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. In United States, health care and associated costs related to eye diseases are estimated at almost $500 M [14]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over 300 M people worldwide by 2025 [19]. 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 have been developed focusing on the automatic analysis of fundus images [1, 17]. 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 [9]. 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 of $3.5\,\mu m$ to $6\,\mu m$ [7].

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, 10]. Few works have addressed the specific problem of DME and its associated features detection from OCT images. Quellec *et al.* [13] proposed a method for the identification of fluid-filled regions in SD-OCT images of the macula based on texture features extracted in the segmented retinal layers. The authors in [16] proposed a classification method for the detection of DME versus Age-related Macular Degeneration (AMD) and normal OCT images. 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. Then, Histogram of Oriented Gradients (HOG) are extracted in each image of a volume and a linear Support Vector Machines (SVM) is used for classification. On a dataset of 45 patients containing 15 normal subjects, 15 DME patients and 15 AMD patients, the methods achieved a correct classification of 100%, 100% and 86.67% for AMD, DME and normal cases respectively. Venhuizen2015 *et al.* [18] also proposed a method for OCT images classification using the Bag-of-Words (BoW) models. The method starts with the selection of interest points in each individual B-scan by keeping the points corresponding to the top 3% vertical gradient values. Then, a 9×9 patch of intensity values is extracted around each selected interest point, and PCA is applied to reduce the dimension of every patch from 81 (9×9) to 9. All extracted patches are used to create a codebook using k-means clustering, and the obtained codebook from training is used to represent each OCT volume as a patch 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.* [11] who proposed a method for macular pathology detection in OCT images using Local Binary Patterns (LBP) as features. The method starts by aligning and flattening the images, then a 3-level multi-scale spatial pyramid is created and LBP histograms are extracted in each block at every level of the pyramid. All obtained LBP histograms are concatenated into a global descriptor whose dimension is reduced using Principal Component Analysis (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 approach [15]. However, we do not based on interest points selection as opposed to the work of Venhuizen *et al.* [18] who also employed the BoW approach. 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.* [11] who classified only the foveal scan for each patient. We will show in the experiments, Section 4, that using the 3D-LBP descriptor provides better classification performances than extraction LBP in each individual B-scan.

This paper is organized as follows. In Section 2, we describe the features extraction methodology and the classification approach based on the BoW method. Section 3 shows experimental results using two different datasets and comparison with another approach. Finally, the paper ends with concluding remarks in Section 4.

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 Mean (NL-mean) algorithm [5]. 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
- our dataset

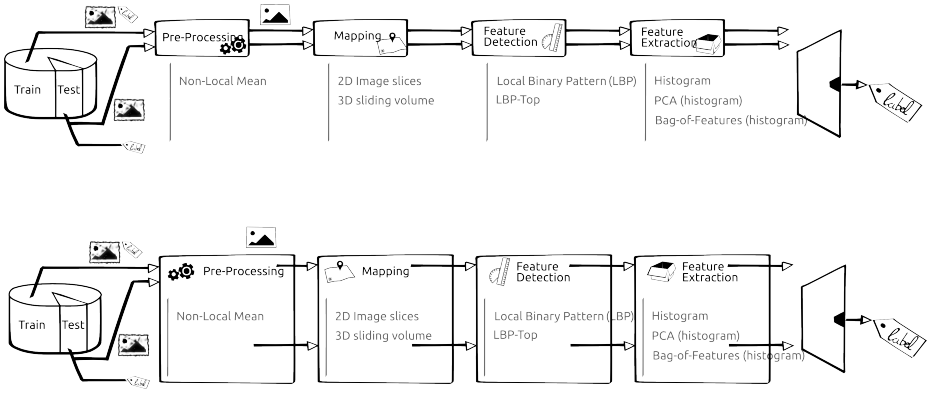

Fig. 1: Machine learning classification basic scheme

– 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

The pre-processing stage in the proposed methodology applies an image denoising method to reduce the speckle noise in OCT images. Since image details and texture of the original image are needed by the following stages in the method, NL-mean algorithm [5] is used. NL-mean algorithm has the advantage to use all the possible self-predictions that the image can provide [5] rather than local or frequency filters such as Gaussian, anisotropic or Wiener filters. Figure .. shows an OCT slice before and after denoising

2.3 Features extraction

Need to write something here !!!

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. [12]. 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.

The LBP features are extracted from each slice of the volume and their histograms are concatenated to build the first low-level descriptor. The second low-level descriptor is defined in a similar manner as the first one. However PCA is applied to the concatenated histograms in order to reduce the dimensions. The first and second low-level descriptors are obtained using the 2D LBP descriptor. However the third low-level feature is obtained using 3D-LBP (Local Binary Pattern histogram from Three Orthogonal Planes (LBP-TOP)). Zhao et al. [20] proposed LBP-TOP as a dynamic texture descriptor. This descriptor is an extension to normal LBP while it considers texture descriptors along the temporal domain. LBP-TOP considers the LBP pattern in three orthogonal planes (see Fig. 2), XY, XT and YT. The obtained LBP patterns from the three planes are concatenated to form the final descriptor. The three low-level descriptors are calculated with the P number of 8, 16 and 24 for the radius if 1, 2 and 3 respectively

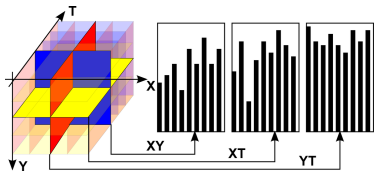

Fig. 2: LBP-TOP framework, the image is taken from [3]

High-level features - are extracted using Back-of-Features (BoF) approach which is illustrated in Fig. **??**. After de-noising the images by non-local-mean approach, a patch detection step is carried out by identifying informative regions of the data. These patches can either be sampled densely or sparsely. The dense sampling extracts more information regarding the object appearance. However, it might retain redundant features. In the contrary sparse sampling is based on detecting salient key points from the most informative regions of the volume.

Our proposed BoF approach is based on the first strategy. In this stage 2D-LBP (7×7) and 3D-LBP-TOP patches (7×7×7) are extracted for each patient. Defining N as the number of selected patches for each volume and d as the number of feature dimensions, then each volume is characterized by a $N \times d$ feature matrix (see Fig. **??**, "feature extraction"). The next step in bag of features consists of building the dictionary of "visual-words". All the feature matrices of the training set are concatenated together and k-means clustering method (see Fig.**??**, "clustering") is used to define the "visual-words". K-means is an iterative algorithm which finds k centroids by alternating assignment and update steps. The assignment steps is based on L_2 norm (Euclidean) distance. Different initialization methods can be used in order to assign the initial k clusters [6], here the initial k clusters are selected based on greedy k-means++ method [2]. **Depending on the framework and application, different choices of the number of "visual-words" (number of k clusters) can be made. In our framework, the number of clusters are varying in the range of [2 4 8 16 32 64 100]. SHOULD CHNAGE** Finally, the probability distribution (e.g., histogram) of the "visual-words" of each feature matrix is computed (feature quantization) and is used to feed the classifier to be trained. In the prediction stage, the histogram of the feature matrix corresponding to the new volume is computed using the previously learned dictionary and, finally, it is classified by the previously trained classifier.

2.4 Classification

Random Forest is an ensemble of decision trees and was introduced by [4]. 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 Data sets

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

Duke - data sets published by Srinivasan et al. [16] 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 dataset consists 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.

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-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 Planes (SE) and Specificity Planes (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 Planes (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 [16]. 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.

3.4 Results

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+SW features 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	
	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	-	-

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

Table 3: Obtained results for LBP+BoW+SW, Lbp-Top+BoW+SW 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	
	SE	SP	SE	SP
8^{riu2}	80.00	86.67	80.00	86.67
16^{riu2}	86.67	100.00	86.67	86.67
24^{riu2}	93.33	86.67	60.00	80.00

3. B. Jiang, M. F. Valstar, B.M., Pantic, M.: A dynamic appearance descriptor approach to facial actions temporal modelling. IEEE Transactions on Cybernetics 44(2), 161–174 (2014)
4. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
5. 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)
6. Celebi, M.E., Kingravi, H.A., Vela, P.A.: A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Systems with Applications 40(1), 200–210 (2013)
7. 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)
8. 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)
9. 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)
10. 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)
11. 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)
12. 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)
13. 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)
14. 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)
15. Sivic, J., Zisserman, A.: Video google: a text retrieva approach to object matching in videos. In: IEEE ICCV. pp. 1470–1477 (2003)

16. 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)
17. 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)
18. 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)
19. 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)
20. 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)