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. However, 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 $3.5\,\mu m$ to $6\,\mu m$ [7].

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

Many 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 [14], 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 [10]. 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.* [11] 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 [12]. 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, the obtained codebook from the training is used to represent each OCT volume as a feature vector occurrence histogram. Finally, this histogram is 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 [13]. 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.

In our later study, Lemaître *et al.* proposed a standard classification procedure to differentiate between DME and normal SD-OCT volumes. 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 LBP, or LBP from LBP from Three Orthogonal Planes (LBP-TOP), and when necessary to

represent each volume as a single feature vector, these structures are encoded either using histogram, PCA or BoW, to present the volumes to a RF classifeir. This methodology was tested against Venhuizen *et al.* [12] using public and non-public datasets showing an improvement within the results achieving a Sensitivity (SE) of 87.5% and a Specificity (SP) of 100%.[sik]

In our latest study [14], we proposed a classification framework for differentiating DME and normal SD-OCT volumes. In this work, we proposed to pre-process the volumes using NL-means filtering and mapped them either into a set of structures, corresponding to patches from slices or volumes (local) or a single structure per slice or volume (global). From the considered structures we proposed to extract LBP and LBP-TOP features and represent them using histogram, PCA or BoW algorithm. The represented features were then classified using RF classifier. This methodology was tested on public and non-public dataset and the obtained results revealed improvement against Venhuizen *et al.* [12] algorithm by achieving the SE and SP of 87.5% and 100%, respectively. The obtained results from 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 [moj]

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

Ref	Task			Data size	Pre-processing				Features	Classifier	Evaluation		
	AMD	DME	Normal		De-noise	Flatten	Aligning	Cropping			SE	SP	AUC
Srinivansan *et al.* [11]	✓	✓	✓	45	✓	✓		✓	HOG	SVM	86.7%,100%,100%		
Venhuizen *et al.* [12]	✓		✓	384					texton	RF			0.984
Liu *et al.* [13]	✓	✓	✓	326		✓	✓		Edge, LBP	SVM			0.93
Lemaître *et al.* [14]		✓	✓	32	✓				LBP-LBP-TOP	RF	87.5%	75%	

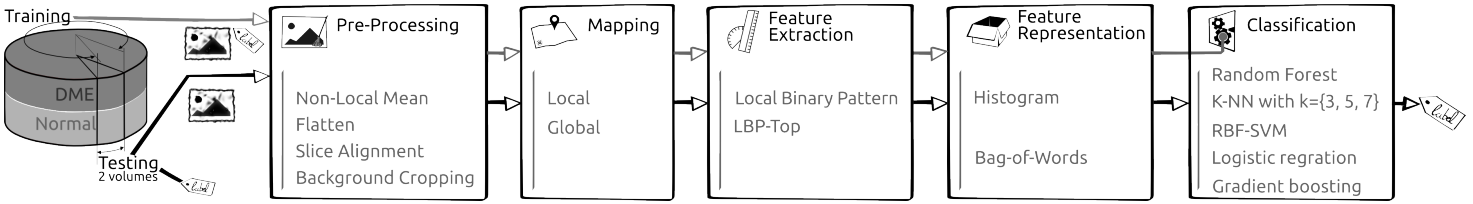


Figure 1: Machine learning classification basic scheme

3. Materials and Methods

Mention texture here, rework where the sections are called

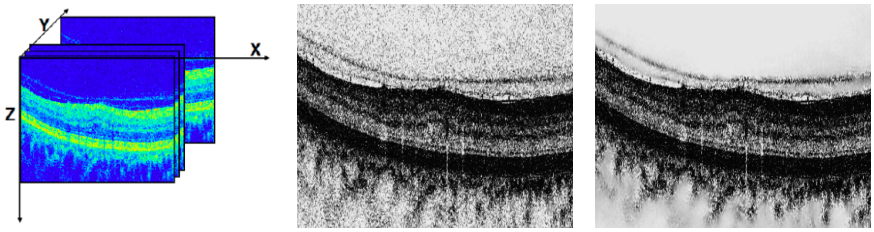
The proposed method, as well as, its experimental set-up for OCT volume classification are outlined in Fig. 1. The methodology is formulated as a standard classification procedure, through 5 steps. First, the OCT volumes are pre-processed as presented in details in Sect. 3.1. The mapping stage is used to determine a discrete set of elements (or structures) which is used for representing the OCT volume. For this work two mapping strategies are defined: (i) *global* and (ii) *local* mapping. In the global mapping approach, a single structure is computed for the image/volume while in the local mapping, a set of structures is defined by sliding a window through the image/volume. Then, a descriptor is computed for each structure. The feature extraction and representation are presented in depth in Sect. 3.2 and Sect. 3.3, respectively. Finally the classification step is presented in Sect. 3.4.

3.1. Image pre-processing

This section describes set of pre-processing techniques which are provided by our framework. These approaches can be arbitrarily combined. Combinations of these technologies and the influence of these combinations are reported in experimentation section 4.

3.1.1. Non-Local Means (NL-means)

OCT images are known to be affected by a speckle noise [15]. Subsequently, NL-means [16] filtering has been successfully used in Ultra-Sound (US) images to filter similar noise [17] and is used in our framework to denoise each B-scan (i.e. each $x - z$ slice) of the OCT volumes (see in Fig. 2(a)). In this regard (For this reason, Due to this reason), we use the NL-means filtering which has been successfully used on US images to remove similar noise [17]. This filter is applied on each B-scan (i.e. each $x - z$ slice) for the OCT volumes (see in Fig. 2(a)).[moj] NL-means filtering offers the advantage to use all the possible self-predictions that the image can provide rather than local or frequency filters



(a) (b) (c)

Figure 2: OCT: (a) Organization of the OCT data - (b) Original image - (c) NL-means filtering.

such as Gaussian, anisotropic or Wiener filters [16]. An example of filtering using NL-means filter on OCT image is depicted in Fig 2(b) and Fig. 2(c).

3.1.2. Flatten

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.
- 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.* [13], as illustrated in Figure 3. 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.

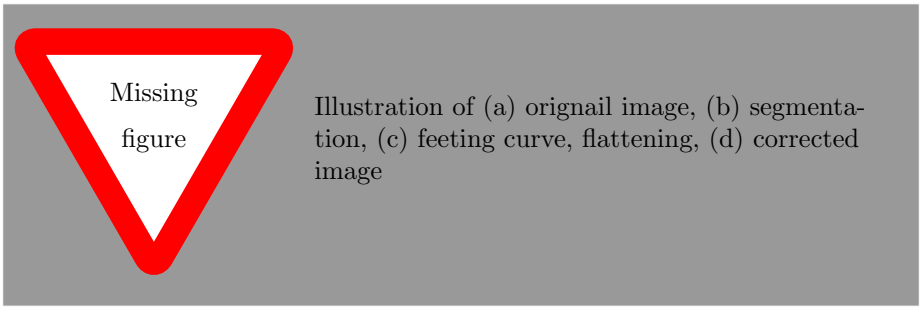


Illustration of (a) orignail image, (b) segmentation, (c) feeting curve, flattening, (d) corrected image

Figure 3:

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.

3.1.4. Background cropping

Cropping out the background allows for less texture to be encoded

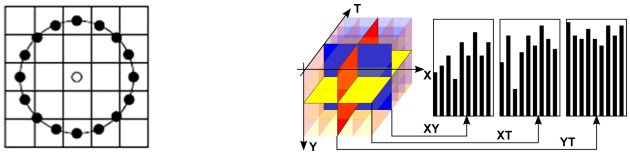
3.2. Features extraction

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.[moj] 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 4(a) illustrates the meaning of P and R.



(a) (b)

Figure 4: The different LBP descriptors: (a) LBP with $(R = 2, P = 16)$ - (b) LBP-TOP [19].

Are we doing Rotation Invariant version??

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].

3.3. Feature representation

give the idea of a single descriptor per volume

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 patch (P) (*local*-mapping).

fab got confused with the Rotation invariant, and would like more detail(or reworking) of the resulting descriptor.

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. [moj]

In *local* mapping, the features are extracted from a set of 2D patches for LBP and a set of sub-volumes for LBP-TOP. Considering a $(m \times m)$ 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 $(d \times N \times LBP_{hist})$ while *local*-LBP-TOP results in a final descriptor of size $\frac{d}{m} \times N \times 3 \times LBP_{hist}$. Here N is the total number of elements in each B-scan, and 3D volume, respectively. [moj]

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 [10]. This model represents the features by creating a visual dictionary or "codebook", 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.

3.4. 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.

k-Nearest Neighbor (NN) is a method in which the labelling decision depends of its is one the simplest supervised machine learning classification methods. In this method a new unlabeled vector is assigned to the most represented class from k nearest-neighbors in the features space. To avoid

a tie case, the parameter k is set to an odd number.

Logistic Regression (LR) [20] is another supervised learning which can provide associated probability of each prediction. As the name suggests, this classifier uses a logistic function, to estimate the probabilities. By defining the posterior probability for one of the classes, using logistic function (see Eq. 2), the probability of other class is defined as $p(c_2|x_i) = 1 - p(c_1|x_i)$.

$$p(c_1|x_i) = \frac{1}{1 + \exp(-w^T x_i)} \tag{2}$$

Here w is a vector of regression parameters, which allows to obtain a linear combination of the input feature vector x_i. Using this model the unlabeled sample is assigned to the class which maximizes the posterior probability.

$$C(x_i) = \arg\max_k p(C = k|x_i) \tag{3}$$

In this method, finding an optimal set of parameters for w is essential. The vectors of parameters w can be inferred by finding the maximum likelihood estimates via optimization methods such as quasi-Newton method [21].

Random Forest (RF) is an ensemble of decision trees and was introduced by [22]. The ensemble uses each tree to predict an output and finalizes the ultimate prediction by aggregating the outputs of all tress. This classifier learns the data by training multiple decision trees on bootstrap samples of the original data. Each bootstrap of D dimension is used for training one decision tree and at each node, the best split among randomly ($d << D$) selected subset of descriptors is chosen. Each tree is grown to its maximum length without any pruning. In the prediction stage a sample is voted by each tree and it is labeled by considering the majority of the votes.

Gradient Boosting (GB) is a generalization form of AdaBoost, which is able to use real-value weak learners and minimizes different loss functions [23].

Gradient Boosting (GB) builds the ensemble in a greedy manner. It iteratively selects the best pair of real-valued weak learners and adjust their weights so that they minimize a given differentiable loss function. Common choice for the weak learner is decision stumps or regression trees while the loss function is generally an exponential loss or a logarithmic loss [24]. This minimization is carried out via gradient descent or quadratic approximation.

Support Vector Machines (SVM) [25] is a sparse kernel method which aims to separate two classes by finding the best hyperplane which maximizes the margin between the two classes. Maximizing the margin is equivalent to minimizing the norm of the normal vector of the hyperplane:

$$\min_{\mathbf{w},\omega_0} \frac{1}{2}\|\mathbf{w}\|^2 \qquad \text{s.t.} \quad y_i(\mathbf{w}^T\mathbf{x_i} + \omega_0) \geq 1, i = 1 : N \tag{4}$$

This constraint intends to force all the point to be in the correct side of the decision boundary (hyperplane) with a minimum distance of 1. This assumption is only valid if the data is linearly separable. Thus for general cases a slack variable $\xi_i \geq 0$ is introduced, which is $\xi_i = 0$, if the points are on/or inside the correct margin boundary, is $0 < \xi_i \leq 1$ if the points are inside the margin but in the correct side of the decision boundary and otherwise if they lie in the wrong side of decision boundary is $\xi_i > 1$. This assumption introduces the *soft margin constraints*. Therefore the optimization problem of SVM classifier is presented by:

$$\begin{aligned} &\min_{\mathbf{w},\omega_0,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{s.t.} \quad \xi_i \geq 0, \quad y_i(x_i^T\mathbf{w} + \omega_0) \geq 1 - \xi_i, i = 1 : N \end{aligned} \tag{5}$$

In Eq. 5, the $\sum_i \xi_i$ term, describes the upper bound on the number of misclassified points and C is the regularization parameter that controls

the tolerance of the classifiers on the number of errors [26].

4. Experiments and Validation

To evaluate the effects and influence of different stages of our framework, we perform various tests using two datasets. A detailed description of the used datasets can be found in Section 4.1. 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.

The *global* and *local* extracted features are then presented in low or high level representation. As previously mentioned, BoW approach is used for high-level representation.

Four experiments are presented in the reminder of this section. Experiment #1 presents the obtained results from our primary study [14] where we the framework was tested with one pre-processing, one classifier, and fixed number of visual words. In this regard in the next experiment #2, we perform a preliminary test in a search for the optimal number of "visual-words" (k). In experiment #3, using the obtained number of words, we test the performance of different classifiers with regards to different pre-processing steps for high-level represented features and finally in experiment #4 we compare the performance of different classifiers with respect to different pre-processing for low-level represented features.

Beside experiment #1, where the primary pipeline is tested on SERI and Duke dataset, the rest of the experiment are performed only on SERI dataset and for the sake of the comparison, these three experiments are performed using all the aforementioned classifiers. However only the relevant part and results related to our experiment is represented within the paper, while the rest are mentioned in online repository.

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

Figure 5: 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.

The pipeline in each experiment is 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 #1 is represented in terms of SE and SP. These statistics are driven from the confusion matrix (see Fig. 5).

The former evaluate the performance of the classifier with respect to the positive class, while the later evaluate it's performance with respect to negative class. These measurements are formulated as:

$$SE = \frac{TP}{TP + FN} \qquad SP = \frac{TN}{TN + FP} \tag{6}$$

In experiment #1 we use Accuracy (ACC) and F1-score (F1) instead. 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. 7 shows the formulation of these measurements.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad F1 = \frac{2TP}{2TP + FP + FN} \tag{7}$$

4.1. Datasets

The majority of the development of our framework has been carried out using our own dataset (SERI) and to facilitate further comparison we use the

Duke public dataset to evaluate our optimal configurations.

SERI - datasets were 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 $\times$ 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.

Duke - datasets published by Srinivasan et al. [11] were 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.

4.2. Experiment #1

As previously stated, this experiment was performed with our primary framework which is presented in [14]. This experiment was performed to evaluate the effects of different features representation. In this experiment, the OCT volumes of our own dataset, were only de-noised using NL-means method and BoW approach was used with fix number of words ($k = 32$). Then the low or high level represented features were classified using RF classifier with 100 un-pruned trees. For the sake of comparison, the same classification were performed for the extracted *global* and *local* LBP-TOP, and *local*-LBP features from Duke dataste. As previously mentioned, this dataset is provided with cropped volumes, which vary in sizes. Thus it was impossible to create a descriptor based on the *global*-LBP features for this dataset. The obtained results from this experiment is listed in Table. 2.

Figure 6:

Once again in order to evaluate the proposed pipeline the proposed method by Venhuizen *et al.* [12] was developed and tested on the two dataset. The comparison of the obtained results with the highest performance achieved by our framework is shown in Table. **??**

4.3. Experiment #2

This experiment, conducted on SERI dataset, is performed to find the optimal number of words for BoW high-level feature representation. *Global* and *local*-LBP and *local*-LBP-TOP feature descriptors are re-represented using BoW approach. In this experiment, the BoW algorithm is performed using various number of words in the range of $\{10, 20, 30, \cdots, 100, 200, \cdots, 500, 1000\}$. The words are randomly selected using k-means++ algorithm and the features are mapped to their nearest word to create the final clusters. In order to asses the effect of number of words, simple linear classifier such as Logistic Regression (LR) is used. The number of words associated with highest ACC and F1 score, is selected as the optimum number of words. Table 4 shows the obtained results of this experiment. The optimum number of words and the achieved ACC and F1 of LR classifier for different configurations are listed in this table.

Figure. 6 shows the obtained graphs for some of the configurations.

4.4. 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-Nearest Neighbor (NN), RF, GB, and SVM. The k-means 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 Radial Basis Function 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 5 shows the obtained results from this experiment.

Table 2: Obtained results using SERI 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 3: Comparing the proposed method by [12] on SERI and Duke datasets.

Data sets	SERI		Duke	
	SE	SP	SE	SP
Venhuizen *et al.* [12]	61.5	58.8	71.4	68.7
{*local*-LBP},16^{riu2}	81.2	75.0	86.6	100.0
{*global*-LBP-TOP},16^{riu2}	75.0	87.5	80.0	86.6

Table 4: The obtained results of experiment #1. 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

Table 5: The obtained results of experiment #2. 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	93.75	75.0	68.7	68.7	68.7
local-LBP-TOP													
	NF												
	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												
	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

4.5. 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. 6.

Table 6: The obtained results of experiment #3. 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	25	50	37.5	68.7	56.2	62.5	56.2	43.7	56.2	68.7
	F	62.5	50	56.2	75	62.5	68.7	75	68.7	62.5	62.5	62.5	68.7
	FA	56.2	50	56.2	75	62.5	68.7	75	68.7	62.5	62.5	62.5	68.7
global-LBP-TOP													
	NF	31.2	93.7	37.5	100	37.5	81.2	62.5	75	62.5	93.7	56.2	87.5
	F	50	56.2	56.2	75	56.2	62.5	68.7	75	43.7	68.7	68.7	56.2
	F+A	75	43.7	56.2	43.7	68.7	50	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	43.7	25	43.7
	F+A	65.2	56.2	50	50	56.2	68.7	43.75	62.5	62.5	50	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	50	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	68.7	75	81.2	56.2	43.7	62.5	62.5	75	75

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

[11] 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.

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

[13] 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.

[14] 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.

[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] Z. Zheng, H. Zha, T. Zhang, O. Chapelle, K. Chen, G. Sun, A general boosting method and its application to learning ranking functions for web search, in: Advances in neural information processing systems, 2007, pp. 1697–1704.

[24] 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.

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

[26] K. P. Murphy, Machine learning: a probabilistic perspective, MIT press, 2012.

[27] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.