56.2	75.0	43.7	62.5
	FA	68.7	50.0	56.2	62.5	62.5	56.2	56.2	50.0	68.7	50.0	43.7	75.0
local-LBP													
	NF	**81.2**	**81.2**	62.5	56.2	56.2	56.2	75.0	62.5	68.7	87.5	50.0	75.0
	F	56.2	81.2	62.5	68.7	68.7	62.5	68.7	75.0	50.0	75.0	50.0	62.5
	FA	68.7	62.5	62.6	68.7	43.7	43.7	56.2	50.0	68.7	56.2	50.0	50.0
local-LBP-TOP													
	NF	68.7	62.5	**68.7**	**81.2**	68.7	68.7	37.5	68.7	62.5	81.2	62.5	50.0
	F	50.0	62.5	62.5	62.5	43.7	75.0	50.0	56.2	43.7	62.5	50.0	62.5
	F+A	50.0	62.5	**81.2**	**87.5**	50.0	68.7	56.2	62.5	81.2	68.7	75.0	68.7

4.6. Experiment #4

This experiment is conducted for low-level features. The *global* LBP and LBP-TOP features are classified using the same classifiers as previous experiments with the same configurations. The obtained results from this experiment is listed in Table. 9.

Table 9: Experiment #4 - Classification results obtained from low-level representation of global LBP and LBP-TOP features with different pre-processing. Pre-processing steps include: NF, F, F+A. Different classifiers such as RF, GB, SVM, and k-NN are used.

Features	Pre-processing	k-NN						k-SVM					
		8^{riu2}		16^{riu2}		24^{riu2}		8^{riu2}		16^{riu2}		24^{riu2}	
		SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%
global-LBP													
	NF	37.5	50.0	25.0	50.0	37.5	68.7	56.2	62.5	56.2	43.7	56.2	68.7
	F	62.5	50.0	56.2	75.0	62.5	68.7	75.0	68.7	62.5	62.5	62.5	68.7
	FA	56.2	50.0	56.2	75.0	62.5	68.7	75.0	68.7	62.5	62.5	62.5	68.7
global-LBP-TOP													
	NF	31.2	93.7	37.5	100.0	37.5	81.2	62.5	75.0	**62.5**	**93.7**	56.2	87.5
	F	50.0	56.2	56.2	75.0	56.2	62.5	68.7	75.0	43.7	68.7	68.7	56.2
	F+A	75.0	43.7	56.2	43.7	68.7	50.0	68.7	62.5	62.5	56.2	56.2	68.7
Features	Pre-processing	RF						GB					
		8^{riu2}		16^{riu2}		24^{riu2}		8^{riu2}		16^{riu2}		24^{riu2}	
		SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%
global-LBP													
	NF	43.7	62.5	43.7	62.5	56.2	75	43.7	43.7	43.7	37.5	37.5	31.25
	F	56.2	56.2	68.7	62.5	62.5	68.7	25	56.2	50.0	43.7	25.0	43.7
	F+A	65.2	56.2	50.0	50.0	56.2	68.7	43.75	62.5	62.5	50.0	31.2	31.2
global-LBP-TOP													
	NF	56.2	68.7	**68.7**	**87.5**	**68.7**	**81.2**	68.7	68.7	75.0	50.0	56.2	43.7
	F	56.2	62.5	81.2	68.7	**81.2**	**81.2**	56.2	62.5	62.5	68.7	68.7	81.2
	F+A	68.7	62.5	75.0	68.7	**75.0**	**81.2**	56.2	43.7	62.5	62.5	**75.0**	**75.0**

5. Conclusions

The work presented here addresses the automatic classification of SD-OCT data to identify subjects with DME versus normal. Based on the reported results, the low level volume 3D features and high level 2D features using patches achieve the most desirable results in the experimental setup presented here. The comparison against different datasets and methodologies, highlights that: regardless of using low or high level representations, volume signatures derived from LBP texture show high discriminative power for distinguishing DME vs normal volumes.

6. Future work

TOMORROW THE MOON !!

References

[1] G. Lemaître, M. Rastgoo, J. Massich, retinopathy: Miccai-omia-2015 (Jul. 2015). doi:10.5281/zenodo.22195.
URL `http://dx.doi.org/10.5281/zenodo.22195`

[2] S. Sharma, A. Oliver-Hernandez, W. Liu, J. Walt, The impact of diabetic retinopathy on health-related quality of life, Curr. Op. Ophtal. 16 (2005) 155–159.

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

[4] M. D. Abramoff, M. K. Garvin, M. Sonka, Retinal image analysis: a review, IEEE Review Biomed. Eng. 3 (2010) 169–208.

[5] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J. Hubschman, B. al Diri, C. Cheung, D. Wong, M. Abramoff, G. Lim, D. Kumar, P. Burlina, N. M. Bressler, H. F. Jelinek, F. Meriaudeau, G. Quellec,

T. MacGillivray, B. Dhillon, Validation retinal fundus image analysis algorithms: issues and proposal, Investigative Ophthalmology & Visual Science 54 (5) (2013) 3546–3569.

[6] Early Treatment Diabetic Retinopathy Study Group, Photocoagulation for diabetic macular edema: early treatment diabetic retinopathy study report no 1, Arch. Ophtalmol. 103 (12) (1985) 1796–1806.

[7] T. C. Chen, B. Cense, M. C. Pierce, N. Nassif, B. H. Park, S. H. Yun, B. R. White, B. E. Bouma, G. J. Tearney, J. F. de Boer, Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophtalmic imaging, Arch. Ophtalmol. 123 (12) (2005) 1715–1720.

[8] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, S. Farsiu, Automatic segmentation of seven retinal layers in sd-oct images congruent with expert manual segmentation, Optic Express 18 (18) (2010) 19413–19428.

[9] R. Kafieh, H. Rabbani, M. D. Abramoff, M. Sonka, Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map, Medical Image Analysis 17 (2013) 907–928.

[10] G. Lemaître, M. Rastgoo, J. Massich, S. Sankar, F. Mériaudeau, D. Sidibé, Classification of sd-oct volumes with lbp: Application to dme detection, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Ophthalmic Medical Image Analysis Workshop (OMIA), 2015.

[11] J. Sivic, A. Zisserman, Video google: a text retrieva approach to object matching in videos, in: IEEE ICCV, 2003, pp. 1470–1477.

[12] P. P. Srinivasan, L. A. Kim, P. S. Metttu, S. W. Cousins, G. M. Comer, J. A. Izatt, S. Farsiu, Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, Biomedical Optical Express 5 (10) (2014) 3568–3577.

[13] F. G. Venhuizen, B. van Ginneken, B. Bloemen, M. J. P. P. van Grisven, R. Philipsen, H. C., T. Theelen, C. I. Sanchez, Automated age-related mac-

ular degeneration classification in oct using unsupervised feature learning, in: SPIE Medical Imaging, Vol. 9414, 2015, p. 94141l.

[14] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, R. J. M., Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding, Medical Image Analysis 15 (2011) 748–759.

[15] J. M. Schmitt, S. Xiang, K. M. Yung, Speckle in optical coherence tomography, Journal of biomedical optics 4 (1) (1999) 95–105.

[16] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, IEEE, 2005, pp. 60–65.

[17] P. Coupe, P. Hellier, C. Kervrann, C. Barillot, Nonlocal means-based speckle filtering for ultrasound images, IEEE TIP (2009) 2221–2229.

[18] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, Pattern Analysis and Machine Intelligence, IEEE Transactions on 24 (7) (2002) 971–987.

[19] G. Zhao, T. Ahonen, J. Matas, M. Pietikäinen, Rotation-invariant image and video description with local binary pattern features, Image Processing, IEEE Transactions on 21 (4) (2012) 1465–1477.

[20] D. R. Cox, The regression analysis of binary sequences, Journal of the Royal Statistical Society. Series B (Methodological) (1958) 215–242.

[21] R. H. Byrd, J. Nocedal, R. B. Schnabel, Representations of quasi-newton matrices and their use in limited memory methods, Mathematical Programming 63 (1-3) (1994) 129–156.

[22] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[23] J. H. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis 38 (4) (2002) 367–378.

[24] G. Lemaitre, J. Massich, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, D. Sidibe, F. Meriaudeau, A boosting approach for prostate cancer detection using multi-parametric mri, in: International Conference on Quality Control and Artificial Vision (QCAV2015), SPIE, 2015.

[25] C. Becker, R. Rigamonti, V. Lepetit, P. Fua, Supervised feature learning for curvilinear structure segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, Springer, 2013, pp. 526–533.

[26] V. Vapnik, A. Lerner, Generalized portrait method for pattern recognition, Automation and Remote Control 24 (6) (1963) 774–780.

[27] A. Aizerman, E. M. Braverman, L. I. Rozoner, Theoretical foundations of the potential function method in pattern recognition learning, Automation and Remote Control 25 (1964) 821–837.

[28] E. Nowak, F. Jurie, B. Triggs, Sampling strategies for bag-of-features image classification, in: Computer Vision–ECCV 2006, Springer, 2006, pp. 490–503.