

An anomaly detection approach for the identification of DME patients using spectral domain optical coherence tomography images

Désiré Sidibé^{*,a}, Shrinivasan Sankar^a, Guillaume Lemaître^a,
Mojdeh Rastgoo^a, Joan Massich^a, Carol. Y. Cheung^{b,d}, Gavin S. W. Tan^b,
Dan Milea^b, Ecosse Lamoureux^b, Tien Y. Wong^b, Fabrice Mériaudeau^{a,c}

^a*LE2I, UMR6306, CNRS, Arts et Métiers, Université de Bourgogne-Franche Comté,
F-21000 Dijon, France*

^b*Singapore Eye Research Institute, Singapore National Eye Center, Singapore*

^c*Center for Intelligent Signal and Imaging Research (CISIR), EEE Department,
Universiti Teknologi Petronas, 32610 Seri Iskandar, Perak, Malaysia*

^d*Department of Ophthalmology and Visual Sciences, The Chinese University of Hong
Kong*

Abstract

This paper proposes a method for automatic classification of spectral domain OCT data for the identification of patients with retinal diseases such as Diabetic Macular Edema (DME). We address this issue as an anomaly detection problem and propose a method that not only allows the classification of the OCT volume, but also allows the identification of the individual diseased B-scans inside the volume. Our approach is based on modeling the appearance of normal OCT images with a Gaussian Mixture Model (GMM) and detecting abnormal OCT images as outliers. The classification of an OCT volume is based on the number of detected outliers. Experimental results with two different datasets show that the proposed method achieves a

*Corresponding author

Email address: dro-desire.sidibe@u-bourgogne.fr (Désiré Sidibé)

sensitivity and a specificity of 80% and 93% on the first dataset, and 100% and 80% on the second one. Moreover, the experiments show that the proposed method achieves better classification performance than other recently published works.

Key words: Diabetic retinopathy, Diabetic macular edema, SD-OCT, Classification, Anomaly detection.

1. Introduction

1.1. Clinical motivation

Diabetic retinopathy (DR) and diabetic macular edema (DME) are leading causes of vision loss worldwide [1]. As the number of people affected by diabetes is expected to grow exponentially in the next few years [2, 3], early detection and treatment of these retinal diseases is becoming a major health issue in developed countries. Currently, the two main tools used for the screening of retinal diseases are fundus photography and optical coherence tomography (OCT) devices. The former type of device allows capturing the inner surface of the eye, including retina, optic disc, macula, and other retinal tissues using low power microscopes equipped with an embedded camera system [4]. The latter is based on infrared optical reflectivity and produces cross-sectional and three-dimensional images of the central retina. The new generation of OCT imaging, namely Spectral Domain OCT (SD-OCT) offers higher resolution and faster image acquisition over conventional time domain OCT [5]. SD-OCT can produce from 27,000 to 40,000 A-scans/second with an axial resolution ranging from $3.5\mu m$ to $6\mu m$ [6].

While many previous works in literature are dedicated to the analysis of

fundus images [7, 8], in this work we focus on the automatic analysis and classification of SD-OCT images. Indeed, the use of fundus photography is limited to the detection of signs which are visible in the retinal surface, such as hard and soft exudates, or micro-aneurysms, and fundus photography cannot always identify the clinical signs such as cysts not visible in the retinal surface. Moreover, contrary to fundus photography, SD-OCT provides quantitative measurements of retinal thickness and information about cross-sectional retinal morphology.

1.2. Related work

The main, common, feature associated with DME is an increase of the macular thickness in the central area of the retina [9]. Therefore, several methods have been proposed which are based on the segmentation of retinal layers and the identification of signs of the disease such as intraretinal and subretinal fluid [10, 11]. For example, Quellec *et al.* [10] proposed a method that starts with the segmentation of 10 retinal layers and the extraction of texture and thickness properties in each layer. Then, local retinal abnormalities are detected by classifying the differences between the properties of normal retinas and the diseased ones. Chen *et al.* [11] proposed a method for 3D segmentation of fluid regions in OCT data using a graph-based approach. The method is based on the segmentation of retinal layers and the identification of potential fluid regions inside the layers. Then, a graph-cut method is applied to get the final segmentation using the probabilities of an initial segmentation based on texture classification as constraints. Although methods based on a prior segmentation of retinal layers have reported good detection accuracy, this first step is still difficult and subject to errors [12, 13]. More-

over, as mentioned by Lee *et al.*, retinal thickness measurement differences produced by different algorithms are important and it is not always possible to compare retinal thicknesses among eyes in which thickness measurements have been obtained by different systems [14].

In this paper, we focus on methods based on direct classification of SD-OCT data without prior segmentation of retinal layers. Several methods have recently been proposed in literature for the classification of SD-OCT data and the identification of patients with retinal diseases versus normal patients. Liu *et al.* [15] proposed a method based on local binary patterns (LBP) and gradient information for macular pathology detection in OCT images. The method uses a multi-resolution approach and builds a 3-level multi-scale spatial pyramid of the foveal B-scan for each patient. LBP and gradients are then computed from each block at every level of the pyramid, and represented as histograms. The obtained histograms are concatenated into a global descriptor whose dimension is reduced using principal component analysis (PCA). Finally a support vector machines (SVM) classifier is used for classification. With a dataset of 326 OCT volumes, the method achieved good results in detecting OCT scans containing different pathologies such as DME or age-related macular degeneration (AMD), with an area under the ROC curve (AUC) value of 0.93.

Another method using gradient information and SVM is proposed by Srinivasan *et al.* [16] to distinguish between DME, AMD, and normal SD-OCT volumes. The method is based on pre-processing to reduce the speckle noise in OCT images and flattening of the images to reduce the variation of retinal curvature among patients. Histograms of oriented gradients (HOG)

features are then extracted from each B-scan of an OCT volume and a linear SVM is used for classification. Note that the method classifies each individual B-scan into one of the three categories, i.e. DME, AMD and normal, and classifies an OCT volume based on the number of B-scans in each category. On a dataset of 45 patients containing 15 normal subjects, 15 DME patients and 15 AMD patients, the method achieved a correct classification of 100%, 100% and 86.67% for AMD, DME and normal cases, respectively. Note that no specificity or sensitivity values are reported in this study.

Anantrasirichai *et al.* [17] proposed a method for OCT images classification using texture features such as LBP, grey-level co-occurrence matrices and wavelet features, together with retinal layer thickness. The method relies on SVM to classify retinal OCT images of normal patients against patients with glaucoma. A classification accuracy of about 85% is achieved with a dataset of 24 OCT volumes. Albarrak *et al.* [18] first decomposed the OCT volume into sub-volumes, and extracted both LBP and HOG features in each sub-volume. The features from the sub-volumes are concatenated into a single feature vector per OCT volume, and PCA is applied for dimensionality reduction. Finally, a Bayesian network classifier is used and the method is tested with 140 OCT volumes of normal and AMD patients. The method achieved an AUC value of 94.4%.

The authors in [19] proposed a method based on the bag-of-words (BoW) approach for OCT images classification. The method starts with the detection and selection of few keypoints in each individual B-scan. This is achieved by keeping the most salient points corresponding to the top 3% of the vertical gradient values. Then, an image patch of size 9×9 pixels is extracted

around each keypoint, and PCA is applied to transform every patch into a feature vector of dimension 9. These feature vectors are then used to create a visual vocabulary using k -means clustering algorithm, which is used to represent each OCT volume as a histogram of the words occurrences. Finally, this histogram is used as feature vector to train a random forest classifier. On the task of classifying OCT volumes between AMD and normal cases, the method achieved an AUC of 98.4% with a dataset of 384 OCT volumes. Another approach based on BoW model is proposed by Lemaître *et al.* [20] for automatic classification of OCT volumes. The method is based on LBP features to describe the texture of OCT images and dictionary learning using the BoW approach. In this work, the authors extracted both 2D and 3D LBP features to describe OCT volumes, and show that 3D LBP features perform better than 2D LBP features from each B-scan. The features are used to create a visual vocabulary using the BoW approach, and a random forest classifier is employed for classification. The method achieved a specificity and a sensitivity of 75% and 87.5% respectively, with a dataset of 32 OCT volumes.

1.3. Contributions

In this paper, we propose a novel method for automatic identification of patients with retinal diseases, such as DME, versus normal subjects. Our approach is based on modeling the appearance of normal OCT images with a Gaussian Mixture Model (GMM) and detecting abnormal OCT images as outliers. Then, the classification of an OCT volume as normal or abnormal is based on the number of detected outliers in the volume. The main contributions of our paper are as follows:

- we propose a method to model the global appearance of normal OCT volumes using a Gaussian mixture model (GMM).
- we use an anomaly detection approach to identify abnormal B-scans as outliers to the GMM model, and finally detect unhealthy OCT volumes based on the number of abnormal B-scans.

Our approach differs from previous works in that our abnormal B-scan detection method does not require a training set with manually identified normal and abnormal B-scans which is a tedious and time consuming task.

The rest of this paper is organized as follows. In Section 2, we describe the proposed anomaly detection approach in details. Experiments and results are discussed in Section 3. Finally, concluding remarks are drawn in Section 4.

2. Methodology

This section describes the proposed method for SD-OCT volumes classification and the identification of abnormal B-scans inside each volume. As mentioned in Section 1.3, the proposed method is based on an anomaly detection approach, in which we model the appearance of normal OCT images with a Gaussian Mixture Model (GMM) and detect abnormal OCT images as outliers. The method comprises four main steps: i) pre-processing to remove noise and align the B-scans; ii) the computation of the GMM model; iii) the detection of individual abnormal B-scans; and iv) the classification of OCT volumes. Each of these steps is further described in the next subsections.

2.1. Pre-processing

The SD-OCT data is organized as a 3D cube, i.e. a series of slices called B-scans as shown in Fig. 1(a). Each B-scan is a cross-sectional image of the retina. The first pre-processing stage of the proposed methodology applies an image denoising method to reduce the speckle noise in each B-scan, since SD-OCT images are known to be corrupted by a speckle noise [21]. In our work, we used the Non-Local Means (NL-means) algorithm [22] which offers the advantage to reduce the noise while preserving the details and texture of the original image. An example of filtering using NL-means filter on an OCT image is depicted in Fig. 1(c).

Furthermore, because SD-OCT images of the retina show natural curvatures across scans, we also apply a flattening procedure described in [15] to ensure that the dominant features are aligned in a nearly horizontal line. This is done by first finding the retina area in the B-scan using image thresholding and morphological operations, and then fitting a curve to the found region using a second-order polynomial. The entire retina region is finally warped so that it is approximately horizontal. The reader is referred to [15] for details. An example of a flattened OCT B-scan is shown in the image of Fig. 1(c).

2.2. Training and model creation

In this section, we describe our approach for modeling the global appearance of normal SD-OCT images. More specifically, we assume that we have a set of N SD-OCT volumes which are known to be from healthy patients. As shown in Fig. 1(a), each OCT volume is organized as a series of B-scans

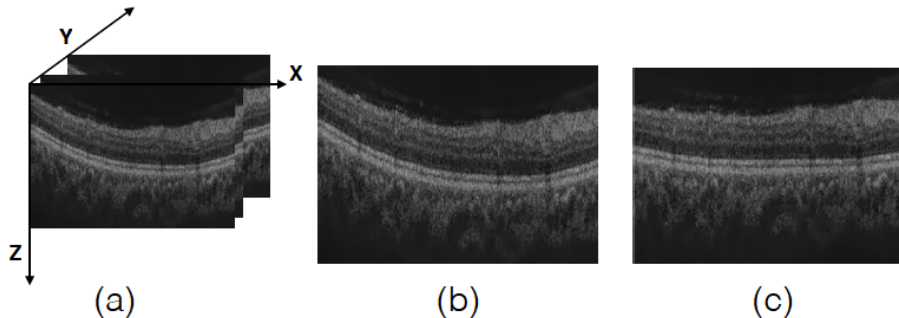


Figure 1: SD-OCT volume: (a) Organization of the OCT data; (b) an original B-scan; (c) denoised and flattened B-scan.

along the y -axis, and each B-scan is a 2D image of size $W \times H$ in the x - z plane.

Let V be one of the N normal SD-OCT volumes. We assume that each volume V contains N_Y B-scans, where the value N_Y depends on the specific resolution of the OCT device used for acquisition, and we represent each B-scan as a vector b of size $d = WH$ obtained by concatenation of the rows of the $W \times H$ 2D image. Then, we can represent the whole SD-OCT volume as a set of N_Y vectors in \mathbb{R}^d each corresponding to one B-scan of the volume: $V = \{b_1, b_2, \dots, b_{N_Y}\}$, $b_i \in \mathbb{R}^d$ for all $i = 1, \dots, N_Y$.

By putting together all B-scans from the N normal SD-OCT volumes, we create a large data matrix \mathbf{X} whose columns are the B-scans from all the volumes: $\mathbf{X} = [b_1, b_2, \dots, b_M]$, with $M = NN_Y$ the total number of normal B-scans. In this formulation, we consider each B-scan as a vector in a d -dimensional space. Since $d = WH$, where $W \times H$ is the spatial resolution of the B-scans, we have a very high dimensional feature space. For example, if each B-scan is a 400×512 image, then $d = 204800$.

We apply principal components analysis (PCA) to reduce the dimensionality of the data. We first compute the mean normal B-scan as $\bar{b} = \frac{1}{M} \sum_{j=1}^M b_j$, and center the data as follows: $\mathbf{X} = [b_1 - \bar{b}, b_2 - \bar{b}, \dots, b_M - \bar{b}]$. We then compute the covariance matrix of the data as $\mathbf{C} = \frac{1}{M} \mathbf{X} \mathbf{X}^T$, and perform the eigen-decomposition of this symmetric matrix: $\mathbf{C} = U \Lambda U^T$. The eigenvectors of \mathbf{C} , i.e. the columns of matrix U , form the set of principal components of the data. The corresponding eigenvalues $\lambda_i, i = 1, \dots, d$, give the importance of each axis. For dimensionality reduction, we keep the first p dominant principal components corresponding to the p largest eigenvalues. We set the value of p such that at least 95% of the total variance of the data is retained. So, we set p such that $\sum_{i=1}^p \lambda_i / \sum_{i=1}^d \lambda_i \geq 0.95$. PCA typically reduces the dimension of the data from $d = 204800$ to few hundreds, i.e. $p = 300$ or $p = 500$ as shown in Section 3.2.

In the reduced p -dimensional space, we represent the distribution of normal B-scans' appearance with a Gaussian mixture model (GMM). We adopt a GMM to account for the inter- and intra-patient variabilities of the intensity appearance of the B-scans. A GMM is a parametric probability density function represented as a weighted sum of K Gaussian component densities [23]:

$$p(\mathbf{x}|\theta) = \sum_{i=1}^K w_i g_i(\mathbf{x}|\mu_i, \Sigma_i), \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^p$, $w_i, i = 1, \dots, K$ are the mixture weights which satisfy $\sum_{i=1}^K w_i = 1$, and $g_i(\mathbf{x}|\mu_i, \Sigma_i)$ is the i -th component of the mixture model. Each component of the model is a p -dimensional multivariate Gaussian de-

defined by:

$$g_i(\mathbf{x}|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i)\right\}, \quad (2)$$

where $\mu_i \in \mathbb{R}^p$ is the mean vector of the Gaussian, and Σ_i is the $p \times p$ covariance matrix.

The parameters of the model $\theta = \{w_i, \mu_i, \Sigma_i, i = 1, \dots, K\}$, are learned from the set of training data using the expectation-maximization (EM) algorithm which iteratively estimates the parameters of the model [23]. In practice, a single global parameter is set to define the GMM, which is the number K of components. Once K is fixed, the EM algorithm estimates the set of parameters $\theta = \{w_i, \mu_i, \Sigma_i, i = 1, \dots, K\}$. We choose the best value for K by cross-validation as discussed in Section 3.2. Moreover, for a good initialization of the GMM we first use k -means clustering algorithm to create K clusters from the set of training data, and we fit a Gaussian density to each of the K clusters to initialize the EM algorithm.

A flowchart of the method for representing the appearance of normal B-scans as a GMM is shown in Fig. 2, where the final model is illustrated for the case $p = 2$.

2.3. Abnormal B-scans identification

Our approach is based on the identification of abnormal B-scans, i.e. B-scans showing visible signs of retinal diseases, using the learned GMM. Given a B-scan \mathbf{b} of size $W \times H$, we first represent it as a vector in \mathbb{R}^d , $d = WH$, and project this d -dimensional vector onto the set of principal components found in Section 2.2, to obtain $\tilde{\mathbf{b}} \in \mathbb{R}^p$, where $p \ll d$.

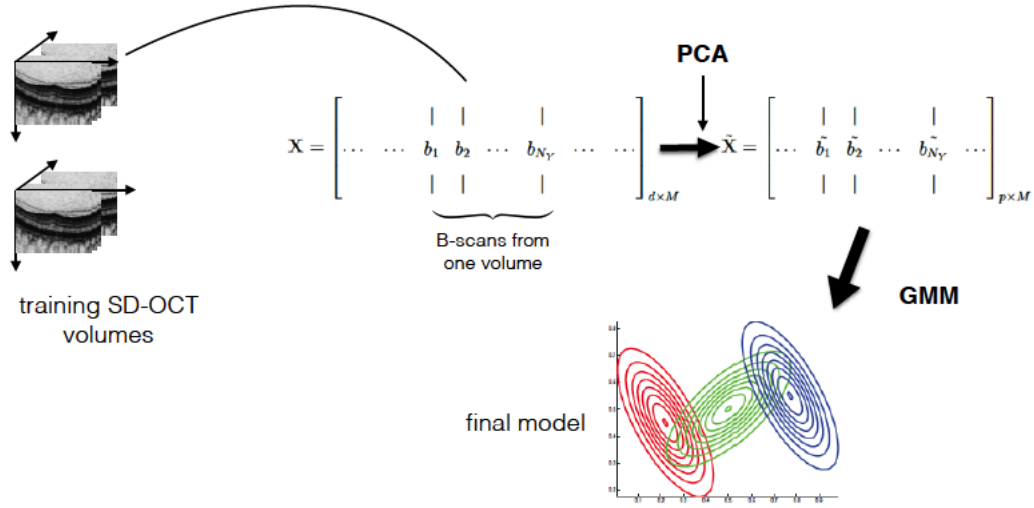


Figure 2: Flowchart of the Gaussian mixtures model creation. The final model is here illustrated for $p = 2$.

Then, we compute the Mahalanobis distance from $\tilde{\mathbf{b}}$ to the GMM as:

$$\Delta_{GMM}(\tilde{\mathbf{b}}) = \underset{i}{\operatorname{argmin}} \Delta_i(\tilde{\mathbf{b}}), \quad (3)$$

where $\Delta_i(\tilde{\mathbf{b}})$ is the distance from $\tilde{\mathbf{b}}$ to the i -th component of the GMM defined by:

$$\Delta_i(\tilde{\mathbf{b}}) = (\tilde{\mathbf{b}} - \mu_i)^T \Sigma_i^{-1} (\tilde{\mathbf{b}} - \mu_i). \quad (4)$$

In order to identify normal and abnormal B-scans, we need to set a threshold value on the distance Δ_{GMM} . That is, we consider as abnormal, B-scans with distances from the GMM above a certain threshold value.

To explain how we find this threshold value, let us first consider the case of 1D Gaussian distributions. If \mathbf{x} is a 1D random variable normally distributed as $\mathbf{x} \sim \mathcal{N}(\mu, \sigma)$, then $P(|\mathbf{x} - \mu| \leq 2\sigma) = 0.95$. Which means that, with probability 0.95, the data point \mathbf{x} lies within a distance of 2σ from the

mean value μ . In the context of anomaly detection, data points which do not satisfy this constraint are considered outliers or abnormal data points. That is, \mathbf{x} does not agree with the Gaussian model if $|\mathbf{x} - \mu| > 2\sigma$, and \mathbf{x} is therefore considered as an anomaly. This is the well known 2σ -rule for 1D normal distributions.

This idea can be extended to the case of multivariate p -dimensional Gaussian distributions. First, it can be shown that the Mahalanobis distance defined in Equation (4) follows a χ^2 -distribution with p degrees of freedom, since it is the sum of p Gaussian random variables [23, 24]. Therefore, $\Delta_{GMM} \sim \chi_p^2$. We can then detect outliers as data points in \mathbb{R}^p with a Mahalanobis distance that does not fit the χ_p^2 -distribution. We set the threshold value to be equal to the 95% quantile of χ^2 -distribution with p degree of freedom, i.e. $\delta = \chi_{p:0.95}^2$. Finally, we consider a B-scan represented as a vector \mathbf{b} , as an abnormal B-scan using the following rule:

$$\mathbf{b} \text{ is abnormal if } \Delta_{GMM}(\tilde{\mathbf{b}}) > \delta. \quad (5)$$

Figure 3 summarizes the abnormal B-scans detection procedure using a GMM.

2.4. SD-OCT volumes classification

The final step of our methodology is the identification of patients with retinal diseases, such as DME, versus normal subjects. Given a SD-OCT volume with N_Y B-scans, $V = \{b_1, b_2, \dots, b_{N_Y}\}$, we first detect the number of abnormal B-scans in V using the approach described in Section 2.3, and Equation (5). The SD-OCT volume is then classified as normal or abnormal based on the number of detected abnormal B-scans. Ideally, a normal OCT

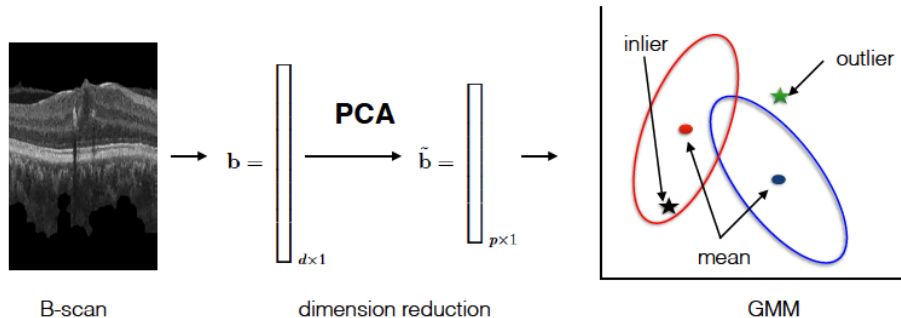


Figure 3: Abnormal B-scans detection using a Gaussian mixture model.

volume should contain no abnormal B-scan, while an abnormal one should contain many. In practice, we set a threshold value on the number of detected abnormal B-scans, in order to identify diseased SD-OCT volumes. In order to increase the robustness of the method, we observe that a clinical signs of a disease, such as hard exudates or cystoid spaces, should be visible in more than one single B-scan. In other words, if a B-scan $b_i \in V = \{b_1, b_2, \dots, b_{N_Y}\}$ is detected as abnormal, then at least one of its neighboring B-scans in the y -axis, i.e. b_{i-1} or b_{i+1} , should also be detected as abnormal. This empirical observation reduces the number of false detections. This threshold parameter is empirically set, and the values used in our experiments are given in Section 3.3.

The two main parameters of the proposed method are the number of Gaussian components K and the threshold, i.e. the number of abnormal B-scans, used to detect diseased SD-OCT volumes. The former parameter is set during training of the GMM by cross-validation, Section 3.2, and the selected parameter is used for testing the model. The latter is chosen such that the classification accuracy is optimized, and we performed experiments

with different values in Section 3.3.

3. Experiments and results

In this section, we evaluate the performance of the proposed method in detecting abnormal SD-OCT volumes. We use two different datasets for the experiments and we also compare the proposed anomaly detection based approach with two existing methods based on features detection and supervised classification [19, 20].

3.1. Datasets

We use two different datasets in our experiments. The first dataset, which we name SERI, consists of 32 SD-OCT volumes (16 DME and 16 normal cases), and was provided by Singapore Eye Research Institute (SERI). This data is acquired using CIRRUS TM (Carl Zeiss Meditec, Inc., Dublin, CA) SD-OCT camera, and each OCT volume contains 128 B-scans each of size 400×512 pixels. All 32 volumes were read and assessed by trained graders and classified as normal or DME based on the presence of specific signs such as retinal thickening, hard exudates, and intra-retinal cystoid space.

The second dataset, which we refer to as Duke, is provided by Duke University [16]. The dataset consists in 45 SD-OCT volumes from 15 DME patients, 15 AMD patients and 15 normal subjects, respectively. Note however that we will only use the DME and normal OCT volumes in our experiments, i.e. 30 SD-OCT volumes. The data is acquired using Spectralis (Heidelberg Inc., Heidelberg, Germany) camera, and the number of B-scans per volume varies from 31 to 97 scans, each of size 512×496 pixels.

3.2. GMM training

The proposed SD-OCT volumes classification method is based on the detection of abnormal B-scans using a GMM as described in Section 2. Therefore, the first step of the approach is the learning of a GMM which represents the global intensity appearance of normal SD-OCT images.

For a given dataset, we use part of the normal SD-OCT data for training, i.e. for learning the GMM. For example, for the SERI dataset, we randomly select 11 of the 16 normal volumes for training. We first apply PCA to reduce the dimensionality of the data from $d = 400 \times 512 = 204800$ to $p = 500$. The value 500 is selected such that 95% of the total variance of the the data is preserved during projection onto the lower dimensional space. Then, we fit a mixture of Gaussian model to the data in the reduced space of dimension 500 as described in Section 2.2. The key parameter of the model is the number K of Gaussians used in the GMM. This parameter is selected such that the created GMM well represents the appearance of normal B-scans. We set its value by cross-validation. More precisely, we use 8 of the 11 training volumes for fitting a GMM with K components, and we use the remaining 3 volumes for validation. This is done by checking each B-scan of the validation OCT volumes against the learned GMM, and detecting outliers using Equation 5. For each value of K , we repeat the cross-validation experiment, i.e. training with 8 volumes and testing with 3, 10 times and report the average classification accuracy.

A similar training procedure is used for the Duke dataset. However, since this dataset contains 15 normal patients, we use 10 normal SD-OCT volumes to learn the GMM.

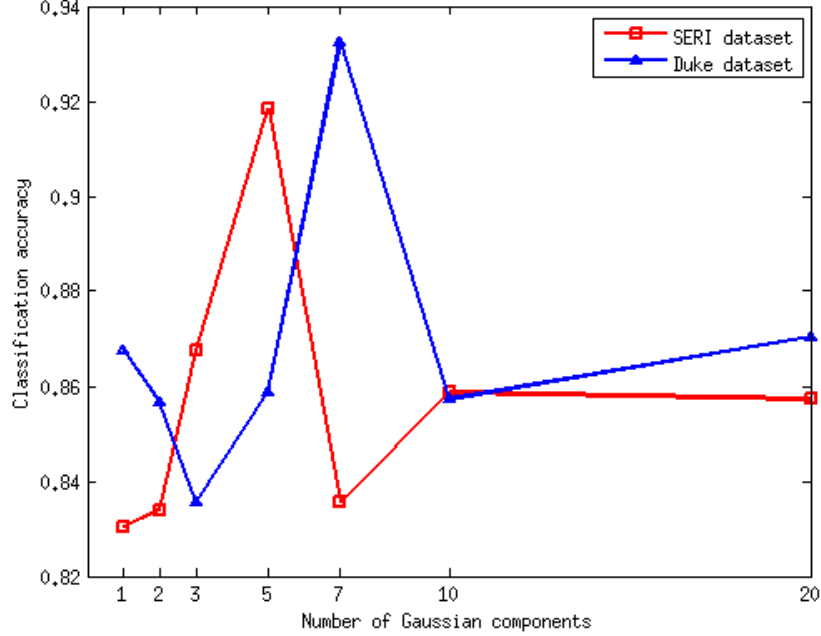


Figure 4: Variation of B-scans classification w.r.t. the number of Gaussian components. Cross-validation results for SERI dataset.

The obtained average classification accuracies are shown in Fig 4. The best classification accuracy is achieved by $K = 5$, which is the GMM giving the least number of misclassified B-scans, for SERI dataset. For Duke dataset, the best value is $K = 7$. Once, the number of Gaussian components is selected, $K = 5$ and $K = 7$ for SERI and Duke dataset, respectively, the final GMM is fitted using all training volumes, i.e. 11 for SERI dataset and 10 for Duke dataset.

One should note that we train one GMM for each dataset. Since the datasets were acquired in different locations, in different conditions, and with different machines, the intensity distribution and the obtained GMM

are different. Hence, the best number of Gaussian components is also dataset dependent.

3.3. Classification results

For SERI dataset, we test the proposed method using the 5 normal SD-OCT volumes not used for training the GMM, and all 16 DME cases. Similarly, for Duke dataset, we test the method with the 5 normal volumes not used for training, and all 15 DME cases. The performance of the method is evaluated in terms of sensitivity and specificity values.

A test SD-OCT volume is classified as normal or abnormal based on the number of detected abnormal B-scans. Ideally, a normal SD-OCT volume should contain no abnormal B-scan, while an abnormal one should contain many. So, a volume is classified as abnormal if it contains a number of abnormal B-scans greater than or equal to a predefined threshold value. We have tried different threshold values with the SERI dataset and the results are shown in Table 1. As expected, a low threshold gives high sensitivity at the cost of low specificity, and as the threshold increases, the specificity increases while the sensitivity decreases. Indeed, with a low threshold value many normal volumes are classified as abnormal, while with a large threshold value many abnormal volumes are classified as normal.

The best results for SERI dataset are obtained with a threshold value of 4, which yields a sensitivity and a specificity of 93.75% and 80%, respectively. Note that we will use the same threshold value for both datasets, i.e. we use the threshold chosen with SERI dataset to classify the OCT data from Duke dataset.

We also compare the proposed SD-OCT volumes classification method

Threshold	2	4	6	8	10
Sensitivity	100	93.75	87.50	75.00	62.50
Specificity	60.00	80.00	80.00	100	100

Table 1: Variation of SD-OCT volumes classification results w.r.t. the threshold (i.e. the # of detected abnormal B-scans). Results for SERI dataset.

with two existing methods based on features detection and supervised classification [19, 20]. Venhuizen *et al.* [19] used intensity features and random forest classifier, and Lemaître *et al.* [20] used LBP features with random forest for classification. Note that for these two methods, the classification performances are computed using a *leave-one-out* strategy. That is, a pair of normal-DME volumes is selected for testing while the rest of the data is used the training the classifier. This procedure is repeated 16 times and 15 times for SERI and Duke dataset, respectively.

Table 2 shows the classification results obtained with the three methods using both datasets. We can observe that the proposed method achieves better performances than the other two methods using the SERI dataset. For this dataset, the proposed method achieves a sensitivity and a specificity

		Method		
		Proposed	Venhuizen <i>et al.</i> [19]	Lemaître <i>et al.</i> [20]
SERI dataset	Sens	93.75	61.53	87.50
	Spec	80.00	58.82	75.00
Duke dataset	Sens	80.00	71.42	86.67
	Spec	100	68.75	100

Table 2: SD-OCT volumes classification results.

of 93.75% and 80%, while our implementation of the method of Venhuizen *et al.* [19] achieves a sensitivity and a specificity of 61.53% and 58.82%, and the method of Lemaître *et al.* [20] achieves 87.50% and 75% respectively. The difference in performances between Venhuizen *et al.* [19] and Lemaître *et al.* [20] indicates that for supervised classification, texture features such as LBP used in [20] are better than intensity features used in [19].

For the second dataset, Duke dataset, the proposed approach achieves a sensitivity of 80% and a specificity of 100%. The method of Lemaître *et al.* [20] achieves the same specificity and a better sensitivity of 86.67%. However, this method classifies the OCT volume as a whole and does not allow the identification of specific B-scans. As mentioned in Section 1.3, one main advantage of the proposed approach is the ability to identify abnormal B-scans inside a volume. This is useful, since it avoids a visual inspection of all B-scans in the volume. Furthermore, once abnormal B-scans are detected, one can focus on the segmentation of specific features such as exudates or cysts in that B-scan. Our approach differs from previous works which require a training set with manually annotated normal and abnormal B-scans, which is a tedious and time consuming task. Figure 5 shows some of the abnormal B-scans detected by the proposed method. As can be seen, B-scans with strong DME features, such as subretinal and intraretinal fluids and cysts, are identified by our approach, as well as B-scans with small changes from healthy condition. Note that the B-scans shown in this figure are raw images from both datasets, without pre-processing. Hence, the visible noise reduction corresponds to the acquisition process inherent to the respective machine.

Figure 6 shows examples of abnormal B-scans which are not detected by

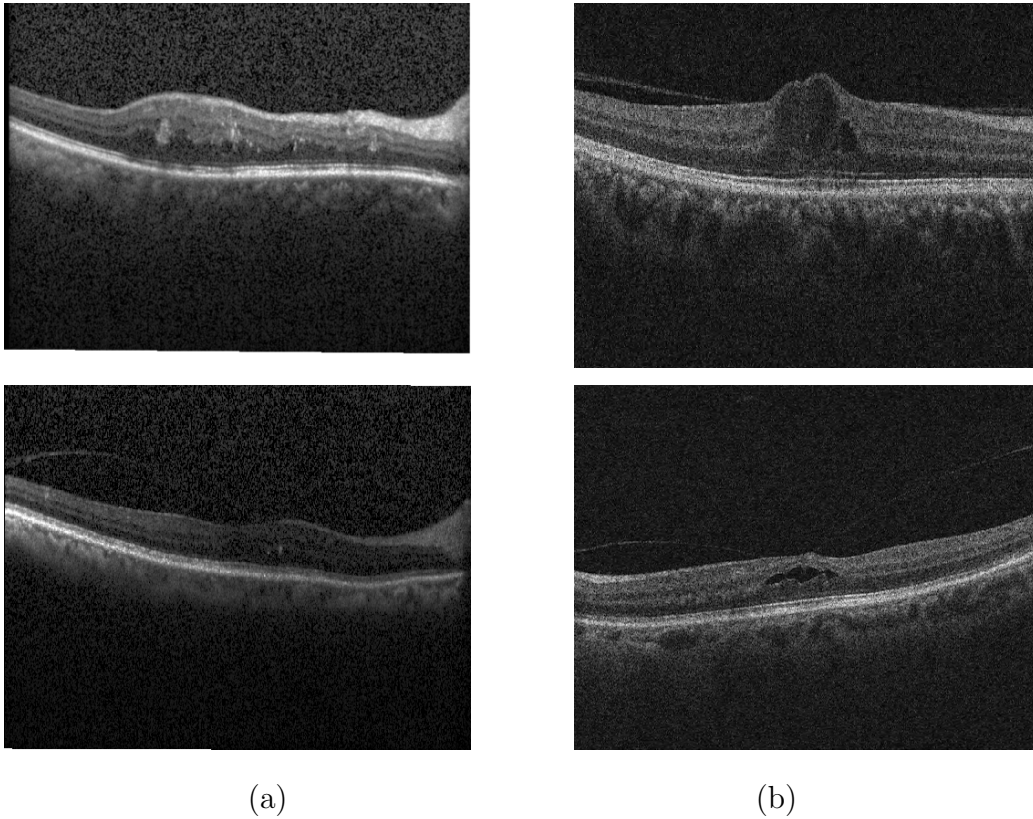


Figure 5: Examples of detected abnormal B-scans by the proposed method: (a) B-scans from Duke dataset; (b) B-scans from SERI dataset.

our method. These are mostly B-scans showing small changes from healthy condition such as few exudates.

4. Conclusion

In this paper, we have proposed a novel method for automatic classification of spectral domain OCT (SD-OCT) data, for the identification of patients with diseases such as Diabetic Macular Edema (DME) versus healthy patients. Our method is based on an anomaly detection approach in which we

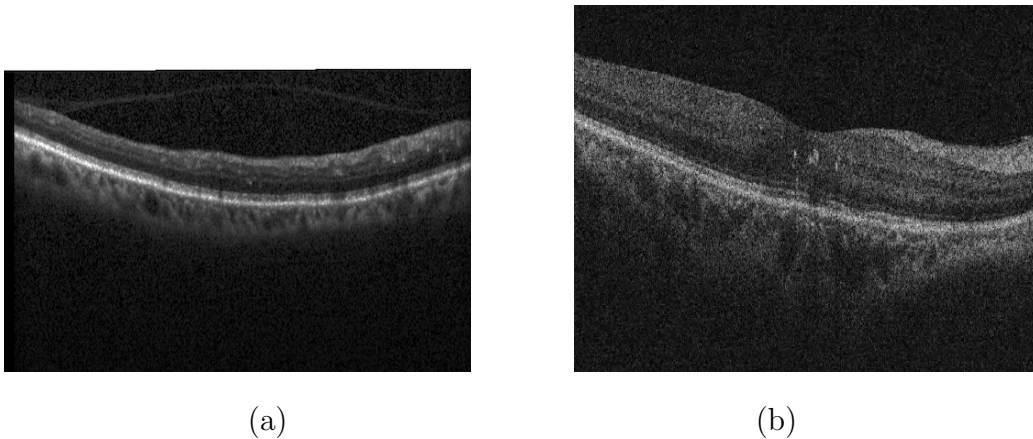


Figure 6: Examples of abnormal B-scans not detected by the proposed method. (a) B-scan from Duke dataset; (b) B-scan from SERI dataset.

model the appearance of normal SD-OCT images with a Gaussian Mixture Model (GMM) and we detect abnormal SD-OCT images as outliers. Finally, we classify a SD-OCT volume as normal or abnormal based on the number of detected outliers in the volume. Experimental results with two different datasets show that the proposed method achieves excellent results and outperforms other recently published works based on feature extraction and supervised classification. In particular, our method achieves a sensitivity and a specificity of 80% and 93% on the first dataset, and 100% and 80% on the second one. Furthermore, the proposed method not only allows the classification of the SD-OCT volume, but also allows the identification of the individual abnormal B-scans inside each volume. So, it can be used as a first step towards more specific detection of DME features from the identified abnormal B-scans.

The proposed framework can easily be extended to include other types of features, such as texture or contrast features. In future works, we will

investigate the detection of other retinal diseases such as AMD, and the possibility to discriminate between different diseases.

Conflict of interest statement

The authors declare no conflicts of interest related to this research work.

Aknowledgement

This work was supported by Institut Français de Singapour (IFS) and Singapore Eye Research Institute (SERI) through the PHC Merlion program (2015-2016).

References

- [1] Cheung, N., Mitchell, P., Wong, T.Y., “Diabetic retinopathy”, *Lancet*, vol. 376 (9735), pp. 124-36, 2010.
- [2] Fong D.S., Aiello L.P., Ferris F.L., and Klein P., “Diabetic retinopathy”, *Diabetes Care*, vol. 27 (10), pp. 2540-53, 2004.
- [3] Wild, S., Roglic, G., Green, A., Sicree, R. and King, H., “Global prevalence of diabetes estimates for the year 2000 and projections for 2030”, *Diabetes Care*, vol. 27(5), pp.1047-1053, 2004.
- [4] Saine P.J., “What is a fundus camera”, *Ophthalmic Photographers Society*, 2006.
- [5] Leung, C.K., Cheung, C.Y., Weinreb, R.N., Lee, G., Lin, D., Pang, C.P., Lam, D.S., “Comparison of macular thickness measurements between time domain and spectral domain optical coherence tomography.”.

- Investigative Ophthalmology & Visual Science*, vol. 49(11), pp. 4893-7, 2008.
- [6] Chen, T.C., Cense, B., Pierce, M.C., Nassif, N., Park, B.H., Yun, S.H., White, B.R., Bouma, B.E., Tearney, G.J., de Boer, J.F., “Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging”. *Arch. Ophthalmol.*, vol. 123(12), pp. 1715-1720, 2005.
- [7] Abramoff M.D, Garvin M.K., and Sonka M., “Retinal image analysis: a review”, *IEEE Review Biomed. Eng.*, vol. 3, pp. 169-208, 2010.
- [8] Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J.P., al-Diri, B., Cheung, C.Y., Wong, D., Abramoff, M., Lim, G., Kumar, D., Burlina, P., Bressler, N.M., Jelinek, H. F., Meriaudeau, F., Quellec, G., MacGillivray, T., and Dhillon, B., “Validation retinal fundus image analysis algorithms: issues and proposal”, in *Investigative Ophthalmology & Visual Science*, vol. 54(5), pp. 3546-3569, 2013.
- [9] Costa, R.A., Skaf, M., Melo Jr, L.A.S., Calucci, D., Cardillo, J.A., Castro, J.C., Huang, D., Wojtkowski, M., “Retinal assessment using optical coherence tomography”, *Progress in Retinal and Eye Research*, 25(3), pp. 325-353, 2006.
- [10] Quellec, G., Lee, K., Dolejsi, M., Garvin, M.K., Abramoff, M.D., Sonka, M., “Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in sd-oct of the macula”, *IEEE Trans. on Medical Imaging*, vol. 29, pp. 1321-1330, 2010.

- [11] Chen, X., Niemeijer, M., Zhang, L., Lee, K., Abramoff, M.D., Sonka, M., “3D segmentation of fluid-associated abnormalities in retinal OCT: Probability constrained graph-search-graph cut”, *IEEE Trans. on Medical Imaging*, 31(8), pp. 1521-1531, 2012.
- [12] Ghorbel, I., Rossant, F., Bloch, I., Tick, S., Paques, M., “Automated segmentation of macular layers in oct images and quantitative evaluation of performances”, *Pattern Recognition*, vol. 44(8), pp.1590-1603, 2011.
- [13] Kafieh, R., Rabbani, H., Kermani, S., “A review of algorithms for segmentation of optical coherence tomography from retina”, *Journal of Medical Signals and Sensors*, vol. 3(1), pp. 45-60, 2013.
- [14] Lee, J.Y., Chiu, S.J., Srinivasan, P.P., Izatt, J.A., Toth, C.A., Farsiu, S., Jaffe, G., “Fully automatic software for retinal thickness in eyes with diabetic macular edema from images acquired by Cirrus and Spectralis systems”, *Investigative Ophthalmology & Visual Science*, vol. 54(12), pp. 7595-7602, 2013.
- [15] Liu, Y.Y., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J.S., Rehg, 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, pp. 748-759, 2011.
- [16] Srinivasan, P.P., Kim, L.A., Mettu, P.S., Cousins, S.W., Comer, G.M., Izatt, J.A., Farsiu, S., “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence

- tomography images”, *Biomedical Optical Express*, 5(10), pp. 3568-3577, 2014.
- [17] Anantrasirichai, N., Achim, A., Morgan, J.E., Erchova, I., Nicholson, L., “SVM-based texture classification in optical coherence tomography”, *In: IEEE Symposium on Biomedical Imaging*, pp. 1332-1335, 2013.
- [18] Albarrak, A., Coenen, F., Zheng, Y., “Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction”, *In: 17th Annual Conference in Medical Image Understanding and Analysis*, pp. 59-64, 2013.
- [19] Venhuizen, F.G., van Ginneken, B., Bloemen, B., van Grisen, M.J.P.P., Philipsen, R., Hoyng, C., Theelen, T., Sanchez, C.I., “Automated age-related macular degeneration classification in oct using unsupervised feature learning”, *In: SPIE Medical Imaging*, vol. 9414, pp. 941411, 2015.
- [20] Lemaître, G., Rastgoo, M., Massich, J., Sankar, S., Mériaudeau, F. Sidibé, D., “Classification of SD-OCT volumes with LBP: application to DME detection”, *In: Ophthalmic Medical Image Analysis Workshop (MICCAI)*, 2015.
- [21] Schmitt, J.M., Xiang, S., Yung, K.M., “Speckle in optical coherence tomography”, *Journal of Biomedical Optics*, vol. 4(1), pp. 95-105, 1999.
- [22] Buades, A., Coll, B., Morel, J.M., “A non-local algorithm for image denoising”, *In: IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 60-65, 2005.

- [23] Murphy, K.P., “Machine learning: a probabilistic perspective”, *MIT Press, Cambridge, MA.*, 2012.
- [24] Bishop, C., “Pattern recognition and machine learning”, *Springer*, 2006.