Volumetric Image Mining Based on Decomposition and Graph Analysis: An Application to Retinal Optical Coherence Tomography

Abdulrahman Albarrak*, Frans Coenen*, Yalin Zheng† and Wen Yu*
* Department of Computer Science, University of Liverpool, Liverpool, UK
† Department of Eye and Vision Science, University of Liverpool, Liverpool, UK

***Abstract*—In this paper, we introduce a method for classifying volumetric images using a decomposition and graph analysis based method. Volume decomposition plays an important role in simplifying Three-Dimensional (3-D) volumes so as to enhance their analysis.**

Hierarchical decomposition techniques incrementally divide a given 3-D volume into sub-volumes according to some critical function and then (typically) represent the decomposition as a tree. The issue is the nature of the critical function which dictates when the decomposition should be stopped. Broadly the decomposition should be stopped whenever homogeneous sub-volumes are reached. The question is how we define homogeneity in this context. In this paper a number of different critical functions are evaluated. The evaluation was conducted by considering the classification of 3-D Optical Coherence Tomography (OCT) retinal images according to whether they feature Age-related Macular Degeneration (AMD) or not. The OCT volumes were encoded using the proposed hierarchical tree decomposition coupled with a number of different critical functions. A frequent sub-graph mining algorithm was then applied to the tree representations and the resulting identified frequent sub-graphs used to define a feature vector encoding which was then fed into a standard classifier.

***Index Terms*—volume decomposition, subgraph mining, volume classification, OCT, AMD.**

I. Introduction

Image mining is concerned with the extraction of knowledge, patterns and relationships from within collections of images. Three-Dimensional (3-D) volumetric image mining is an important, but understudied element within the domain of image mining, despite the widespread use of volumetric data in real applications. A typical image mining task is image classification. The work in this paper is directed at the representation of 3-D volumes so as to permit the effective application of classification techniques. Effective in this context is defined in terms of classification performance. More specifically the work described is directed at 3-D volumes generated using Spectral Domain Optical Coherence Tomography (SD-OCT), particularly retinal SD-OCT volumes.

SD-OCT is a 3-D imaging technique that is used to digitally capture, with a high level of resolution, the inner structure of objects [1]. Light and ultrashort laser pulses are the medium for the scanning. In the context of ophthalmology, due to the transparent nature of the eye, SD-OCT provides an ideal method for viewing the retina. Retinal volumetric data is comprised of a set of Two-Dimensional (2-D) “B-scans”, 2-D cross-sectional slices. Ophthalmologists have recently utilized OCT images to diagnosis retinal disorders. One example retinal disorder is Age-related Macular Degeneration (AMD). AMD is a condition, typically contracted in old age, which causes irreversible vision loss at its advanced stage [2], [3]. At present OCT is the only means whereby 3-D details of the retina and choroid, where most of the potential AMD indicators are likely to appear, can be revealed. AMD is typically identified in retinal SD-OCT images by visual inspection. Figure 1(a) shows a normal retina OCT volume; the retina has a smooth contour and a regular arrangement of individual retinal layers. Figure 1(b) shows a retina with AMD; here as a result of the presence of AMD, fluid within the retinal layer and detachment of the retina causes the layers of the retina to separate.

The motivation for the work described in this paper is to design a technique to assess the presence of AMD in 3-D OCT retinal volumetric data. Given the rapid growth in the global “senior” population, a growth that is expected to increase, it is anticipated that AMD will become commonplace. Although AMD cannot be cured, its progress can be reduced given early detection. AMD screening programmes are therefore desirable. However, due to the large number of expected “patients” clinicians will have to examine large amounts of image data, which in turn will entail a significant resource. Automated support for AMD detection is therefore desirable. At present there is a lack of such automated AMD detection tools. Most of the current retinal SD-OCT analysis tools either: simply measure some features of the retina such as thickness, or apply some form of retinal layer segmentation [4] or drusen segmentation [5].

There are a variety of ways that 3-D volumes can be represented so that classification techniques can be applied. Typically such techniques operate using a feature vector representation. Thus the 3-D volumes need to be translated into such a representation. This can be done, for example, by using some statistical representation of the space, where the space is somehow formatted in terms of a sequence of measures. The approach presented in this paper is to first represent the volumes of interest using a tree representation, one tree per volume, generated through a hierarchical decomposition process. Next a frequent sub-graph mining technique is applied to the tree representation so as to identify frequently occurring sub-graphs which then define the elements in a binary valued feature space (which can be used to define feature vectors, one per volume).

The main issue with hierarchical decomposition techniques is when to stop; we want a tree representation that is both expressive and economic, one is always maximized at the expense of the other, hence a balance has to be sought. In this paper, the decomposition stops when either: (i) some maximum level of decomposition is reached or (ii) homogeneous volumes are arrived at (note that the tree does not have to be balanced). The latter is defined by some critical functions.

Given the above, the study described in this paper provides two significant contributions. The first is a novel method for volume classification using a tree decomposition. The second is a comparison between seven different methods likely the described critical function can be implemented. Further details of these are given in section III.

The rest of this paper has been organised in the following manner. Section II reviews some previous work related to retinal diseases diagnosis, 3-D retinal image classification and graph analysis. The design of the proposed technique is then described in section III. Section IV assesses the performance of the proposed approach. A discussion is then presented in Section V. Finally, this paper is concluded in Section VI with a review of the main findings.

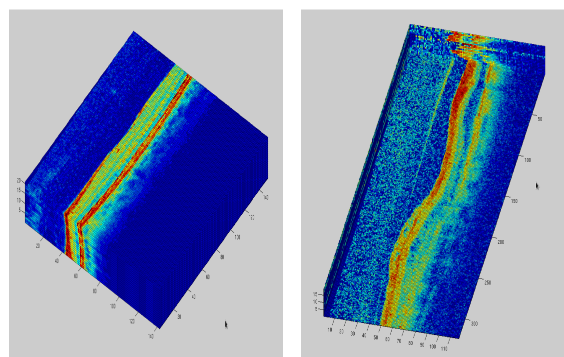
II. Related Work

Most of the current macular disease diagnosis tools are focused on 2-D image data. Two examples can be found in [6] and [7] (used for evaluation purposes later in this paper in Section IV). In [6] Liu et al. implemented an algorithm for classifying retinal diseases including AMD. The image was represented using a Multi-Scale Spatial Pyramid (MSSP) with different levels. The local descriptors of each sub-section of the MSSP, in each level, were generated using the histograms of the Local Binary Patterns (LBPs). Then the dimensionality was reduced using Principal Component Analysis (PCA). All the LBP's were concatenated together forming a global feature descriptor. The Radial Basis Function (RBF) kernel based Support Vector Machine (SVM) classifier was then applied to the global descriptors in order to categorise the descriptors according to retinal diseases.

In [7] a texture based method was used. Spatial Gray-Level Dependence Matrix (SGLDM) was used to represent the OCT image feature. The basic idea of the SGLDM matrix is to count the number of pixel pairs, that display the same attributes, located within a certain distance and a direction of each other. Then a statistical method was used to extract the features of the SGLDM such as energy, entropy, correlation, local homogeneity, and inertia. In addition, the Discrete Fourier Transform (DFT) was computed on the frequency of the image. A Mahalanobis distance based method was applied to measure the similarities between image features and a Bayesian classifier used to differentiate between features. In SGLDMs, different matrices are typically computed according to selected direction, which leads to long feature vectors, which may in turn decrease the accuracy of the classification. When applying SGLDM to 3-D image, we can identify 13 direction, so 13 matrices are generated for every image.

In the context of general 3-D classification, as in the case of OCT volumes, there is less reported work compared to work on 2-D classification. The research to date has tended to focus on extracting image features rather than classifying them. For instance in Ankerst et al. [8] the object space was divided into many partitions. The partitioning was performed using two models: (i) the *shell model* and (ii) the *sector model*. Using the shell model the space was divided into "shells" around the object's centre of mass. Using the sector model equal sized "sections" were constructed. The partitionings were then combined to form what the authors refer to as the *spiderweb model*. For each element of this model histograms were generated and a quadratic distance function used to measure the similarity between histogram bins so that classification could be performed.

One way to extract (represent) image features is to hierarchically decompose the object down to a certain level and represent this decomposition as a tree or a graph. Then a graph mining technique, such as gSpan, may be applied to extract the most important features of the graph [9].



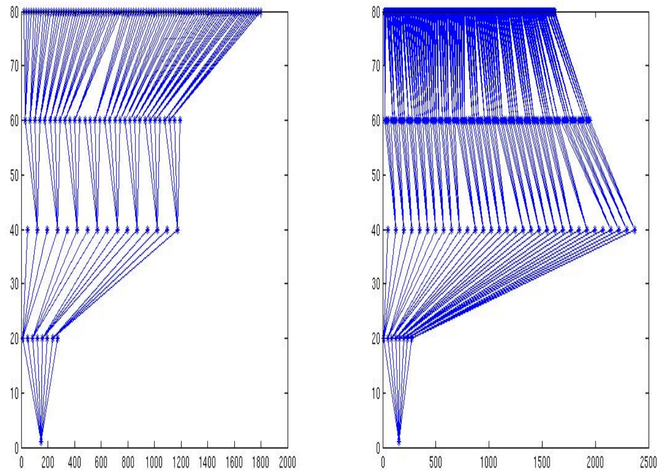
(a) (b)

Fig. 1. Examples of 3-D OCT images illustrating the difference between an AMD and a normal retina: (a) normal retina, (b) AMD retina.

III. Proposed Technique

The proposed method involves three steps. Firstly, the retinal OCT volumetric data is decomposed into a tree. Secondly, the commonly occurring sub-trees are extracted using a sub-graph mining technique (a commonly occurring tree in this sense equates to a frequent sub-graph). The identified frequent sub-trees are then considered to represent binary valued features in a feature space. Thirdly, from this feature space, feature vectors are generated which can then be used for classifier generation and evaluation.

Figure 2(a) shows a tree generated for the normal 3-D retinal volume presented in Figure 1(a), whilst Figure 1(b) shows a tree generated from the volume shown in Figure 2(b). Inspection of Figure 2 clearly demonstrates that there are



(a) (b)

Fig. 2. Tree representation for: (a) the normal retinal image in Figure 1(a), and (b) the AMD retinal image in Figure 1(b).

differences in the two tree structures, these difference can be used to distinguish between classes in a classification setting.

A. *Image Decomposition*

As noted above the proposed method used for decomposing the volumetric data is a tree based technique. We propose to divide a given volume into eight sub-volumes. The whole volume is considered as the root of the tree. The eight sub-volumes at the first level of decomposition are then the children of the root. These children sub-volumes (nodes) are recursively decomposed into further four sub-volumes until the chosen maximum level is reached or homogeneous volumes are arrived at (when the critical function is satisfied). From Figure 1, it can be observed that the depth of the image (z axis) is not of the same size as the width and the height of the image. Thus it does not make sense to continuously divide the sub-volumes into eight partining, after the first partition, and quadrats are considered instead.

The reason for using a decomposition based method is that decomposing the volume in this way leads to the identification of smaller sub-volumes with specific features: which, it was conjectured, would help to distinguish between different 3-D images (AMD and non-AMD with respect to the evaluation presented in this paper). Algorithm 1 describes the tree decomposition process. In addition, the spatial relationships between nodes are maintained. So nodes that share the same parent could be considered for encoding the spatial positions. In the spatial pyramid technique (see Section II), when applied to 2-D images, the image is divided into a certain number of sub-images so that the close neighbours are only considered. In the tree based method, the neighbours can be traced from any position by following the tree connections.

Algorithm 1 Pseudocode for the proposed tree decomposition method

Input: VolumetricData, maxLevel
Output: FeatureVector
root Features $\leftarrow$ VolumeFeatures (VolumetricData)
$SetNodeData(0, 1) \leftarrow VolumetricData$
nodeFeature $\leftarrow$ setNodeFeatures(root Features)
for $level = 1$ to $maxLevel$ **do**
  $nodeNumber \leftarrow GetNumboerOFNode(level - 1)$
  for $n = 1$ to $nodeNumber$ **do**
    $node = GetNode(level - 1, n)$
    $nodeData = NodeData(level - 1, node)$
    if isHomogeneous(nodeData) **then**
      subVolume1-4 $\leftarrow$decomposeVolume(nodeData)
      NodeFeatures $\leftarrow$setNodeFeatures(subVolume1-4)
      EdgeFeatures $\leftarrow$setEdgeFeatures(subVolume1-4)
      node1-4 $\leftarrow$GenerateNodes(subVolume1-4)
      edge1-4 $\leftarrow$setEdges(subVolume1-4,ParentNode)
      SetNodeFeatures(node1-4,NodeFeatures)
      SetEdgeFeatures(edge1-4,NodeFeatures)
    end if
  end for
end for
SubgraphF $\leftarrow$ Miningtree(tree)
FeatureVector $\leftarrow$ GenerateFV(SubgraphF)

1) Sub-Volume Homogeneity: Sub-volume homogeneity is an important aspect of the proposed approach as it determines the eventual shape of the tree. In this study we use the term *Homogeneity*, with respect to 3-D sub-volumes, to describe the situation where a sub-volume's voxels are in some sense uniform. To this end, as already noted, we use a critical function ($isHomogeneous$) in Algorithm 1. Various mechanisms can be used to implement this function.

The design of any adopted critical function should be such that it controls the decomposition in such a way that the number of nodes and edges generated is minimized (thus limiting the overall graph size) while at the same time ensuring that the decomposition allows for the discrimination of the essential features required for classification. Let $G = (V, E)$ be a graph such that V is the set of vertices and E is the set edges. Our aim is to minimize the size of G in terms of its nodes and edges while at the same time ensuring that our final sub-volumes are homogenous. The basic idea is to employ a function that can satisfy the condition that minimises the number of nodes and edges and thus decreases the graph size. While at the same time it should ensure that the 3-D object is divided enough to include all the essential features required for graph recognition. Lets $G = (V, E)$ be a graph with V as the set of vertices in the graph and E as the set of the graph's edges. Our aim is to minimise the number of nodes and maximise homogeneity of individual nodes.

In this paper, we consider seven distinctive critical functions, the aim being to identify the most appropriate. Some of these have been proposed with respect to other applications while some have been developed specifically with repect to the work described. The following illustrates the seven functions that are employed to measure the homogeneity of sub-volumes:

1) **Average Intensity Value (AIV).** AIV was used by Hijazi et al. [9]. AIV is the mean of the intensity of the region. In this method, a distance measure was used to compare the homogeneity of the immediate sub-volumes and the parent volume. In order to apply AIV, the current node is first decomposed and then the relationships between the new sub-nodes and the parent checked. A threshold value is used to decide whether each sub-volume, in comparison with the parent node, is homogenous or not; if so the decomposition is accepted, otherwise the parent node is not decomposed further. The homogeneity ω of the parent volume is calculated using Equation 1, where s is the number of sub-nodes, AIV_p indicates the parent AIV. If ω is greater than the specified threshold then the new nodes are added to the tree (and the decomposition continues).

$$\omega = \frac{1}{s}\sum_{i=1}^{s}\sqrt{(AIV_p - AIV_i)^2} \tag{1}$$

2) **Gray Level Co-occurance Matrix (GLCM).** Use of GLCMs is another option for deciding if a set of sub-volumes, generated by decomposing an immediate parent node, are homogeneous or not. In the 3-D GLCM matrix the number of times that two voxel are neighbours is counted within a certain distance and direction. The desired GLCM matrix is first extracted and then the GLCM diagonal is checked. The maximum value in the diagonal then indicates that the sub-volume is the most homogeneous according to Equation 2 [10] .

$$GLCMhomogeneity = \sum_{i,j}\frac{GLCM(i,j)}{1+|i-j|} \tag{2}$$

If the calculated $GLCMhomogeneity$ value is less than a given threshold, then the sub-volume is homogeneous and so it is not divided further.

3) **Kendall's Coefficient Concordance (KCC).** KCC is used to assess the homogeneity of a space. In Equation 3 W is the KCC value, which ranges from between 0 and 1 [11]. If the value W is close to one, then the space is not homogeneous. In our case, a threshold value (close to zero) is used to determine if the sub-volume should be decomposed further or not.

$$W = \frac{\sum(R_i)^2) - n(\bar{R})^2)}{1/12K^2(n^3-n))} \tag{3}$$

In Equation 3, R_i is the sum of intensity value of the ith row of the image, where $\bar{R} = \frac{(n+1)K}{2}$, K is the height of the image and n is the width of the image.

Using a time series approach histograms of the 3-D image intensity are generated for each of the sub-volumes and compared to the parent. If the maximum distance is less then some threshold, then the decomposition stops. Four critical functions were derived whereby the time series could be computed as follows:

4) **Simple Euclidean Distance (ED)**:

$$ed = \sum(histogram_p - histogram_i)^2 \tag{4}$$

5) **Dynamic Time Warping (DTW) [12]**:

$$LB\_DTW(Q,C) = \sqrt{\sum_{i=1}^{n}\begin{cases}(c_i - U_i)^2 & if c_i > U_i\\ (c_i - L_i)^2 & if c_i < L_i\\ 0 & otherwise\end{cases}} \tag{5}$$

where Q and C are two histograms of length n. U and L are some Upper and Lower bound. n is set to 16 bins.

6) **Longest Common Subsequence (LCS)** [13]:

$$lcs(A,B) = \begin{cases} 0\ if\ A\ or\ B\ is\ empty \\ 1 + LCSS_{\delta,\varepsilon}(Head(A)Head(B)) \\ if \mid a_{x,n} - b_{x,m} \mid < \varepsilon \\ if \mid a_{y,n} - b_{y,m} \mid < \varepsilon \\ and \mid n - m \mid < \delta \\ max\{LCSS_{\delta,\varepsilon}(Head(A),B), \\ LCSS_{\delta,\varepsilon}(A, Head(B)), \\ otherwise \end{cases} \tag{6}$$

where A and B are two histograms of the size n and m respectively, δ and ε are real numbers < 0.

7) **Kullback-Leibler divergence (KLD)** [14].

$$\begin{gathered} KLD(P1(x), P2(x)) = \\ sum[P1(x) \cdot log(P1(x)/P2(x))] \end{gathered} \tag{7}$$

Where $P1(x)$ and $P2(x)$ are the histograms of the sub-volume.

2) Node Features: We have already noted that the fundamental idea is to break the given SD-OCT volume down into sub-volumes and store these sub-volumes in a tree structure. The nodes in the tree represent volumes. The simplest way to assign a value to a node is to use the mean value of the sub-volume intensity. We refer to the different types of value that may be associate with nodes as *node features*. The selection of appropriate node features plays an important role in the description of sub-volumes.

3) Edge Features: Edge features are defined in terms of the similarity distance between a parent node feature and the node feature. Edge features are essential for the envisioned volumetric tree classification as they are used as one of the basic feature for distinguishing between trees.

B. Classification

A Feature Vector (FV) is used to describe each data volume (retinal volume with respect to the motivation for this paper). The FV are constructed according to the identified frequent sub-graphs (features) extracted using a subgraph mining technique (Gspan [15] was used with respect to the evaluation described in this paper). In subgraph mining, the most frequently occurring sub-graphs are identified using some appropriate search method such as Depth First Search (DFS). The identified frequent sub-graphs were then used to define a

TABLE I
RESULTS OF EVALUATING THE PROPOSED APPROACH WITH DIFFERENT FUNCTIONS.

Method	Level	AUC	FNR
1. AIV [9]	4	0.818	0.333
	5	0.839	0.22
2. GLCM	4	0.900	0.25
	5	0.897	0.158
3. KCC	4	0.953	0.182
	5	0.912	0.10
4. ED	4	0.868	0.176
	5	0.812	0.154
5. DTW	4	0.918	0.19
	5	0.935	0.118
6. LCS	4	0.912	0.11
	5	**0.959**	**0.067**
7. KLD	4	0.897	0.167
	5	0.885	0.158
8. MSSP [6]		0.894	0.1
9. Texture [7]		0.906	0.117

feature space which was, in turn, used to define feature vectors (one per volume). For training purposes each feature vector was combined with a class label (in our case +1 for AMD, -1 for normal) informed by medical retina experts. With respect to the evaluation described in Section IV a boosting based method (gBoost) was adopted to classify feature vectors [16]. The results will be presented in the next section.

IV. EVALUATION

To evaluate the effectiveness of the proposed approach and compare the different critical functions considered, experiments using volumetric retinal data set comprising 77 3-D SD-OCT images were conducted. The size of each volume was approximately $1024 \times 496\ pixels \times 20\ slices$ describing a $6 \times 6 \times 2$ mm retinal volume. We split the dataset into half (half for training and half for testing). The experiments were conducted using two levels of maximum decomposition, 4 and 5.

Table I shows the performance of the methods. Two performance metrics were recorded to measure the performance of the proposed mechanism: the Area Under receiver operator characteristic Curve (AUC) and the False Negative Rate (FNR), see equation 8. In addition, we have compared with the 2-D methods described in [6] and [7] using the section crossing the fovea [1]. Results using MSSP [6] and Texture [7] are included in the Table I.

$$FNR = \frac{FN}{TP + FN} \tag{8}$$

V. DISCUSSION

The aim of the study described in this paper was to devise a mechanism for classifying 3-D SD-OCT retinal volumes. Various critical functions for measuring the homogeneity of a sub-volume were proposed. In general, medical applications require a high level of accuracy.

[1]Fovea is the centre of the retina.

The present study was designed to determine the effect of the proposed volume decomposition and graph based method when applied to OCT retinal images. The results of using the proposed technique demonstrated a good performance with respect to some of the considered critical functions. With reference to Table I it can be observed that the best recorded accuracy was obtained using the LCS critical function with a maximum decomposition of 5 levels. The most interesting finding was that time series methods demonstrated a good overall performance (for both levels of decomposition).

We can conclude that volume decomposition improves the accuracy of the classification when compared to the 2-D methods described in [6] and [7]. A possible explanation for this might be that features are selected from small volumes and so similar sub-volumes with the same attributes will be matched through application of the sub-graph mining process.

VI. CONCLUSION

This paper has given an overview of a volume decomposition and graph analysis based approach for recognising AMD diseases in OCT volumetric data. In this investigation, the aim was to design and assess a method for identifying disorders in the retina (specifically AMD). Returning to the question posed at the beginning of this paper, it is now possible to state that the proposed decomposition combined with the LCS critical function for measuring the sub-volume homogeneity results in a good classification. One of the more significant findings to emerge from this study is that volume decomposition provides an accurate shape description as it divides the volume image into small sub-volumes where the features can be computed in a way that ensures traceability.

A number of possible future studies using the same experimental set up are envisioned. It would be interesting to assess the effects of applying different mechanisms for decomposing volumetric data taking into consideration the required homogeneity of sub-volumes. In addition, the "overlapped" issues with respect to the decomposition should be considered, which may in turn serve to improve in the overall accuracy of the proposed volumetric decomposition and graph based approach.

REFERENCES

[1] J. Ho, A. C. Sull, L. N. Vuong, Y. Chen, J. Liu, J. G. Fujimoto, J. S. Schuman, and J. S. Duker, "Assessment of artifacts and reproducibility across spectral- and time-domain optical coherence tomography devices," *Ophthalmology*, vol. 116, no. 10, pp. 1960 – 1970, 2009.

[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*, vol. 144, no. 4, pp. 592 – 599, 2007.

[3] R. D. Jager, W. F. Mieler, and J. W. Miller, "Age-related macular degeneration," *New England Journal of Medicine*, vol. 358, no. 24, pp. 2606–2617, 2008.

[4] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation," *Optical Express*, vol. 18, no. 18, pp. 19 413–19 428, Aug 2010.

[5] N. Jain, S. Farsiu, A. A. Khanifar, S. Bearelly, R. T. Smith, J. A. Izatt, and C. A. Toth, "Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs," *Investigative Ophthalmology and Visual Science*, 2010.

[6] 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*, vol. 15, no. 5, pp. 748–759, 2011.

[7] 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*, vol. 8, no. 3, pp. 570–575, 2003.

[8] M. Ankerst, G. Kastenmüller, H.-P. Kriegel, and T. Seidl, “3d shape histograms for similarity search and classification in spatial databases,” in *Proceedings of the 6th International Symposium on Advances in Spatial Databases*, ser. SSD ’99. London, UK, UK: Springer-Verlag, 1999, pp. 207–226.

[9] M. 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*, ser. Lecture Notes in Computer Science, D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis, Eds. Springer Berlin Heidelberg, 2011, vol. 6912, pp. 65–80.

[10] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1992.

[11] Y. Zang, T. Jiang, Y. Lu, Y. He, and L. Tian, “Regional homogeneity approach to FMRI data analysis,” *NeuroImage*, vol. 22, no. 1, pp. 394–400, 2004.

[12] E. Keogh, “Exact indexing of dynamic time warping,” in *Proceedings of the 28th international conference on Very Large Data Bases*, ser. VLDB ’02. VLDB Endowment, 2002, pp. 406–417.

[13] 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*, ser. KDD ’03. New York, NY, USA: ACM, 2003, pp. 216–225.

[14] D. H. Johnson and S. Sinanovic, “Symmetrizing the kullback-leibler distance,” *IEEE Transactions on Information Theory*, Mar. 2001.

[15] X. Yan and J. Han, “gspan: Graph-based substructure pattern mining,” in *Proceedings of the 2002 IEEE International Conference on Data Mining*, ser. ICDM ’02. Washington, DC, USA: IEEE Computer Society, 2002, pp. 721–724.

[16] H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda, “gboost: a mathematical programming approach to graph classification and regression,” *Machine Learning*, vol. 75, no. 1, pp. 69–89, Apr. 2009.