An optimization approach to segment breast lesions in ultra-sound images using clinically validated visual cues

**** ********[*1], ********* ********[1,2], **** *****[2], and ******* **********[1]

[1] **********, ********** ** *********, ** *** ** ** ********,
******* ** *******, ******;
****.*******@***********.**
[2] *******, *********** ** ******, ****** *********, ******* **, ******* ******,
*****

Abstract. THIS IS THE ABSTRACT

Keywords: kw1, kw2

1 Introduction

Eye diseases such as Diabetic Retinopathy (DR) and Diabetic Macula 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 [13]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over 300 M people worldwide by 2025 [18]. 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 (CAD) systems devoted to ophthalmology, have been developed focusing on the automatic analysis of fundus images [1, 16]. 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 [7]. 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 $6\,\mu m$ [4].

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 [5, 8]. Few works have addressed the specific problem of DME and its associated features detection from OCT images. Quellec *et al.* proposed a method for the identification of fluid-filled regions in SD-OCT images of the macula based on texture features extracted in the pre-segmented retinal layers [11].

The authors in [15] 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.* also proposed a method for OCT images classification using the Bag-of-Words (BoW) models [17]. 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, and 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.

The most similar work to ours is the work of Liu *et al.* who proposed a method for macular pathology detection in OCT images using Local Binary Patterns (LBP) and gradient information as attributes [9]. The method starts by aligning and flattening the images, then a 3-level multi-scale spatial pyramid is created and edge and LBP histograms are extracted in each block at every level of the pyramid. All 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 this paper, we propose a method for automatic identification of patients with DME versus normal subjects by classifying the OCT volumes. Our method is based on LBP features to describe the texture of OCT images and dictionary learning using the BoW models [14]. However, our method do not rely on key-points detection as opposed to the work of Venhuizen *et al.* who also employed the BoW models [17]. We rather divide the images into local patches and extract a dense set of LBP descriptors. We also use the entire OCT volume and extract 3D-LBP features to describe the volume, which is different from the work of Liu *et al.* who classified only the foveal scan for each patient [9].

This paper is organized as follows. Section **??** describes the features extraction methodology and the classification approach based on the BoW models. Experiments and results are discussed in Sect. **??** and Sect. **??**, respectively. Conclusions and avenue for future directions are drawn in Sect. **??**

2 Materials and Methods

This section offers a general description of the methodology proposed for OCT volume classification, whereas further details of some elements involved in the methodology are found as subsections.

The proposed method, as well as, its experimental set-up are outlined in Fig. 1. The methodology is formulated as a standard classification procedure. The available dataset with its acompaining Ground Truth (GT) are divided into training $(S1, l1)$ and testing $(S2, l2)$. The final goal is to represent $S1$ and $S2$ in the feature space F by supplying $(sxF, l1)$ as a training to a classifier, using the trained classifier to estimate $l2$ from $S2xF$ and comparing the estimation with the GT. To do so, the images forming the OCT volumes are preprocessed using Non-Local Means (NL-means) algorithm [3]. This algorithm preserve important details and textures of the original image, while reducing the noise. The mapping stage is used to determine a discrete set of elements (or structures) Z which is used for representing the volume $sinS$. The feature detection stage correspond to measurements done in $G(Z)$ used for representing s in terms of ZxG. This mapping and feature detection steps can be found as a single-steps in the literature. The feature extraction procedure combines the elements in Z and its measurements $G(Z)$ to create the final feature space F and project s on it.

The design choices are all illustrated in Fig. 1 and discussed further in this section. The work here presented does not discuss in detail neither the mapping, nor the adopted classifier, further than this lines. As a possible mappings, for representing the volumes, 2D image slices of the volume and 7x7x7sliding volumes, have been considered. As a classifier, a Random Forestusing 100 trees, has been considered.

2.1 Data

– cross-validation

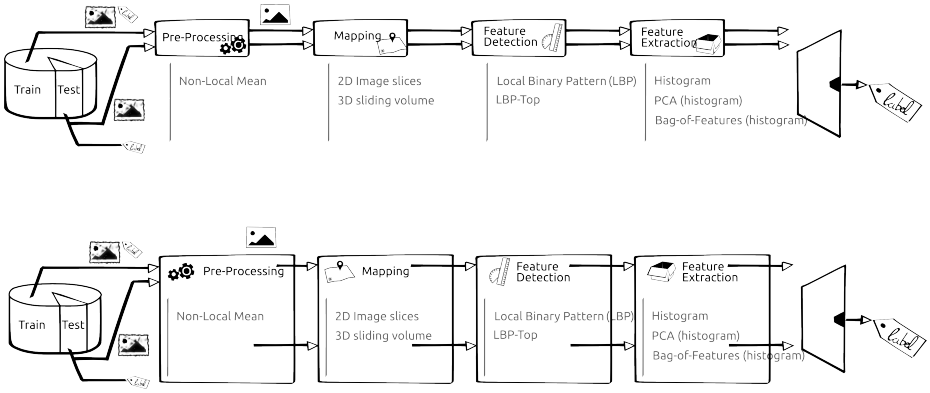


Fig. 1. Machine learning classification basic scheme

- our dataset
- DUC dataset

For evaluation purposes, the results have been cross-validated, by splitting the data in training and testing using a Leave-One-Patient Out (LOPO) strategy. In this manner for each round a pair dce,normal has been selected to be used as the round test set, while the rest of the dataset has been used as a training. Doing the cross validation in this manner, has the limitation that despite the fact that the results are robust due to the cross validation, no results variance can be reported. However, and despite this limitation, LOPO has been choose due to the reduced amount of OCT volumes available.

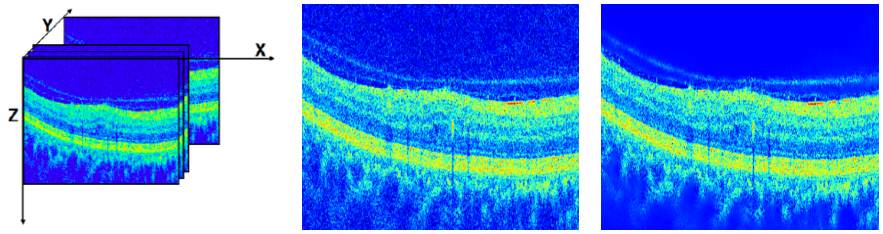
The dataset blablablabal...The duc dataset blabla bla...

2.2 Image pre-processing

OCT images are known to be affected by a speckle noise [12]. Subsequently, NL-means [3] filtering has been previously successfully used in Ultra-Sound (US) images to filter similar noise [6] and is used in our framework to denoise each B-scan (i.e. y axis) of the OCT volumes (see in Fig. 2(a)). NL-means filtering offers the advantage to use all the possible self-predictions that the image can provide rather than local or frequency filters such as Gaussian, anisotropic or Wiener filters [3]. The different parameters were empirically tested and fixed such that the patch size, the search window and the filtering parameter h were set to (15×15), (35×35) and 0.4, respectively. An example of filtering using NL-means filter on OCT image is depicted in Fig 2(b) and Fig. 2(c).

2.3 Features extraction

Need to write something here !!!



(a) (b) (c)

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

Low-level features are extracted considering the whole volume using LBP and 3D-LBP descriptors. LBP is a discriminative rotation invariant feature descriptor proposed by Ojala et al. [10]. LBP descriptor encodes the intensity differences of a central pixel (g_c) with its neighboring pixels (g_p), within in a defined neighborhood of radius R. The differences are encoded in terms of binary patterns as in Eq. 1:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{1}$$

where $s(a) = 1$ if $a \geq 0$, and $s(a) = 0$ otherwise. P is the number of sampling points in the circle of radius R.

The binary patterns are calculated for each pixel in the given image and their histogram defines the final descriptor. The LBP histograms are computed for each slice of the volume and are concatenated into a single histogram. This forms the first low-level feature. The second low-level descriptor is defined in a similar manner as the first one. However principal component analysis (PCA) is applied to the concatenated histograms in order to reduce the dimension.

For the third low-level descriptor, since the OCT data is a 3D volume, following the approach of Zhao *et al.* [19], we extract 3D-LBP by considering three orthogonal planes, XY, XZ and YZ. Note that X, Y, and Z are respectively the horizontal, vertical and depth direction of the OCT volume as shown in Figure **??**(a). LBP patterns are computed for each of the three planes, and the obtained three histograms are concatenated into a final 3D-LBP descriptor.

High-level features - are extracted using bag of words (BoW) approach which is a feature representation technique based of creating a visual dictionary, or codebook, from a set of low-level features [14]. To do so, the OCT images are divided into local patches and LBP histograms are computed for every local patch. This set of LBP histograms is then used to create a codebook using K-means clustering. If we define K clusters in the feature space, then the visual dictionary will contain K words each one being the center of one cluster. After

creating the codebook, each of the training example is represented as a histogram of size K obtained by calculating the frequency of occurrences of each of the K words in the features extracted from the training example. Note that in the 2D case, each slice is divided into patches of size $N \times N$ and we extract 2D-LBP from each patch, while in the 3D case, the volume is divided into $N \times N \times N$ patches and 3D-LBP histograms are computed. In our experiments in Section 3, we set $N = 7$, and vary the size of the codebook K in the range $\{2, 4, 8, 16, 32, 64, 100\}$.

2.4 Classification

Random Forest is an ensemble of decision trees and was introduced by [2]. The ensemble uses each tree to predict an output and finalize 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.

3 Experiments and Validation

3.1 Datasets

SERI - dataset contains 32 OCT volumes (16 DME and 16 normal). This dataset was acquired in **Institutional Review Board-approved protocols** using CIRRUS TM (Carl Zeiss Meditec, Inc, Dublin, CA) SD-OCT device. 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. This dataset was acquired by our colleagues from Singapore Eye Research Institute (SERI).

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

3.2 Validation

For evaluation purposes, the results have been cross-validated, by splitting the data in training and testing using a LOPO strategy. In this manner for each round a pair DME, normal has been selected to be used as the round test set,

while the rest of the dataset has been used as a training. Doing the cross validation in this manner, has the limitation that despite the fact that the results are robust due to the cross validation, no results variance can be reported. However, and despite this limitation, LOPO has been choose due to the reduced amount of OCT volumes available.

3.3 Experiment

The SERI dataset is provided in complete OCT volumes by 512×1024×128 dimensions. Using this dataset, first the three low-level features such as LBP, LBP+PCA and LBP histogram from Three Orthogonal Planes (LBP-TOP) are extracted. The rotation invariant uniform ($riu2$) descriptors are calculated with the P number of 8, 16 and 24 for the radius if 1, 2 and 3 respectively. The features are classified using RF with 100 tress. Table 1 shows the relative results for $8riu2$, $16riu2$, $24riu2$ and their combination $8riu2 + 16riu2 + 24riu2$. The results are presented in terms of Sensitivity (SE) and Specificity (SP) percentages.

The second experiment is carried out using high-level features and BoW approach, on SERI dataset. The first high-level feature LBP+BoW is obtained by applying BoW with 32 visual-words on the previously low-level LBP features (applied on each B-scan). The second and third high-level descriptors are obtained using a dense approach by applying the Sliding Window (SW) of size (7×7) on each B-scan and SW of size (7×7×7) to the whole volume respectively. LBP+BoW+SW represent the second high-level feature where the 2D-LBP features are extracted for each sliding window on each B-scan and the visual-words are selected from the pool, consisting of their histograms. The third high-level feature, LBP-TOP+BoW+SW, is defined using LBP-TOP. By using the sliding window the 3D-LBP features are extracted for each patch. Same as previous experiment with low-level features, the descriptors are calculated with the P number of 8, 16 and 24 for the radius if 1, 2 and 3 respectively. The obtained results of this experiment are illustrated in Tab. 2.

In order to compare our proposed framework the third experiment is carried out using the subsection of Duke dataset [15]. The OCT volumes provided by this dataset are of different volume size, cropped and denoised by the method of authors choice. Subsequently only the second experiment with high-level features and low-level LBP-TOP features comply with these requirements. The number of visual-words and the size of SW for 2D and 3D features are the same than the previous experiment. The 2D and 3D LBP features are extracted with P number of 8, 16 and 24 for the radius if 1, 2 and 3 respectively. The obtained results for this experiment are shown in Tab. 3.

Table 1. Obtained results with LBP, LBP+PCA and LBP-TOP features and RF with 100 trees on SERI dataset

Features	LBP		LBP+PCA		LBP-TOP	
	SE	SP	SE	SP	SE	SP
8^{riu2}	43.75	43.75	50.00	68.75	56.25	62.50
16^{riu2}	37.50	50.00	68.75	56.25	**87.50**	**75.00**
24^{riu2}	50.00	62.50	56.25	37.50	68.75	68.75
$\{8, 16, 24\}^{riu2}$	37.50	56.25	**68.75**	**68.75**	**81.25**	**81.25**

Table 2. Obtained results for LBP+BoW, LBP+BoW+SW, LBP-TOP+BoW+SWfeatures and RF with 100 trees on SERI dataset. The BoW is computed with 32 visual words for all the experiments

Features	LBP+BoW		LBP+BoW+SW		LBP-TOP+BoW+SW	
	SE	SP	SE	SP	SE	SP
8^{riu2}	50.00	81.25	75.00	87.50	62.50	68.75
16^{riu2}	57.50	68.75	81.25	75.00	56.25	37.50
24^{riu2}	50.00	50.00	68.75	62.5	37.50	43.75

Table 3. Obtained results for LBP+BoW+SW, LBP-TOP+BoW+SW, LBP-TOP features and RF with 100 trees on Duke dataset. The BoW is computed with 32 visual words for all the experiments

Features	LBP+BoW+SW		LBP-TOP+BoW+SW		LBP-TOP	
	SE	SP	SE	SP	SE	SP
8^{riu2}	80.00	86.67	80.00	86.67	80.00	93.33
16^{riu2}	86.67	100.00	86.67	86.67	73.33	86.67
24^{riu2}	93.33	86.67	60.00	80.00	73.33	86.67

3.4 Results

4 Conclusions

References

1. Abramoff, M.D., Garvin, M.K., Sonka, M.: Retinal image analysis: a review. IEEE Review Biomed. Eng. 3, 169–208 (2010)
2. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
3. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 2, pp. 60–65. IEEE (2005)
4. 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 ophtalmic imaging. Arch. Ophtalmol. 123(12), 1715–1720 (2005)
5. Chiu, S.J., Li, X.T., Nicholas, P., Toth, C.A., Izatt, J.A., Farsiu, S.: Automatic segmentation of seven retinal layers in sd-oct images congruent with expert manual segmentation. Optic Express 18(18), 19413–19428 (2010)
6. Coupe, P., Hellier, P., Kervrann, C., Barillot, C.: Nonlocal means-based speckle filtering for ultrasound images. IEEE Trans Image Process 18(10), 2221–2229 (Oct 2009)
7. Early Treatment Diabetic Retinopathy Study Group: Photocoagulation for diabetic macular edema: early treatment diabetic retinopathy study report no 1. Arch. Ophtalmol. 103(12), 1796–1806 (1985)
8. Kafieh, R., Rabbani, H., Abramoff, M.D., Sonka, M.: Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map. Medical Image Analysis 17, 907–928 (2013)
9. Liu, Y.Y., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J.S., M., R.J.: 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, 748–759 (2011)
10. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(7), 971–987 (2002)
11. 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 29, 1321–1330 (2010)
12. Schmitt, J.M., Xiang, S., Yung, K.M.: Speckle in optical coherence tomography. Journal of biomedical optics 4(1), 95–105 (1999)
13. Sharma, S., Oliver-Hernandez, A., Liu, W., Walt, J.: The impact of diabetic retinopathy on health-related quality of life. Curr. Opin. Ophtalmol. 16, 155–159 (2005)
14. Sivic, J., Zisserman, A.: Video google: a text retrieva approach to object matching in videos. In: IEEE ICCV. pp. 1470–1477 (2003)
15. Srinivasan, P.P., Kim, L.A., Metttu, 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), 3568–3577 (2014)

16. Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J., al Diri, B., Cheung, C., Wong, D., Abramoff, M., Lim, G., Kumar, D., Burlina, P., Bressler, N.M., Jelinek, H.F., Meriaudeau, F., Quellec, G., MacGillivray, T., Dhillon, B.: Validation retinal fundus image analysis algorithms: issues and proposal. Investigative Ophthalmology & Visual Science 54(5), 3546–3569 (2013)
17. Venhuizen, F.G., van Ginneken, B., Bloemen, B., van Grisven, M.J.P.P., Philipsen, R., C., H., Theelen, T., Sanchez, C.I.: Automated age-related macular degeneration classification in oct using unsupervised feature learning. In: SPIE Medical Imaging. vol. 9414, p. 94141l (2015)
18. Wild, S., Roglic, G., Green, A., Sicree, R., King, H.: Global prevalence of diabetes estimates for the year 2000 and projections for 2030. Diabetes Care 27(5), 1047–1053 (2004)
19. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29(6), 915–928 (2007)