AMD Classification in Choroidal OCT using Hierarchical Texton Mining

Dafydd Ravenscroft[1] Jingjing Deng[1] Xianghua Xie[1]
Louise Terry[2] Tom H. Margrain[2] Rachel V. North[2] Ashley Wood[2]

1. Department of Computer Science, Swansea University, United Kingdom
http://csvision.swansea.ac.uk
2. School of Optometry and Vision Sciences, Cardiff University, United Kingdom

Abstract. In this paper, we propose a multi-step textural feature extraction and classification method, which utilizes the feature learning ability of Convolutional Neural Networks (CNN) to extract a set of low level primitive filter kernels, extracts spatial information using clustering and Local Binary Patterns (LBP) and then generalizes the discriminative power by forming a histogram based descriptor. It integrates the concept of hierarchical texton mining and data driven kernel learning into a uniform framework. The proposed method is applied to a practical medical diagnosis problem of classifying different stages of Age-Related Macular Degeneration (AMD) using a dataset comprising long-wavelength Optical Coherence Tomography (OCT) images of the choroid. The results demonstrate the feasibility of our method for classifying different AMD stages using the textural information of the choroidal region.

1 Introduction

AMD is a progressive eye disease which is the leading cause of vision loss in the developed world [9]. It is a progressive disease; partial vision loss is minimal at an early stage but it can develop to one of two end stages: dry (geographic atrophy) or wet (neovascular AMD) [16]. Early and accurate diagnosis with effective treatment can prevent developing into an irreversible AMD stage and minimize the damage to retina layer and choroidal region. Fig. 1 shows examples of the choroid OCT images in different AMD categories. It is our hypothesis that the pathological progression of AMD has an effect on the shape and texture of the choroidal region due to changes in the choroidal vascular structure and this information is embedded in the OCT images. However, this hypothesis has not been fully studied due to the fact that the images obtained at the choroidal regions are very noisy and exhibit large variations from patient to patient. Acquiring a large cohort of patient data with labeled choroidal region is also a challenge, see examples in Fig. 1. As the first step towards automated diagnosis, in this work we study the feasibility of applying texture analysis to AMD classification using only choroidal regions. Texture analysis is an important problem in computer vision with applications in surface defection discovery [21], segmentation [8], and image-based medical diagnosis [1]. It involves extracting features

from an image, based on its textural appearance, which can then be used for classification. Hand-crafted features are designed for the specific problem which highlight the discriminative pattern for visual recognition task. For example, Gabor filter banks have been successfully used in medical imaging [3, 15] outperforming other methods by extracting impulse responses from different scales and orientations. There are similarities between its feature descriptors and the stimulation mechanism of human visual system [4].

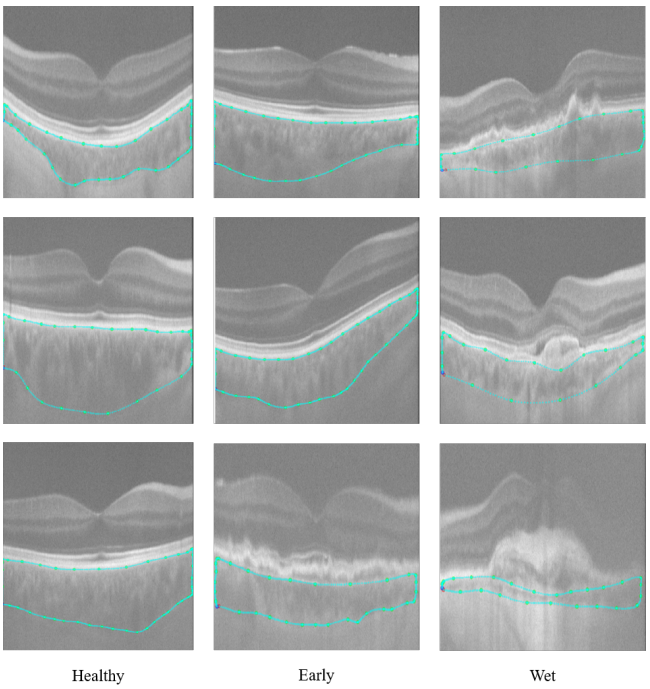

Fig. 1. Examples of choroidal OCT scans for healthy, early AMD and wet AMD.

Priya et al. [13] proposed a machine learning approach for classifying AMD using color retinal photographs, where the hand-crafted features were extracted, such as retinal vessel density and average retinal vessel thickness. In [17] the abnormality measurements of Retinal Pigment Epithelium (RPE) layer, bubbles in Retinal Nerve Fiber Layer (RNFL) complex region and outer RNFL region near RPE layer were used to construct a binary discriminative model that classifies the images into AMD and Diabetic Macular Edema (DME). Farsiu et al. [7] used the thickness measurement of RPE Drusen Complex (RPEDC) and Total Retina (TR) as features to build a generalized linear model for AMD classification. Similarly, Koprowski et al. [10] proposed a random forests based method to classify choroidal OCT images into predefined clinical conditions by extracting high level features, such as number of detected objects and average position of

the centre of gravity, from low level texture information. These high level features heavily rely on high quality detection and segmentation results of blood vessel and other anatomical structures, which normally requires extra human resource. [2] uses hand crafted Gabor filters for feature extraction with machine learning techniques being used to classify stages of AMD. Designing hand-crafted filters is a time consuming and challenging task. More often than not, such techniques do not adapt well with data and also can not readily be implemented when input images are of inconsistent size and shape. There are three major difficulties of applying traditional hand-crafted filters to AMD classification problem using choroidal OCT images. Firstly, variations of local textural appearance within the choroid are very subtle and nearly random in high frequency bands, while such variations change across slices in low frequency bands, i.e. designing feature extractors that are able to capture the representative patterns is a non-trivial task. Secondly, pathological effects of AMD are not homogenous in the choroidal regions. Textural features are thus highly non-uniform. Thirdly, the choroid sections are irregular in size and shape across different subjects resulting in feature descriptors of arbitrary length.

Current advances in Deep Neural Networks (DNN) demonstrate superior performances in visual recognition tasks. These methods have shown effective in joint end-to-end learning for both feature representation and decision making. In this paper, we present a hierarchical texton mining method in order to classify choroidal OCTs. CNNs are used to learn low level textural features, which are then generalized as textons. The local binary patterns (LBPs) of generalized textons represent the spatial distribution of the primitive textural information and create higher level feature descriptors. Given these automatically extracted texton descriptors, the OCT choroidal images are then classified into three AMD disease stages of increasing severity. Our major contributions are two fold as follows: (1) We combine hierarchical texton learning with data driven kernel learning. Textural feature mining is a hierarchical process that involves from the low level primitive feature extraction to mid level cognitive representation and high level generalization. The primitive pattern generally is subtle and difficult to hand-craft whereas in our work we propose an automatic method that learns those image level patterns through a supervision task which can then be further generalized with a clustering method to retain the commonality. For the mid-level feature, we believe that the spatial arrangement of those learned at low level textons reflects the localized structural information, therefore LBPs are used. Furthermore, the distribution of the LBP response is computed to represent the distribution over the whole region of interest which allows generalization for the classification task. (2) We present an AMD dataset consisting of 75 scans split into 3 pathological categories, each with 25 patients. Long-wavelength OCT imaging technique that produces the scans, and the annotation of the choroidal regions. The experimental results show it is possible to classify AMD into different stages for individual patients using the textural information from the choroidal regions in OCT. The rest of the paper is organized as follows: Section 2 presents our proposed method for data driven filter training,

feature extraction, texton generalization, and classification; Section 3 presents and discusses the dataset and experimental results. Our concluding remarks and discussions on future work are presented in Section 4.

2 Proposed Method

Mining discriminative feature descriptors is the key component of designing an efficient visual recognition model for AMD stage classification using choroidal OCT images. The primitive low-level features are automatically learned using a CNN, where the convolutional filter kernels are learned via a supervised discriminative training procedure. The textons can then be inferred by clustering the image responses of learned filter kernels, where the cluster centers form the texton dictionary. The spatial distribution of mined textons is extracted using LBPs. Patch-based local textural features are then generalized to regional feature descriptors using histograms over the Region of Interest (RoI), which provide high level features as a representation of the local distribution of the LBPs. Supervised classification is then carried out using a range of machine learning techniques to classify input images into different AMD stages based on the texton feature descriptors that are mined hierarchically.

2.1 Data Driven Primitive Textural Feature

CNNs combine both feature representation mining and supervised discriminative learning into a unified end-to-end training framework, which has been widely adopted in recent years and produced some of the top results for many machine vision problems [11, 14, 18]. The convolutional filter kernels that are learned through the supervised training procedure can be considered as discriminative features and can be generalized further. In this work, a CNN is used to hierarchically learn textural features in a texton mining framework and the CNN kernels produce low level data driven textural features. In convolutional layers, a bank of locally receptive filters convolve across the input image to form visual evidences for prediction layers at the forward pass stage. At the backward pass stage, these filters are automatically optimized via back-propagating the prediction error that is calculated in the previous round of forward pass. Fully connected layers are also included in the network, where all nodes from one layer are connected to all nodes in the next with weightings updated in the same way. This allows pertinent localized features to be more easily identified.

Fig. 2 and Table 1 show the architecture details of the proposed CNN. Due to the irregular shape of the choroidal region (see Fig. 1), it is difficult to extract large local patches without including other structures. As such, the local patches with size of 48×48 pixels are cropped randomly from the choroidal regions with overlap. In order to interpret the low level textural features learned through the discriminative task, only one convolutional layer is used. The networks that are used for natural image recognition tasks generally have smaller kernel sizes, such as 3×3, as natural images have much sharper corners and higher contrast

compared to medical imaging. In our case, 40 filter kernels with size of 9×9 are used in order to identify the discriminative patterns in low frequency bands.

Table 1. The parameters of the proposed CNN architecture.

No	Type	Parameter
0	Input	48×48×3 images scaled to [0,1]
1	Conv.	40 9×9 filters with stride 1
2	ReLU	Rectified linear unit
3	F.C.	Fully connected with 128 outputs
4	ReLU	Rectified linear unit
5	F.C.	Fully connected with 128 outputs
6	ReLU	Rectified linear unit
7	F.C.	Fully connected with 3 outputs
8	Softmax	Softmax probability for multi-classes

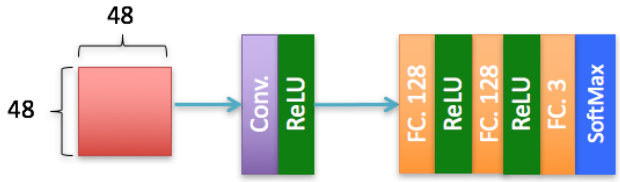

Fig. 2. The network architecture of the proposed CNN.

2.2 Spatial Texton Descriptor

In this work, we introduce additional steps to explore both statistical and spatial distribution of the primitive texture feature that are produced from CNN. Fig. 3 shows the examples of learned filter kernels from the convolutional layer. The bank of learned filters are, in turn, convolved across the images of the extracted choroidal regions. Convolving the filters across the images is a form of linear filtering which produces a map of filter responses of the same dimension as the image. The texture is modeled by the distribution of filter responses, these can be represented by textons (cluster centers) which can be used to create a texture model [20]. K-Means clustering is used to develop the set of textons,

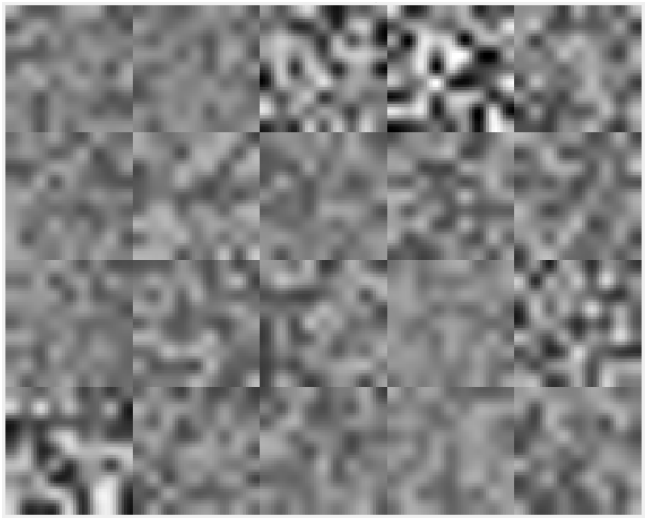

Fig. 3. Examples of self-learned filter kernels using CNN.

which can be used to label all filter responses with each observation assigned to the partition with the closest mean. The textons group the textural features into a compact representation via examining the statistical distribution of filter response, which removes the subtle variations at high frequency bands. These textons are more robust than the raw filter responses. However, for OCT retina image (See Fig. 1), the primitive texture appearances of choroidal region learned from CNN are rather noisy and do not well form structural patterns (See Fig. 3). In order to overcome this difficulty, the spatial distribution of these mined textons is introduced which represents the patterns of local arrangement of mined texton. In the spatial domain, the local correlation of those textons can be further genealized in a hierarchical manner, which are more representative and informative for classification task. In this work, LBPs are used to represent the spatial distribution of textons. Spatial features look for texture elements, known as texture primitives, which are extracted to create a representation that maps their regional locations. It looks for regular or repeated patterns of texture elements in the image, and learn spatial information by comparing each pixel to its neighbors and assigning each a binary value [12]. In this work, a texture unit is the central value in a 3×3 neighborhood and is represented by the 8 elements that surround it. Each is assigned a binary value with the centre pixel acting as a threshold and are multiplied by predefined weightings based on the pixel location. The results of the eight neighboring pixels are summed and this value is assigned to the texture unit. A value for each pixel is calculated meaning the response output has the same dimensions as the input.

2.3 Regional Texton Generalization

As the CNNs are trained only on relatively small patches extracted from the choroidal regions, a higher level descriptor is required to make predictions on image level. We thus convolve the learned CNN filters across the entire choroidal regions. Note that this is different to conventional texton learning, where kernel filters are pre-defined and static. The kernels in our method is data driven and dynamic. It is also worth noting that the choroidal regions vary in size and shape, cf. Fig 1, which leads to varied length of LBP feature vectors. In order to train discriminative classifiers, it is desirable to obtain feature vectors of uniform length. Thus, a regional texton generalization is carried out via computing the histogram of LBP feature vectors of the annotated region. The histogram based descriptors produce a representation of the distribution of responses which also improves the generalization ability. For each of the filters a histogram is calculated with each LBP response being grouped into one of 59 bins depending on its value. The number of bins was calculated using the formula $P \times (P-1)+3$ where P is the number of neighbors, 8. The histogram descriptor is calculated for each of the different filters independently, and the results are concatenated to produce one feature vector for each image. Fig. 4 shows examples of histogram descriptors of different AMD classes produced by the top 3 filter kernels. In Fig. 4, it is clear that the differences between 3 AMD classes are distinct, although healthy and early AMD classes show some similarities between the histogram distribution. This is consistent with the clinical interpretation as visual appearance change is more gradual between healthy and early than between early and wet.

2.4 Supervised Classification

To evaluate the discriminative power of proposed regional texton descriptors, both two classifiers are employed to distinguish different AMD stages, such as: Neural Networks (NN), and Random Forests (RF). The traditional fully connected NN is used in this work to build a supervised classifier using the mined texton feature descriptor. RF is an ensemble method which combines a number of weak classifiers to create an accurate predictive model. It averages the results of multiple decision trees, each of which consists of a set of recursive binary splits with leaf nodes assigning a probability of the training sample belonging to each class. The variable importances are evaluated during the training process through permutation, which ranks the discriminative power of learned filter kernels.

3 Experimental Result

3.1 Dataset

The dataset consists of 25 healthy eye scans from the control group, and 50 scans from AMD patients classified into one of two categories: early AMD and

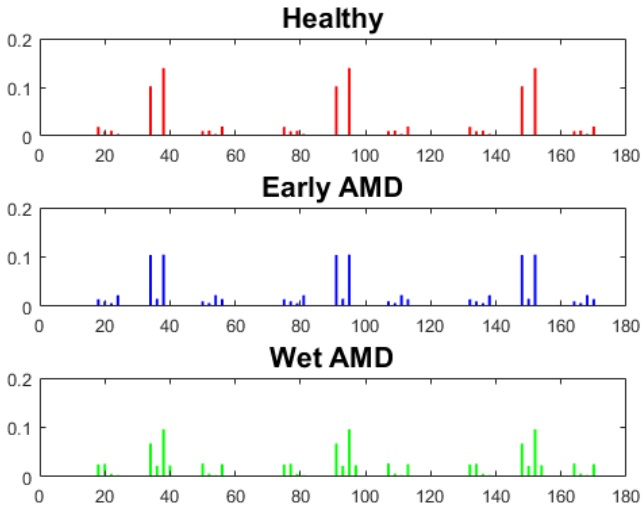

Fig. 4. Examples of **texton** descriptors of different AMD classes from top 3 filter kernels.

wet AMD. Therefore, for each category the dataset contains 25 eye scans. In order to obtain high quality images, the long-wavelength (1040nm) OCT imaging technique is used to provide sufficient light penetration into the choroid structure. For each eye, a volume of 512×1024×512 pixels is produced. Each eye has its axial eye length (AEL) measured, and the images were scaled accordingly; this was done to control for errors in image scaling [19]. All samples were collected by the same operator and classified by three experienced optometrists into the pathological categories. Classifications were made by examining the shape and appearance of the retina based on an adapted version of an accepted and widely used clinical classification system. We take these classifications to be the ground truth. In preprocessing, for each eye, the outline of the choroidal region was manually labelled on every tenth slice, hence the dataset consisted of over 3,800 labelled slices. Automatic image segmentation has had useful applications in medical applications [5, 6], but we chose manual segmentation for accuracy and consistency. Fig. 5 shows examples of labelled OCT scans of the three categories. From each image the closed curve created by the labels was extracted leaving just the choroidal layer for each slice.

3.2 Evaluation

The CNN was trained with weight decay of 5×10^{-4}, a batch size of 128 and was trained for 20 epochs with learning rates logarithmically spaced vectors between 10^{-2} and 10^{-5}. Patches of consistent dimension are extracted from

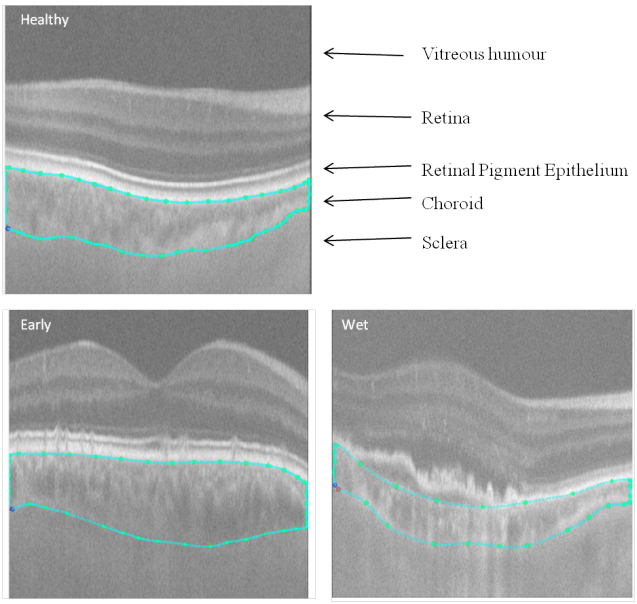

Fig. 5. Examples of labelled OCT scans for each of the three classes with visible signs of pathology within the retina.

the slices to train the network. Ten patches of 48×48 pixels are extracted from each annotated slice, providing over 500 patches per eye. Each patch is given the same classification as the slice to which it belongs. K-means clustering was computed using 10 cluster centres. LBP used a neighbourhood of 8 pixels for value calculations. The number of bins for the histogram of LBP responses is calculated as $(P \times (P-1)+3)$, where P is the number of of neighbours, resulting in 59 bins. A histogram is calculated for each of the 40 filters with the results concatenated to produce a feature vector of 2360 values for each image. Then each of the classifiers were applied independently. The random forest consisted of 50 random decision trees, and the neural networks contained two hidden layers with 200 and 40 nodes respectively. For each method of validation the training and testing process was iterated 10 times with the demonstrated results the combination of these.

To perform AMD classification, 10-fold and 2-fold cross validations were used. For N-fold cross validation the whole dataset was split into N randomly sampled, evenly sized groups with an equal numbers of slices from each eye. One subset was held for testing whilst the other nine were used for training. This training set was used for learning the filters in the CNN and to train the classifiers. Table 2 shows the result of using the kernel feature learned through CNN only, where on average 33.6% and 33.3% are achieved for 10-fold and 2-fold respectively, where the classification is dominated by the control group, and the prediction is nearly selected by random. Therefore, it strongly suggests that using the fea-

ture learned from CNN only is unable to distinguish between different stages of AMD. Table 3 shows the results of 10-fold classification for three classifiers. NNs and RFs achieved correct classification accuracies of 78.5% and 87.8% respectively. There was a significant difference between the accuracy achieved using NNs as the classifier compared to RFs. NNs perform well when learning hierarchical structures of features directly from the raw input image. However, we develop the feature descriptors through learned filters, spatial descriptors and histograms. As such, discriminative models which find a separation boundary between classes can be expected to outperform generalization models. The results of 2-fold cross validation of our proposed method are summarized in Table 4, where the respective prediction accuracies for the NNs and RFs were 75.0% and 85.2% . The accuracy was expected to decline across all classifiers due to the relative decrease in the size of the training set. However, a similar pattern occurs in which using the random forest as a classifier produces greater accuracy than the neural network. The hierarchical texton mining produces a more compact feature descriptor which enables a separable boundary to be found. In addition, the distinct accuracy differences between Tables 3, 4 and Table 2 show that the proposed feature descriptor improves the discriminative power by a large margin. From feature selection perspective, the primitive filter kernels shown in Fig. 3 is rather noisy and tends to appear random, however, in Fig. 4, the distributions of their responses are far more discriminative. The results suggest the feasibility of our approach for detecting textural changes in the choroid from which stages of AMD can be classified.

Table 2. Confusion matrix of 10-fold and 2-fold CNN **without proposed texton generalization** (%)

		Healthy	**Early AMD**	**Wet AMD**
10-fold	**Healthy**	**80.7**	**80.0**	**80.4**
	Early AMD	19.2	20.0	19.4
	Wet AMD	0.06	0	0.20
2-fold	**Healthy**	**100**	**100**	**100**
	Early AMD	0	0	0
	Wet AMD	0	0	0

4 Conclusions

In this paper we propose a machine learning approach for the classification of the stages of AMD using the textural appearance of OCT choroidal images. In traditional texture recognition techniques, hand-picked feature extractors such as Gabor filters or wavelets are used for feature extraction with the resultant feature descriptor being passed directly onto machine learning classifiers. This

Table 3. Confusion matrices of 10-fold cross validation **with the proposed feature descriptors** (%)

		Healthy	**Early AMD**	**Wet AMD**	**Avg.**
NN	**Healthy**	**74.3**	14.6	10.2	**78.5**
	Early AMD	16.3	**78.2**	6.8	
	Wet AMD	9.4	7.1	**83.0**	
RFC	**Healthy**	**84.2**	8.2	4.6	**87.8**
	Early AMD	9.9	**87.5**	3.5	
	Wet AMD	5.8	4.3	**91.8**	

Table 4. Confusion matrices of 2-fold cross validation **with the proposed feature descriptors** (%)

		Healthy	**Early AMD**	**Wet AMD**	**Avg.**
NN	**Healthy**	**65.4**	19.1	11.3	**75.0**
	Early AMD	22.5	**75.8**	4.8	
	Wet AMD	12.1	5.1	**83.9**	
RFC	**Healthy**	**81.9**	10.4	6.2	**85.2**
	Early AMD	10.7	**84.2**	4.4	
	Wet AMD	7.5	5.4	**89.4**	

study presents a method of texture recognition using learnable feature extractors rather than hand-picked ones. The spatial arrangement of the features are also learned and used for classification. A CNN is used to automatically train the filters to be used for feature extraction rather than using predefined filters. This allows the development of a set of filters which are best suited to the data rather than having to make assumptions about what features exist and choosing filters accordingly. The set of learned filters are convolved across the input images to produce a map of responses, where the textons can be further mined via clustering. The spatial arrangement of the features are examined using clustering and LBPs. Histograms are computed from the resultant output to produce the feature descriptors. These are passed onto the machine learning techniques for supervised classification. The method, applied on an OCT dataset to distinguish between different stages of AMD, produced promising quantitative results demonstrating its feasibility. Future work includes establishing a much larger patient dataset to study performance of the proposed method on leave one patient out prediction and classification.

Bibliography

[1] Castellano, G., et al.: Texture analysis of medical images. Clinical radiology 59(12), 1061–1069 (2004)
[2] Deng, J., et al.: Age-related macular degeneration detection and stage classification using choroidal oct images. In: ICIAR (2016)
[3] Doyle, S., et al.: A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE TBE 59(5) (2012)
[4] Dunn, D., Higgins, W.: Optimal gabor filters for texture segmentation. IEEE T-IP 4(7), 947–964 (1995)
[5] Essa, E., et al.: Shape prior model for media-adventitia border segmentation in ivus using graph cut. In: MICCAI-MCV. pp. 114–123. Springer (2012)
[6] Essa, E., et al.: Minimum s-excess graph for segmenting and tracking multiple borders with hmm. In: MICCAI. pp. 28–35. Springer (2015)
[7] Farsiu, S., et al.: Quantitative classification of eyes with and without intermediate amd using oct. Ophthalmology 121(1), 162–172 (2014)
[8] Jones, J., Xie, X., Essa, E.: Combining region based and imprecise boundary based cues for interactive medical image segmentation. IJNMBE 30(12), 1649–1666 (2014)
[9] Kanagasingam, Y., et al.: Progress on retinal image analysis for amd. Progress in retinal and eye research 38, 20–42 (2014)
[10] Koprowski, R., et al.: Automatic analysis of selected choroidal diseases in oct images of the eye fundus. BioMedical Engineering 12, 117 (2013)
[11] Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems (2012)
[12] Ojala, T., et al.: A comparative study of texture measures with classification based on featured distributions. Pattern recognition 29(1), 51–59 (1996)
[13] Priya, R., Aruna, P.: Automated diagnosis of amd from color retinal fundus images. In: ICECT (2011)
[14] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
[15] Soares, J.V., et al.: Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. IEEE TMI 25(9), 1214–1222 (2006)
[16] Spraul, C.W., et al.: Morphometric analysis of the choroid, bruch's membrane, and rpe in eyes with amd. IOVS 37(13), 2724–2735 (1996)
[17] Sugmk, J., et al.: Automated classification between age-related macular degeneration and diabetic macular edema in oct image using image segmentation. In: BMEiCON (2014)
[18] Szegedy, C., et al.: Going deeper with convolutions. In: CVPR (2015)
[19] Terry, L., et al.: Automated retinal layer segmentation using spectral domain oct. PloS one 11(9) (2016)
[20] Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. IJCV 62(1), 61–81 (2005)
[21] Xie, X.: A review of recent advances in surface defect detection using texture analysis techniques. ELCVIA 7(3), 1–22 (2008)