Classification of Volumetric Retinal Images Using Overlapping Decomposition and Tree Analysis

Abdulrahman Albarrak
Dept of Computer Science
University of Liverpool
United Kingdom
aabkb@liverpool.ac.uk

Frans Coenen
Dept of Computer Science
University of Liverpool
United Kingdom
coenen@liverpool.ac.uk

Yalin Zheng
Dept of Eye and Vision Science
University of Liverpool
United Kingdom
yzheng@liverpool.ac.uk

Abstract

Methods for the classification of volumetric three-dimensional (3D) volumes (images) play an important role in the context of medical applications. In this paper, a dedicated tree based 3D representation is proposed that serves to directly capture 3D image features in such a way that classification techniques can be applied. More specifically an Overlapping Hierarchical Decomposition (OHD) technique is presented to generate a tree representation of a given 3D volume. The OHD method recursively decomposes a given 3D volume into sub-volumes forming a tree. Once the tree has been generated, a frequent sub-graph mining algorithm is applied to mine the tree representation so as to generate sub-graphs. These sub-graphs are then used to define a feature space from which feature vectors representing 3D images (one per 3D volume) can be extracted and fed into a classifier generator. To demonstrate the applicability of the proposed method a 3D Optical Coherence Tomography (OCT) retinal image screening application is considered directed at the identification of Age-related Macular Degeneration (AMD). The results show a promising performance with a best Area Under the receiver operating Curve (AUC) value of 98.7%.

1 Introduction

Image mining is concerned with the application of data mining techniques to image data, it can be thought of as a form of "deep" image analysis. One important type of image mining is image classification, the automated categorisation of images using a classifier generated from a pre-labelled training set. Image classification algorithms typically rely on some forms of image feature extraction; the features are then stored in a manner that is compatible with the application of a classifier generator of some kind [13, 14]. The most significant challenge with respect to the design of feature extraction algorithms for 3D volumetric image mining is how they can best identify the most significant features so that volumetric 3D data of interest can be appropriately classified. Most of the current 3D approaches have focused on extending 2D based methods, typically statistical based methods founded on concepts such as Local Binary Patterns (LBPs) [19].

There are a number of application domains where 3D images are regularly used. One important domain is the medical domain. This paper considers the detection of a retinal disease, Age-related Macular Degeneration (AMD), from 3D volumes produced using Optical Coherence Tomography (OCT). Analogue to ultrasound, OCT is a relatively new imaging technique that can show the cross-sectional view of the retina at a high level of resolution and speed. AMD is a condition, typically contracted in old age, which leads to irreversible vision loss at its advanced stage [2, 8]. There has been little reported work concerning the development of 3D OCT retinal image classifiers.

In this paper we propose a tree representation aimed at effectively capturing 3D feature information. More specifically we proposed a new Overlapping Hierarchical Decomposition (OHD) to construct Non-trees (trees where each node, except the leaf nodes, typically have nine branches) and demonstrate that this representation serves to succinctly and accurately capture the salient features within volumetric data collections in such a way that effective classifies can be constructed. The proposed technique iteratively divides a given 3D image into a hierarchy of sub-volumes that can be stored in a Non-tree structure where each node in the tree represents a specific sub-volume. The decomposition is controlled using a critical function supported by a predefined maximum level of decomposition.

The main contribution of the work presented in this paper is the proposed OHD representation coupled with the use of frequent sub-graph mining to produce effective AMD clas-

sifiers for OCT image data. It is envisaged that by taking the overlapping areas into consideration the likelihood of missing important information (features) is reduced.

The rest of this paper is organised as follows. A brief overview of the related literature concerning classification, feature extraction methods and OCT image analysis is presented in Section 2. The AMD screening application domain which provides the motivation for the work described in this paper is presented in Section 3. The proposed algorithm is then detailed in Section 4. The evaluation of the proposed approach is reported in Section 5 where the performance of the approach is compared with other related methods. Finally, this paper is concluded in Section 6 with a summary and a review of the main findings.

2 Background and related work

In this section we overview the related work relevant to this paper. This section commences with consideration of some existing statistical techniques used to represent (and classify) image data. Next tree representations are considered with respect to the proposed overlapping Non-tree representation. The section is concluded with a brief overview of frequent sub graph mining algorithms and a short review of some related existing work on retina image analysis.

Examples of two well established statistical approaches are: (i) Local Phase Quantization (LPQ) and (ii) the Local Binary Patterns (LBP). The concept of LPQ is based on the local Fourier transform at low frequency where by a histogram of the quantized Fourier transform on each sub-volume is computed [15]. LBPs have been widely used in 2D representations to compute the relationship between a pixel and its immediate neighbours. The same idea can be applied to 3D, however, the generation of 3D rotation invariant LBPs is computationally expensive. Zhao and Pietikainen [19] used Three Orthogonal Planes LBP (LBP-TOP) so only voxels located in the XY, XZ and YZ planes were considered.

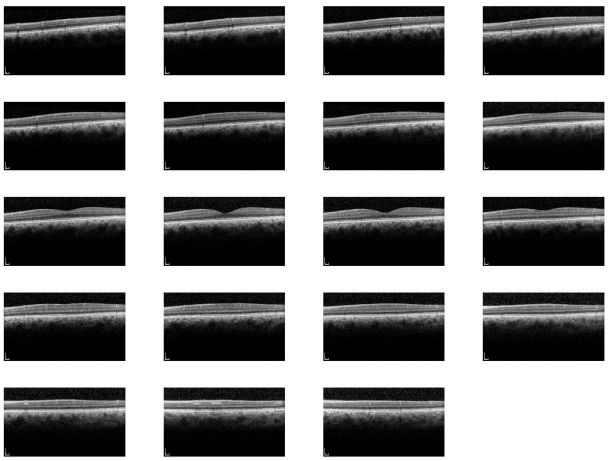
In general, image hierarchical decomposition can be defined as a way of dividing an image into sub-regions so that the identified sub-regions are in some sense similar. A parent-child relation is said to exist between a region and its immediate sub-regions [1]. Hierarchical decomposition has been used in many contexts such as image segmentation [11] and image clustering [3]. The work described in [7] is applied to 2D retinal images whereas a similar approach is applied to 2-D MRI "slices" in [5]. Neither [5] nor [7] have addressed the boundary problem associated with tree decomposition based representations.

To analyse our Non-tree representation (one per OCT image) we propose to identify frequently occurring subtrees and then use these as the input to a classification system. This approach is inspired in part by the work of [5] and [7]. To identify the desired frequent sub-graphs Frequent Sub-Graph (FSG) mining techniques can be used. FSG mining is a separate topic within the domain of data mining. One of the most well-known FSG mining algorithms is the gSpan algorithm which uses depth-first search to identify FSGs [18].

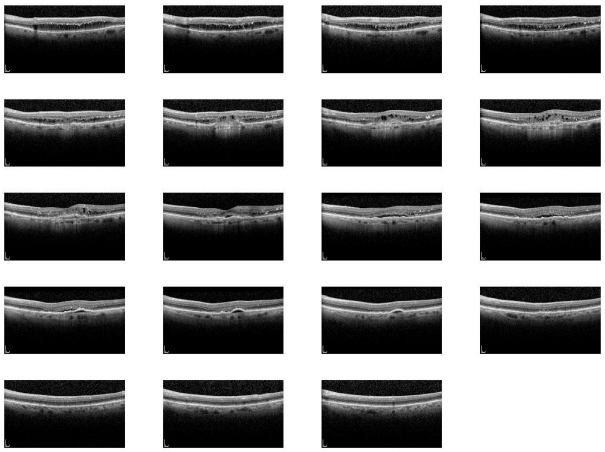
As already noted the focus for the work described in this paper is AMD screening. Most of the current macular disease diagnosis tools are founded on 2D retinal image analysis techniques because, up until recently, this was the only type of image commonly available. Two instances using OCT images can be found in [6] and [12]. In [6] a texture based method is employed using a Spatial Gray-Level Dependence Matrix (SGLDM) and the Discrete Fourier Transform (DFT) for capturing salient OCT image data. The traditional statistical approach is used to extract various metrics from the SGLDM such as energy, entropy, correlation, local homogeneity and inertia. A Bayesian classifier is used to categorise images according to these metrics. In [12] a method for automatic detection of retinal diseases, including AMD, is presented using a Multi-Scale Spatial Pyramid (MSSP) representation. Histograms of LBPs generated from each sub-block of the MSSP are then produced. Dimensionality reduction is achieved using Principal Component Analysis (PCA). All the selected LBPs are then concatenated together to build a single feature vector description (one per image). The Radial Basis Function (RBF) kernel based Support Vector Machine (SVM) classifier is then used for categorising these feature vectors.

3 Application Domain

The motivation and focus for the work described in this paper, as already noted, is OCT image analysis to support AMD screening. There are some distinct photographic features of AMD that can be found in OCT images such as: (i) disturbance of the Retinal Pigment Epithelium (RPE) layer underneath the neuro-retina due to the presence of drusen, pigment epithelium detachment, geographic atrophy or membrane, (ii) disruption of layered neuro-retinal tissue, (iii) presence of intra- and sub-retinal fluid and (iv) retinal thickening. Figure 1 shows two sequences of slices of two OCT volumes. Figure 1(a) shows slices of a normal OCT volume and Figure 1(b) slices of an OCT volume that features AMD. From these figures it can be seen that there are notable distinctions between the normal and the AMD volumes. The normal volume features smooth and connected layers. However, the AMD volume features thickening of the RPE layer, intra-retinal fluid, pigment epithelium detachment and some unusual texture patterns.



(a) A set of 3D OCT slices for a normal eye



(b) A set of 3D OCT slices for an AMD eye

Figure 1: Examples of 3D OCT volumes illustrating the difference between normal and AMD retina volumes. A normal retina volume has smooth and connective layers while an AMD retina volume has disrupted layers and other abnormal patterns from the normal.

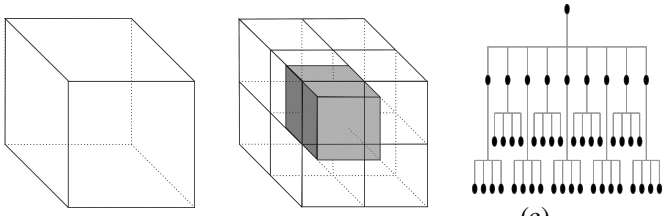
4 Proposed Approach

Hierarchical decomposition based approaches have a number of attractive aspects: (i) regions with similar features are grouped, (ii) the decomposition can be represented as a tree that maintains the relationships between volumes and sub-volumes and (iii) it can be combined with statistical feature representations. In the proposed approach, the 3D image is decomposed down to some maximal level or when homogeneous regions are arrived at. The latter is determined using what is known as a *critical function*. The proposed decomposition also features an overlap region (to address the boundary problem). The outcome of the decomposition is a tree (a graph). The resulting tree represented images are then processed so that a feature vector representation is obtained which can be input to a classifier generator. The rest of this section is organised as follows. The tree decomposition mechanism is discussed in Sub-section 4.1, and the feature extraction and classification mechanism in Sub-section 4.2.

4.1 Tree Generation

The aim of the proposed decomposition is to divide a given volume into homogeneous sub-volumes so that regions that feature a common structure are grouped together. The decomposition also features an overlap region. Most of the hierarchical decomposition based methods, such as that described in [7], do not consider overlapping regions.

The generated tree is defined in a form of a 4-tuple $T = (N, E, nl, el)$, where N represents the set of nodes, $E \subseteq N \times N$ the set of tree edges, $f(N) : X \rightarrow nl$ is the set of node labels, and $f(N, E) : X :\rightarrow el$ is the set of edge labels, which is typically a mapping between the nodes and the edges. In the tree, every new generated node is connected to an existing parent node with an edge having a label that indicates the relationship between the parent node and the child node. The decomposition commences by dividing the initial volume into eight sub-volumes plus an overlap volume. The OCT volumes of interest are "flat", in the sense that the X and Y axes are longer than the Z axis. Thus the following levels of decomposition comprise a quad-decomposition. This process continues until the chosen maximum level of decomposition is reached or homogeneous volumes are arrived at (when the critical function is satisfied). On completion each node in the tree will describe a sub-volume in terms of its intensity values (except the root node which describes the entire volume).



(a) Root (b) Level1 (c) Non-tree

Figure 2: 3D volume decomposition with overlap sub-volume

4.1.1 Critical Function

As noted above the objective of a critical function is to determine whether homogeneous sub-volumes have been reached. In this paper the similarity between the parent and child intensity values are used. We represent these in the form of histograms. The Longest Common Subsequence

(LCS) time series analysis technique [17] is used to establish the similarity between histograms. The advantage offered by the LCS mechanism, compared to other time series mechanisms, is that it reduces the complexity of the problem by using an upper and lower Minimum Bounding Envelope (MBE) so as to estimate the similarities within the MBE. For every sub-volume, the child histogram is compared with its parent histogram. For every histogram the normalisation function, s, presented in equation 1, where h is the given histogram, is applied. The LCS method is then used to measure the similarity between a parent volume and each of its child volumes. If the maximum LCS distance is less than a threshold the decomposition will terminate.

$$s(h) = (h - mean(h))/std(h) \quad (1)$$

4.1.2 Node Labels

It is desirable to use node and edge labels to distinguish between trees. It is important that the nodes and edges in the generated trees are allocated sufficiently descriptive labels so that good features (features that significantly contribute to the accurate classification of volumes) can be extracted. In [7] the Average Intensity Values (AIVs), the mean of the intensity values of an image region, were used to label the nodes. However, it is conjectured here that such mean values for sub-volumes do not provide a sufficient description of content. Instead, in this paper, the *Kurtosis* of each sub-volume is used as the node label. Kurtosis measures the "peakedness" of a distribution (an intensity histogram in our case), in other words Kurtosis describes the shape of a distribution [9]. The following equations demonstrate how Kurtosis is calculated (Equations 2 to 4).

$$P = h/size(h),\ m = \sum P * h \quad (2)$$

In Equation 2, h is the histogram vector of the sub-volumes.

$$v = \sum (P * (h - m)^2),\ t = P * (h - m)^4 \quad (3)$$

$$s = \sqrt{v},\ Kurtosis = s^{-4} * \sum t \quad (4)$$

4.1.3 Edge Labels

In the proposed tree based representation for 3D images (volumes) edge labels are defined in terms of the "similarity distance" between the corresponding parent and child node. In this paper the Kullback-Leibler Divergence (KLD) [10] is used to indicate the difference between the parent and child nodes. The following equations (Equations 5 to 8) demonstrate how the KLD function is used to generate edge labels.

$$np = \frac{p}{\sum p},\ nc = \frac{n}{\sum n} \quad (5)$$

In Equation 5, p is the histogram for the parent node and n is the histogram for the child node.

$$KL1 = \sum (np * (log2(np) - log2(nc))) \quad (6)$$

$$KL2 = \sum (nc * (log2(nc) - log2(np))) \quad (7)$$

$$KLD = (KL1 + KL2)/2 \quad (8)$$

4.2 Feature Extraction and Classification

After the Non-trees have been generated, one tree per image, the next stage is to generate the feature vector representation so that the salient features of the tree collection are preserved. In this case the salient features of interest are frequently occurring sub-trees, sub-trees that represent common structures that occur across the data set as whole. The extraction of the desired distinctive sub-tree patterns is one of the most significant aspects of the proposed approach.

One way of identifying frequently occurring sub-trees in our collection of trees (forest) is to apply some frequent sub-graph mining technique (a tree is a type of graph). With respect to the work described in this paper the well-known gSpan [18] algorithm was employed although alternative frequent sub-graph mining algorithms could equally well have been employed. The sub-graphs are then used to define a feature space where each identified sub-graph represents a dimension within that space and each dimension can have the values 0 or 1 (exists or does not exist). Each Non-tree of a 3D image can then be defined in terms of this feature space so that each tree generated by the 3D image is represented as a feature vector. More specifically:

Definition Given two graphs (trees) $G' = (N', E', nl', el')$ and $G = (N, E, nl, el)$, G' is a subgraph of G ($G' \subseteq G$) if $N' \subseteq N$, $E' \subseteq E$, $nl' \subseteq nl$ and $el' \subseteq el$.

The extracted set of feature vectors (one per 3D image) can then be used as input to a supervised learning algorithm provided that we have a known class label for each vector, in other words we need to have a *training set*. In the case of the AMD application considered in this paper a labelled set of OCT 3D images was used for evaluation purposes (see Section 5 below for detail). The objective of the classification is then to map sequences of features (frequent sub-trees) to specific labels.

To conduct the desired classification, gboost [16] was adopted to classify the generated graphs. Gboost uses a Linear Programming (LP) classifier designed by Demiriz et al. [4]. LP is used to find the optimal frequent subgraphs extracted by Gspan [18].

5 Evaluation

To assess the performance of the proposed overlapping non-tree representation and the associated frequent sub-tree feature space formalism in comparison to alternative approaches and methods, a data set comprising 140 3D OCT volumes was used, 68 "normal" and the remainder wet-AMD. The size of each volume was approximately (1024×496 pixels $\times 19$ slices) representing a $6 \times 6 \times 2$ mm retinal volume. The dataset was split $50/50$ (half for training and half for testing). Two levels of maximum decomposition, 4 and 5, were used in the experiments. The Area Under the receiver operator characteristic Curve (AUC), accuracy, sensitivity and specificity were used to measure the performance (see Equations 9, 10).

Two sets of experiments were conducted. The first set was directed at analysing the proposed non-tree representation with respect to: (i) a more standard oct-tree representation, (ii) the 2D methods described in [12] and [6] (namely the MSSP method described in [12] and the texture method described in [6]) using the OCT slice passing through the centre of the fovea (the fovea is located at the centre of the retina), and (iii) 3D methods LPQ and LBP-TOP presented in [15] and [19] respectively. The second set of experiments was directed at investigating the time complexity of the proposed approach (using both the LCS and AIV critical functions) in comparison with the other approaches considered in the first set of experiments.

The results obtained with respect to the first set of experiments are presented in Table 1. From the table it can be seen that the 3D decomposition based methods outperform the 2D methods described in [6] and [12], thus providing clear evidence that using 3D data provides for more effective classification than when using 2D data (at least in the context of the retina screening application considered in this evaluation). If we compare the proposed critical function LCS with the AIV critical function used [7], the LCS function seems to provide better performance. There is also some evidence that a level of decomposition of 5 produces better results than a level of 4, probably because a greater degree of detail is captured using the higher level of decomposition. Finally, with respect to the overlapping principle, better results tend to be obtained using the non-tree (oct-tree with overlap) than the oct-tree on its own.

The results obtained with respect to the second set of experiments are presented in Table 2. The table lists both Average Feature Extraction Time (AFET) and Average Classification Time (ACT) as well as the Total Execution Time (TET). AFET is the time required to decompose the image and generate feature vectors. ACT is the time required to conduct the desired classification. From the table it can be seen that there is a trade-off between performance and time complexity. The proposed overlapping methods require more processing time than the other methods, especially when the size of the graph (the level of decomposition) is increased.

Table 1: Comparison of proposed non-tree approach with respect to the oct-tree, MSSP and texture approaches in terms of (i) critical function, (ii) level of decomposition (L), (iii) Accuracy (Acc), (iv) Sensitivity (Sen), (v) Specificity (Spec) and (vi) Area Under the Curve of the receiver operating characteristic (AUC) computed using a confusion matrix.

Method	L	Acc	Sen	Spec	AUC
oct-tree LCS	4	84.2	82.1	87.1	92.8
	5	87.1	86.5	87.9	94.2
non-tree LCS	4	85.7	84.2	87.5	95.7
	5	94.2	92.1	96.0	98.7
oct-tree AIV [7]	4	71.4	73.5	69.4	81.0
	5	81.4	84.8	78.4	89.9
non-tree AIV	4	77.1	85.7	71.4	86.1
	5	87.1	86.5	87.9	96.4
Texture [6]	-	90.0	88.6	91.4	90.0
MSSP [12]	-	81.4	74.4	92.6	81.8
LPQ [15]	-	91.4	91.2	91.7	91.4
LBP-TOP [19]	-	82.8	92.3	77.3	82.5

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Sen = \frac{TP}{TP + FN},\ Spec = \frac{TN}{TN + FP}. \tag{10}$$

6 Discussion and Conclusions

A comparison between the usage of two critical functions, LCS and AIV, for measuring the homogeneity of the sub-volume was conducted in the experiments. From the reported evaluation it was observed that the best recorded accuracy was obtained using the proposed overlapped decomposition (regardless of whether the LCS or AIV critical function was adopted). In addition, the results, as shown in Table 1, indicate that using the LCS function provides better results than using the AIV function. The reported results also demonstrated that the proposed 3D method performed better than the 2D methods described in [6] and [12] and the 3D methods described in [15] and [19] with respect to our dataset. It was argued that this was because the features obtained using the proposed non-tree representation were more descriptive and took into account the boundary issue.

In conclusion, a mechanism, founded on the use of non-trees, for classifying 3D retinal OCT volumes has been de-

Table 2: Running time (seconds) of the programs for Average Feature Extraction Time (AFET), Average Classification Time (ACT) and Total Execution Time (TET).

Method	Level	AFET	ACT	TET
oct-tree LCS	4	5.6	8.6	14.2
	5	5.8	4.8	10.6
non-tree LCS	4	9.1	45.0	54.1
	5	9.7	53.1	62.8
oct-tree AIV [7]	4	0.8	0.4	1.3
	5	0.8	0.5	1.4
non-tree AIV	4	6.4	0.3	6.7
	5	7.3	1.0	8.3
Texture [6]	-	4.3	0.2	4.5
MSSP [12]	-	2.1	0.7	2.8
LPQ [15]	-	5.9	0.04	5.9
LBP-TOP [19]	-	52.7	0.1	52.8

scribed. More specifically an overlapping volume decomposition, coupled with a graph based feature identification method, was presented. A number of possible future directions for the work, using the same experimental set up, are envisioned. It would be interesting to assess the effects of applying different mechanisms for decomposing the volumetric data such as wavelet or curvelet transforms. It is also conjectured that the use of different frequent sub-graph mining techniques might result in the extraction of more descriptive features (sub-graphs) and/or lead to improvements in the time-complexity. For example stochastic methods might be used for mining graphs in order to improve the time complexity.

References

[1] G. Borgefors, G. Ramella, and G. Sanniti Di Baja. Hierarchical decomposition of multiscale skeletons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1296–1312, November 2001.

[2] F. Coscas, G. Coscas, E. Souied, S. Tick, and G. Soubrane. Optical coherence tomography identification of occult choroidal neovascularization in age-related macular degeneration. *American Journal of Ophthalmology*, 144(4):592 – 599, 2007.

[3] D. Curtis, V. Kubushyn, E. A. Yfantis, and M. Rogers. A hierarchical feature decomposition clustering algorithm for unsupervised classification of document image types. In *Proceedings of the Sixth International Conference on Machine Learning and Applications*, pages 423–428, Washington, DC, USA, 2007. IEEE Computer Society.

[4] A. Demiriz, K. Bennett, and J. Shawe-Taylor. Linear programming boosting via column generation. *Machine Learning*, 46(1):225–254, 2001.

[5] A. El Sayed, F. Coenen, C. Jiang, M. Garcia-Finana, and V. Sluming. Corpus callosum MR image classification. *Knowledge Based Systems*, 23(4):330–336, 2010.

[6] K. W. Gossage, T. S. Tkaczyk, J. J. Rodriguez, and J. K. Barton. Texture analysis of optical coherence tomography images: feasibility for tissue classification. *Journal of Biomedical Optics*, 8(3):570–575, 2003.

[7] M. H. A. Hijazi, C. Jiang, F. Coenen, and Y. Zheng. Image classification for age-related macular degeneration screening using hierarchical image decompositions and graph mining. In *Machine Learning and Knowledge Discovery in Databases*, volume 6912 of *Lecture Notes in Computer Science*, pages 65–80. Springer Berlin Heidelberg, 2011.

[8] R. D. Jager, W. F. Mieler, and J. W. Miller. Age-related macular degeneration. *New England Journal of Medicine*, 358(24):2606–2617, 2008.

[9] R. Jindal, S. Jindal, and N. Kaur. Analyses of higher order metrics for spiht based image compression. *International Journal of Computer Applications*, 1(20):56–59, February 2010. Published By Foundation of Computer Science.

[10] D. H. Johnson and S. Sinanovic. Symmetrizing the kullback-leibler distance. *IEEE Transactions on Information Theory*, March 2001.

[11] S. Katz and A. Tal. Hierarchical mesh decomposition using fuzzy clustering and cuts. *ACM Transactions on Graph*, 22(3):954–961, July 2003.

[12] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. Schuman, and J. M. Rehg. 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(5):748–759, 2011.

[13] D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, January 2007.

[14] R. F. Murray. Classification images: A review. *Journal of Vision*, 11(5), 2011.

[15] J. Paivarinta, E. Rahtu, and J. Heikkila. Volume local phase quantization for blur-insensitive dynamic texture classification. In *Proceedings of the 17th Scandinavian conference on Image analysis*, pages 360–369, Berlin, Heidelberg, 2011. Springer-Verlag.

[16] H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda. gboost: a mathematical programming approach to graph classification and regression. *Machine Learning*, 75(1):69–89, April 2009.

[17] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh. Indexing multi-dimensional time-series with support for multiple distance measures. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 216–225, New York, NY, USA, 2003.

[18] X. Yan and J. Han. gspan: Graph-based substructure pattern mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining*, pages 721–724, Washington, DC, USA, 2002. IEEE Computer Society.

[19] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–928, June 2007.